



亚马逊EMR发布指南

# Amazon EMR



# Amazon EMR: 亚马逊EMR发布指南

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

Amazon 的商标和商业外观不得用于任何非 Amazon 的商品或服务，也不得以任何可能引起客户混淆、贬低或诋毁 Amazon 的方式使用。所有非 Amazon 拥有的其他商标均为各自所有者的财产，这些所有者可能附属于 Amazon、与 Amazon 有关联或由 Amazon 赞助，也可能不是如此。

# Table of Contents

关于 Amazon EMR 发行 .....	1
标准支持 .....	2
了解 Amazon EMR 发布的版本 .....	2
发布版本控制 .....	3
支持策略 .....	3
注意事项 .....	6
亚马逊 EMR 7.x 发布版本 .....	7
亚马逊 EMR 7.x 版本中的应用程序版本 .....	7
emr-7.2.0 .....	7
emr-7.1.0 .....	59
emr-7.0.0 .....	81
亚马逊 EMR 6.x 发布版本 .....	113
Amazon EMR 6.x 发行版中的应用程序版本 .....	115
emr-6.15.0 .....	115
emr-6.14.0 .....	159
emr-6.13.0 .....	203
emr-6.12.0 .....	253
emr-6.11.1 .....	308
emr-6.11.0 .....	358
emr-6.10.1 .....	390
emr-6.10.0 .....	442
emr-6.9.1 .....	479
emr-6.9.0 .....	530
emr-6.8.1 .....	573
emr-6.8.0 .....	624
emr-6.7.0 .....	666
emr-6.6.0 .....	732
emr-6.5.0 .....	799
emr-6.4.0 .....	822
emr-6.3.1 .....	850
emr-6.3.0 .....	873
emr-6.2.1 .....	900
emr-6.2.0 .....	922
emr-6.1.1 .....	950

emr-6.1.0 .....	967
emr-6.0.1 .....	988
emr-6.0.0 .....	1002
亚马逊 EMR 5.x 发布版本 .....	1020
Amazon EMR 5.x 发行版中的应用程序版本 .....	1023
emr-5.36.2 .....	1024
emr-5.36.1 .....	1047
emr-5.36.0 .....	1098
emr-5.35.0 .....	1133
emr-5.34.0 .....	1156
emr-5.33.1 .....	1177
emr-5.33.0 .....	1202
emr-5.32.1 .....	1221
emr-5.32.0 .....	1242
emr-5.31.1 .....	1266
emr-5.31.0 .....	1281
emr-5.30.2 .....	1300
emr-5.30.1 .....	1316
emr-5.30.0 .....	1334
emr-5.29.0 .....	1352
emr-5.28.1 .....	1368
emr-5.28.0 .....	1383
emr-5.27.1 .....	1399
emr-5.27.0 .....	1413
emr-5.26.0 .....	1428
emr-5.25.0 .....	1444
emr-5.24.1 .....	1460
emr-5.24.0 .....	1474
emr-5.23.1 .....	1489
emr-5.23.0 .....	1502
emr-5.22.0 .....	1517
emr-5.21.2 .....	1533
emr-5.21.1 .....	1546
emr-5.21.0 .....	1560
emr-5.20.1 .....	1576
emr-5.20.0 .....	1589

emr-5.19.1 .....	1605
emr-5.19.0 .....	1619
emr-5.18.1 .....	1633
emr-5.18.0 .....	1647
emr-5.17.2 .....	1661
emr-5.17.1 .....	1674
emr-5.17.0 .....	1687
emr-5.16.1 .....	1701
emr-5.16.0 .....	1714
emr-5.15.1 .....	1728
emr-5.15.0 .....	1741
emr-5.14.2 .....	1755
emr-5.14.1 .....	1768
emr-5.14.0 .....	1781
emr-5.13.1 .....	1795
emr-5.13.0 .....	1808
emr-5.12.3 .....	1820
emr-5.12.2 .....	1833
emr-5.12.1 .....	1845
emr-5.12.0 .....	1858
emr-5.11.4 .....	1871
emr-5.11.3 .....	1884
emr-5.11.2 .....	1896
emr-5.11.1 .....	1908
emr-5.11.0 .....	1920
emr-5.10.1 .....	1933
emr-5.10.0 .....	1945
emr-5.9.1 .....	1959
emr-5.9.0 .....	1971
emr-5.8.3 .....	1985
emr-5.8.2 .....	1997
emr-5.8.1 .....	2009
emr-5.8.0 .....	2021
emr-5.7.1 .....	2034
emr-5.7.0 .....	2046
emr-5.6.1 .....	2058

emr-5.6.0 .....	2070
emr-5.5.4 .....	2082
emr-5.5.3 .....	2094
emr-5.5.2 .....	2106
emr-5.5.1 .....	2118
emr-5.5.0 .....	2130
emr-5.4.1 .....	2143
emr-5.4.0 .....	2155
emr-5.3.2 .....	2167
emr-5.3.1 .....	2179
emr-5.3.0 .....	2191
emr-5.2.3 .....	2203
emr-5.2.2 .....	2215
emr-5.2.1 .....	2227
emr-5.2.0 .....	2239
emr-5.1.1 .....	2251
emr-5.1.0 .....	2263
emr-5.0.3 .....	2275
emr-5.0.2 .....	2286
emr-5.0.1 .....	2297
emr-5.0.0 .....	2308
亚马逊 EMR 4.x 发布版本 .....	2320
Amazon EMR 4.x 发行版中的应用程序版本 .....	2322
各发行版之间的差异 .....	2322
emr-4.9.6 .....	2365
emr-4.9.5 .....	2376
emr-4.9.4 .....	2388
emr-4.9.3 .....	2399
emr-4.9.2 .....	2410
emr-4.9.1 .....	2422
emr-4.8.5 .....	2433
emr-4.8.4 .....	2445
emr-4.8.3 .....	2456
emr-4.8.2 .....	2468
emr-4.8.1 .....	2479
emr-4.8.0 .....	2491

emr-4.7.4 .....	2502
emr-4.7.3 .....	2513
emr-4.7.2 .....	2523
emr-4.7.1 .....	2534
emr-4.7.0 .....	2545
emr-4.6.1 .....	2557
emr-4.6.0 .....	2566
emr-4.5.0 .....	2577
emr-4.4.0 .....	2586
emr-4.3.0 .....	2596
emr-4.2.0 .....	2605
emr-4.1.0 .....	2614
emr-4.0.0 .....	2621
2.x 和 3. AMI x 版本 .....	2627
创建集群 .....	2628
安装应用程序 .....	2630
自定义配置 .....	2630
Hive .....	2637
HBase .....	2646
Pig .....	2657
Spark .....	2662
S3 DistCp .....	2665
新增功能 .....	2668
亚马逊 EMR 7.2.0 .....	2668
亚马逊 EMR 6.15.0 .....	2670
亚马逊 EMR 5.36.2 .....	2689
SigV4兼容性 .....	2694
缓解方法 CVE -2021-44228 .....	2694
适用于 Log4j 的亚马逊EMR引导操作解决方案-2021-44228 和 -2021-45046 CVE CVE .....	2695
常见问题 .....	2702
档案 .....	2704
发行版 6.14.0 .....	2704
发行版 6.13.0 .....	2723
版本 6.12.0 .....	2748
版本 6.11.1 .....	2777
版本 6.11.0 .....	2805

版本 6.10.0 .....	2814
发行版 6.9.0 .....	2827
发行版 6.8.0 .....	2846
发行版 6.7.0 .....	2866
发行版 6.6.0 .....	2909
发行版 5.35.0 .....	2955
发行版 5.34.0 .....	2959
发行版 6.5.0 .....	2961
发行版 6.4.0 .....	2962
发行版 5.32.0 .....	2969
发行版 6.2.0 .....	2974
发行版 5.31.0 .....	2981
发行版 6.1.0 .....	2985
发行版 6.0.0 .....	2991
发行版 5.30.1 .....	2995
发行版 5.30.0 .....	2999
发行版 5.29.0 .....	3003
版本 5.28.1 .....	3005
版本 5.28.0 .....	3006
版本 5.27.0 .....	3008
版本 5.26.0 .....	3010
版本 5.25.0 .....	3012
版本 5.24.1 .....	3014
版本 5.24.0 .....	3015
版本 5.23.0 .....	3017
版本 5.22.0 .....	3019
发布版本 5.21.1 .....	3022
版本 5.21.0 .....	3023
版本 5.20.0 .....	3025
版本 5.19.0 .....	3028
版本 5.18.0 .....	3029
发布版本 5.17.1 .....	3030
版本 5.17.0 .....	3031
版本 5.16.0 .....	3032
版本 5.15.0 .....	3033
版本 5.14.1 .....	3034



版本 5.14.0 .....	3034
版本 5.13.0 .....	3036
版本 5.12.2 .....	3037
版本 5.12.1 .....	3037
版本 5.12.0 .....	3037
发布版本 5.11.3 .....	3039
版本 5.11.2 .....	3039
版本 5.11.1 .....	3039
版本 5.11.0 .....	3040
版本 5.10.0 .....	3041
版本 5.9.0 .....	3043
版本 5.8.2 .....	3044
版本 5.8.1 .....	3044
版本 5.8.0 .....	3045
版本 5.7.0 .....	3047
版本 5.6.0 .....	3047
版本 5.5.3 .....	3048
版本 5.5.2 .....	3048
版本 5.5.1 .....	3048
版本 5.5.0 .....	3049
版本 5.4.0 .....	3050
版本 5.3.1 .....	3051
版本 5.3.0 .....	3051
版本 5.2.2 .....	3052
版本 5.2.1 .....	3052
版本 5.2.0 .....	3053
版本 5.1.0 .....	3054
版本 5.0.3 .....	3054
版本 5.0.0 .....	3055
版本 4.9.5 .....	3056
版本 4.9.4 .....	3056
版本 4.9.3 .....	3057
版本 4.9.2 .....	3057
版本 4.9.1 .....	3057
版本 4.8.4 .....	3058
版本 4.8.3 .....	3058

版本 4.8.2 .....	3059
版本 4.8.0 .....	3059
版本 4.7.2 .....	3060
版本 4.7.1 .....	3061
版本 4.7.0 .....	3061
版本 4.6.0 .....	3062
版本 4.5.0 .....	3064
版本 4.4.0 .....	3064
版本 4.3.0 .....	3066
版本 4.2.0 .....	3067
配置应用程序 .....	3069
在创建集群时配置应用程序 .....	3071
在创建集群时，在控制台中提供配置 .....	3071
创建集群 AWS CLI 时使用提供配置 .....	3072
创建集群SDK时使用 Java 提供配置 .....	3072
在正在运行的集群中重新配置实例组 .....	3073
重新配置实例组时的注意事项 .....	3074
在控制台中重新配置实例组 .....	3076
使用重新配置实例组 CLI .....	3077
使用 Java 重新配置实例组 SDK .....	3081
故障排除 .....	3083
在 AWS Secrets Manager中存储敏感配置数据 .....	3085
创建密钥 .....	3085
授予 Amazon 检索密钥的EMR权限 .....	3085
在配置分类中使用密钥 .....	3086
更新密钥值 .....	3087
配置应用程序来使用特定 Java 虚拟机 .....	3087
注意事项 .....	3087
覆盖 JVM .....	3089
服务端口 .....	3091
应用程序用户 .....	3093
使用项目存储库检查依赖项 .....	3095
EMR文件系统 (EMRFS) .....	3097
一致视图 .....	3099
启用一致视图 .....	3102
了解EMRFS一致视图如何跟踪 Amazon S3 中的对象 .....	3104

重试逻辑 .....	3105
EMRFS一致的视图元数据 .....	3106
为 CloudWatch 和 Amazon 配置一致性通知 SQS .....	3109
配置一致视图 .....	3110
EMRFSCLI命令参考 .....	3114
授权访问 Amazon S3 中的EMRFS数据 .....	3124
为 Amazon S3 中的EMRFS数据创建自定义凭证提供程序 .....	3124
管理默认 AWS Security Token Service 端点 .....	3125
使用EMRFS属性指定 Amazon S3 加密 .....	3126
AWS KMS keys 用于EMRFS加密 .....	3127
Amazon S3 服务器端加密 .....	3128
Amazon S3 客户端加密 .....	3130
亚马逊 CloudWatch 代理 .....	3138
创建集群 .....	3140
所需的权限 .....	3140
所需端点 .....	3140
创建集群 .....	3141
默认指标 .....	3142
配置 .....	3144
亚马逊 EMR 7.1.0 .....	3145
亚马逊 EMR 7.0.0 .....	3168
注意事项 .....	3172
历史记录 .....	3173
Delta Lake .....	3175
简介 .....	3175
使用 Delta Lake 集群 .....	3176
将 Delta Lake 与 Flink 结合使用 .....	3176
将 Delta Lake 与 Trino 结合使用 .....	3180
将 Delta Lake 与 Spark 结合使用 .....	3182
将 Delta Lake 与 Spark 和 Glue 结合使用 .....	3188
注意事项 .....	3188
历史记录 .....	3189
Flink .....	3191
使用 Flink 创建集群 .....	3193
配置 Flink .....	3194
Hive 和 Glue .....	3194

Config 文件 .....	3196
多个主节点 .....	3198
内存进程大小 .....	3198
日志输出文件大小 .....	3199
Java 11 .....	3200
Flink 作业 .....	3204
启动 Flink YARN 应用程序，作为长时间运行集群上的步骤 .....	3205
将工作提交到长时间运行集群上的现有 Flink 应用程序 .....	3206
提交临时 Flink 作业 .....	3207
Flink Scala Shell .....	3209
Flink UI .....	3210
Flink Autoscaler .....	3211
概述 .....	3211
注意事项 .....	3212
启用 Autoscaler .....	3212
配置 .....	3213
优化重启时间 .....	3217
任务本地恢复 .....	3218
增量检查点 .....	3218
精细恢复 .....	3219
组合重启 .....	3220
通过 Zeppelin 使用 Flink .....	3221
简介 .....	3221
先决条件 .....	3221
在 EMR 集群上配置 Zeppelin-Flink .....	3221
在 EMR 集群上使用 Zeppelin-Flink 运行 Flink 作业 .....	3223
Flink 发布历史记录 .....	3228
按版本划分的 Flink 发行说明 .....	3263
Ganglia .....	3265
使用 Ganglia 创建集群 .....	3266
查看 Ganglia 指标 .....	3268
Ganglia 中的 Hadoop 和 Spark 指标 .....	3269
Ganglia 发行版历史记录 .....	3270
Hadoop .....	3315
配置 Hadoop .....	3317
任务配置 .....	3317

Hadoop 守护进程配置设置 .....	3678
HDFS配置 .....	3948
Amazon HDFS 上有透明加密 EMR .....	3949
配置HDFS透明加密 .....	3950
HDFS透明加密的注意事项 .....	3952
Hadoop 密钥管理服务器 .....	3952
HDFS在具有多个主节点的EMR集群上进行透明加密 .....	3956
创建或运行 Hadoop 应用程序 .....	3958
使用 Amazon EMR 构建二进制文件 .....	3958
通过流式处理来处理数据 .....	3960
使用自定义处理数据 JAR .....	3965
读取已恢复的对象 .....	3968
示例 .....	3968
注意事项 .....	3969
为容器开启非均匀内存访问感知 YARN .....	3970
Hadoop 版本历史记录 .....	3972
Hadoop 发布说明 ( 按版本分类 ) .....	4018
HBase .....	4024
使用 HBase 创建集群 .....	4027
HBase使用控制台创建集群 .....	4027
HBase使用创建集群 AWS CLI .....	4028
HBase在亚马逊 S3 ( 亚马逊 S3 存储模式 ) 上 .....	4029
HBase在 Amazon S3 上启用 .....	4030
使用只读副本集群 .....	4031
持续HFile跟踪 .....	4032
操作注意事项 .....	4034
Amazon 的预写日志 (WAL) EMR .....	4037
WAL工作空间 .....	4038
所需的权限 .....	4039
正在启用 WAL .....	4040
从中恢复 WAL .....	4042
安全配置 .....	4043
使用 AWS PrivateLink .....	4044
WAL定价和指标 .....	4045
为工作空间添加标签 WAL .....	4046
注意事项和可用性 .....	4047

EMRWALCLI参考 .....	4049
使用外HBase壳 .....	4052
创建表 .....	4052
设置值 .....	4052
获取值 .....	4052
删除表 .....	4053
使用 Hive 访问HBase表 .....	4053
使用HBase快照 .....	4054
使用表创建快照 .....	4055
删除快照 .....	4055
查看快照信息 .....	4055
将快照导出到 Amazon S3 .....	4055
从 Amazon S3 导入快照 .....	4056
从 s HBase hell 中的快照还原表 .....	4057
配置 HBase .....	4058
中对内存分配的更改 YARN .....	4059
HBase端口号 .....	4059
HBase要优化的网站设置 .....	4060
查看HBase用户界面 .....	4062
查看 HBase 日志文件 .....	4063
HBase使用 Ganglia 进行监控 .....	4064
从以前的HBase版本迁移 .....	4065
HBase发布历史 .....	4065
HCatalog .....	4138
使用 HCatalog 创建集群 .....	4139
使用 HCatalog .....	4140
使用时禁用直接写入 HCatalog HStorer .....	4140
使用创建表HCatCLI并在 Pig 中使用该数据 .....	4141
使用 Spark 访问表格 SQL .....	4142
示例：创建HCatalog表并使用 Pig 写入该表 .....	4143
HCatalog发布历史 .....	4144
Hive .....	4213
亚马逊 Hive 的区别和注意事项 EMR .....	4215
亚马逊上的 Apache Hive EMR 和 Apache Hive 之间的区别 .....	4215
亚马逊EMR发布版本 4.x 和 5.x 之间的 Hive 差异 .....	4216
亚马逊 Hive 的其他功能 EMR .....	4217

为 Hive 配置外部元存储 .....	4222
使用 AWS Glue 数据目录作为 Hive 的元数据库 .....	4223
使用外部“我的SQL数据库”或 Amazon Aurora .....	4228
使用 Hive JDBC 驱动程序 .....	4230
改进 Hive 性能 .....	4232
启用 Hive EMRFS S3 优化提交器 .....	4233
使用 S3 Select .....	4234
MSCK优化 .....	4236
使用 Hive LLAP .....	4237
启用 LLAP .....	4237
LLAP在您的集群上启动 .....	4238
查看LLAP状态 .....	4239
启动或停止 LLAP .....	4239
调整LLAP守护程序计数的大小 .....	4239
Hive 中的加密 .....	4240
Hive 中的 Parquet 模块化加密 .....	4240
在传输中加密 HS2 .....	4243
Hive 发行历史记录 .....	4245
Hive 发布说明 ( 按版本分类 ) .....	4314
Hudi .....	4382
Hudi 的工作原理 .....	4383
了解数据集存储类型：写入时复制与读取时合并 .....	4384
将 Hudi 数据集注册到您的元数据仓 .....	4385
注意事项和限制 .....	4385
创建安装了 Hudi 的集群 .....	4386
使用 Hudi 数据集 .....	4387
初始化 Hudi 的 Spark 会话 .....	4391
写入 Hudi 数据集 .....	4391
更新插入数据 .....	4396
删除记录 .....	4397
从 Hudi 数据集读取 .....	4398
使用 Hudi CLI .....	4400
Hudi 发行版历史记录 .....	4401
Hue .....	4404
Hue 版本 .....	4404
Amazon 上支持和不支持的 Hue 功能 EMR .....	4406

注意事项 .....	4407
使用大型 Hue 元数据表时的性能 .....	4407
Hue 版本之间的不兼容 .....	4408
连接到 Hue Web 用户界面 .....	4408
在 Amazon 中将 Hue 与远程数据库配合使用 RDS .....	4409
故障排除 .....	4411
Hue 的高级配置 .....	4411
为LDAP用户配置 Hue .....	4411
Hue 发行版历史记录 .....	4414
Iceberg .....	4461
Iceberg 的工作原理 .....	4462
使用带 Iceberg 的集群 .....	4463
将 Iceberg 集群与 Spark 结合使用 .....	4463
将 Iceberg 集群与 Trino 结合使用 .....	4468
将 Iceberg 集群与 Flink 结合使用 .....	4470
将 Iceberg 集群与 Hive 结合使用 .....	4474
注意事项和限制 .....	4477
将 Iceberg 与 Spark 结合使用的注意事项 .....	4477
将 Iceberg 与 Trino 结合使用的注意事项 .....	4477
将 Iceberg 与 Flink 结合使用的注意事项 .....	4478
将 Iceberg 与 Hive 结合使用的注意事项 .....	4478
Iceberg 发布历史记录 .....	4478
Iceberg 发布说明 ( 按版本分类 ) .....	4480
Jupyter Notebook .....	4482
EMR Studio .....	4482
EMR Notebook .....	4482
JupyterHub .....	4482
使用创建集群 JupyterHub .....	4486
JupyterHub 在 Amazon 上使用时的注意事项 EMR .....	4487
正在配置 JupyterHub .....	4488
在 Amazon S3 中配置笔记本的持久性 .....	4489
连接到主节点和笔记本服务器 .....	4490
JupyterHub 配置和管理 .....	4491
添加 Jupyter notebook 用户和管理员 .....	4492
安装其它内核和库 .....	4502
JupyterHub 发布历史 .....	4506



Livy .....	4543
启用 HTTPS .....	4545
Livy 发行历史记录 .....	4546
MXNet .....	4590
MXNet发布历史 .....	4591
Oozie .....	4620
在亚马逊中将 Oozie 与远程数据库配合使用 RDS .....	4622
为 Oozie 配置 Java 版本 .....	4624
Oozie 发行历史记录 .....	4625
Oozie 按版本划分的发行说明 .....	4682
Phoenix .....	4683
使用 Phoenix 创建集群 .....	4685
自定义 Phoenix 配 .....	4686
Phoenix 客户端 .....	4687
Phoenix 发行历史记录 .....	4691
Pig .....	4764
提交 Pig 工作 .....	4766
使用亚马逊EMR控制台提交 Pig 作品 .....	4766
使用提交 Pig 作品 AWS CLI .....	4767
从 Pig 调用由用户定义的函数 .....	4769
从 JAR Pig 那里调用文件 .....	4769
从 Pig 调用 Python/Jython 脚本 .....	4769
Pig 发行历史记录 .....	4770
Presto 和 Trino .....	4836
在 AWS Glue 数据目录中使用 Presto .....	4839
将 AWS Glue 数据目录指定为元数据库 .....	4839
IAM权限 .....	4225
使用 AWS Glue 数据目录时的注意事项 .....	4843
使用 S3 Select Pushdown .....	4844
S3 Select Pushdown 是否适合我的应用程序？ .....	4844
注意事项和限制 .....	4844
启用 S3 Select Pushdown with PrestoDB 或 Trino .....	4845
添加数据库连接器 .....	4845
使用SSL/TLS和 LDAPS .....	4846
使用LDAP身份验证 .....	4847
激活 Presto 严格模式 .....	4854

注意事项 .....	4855
在 Presto 中处理竞价型实例丢失 .....	4856
容错执行 .....	4857
配置 .....	4857
交换管理器 .....	4858
注意事项和限制 .....	4859
使用采用 Graceful Decommission 的 Presto 自动扩展配置 .....	4859
亚马逊 Presto 的注意事项 EMR .....	4860
Presto 命令行可执行文件 .....	4860
不可配置的 Presto 部署属性 .....	4861
PrestoDB 和 Trino 安装 .....	4862
EMRFS和 Presto FileSystem S3 配置 .....	4862
终端用户模拟的默认设置 .....	4863
Presto Web 界面的默认端口 .....	4863
某些版本中的 Hive 存储桶执行问题 .....	4863
Presto 发行历史记录 .....	4863
各版本的 Trino (PrestoSQL) 发行说明 .....	4920
Spark .....	4923
创建 Spark 集群 .....	4926
在亚马逊 EMR 6.x 上使用 Docker 运行 Spark 应用程序 .....	4929
注意事项 .....	4929
创建 Docker 镜像 .....	4930
使用来自亚马逊的 Docker 镜像 ECR .....	4931
使用 AWS Glue 数据目录作为 Spark 的元数据库 SQL .....	4936
将 AWS Glue 数据目录指定为元数据库 .....	4937
IAM权限 .....	4225
注意事项 .....	4227
配置 Spark .....	4940
Spark 默认值 .....	4941
设置垃圾回收 .....	4943
maximizeResourceAllocation .....	4943
停用行为 .....	4945
Spark ThriftServer 环境变量 .....	4947
更改 Spark 默认设置 .....	4947
从 Apache Log4j 1.x 迁移到 Log4j 2.x .....	4949
优化 Spark 性能 .....	4950

自适应查询执行 .....	4950
动态分区修剪 .....	4952
展平标量子查询 .....	4954
DISTINCT之前 INTERSECT .....	4954
Bloom 筛选条件连接 .....	4955
优化的连接重新排序 .....	4956
结果片段缓存 .....	4956
启用 Spark 结果片段缓存 .....	4957
注意事项 .....	4958
使用RAPIDS加速器 .....	4959
选择实例类型 .....	4959
设置应用程序配置 .....	4959
添加引导操作 .....	4967
启动您的集群。 .....	4967
访问 Spark Shell .....	4968
使用 Amazon SageMaker Spark 进行机器学习 .....	4970
编写 Spark 应用程序 .....	4970
Scala .....	4970
Java .....	4971
Python .....	4972
使用 S3 提高 Spark 性能 .....	4973
使用 S3 Select .....	4974
EMRFSS3 优化的提交者 .....	4977
使用 EMRFS S3 优化的提交协议 .....	4983
重试 S3 请求 .....	4988
添加 Spark 步骤 .....	4991
覆盖 Spark 默认配置设置 .....	4994
查看 Spark 应用程序历史记录 .....	4994
访问 Spark 网站 UIs .....	4994
将 Spark 与亚马逊 Kinesis Data Streams 配合使用 .....	4995
在 Amazon Redshift 上使用 Spark .....	4995
启动 Spark 应用程序 .....	4995
对 Amazon Redshift 进行身份验证 .....	4996
对 Amazon Redshift 进行读取和写入 .....	4999
注意事项 .....	5000
Spark 发行历史记录 .....	5001

Sqoop .....	5065
Sqoop 版本 .....	5065
在亚马逊上使用 Sqoop 的注意事项 EMR .....	5067
将 Sqoop 与集成一起使用 HCatalog .....	5067
Sqoop JDBC 和数据库支持 .....	5067
保护密码 .....	5068
Sqoop 发行历史记录 .....	5069
TensorFlow .....	5108
TensorFlow 按 Amazon EC2 实例类型构建 .....	5109
安全性 .....	5110
使用 TensorBoard .....	5110
TensorFlow 发布历史 .....	5111
Tez .....	5133
使用 Tez 创建集群 .....	5135
配置 Tez .....	5135
示例配置 .....	5136
Tez 异步拆分打开 .....	5137
Tez Web UI .....	5138
时间线服务器 .....	5139
Tez 发行历史记录 .....	5139
Tez 发布说明 ( 按版本分类 ) .....	5183
Zeppelin .....	5189
在 Amazon 上使用齐柏林飞艇时的注意事项 EMR .....	5191
Zeppelin 发行历史记录 .....	5191
ZooKeeper .....	5247
ZooKeeper 发布历史 .....	5248
连接器和实用工具 .....	5288
导出、查询和连接 DynamoDB 中的表格 .....	5288
设置 Hive 表来运行 Hive 命令 .....	5290
用于导出、导入和查询数据的 Hive 命令示例 .....	5297
优化性能 .....	5305
Kinesis .....	5308
我能用亚马逊EMR和亚马逊 Kinesis 的集成做什么？ .....	5308
对 Amazon Kinesis 流进行检查点分析 .....	5309
性能注意事项 .....	5310
与亚马逊一起安排亚马逊 Kinesis 分析 EMR .....	5310

将适用于亚马逊 7.0 的 Spark Kinesis 连接器迁移到 SDK 2.x EMR .....	5310
S3 DistCp (s3-dist-cp) .....	5319
S3 DistCp 选项 .....	5319
将 S3 DistCp 作为集群中的一个步骤添加 .....	5325
在 S3 DistCp 任务失败后进行清理 .....	5326
在集群上运行命令和脚本 .....	5328
提交自定义 JAR 步骤以运行脚本或命令 .....	5328
其他使用 <code>command-runner.jar</code> 的方法 .....	5329
AWS 术语表 .....	5331
.....	5332

# 关于 Amazon EMR 发行

Amazon EMR 版本是一组来自大数据生态系统的开源应用程序。每个版本都包含不同的的大数据应用程序、组件和功能，您在创建集群时选择让 Amazon EMR 安装和配置这些应用程序、组件和功能。应用程序使用基于Apache的系统进行打包 BigTop，[Apache](#) 是一个与Hadoop生态系统相关的开源项目。本指南提供了 Amazon EMR 版本中包含的应用程序的信息。

有关入门和使用亚马逊的更多信息EMR，请参阅 [《亚马逊EMR管理指南》](#)。

启动集群时，您可以从 Amazon 的多个版本中进行选择EMR。这允许您测试和使用满足您解决方案兼容性需求的应用程序版本。您可以使用发行版标注指定版本号。版本标签的格式是 `emr-x.x.x`。例如，`emr-7.2.0`。

从 Amazon EMR 5.18.0 开始，您可以使用 Amazon 构EMR件存储库针对特定亚马逊版本中提供的库和依赖项的确切版本构建任务代码。EMR有关更多信息，请参阅 [使用 Amazon EMR 项目存储库检查依赖项](#)。

要在亚马逊新EMR版本发布时获取更新，请订阅[亚马逊EMR发行说明的RSS提要](#)。

最新版本详情，包括 Amazon EMR 7.x、6.x 和 5.x 系列的应用程序版本、发行说明、组件和配置分类：

- [亚马逊 7.2.0 EMR 版本](#)
- [亚马逊 6. EMR 15.0 版本](#)
- [亚马逊EMR版本 5.36.2](#)

## Note

Amazon 的新EMR版本将在几天内在不同地区上市，从最初发布日期的第一个地区开始。在此期间，您所在区域可能无法提供最新发行版。

Amazon 最新@@ 版本的EMR发行说明和所有版本的历史记录：

- [新增功能](#)
- [Amazon 发行说明EMR档案](#)

每个 Amazon EMR版本中应用程序版本的完整历史记录：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

每个 Amazon EMR 版本的详细信息以及版本系列之间的差异（如适用）：

- [亚马逊 EMR 7.x 发布版本](#)
- [亚马逊 EMR 6.x 发布版本](#)
- [亚马逊 EMR 5.x 发布版本](#)
- [亚马逊 EMR 4.x 发布版本](#)
- [亚马逊 EMR 2.x 和 3. AMI x 版本](#)

## 亚马逊EMR标准支持

### 了解 Amazon EMR 发布的版本

Amazon EMR 版本是提供在EC2、EKS或 Serverless 平台上运行开源应用程序所需的软件的工具。Amazon EMR 版本由三部分组成：运行时环境、核心引擎和额外内容。

- 运行时环境 — 运行时环境包括在您启动集群或容器（例如 Amazon Linux 2023 AMI）时作为亚马逊系统映像（）或其他容器映像一部分包含的操作系统。它还包括语言运行时，例如 Amazon JDK Corretto，以及映像安装的其他相关工具。
- 核心引擎 — 这些引擎包括 Apache Hive 等核心开源软件和 Apache Iceberg 等开放表格格式。
- 额外内容 — 其中包括便利库和 Python 软件包，例如 mariadb-connector-java 以及 Apache Pig 等开源软件。

您选择的亚马逊EMR发行版本将所有这些组件捆绑到基于亚马逊 Linux (AL) 的新版本的亚马逊系统映像 (AMI) 或容器映像中。Amazon 为集群、EKS容器或无服务器应用程序的所有节点EMR维护相同的 AL 版本。亚马逊的EMR目标是在Amazon Linux发布后90天内获得最新的运行时环境，并在上游版本发布后的90天内获得新的开源版本的核心引擎。额外内容将根据需要发布。您可以在每个 Amazon EMR 版本的发行说明中找到该软件的列表。

## 发布版本控制

为了帮助您了解每个版本的变更范围，Amazon EMR 使用语义版本控制。语义版本控制遵循的格式为。`<major>.<minor>.<patch>`举个例子，让我们来看看 7.0.0 版本。

第一个数字表示主要版本或主要版本。主要版本通常会为运行时环境或核心引擎带来不向后兼容的重大更改、改进和新功能。例如，7.x 主要版本使用亚马逊 Linux 2023，默认为亚马逊 Corretto 17 JDK。因此，与6.x主要版本相比，该版本包含一些重大更改，后者在亚马逊Linux 2上发布，默认使用 Amazon Corretto 8。

第二位数字表示次要版本或次要版本。次要版本是不向后兼容的版本，其中包含对核心引擎和额外内容的增量更改、改进和功能以及新功能。例如，6.15 使用 Apache Spark 3.4.1，而 Apache Spark 3.3.2 附带的 6.11.1。

第三位数字表示补丁版本或补丁版本。补丁版本旨在成为向后兼容的版本，其中包含核心引擎和额外内容的修复和补丁更新，但不包含新功能或OSS版本升级。例如，6.11.1 和 6.11.0 都包含相同的 Apache Spark 3.3.2 版本。OSS从 6.6.x 开始，最新的补丁版本提供了可用的最新运行时环境补丁。运行环境不会影响 Amazon 的补丁发布EMR。例如，6.11.1 于 2023 年 12 月 19 日至 2024 年 1 月 10 日在 AL 版本 2.0.20240109.0 上线，但从 2024 年 1 月 11 日起在 AL 版本 2.0.20240124.0 上线。某些 Amazon EMR on EKS 版本允许您在语义版本中添加标签或后缀，以便为您提供替代功能。例如，亚马逊的 `emr-6.15.0-latest` 发布支持亚马逊 Corretto JDK 8 的发布版本，而亚马逊EMR上发布的版本允许你启动支持 Amazon EMR on EKS on Corretto 17 的 `emr-6.15.0-java17-latest` 发布版本。JDK

## 支持策略

### 介绍

Amazon 的EMR目标是至少每 90 天发布一次新的次要版本，并从发布之日起 24 个月内为次要版本提供支持。此支持涵盖运行时环境和核心引擎及其相关依赖项，但不包括额外内容或其关联的依赖项。Runtime Environment APIs 还包括启动、停止和操作集群和应用程序所需的专有平台组件。这种一致的发布时间表确保了可预测的周期，使您可以更方便地规划、测试和过渡到提供支持和安全的版本。

### Standard Support 会带来什么

Standard Support 修复了您在推荐配置下遇到的运行时环境和核心引擎组件的问题。所有修复均视可用性而定。推荐的配置是使用 Amazon EMR 而无需对亚马逊EMR版本中存在的二进制文件和配置进行任何修改、添加或更改，但[亚马逊EMR文档](#)中的二进制文件和配置除外。在我们验证修复后 90 天



内，Amazon EMR 会针对最新补丁、次要版本或主要版本EMR部署修补程序。当您在EC2集群EMR上启动新的 Amazon、在EKS容器上启动新的 Amaz EMR on 或触发新的EMR无服务器任务时，Amazon 会EMR自动应用修复程序。为了方便起见，提供了额外的组件，Amazon EMR 不提供与Extras 相关的修复程序。

## 组成部分

标准支持涵盖运行时环境和核心引擎组件，例如操作系统、语言运行时以及核心开源软件，例如 Apache Hive 和 Apache Iceberg。您可以在 Amazon Linux [关于 Amazon EMR 发行](#) 和 [Amazon Linux](#) 中找到每个版本的支持组件的完整列表FAQs。

以下列表描述了我们在 Standard Support 下为不同组件类型提供的支持：

- **运行时环境组件**：运行时环境组件将收到有关技术支持票证的修复。修复分为 (a) 严重错误、(b) 关键数据损坏问题和 (c) 关键安全问题。符合条件时，Amazon EMR 会将运行时环境的修复向后移植到旧版本。为了与开源 Core Engine 组件保持操作兼容性，某些运行时环境组件必须保持特定版本，以防止破坏客户应用程序。对于这些组件，EMRAmazon 依赖上游开源来提供修复程序。当有开源版本的修复程序可用时，我们将在 Amazon 验证后的 90 天内提供最新的稳定版本EMR。
- **核心引擎组件**：核心引擎为许多开源项目提供了最新版本，每个项目都有数百个传递依赖库。虽然管理这些项目的开源社区经常尝试解决问题和已知的常见漏洞和暴露 (CVE)，但最新版本可能仍包含已知的错误和CVEs。EMRAmazon 依赖上游开源来提供修复程序，并将在亚马逊验证后的 90 天内作为核心引擎组件的一部分提供最新的稳定版本EMR。在某些情况下，Amazon EMR 可能会为需要 CVE在上游开源之前解决的核心引擎组件提供修复程序。Amazon EMR 还为您提供技术支持和针对在推荐配置下在开源之上添加的功能的修复。我们不会将核心引擎组件的修复向后移植到较旧的补丁或次要版本。
- **额外组件**：Amazon EMR 不支持 Extras 组件。额外的组件是为方便起见而提供的开源项目，Amazon EMR 不提供解决它们遇到的问题的修复程序。任何支持请求或修复都可以通过支持这些组件的开源社区来解决。

## 标准 Support 生命周期

以下内容描述了 Standard Support 生命周期中的里程碑：

- **标准支持**：自EMR发布之日起 24 个月内，Amazon 版本有资格获得标准支持。您可以创建技术支持请求单，并期待在这些版本中遇到的问题得到更新。
- **终止支持**：标准支持结束后，Amazon EMR 版本将进入终止支持 (EoS) 阶段，为期 12 个月。EoS 版本没有资格获得技术支持，您将无法为在这些版本上运行的集群、容器或作业创建任何票证。EoS 版本不会收到任何修复、补丁或更新。EoS 版本将从控制台中删除，但将继续通过API和 AWS CLI

提供。您可以随时继续在 EoS 版本上运行工作负载。我们强烈建议您迁移到最新的 Amazon EMR 版本，这样您就可以继续收到安全补丁，保持获得技术支持的资格，并且可以在需要时创建支持工单。

- 生命周期结束：在终止 Support (EoS) 期限之后，版本被视为生命周期结束 (EoL)。尽管您可以继续运行 EoL 集群，但出于安全和运营方面的考虑，Amazon EMR 保留在 API 和 case-by-case 的基础 SDK 上删除 EoL 版本的权利。我们强烈建议您迁移到最新的 Amazon 版本 EMR，因为 SDK 在特殊情况下，EoL 版本可能会从 API 和中删除。

## Bridge Su

亚马逊EMR于2024年7月25日宣布了这项新的支持政策。根据该政策，2022年7月24日当天或之前EMR发布的 Amazon 版本现在被指定为终止支持。但是，为了让您有更多时间规划和迁移到新版本，亚马逊EMR将为在本公告发布前两年内发布的旧版本提供 Bridge Support（相当于标准支持）。

2024年7月25日之后，您可以在发行说明中查看当前状态和支持时间表。

下表显示了2024年7月25日EMR发布该政策时所有现有亚马逊版本的支持状态：

### 版本和支持期限

亚马逊EMR发行版	最初发行日期	标准支持结束日期	终止支持开始日期	生命周期终止开始日期
7.2.0	2024年7月25日	2026年7月24日	2026年7月25日	2027年7月25日
7.1.0	2024年4月23日	2026年4月22日	2026年4月23日	2027年4月23日
7.0.0	2023年12月19日	2025年12月18日	2025年12月19日	2026年12月19日
5.36.x 和 6.6.x — 6.15.x	2022年5月9日至2023年11月13日	桥接支持将持续到2026年1月24日	2026年1月25日	2026年1月25日
• 6.x 系列： 6.5.0 及更低版本	2013年1月1日至2022年3月30日	2025年7月24日之前的桥梁支持	2025年7月25日	2025年7月25日

亚马逊EMR发行版	最初发行日期	标准支持结束日期	终止支持开始日期	生命周期终止开始日期
<ul style="list-style-type: none"> <li>5.x 系列： 5.35.0 及更低版本</li> <li>4.x、3.x 和 2.x 系列</li> </ul>				

## 注意事项

标准支持适用于所有亚马逊EMR部署模式（EMR on EC2、Amazon EMR on EKS 和 EMR Serverless），适用于所有提供亚马逊EMR服务的区域，无需支付额外费用。按照策略中所述，使用推荐配置运行的集群自动有资格获得支持，因此您无需采取任何其他操作即可激活支持。

- Standard Support 仅支持 Amazon EMR 集群所需的组件。当 EMR Core Engine 开源组件到达上游的 EoL，或者依赖项不再有安全更新可用时，Amazon 无法保证安全补丁和修复的可用性。虽然您可以选择安装 Extras，但亚马逊EMR不支持它们或其依赖项。例如，您可以在自定义中安装第三方应用程序AMI以加强集群的安全性，使用引导操作脚本安装其他组件或复制对象，或者将对象复制SSH到集群中并升级默认软件包版本。Amazon EMR 不支持这些组件。Standard Support 不包括客户提供的引导操作、程序包、库、您的自定义代码和 bring-your-own 自定义应用程序，您可以将这些应用程序配置为方便起见，您可以EMR将 Amazon 配置为安装这些应用程序。
- 无论您的现有集群运行哪个 Amazon EMR 版本，都不会受到影响。您可以在不中断的情况下继续运行现有集群。您也可以继续启动新集群，并在任何现有版本和新版本上运行作业。该政策生效时的所有现有版本和新版本均在 Standard Support 的支持范围内，从 Amazon EMR 发行版本的初始发布之日起 24 个月内享有 Standard Support 的保障。在该政策的首次发布期间，Amazon EMR 将为您提供桥梁支持。为了获得不间断的支持，我们建议您测试您的应用程序并立即升级到最新的 Amazon EMR 版本。
- Amazon EMR 不会更改现有版本或集群上的标准支持组件。但是，Amazon EMR 保留在特殊情况下尊重上游生命周期终 case-by-case 止、在新版本中删除此类组件或将现有版本更改为终止支持 (EoS) 或生命周期结束 (EoL) 状态的权利。我们会通过可用渠道将任何移除通知您。
- 新地区推出后，亚马逊EMR将仅支持标准支持下的亚马逊EMR发布版本，该版本将在新区域正式上市之日前六个月内发布。
- Amazon EMR 不会自动将您的现有集群更新到最新版本。但是，如果需要，您可以选择将新集群更新到最新的补丁版本。

- 您不受支持的 Amazon EMR 使用费用将计入您的 AWS 账单。即使您以不受支持的方式使用 Amazon EMR，与该使用量相关的费用仍包含在您的总 AWS 消费中，并将包含在您的支持费用计算中。

如需更多信息，请联系 Developer [AWS Support](#)。

## 亚马逊 EMR 7.x 发布版本

本部分包含每个 Amazon EMR 7.x 发行版本中可用的应用程序版本、发行说明、组件版本和配置分类。

启动集群时，您可以从 Amazon 的多个版本中进行选择EMR。这允许您测试和使用满足您解决方案兼容性需求的应用程序版本。您可以使用发行版标注指定版本号。版本标签的格式是 `emr-x.x.x`。例如，`emr-7.2.0`。

Amazon 的新EMR版本将在几天内在不同地区上市，从最初发布日期的第一个地区开始。在此期间，您所在区域可能无法提供最新发行版。

有关每个 Amazon EMR 7.x 版本中应用程序版本的完整表，请参阅[亚马逊 EMR 7.x 版本中的应用程序版本](#)。

### 主题

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [亚马逊EMR版本 7.2.0](#)
- [亚马逊EMR版本 7.1.0](#)
- [亚马逊EMR版本 7.0.0](#)

## 亚马逊 EMR 7.x 版本中的应用程序版本

要查看列出每个 Amazon EMR 7.x 版本中可用的应用程序版本的完整表格，请在浏览器中[打开 Amazon EMR 7.x 版本中的应用程序版本](#)。

## 亚马逊EMR版本 7.2.0

### 7.2.0 支持的生命周期

下表描述了 Amazon 7.2.0 EMR 版本支持的生命周期日期。

Support 阶段	Date
最初发行日期	2024 年 7 月 25 日
标准支持直到	2026年7月24日
终止支持	2026年7月25日
使用寿命结束	2027年7月25日

有关在推荐的软件和配置下不会影响EMR集群的关键和高度常见漏洞和漏洞 (CVEs) 的列表，请参阅 [7.2.0 已知的核心引擎常见漏洞和漏洞](#)。有关 7.2.0 版本中已修复的CVEs漏洞列表，请参阅 [7.2.0 已修复的常见漏洞和漏洞](#)。

## 7.2.0 常见漏洞和漏洞

本页列出了与 Amazon EMR 7.2.0 相关的所有常见漏洞和漏洞 (CVEs)。

### 7.2.0 核心引擎的已知常见漏洞和暴露情况

下表列出了所有CVEs不会影响在 Amazon EMR 7.2.0 上运行的EMR集群的内容。EMRAmazon 依赖上游开源来提供修复程序，并将在亚马逊EMR验证修复后的 90 天内作为核心引擎组件的一部分提供最新的稳定版本。

CVE身份证	严重性	CVE详情 URL
CVE-2024-1597	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2024-1597">https://nvd.nist.gov/vuln/detail/CVE-2024-1597</a>
CVE-2023-44981	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-44981">https://nvd.nist.gov/vuln/detail/CVE-2023-44981</a>
CVE-2023-31047	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-31047">https://nvd.nist.gov/vuln/detail/CVE-2023-31047</a>
CVE-2023-25668	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-25668">https://nvd.nist.gov/vuln/detail/CVE-2023-25668</a>

CVE身份证	严重性	CVE详情 URL
CVE-2023-25664	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-25664">https://nvd.nist.gov/vuln/detail/CVE-2023-25664</a>
CVE-2022-46337	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2022-46337">https://nvd.nist.gov/vuln/detail/CVE-2022-46337</a>
CVE-2022-42889	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2022-42889">https://nvd.nist.gov/vuln/detail/CVE-2022-42889</a>
CVE-2022-39135	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2022-39135">https://nvd.nist.gov/vuln/detail/CVE-2022-39135</a>
CVE-2022-37865	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2022-37865">https://nvd.nist.gov/vuln/detail/CVE-2022-37865</a>
CVE-2022-26612	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2022-26612">https://nvd.nist.gov/vuln/detail/CVE-2022-26612</a>
CVE-2022-25168	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2022-25168">https://nvd.nist.gov/vuln/detail/CVE-2022-25168</a>
CVE-2022-1471	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2022-1471">https://nvd.nist.gov/vuln/detail/CVE-2022-1471</a>
CVE-2021-37404	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2021-37404">https://nvd.nist.gov/vuln/detail/CVE-2021-37404</a>
CVE-2021-35958	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2021-35958">https://nvd.nist.gov/vuln/detail/CVE-2021-35958</a>
CVE-2021-32798	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2021-32798">https://nvd.nist.gov/vuln/detail/CVE-2021-32798</a>
CVE-2021-32797	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2021-32797">https://nvd.nist.gov/vuln/detail/CVE-2021-32797</a>
CVE-2020-9548	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-9548">https://nvd.nist.gov/vuln/detail/CVE-2020-9548</a>

CVE身份证	严重性	CVE详情 URL
CVE-2020-9547	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-9547">https://nvd.nist.gov/vuln/detail/CVE-2020-9547</a>
CVE-2020-9546	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-9546">https://nvd.nist.gov/vuln/detail/CVE-2020-9546</a>
CVE-2020-8840	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-8840">https://nvd.nist.gov/vuln/detail/CVE-2020-8840</a>
CVE-2019-20445	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2019-20445">https://nvd.nist.gov/vuln/detail/CVE-2019-20445</a>
CVE-2019-20444	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2019-20444">https://nvd.nist.gov/vuln/detail/CVE-2019-20444</a>
CVE-2019-20330	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2019-20330">https://nvd.nist.gov/vuln/detail/CVE-2019-20330</a>
CVE-2019-17531	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2019-17531">https://nvd.nist.gov/vuln/detail/CVE-2019-17531</a>
CVE-2019-17267	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2019-17267">https://nvd.nist.gov/vuln/detail/CVE-2019-17267</a>
CVE-2019-17195	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2019-17195">https://nvd.nist.gov/vuln/detail/CVE-2019-17195</a>
CVE-2019-16943	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2019-16943">https://nvd.nist.gov/vuln/detail/CVE-2019-16943</a>
CVE-2019-16942	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2019-16942">https://nvd.nist.gov/vuln/detail/CVE-2019-16942</a>
CVE-2019-16335	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2019-16335">https://nvd.nist.gov/vuln/detail/CVE-2019-16335</a>
CVE-2019-14893	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2019-14893">https://nvd.nist.gov/vuln/detail/CVE-2019-14893</a>

CVE身份证	严重性	CVE详情 URL
CVE-2019-14892	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2019-14892">https://nvd.nist.gov/vuln/detail/CVE-2019-14892</a>
CVE-2019-14540	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2019-14540">https://nvd.nist.gov/vuln/detail/CVE-2019-14540</a>
CVE-2019-14379	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2019-14379">https://nvd.nist.gov/vuln/detail/CVE-2019-14379</a>
CVE-2019-10202	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2019-10202">https://nvd.nist.gov/vuln/detail/CVE-2019-10202</a>
CVE-2018-7489	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2018-7489">https://nvd.nist.gov/vuln/detail/CVE-2018-7489</a>
CVE-2018-19362	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2018-19362">https://nvd.nist.gov/vuln/detail/CVE-2018-19362</a>
CVE-2018-19361	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2018-19361">https://nvd.nist.gov/vuln/detail/CVE-2018-19361</a>
CVE-2018-19360	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2018-19360">https://nvd.nist.gov/vuln/detail/CVE-2018-19360</a>
CVE-2018-14721	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2018-14721">https://nvd.nist.gov/vuln/detail/CVE-2018-14721</a>
CVE-2018-14720	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2018-14720">https://nvd.nist.gov/vuln/detail/CVE-2018-14720</a>
CVE-2018-14719	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2018-14719">https://nvd.nist.gov/vuln/detail/CVE-2018-14719</a>
CVE-2018-14718	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2018-14718">https://nvd.nist.gov/vuln/detail/CVE-2018-14718</a>
CVE-2018-11307	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2018-11307">https://nvd.nist.gov/vuln/detail/CVE-2018-11307</a>



CVE身份证	严重性	CVE详情 URL
CVE-2017-7658	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2017-7658">https://nvd.nist.gov/vuln/detail/CVE-2017-7658</a>
CVE-2017-7657	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2017-7657">https://nvd.nist.gov/vuln/detail/CVE-2017-7657</a>
CVE-2017-7525	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2017-7525">https://nvd.nist.gov/vuln/detail/CVE-2017-7525</a>
CVE-2017-17485	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2017-17485">https://nvd.nist.gov/vuln/detail/CVE-2017-17485</a>
CVE-2017-15095	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2017-15095">https://nvd.nist.gov/vuln/detail/CVE-2017-15095</a>
CVE-2017-12629	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2017-12629">https://nvd.nist.gov/vuln/detail/CVE-2017-12629</a>
CVE-2016-5018	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2016-5018">https://nvd.nist.gov/vuln/detail/CVE-2016-5018</a>
CVE-2015-1832	重大	<a href="https://nvd.nist.gov/vuln/detail/CVE-2015-1832">https://nvd.nist.gov/vuln/detail/CVE-2015-1832</a>
CVE-2024-29133	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2024-29133">https://nvd.nist.gov/vuln/detail/CVE-2024-29133</a>
CVE-2024-29131	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2024-29131">https://nvd.nist.gov/vuln/detail/CVE-2024-29131</a>
CVE-2024-24680	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2024-24680">https://nvd.nist.gov/vuln/detail/CVE-2024-24680</a>
CVE-2024-21634	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2024-21634">https://nvd.nist.gov/vuln/detail/CVE-2024-21634</a>
CVE-2023-6481	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-6481">https://nvd.nist.gov/vuln/detail/CVE-2023-6481</a>

CVE身份证	严重性	CVE详情 URL
CVE-2023-6378	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-6378">https://nvd.nist.gov/vuln/detail/CVE-2023-6378</a>
CVE-2023-52428	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-52428">https://nvd.nist.gov/vuln/detail/CVE-2023-52428</a>
CVE-2023-4759	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-4759">https://nvd.nist.gov/vuln/detail/CVE-2023-4759</a>
CVE-2023-46695	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-46695">https://nvd.nist.gov/vuln/detail/CVE-2023-46695</a>
CVE-2023-46120	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-46120">https://nvd.nist.gov/vuln/detail/CVE-2023-46120</a>
CVE-2023-43665	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-43665">https://nvd.nist.gov/vuln/detail/CVE-2023-43665</a>
CVE-2023-43642	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-43642">https://nvd.nist.gov/vuln/detail/CVE-2023-43642</a>
CVE-2023-41164	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-41164">https://nvd.nist.gov/vuln/detail/CVE-2023-41164</a>
CVE-2023-36478	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-36478">https://nvd.nist.gov/vuln/detail/CVE-2023-36478</a>
CVE-2023-36053	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-36053">https://nvd.nist.gov/vuln/detail/CVE-2023-36053</a>
CVE-2023-34610	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-34610">https://nvd.nist.gov/vuln/detail/CVE-2023-34610</a>
CVE-2023-34478	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-34478">https://nvd.nist.gov/vuln/detail/CVE-2023-34478</a>
CVE-2023-34455	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-34455">https://nvd.nist.gov/vuln/detail/CVE-2023-34455</a>

CVE身份证	严重性	CVE详情 URL
CVE-2023-34454	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-34454">https://nvd.nist.gov/vuln/detail/CVE-2023-34454</a>
CVE-2023-34453	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-34453">https://nvd.nist.gov/vuln/detail/CVE-2023-34453</a>
CVE-2023-30608	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-30608">https://nvd.nist.gov/vuln/detail/CVE-2023-30608</a>
CVE-2023-2976	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-2976">https://nvd.nist.gov/vuln/detail/CVE-2023-2976</a>
CVE-2023-27579	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-27579">https://nvd.nist.gov/vuln/detail/CVE-2023-27579</a>
CVE-2023-25801	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-25801">https://nvd.nist.gov/vuln/detail/CVE-2023-25801</a>
CVE-2023-25676	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-25676">https://nvd.nist.gov/vuln/detail/CVE-2023-25676</a>
CVE-2023-25675	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-25675">https://nvd.nist.gov/vuln/detail/CVE-2023-25675</a>
CVE-2023-25674	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-25674">https://nvd.nist.gov/vuln/detail/CVE-2023-25674</a>
CVE-2023-25673	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-25673">https://nvd.nist.gov/vuln/detail/CVE-2023-25673</a>
CVE-2023-25672	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-25672">https://nvd.nist.gov/vuln/detail/CVE-2023-25672</a>
CVE-2023-25671	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-25671">https://nvd.nist.gov/vuln/detail/CVE-2023-25671</a>
CVE-2023-25670	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-25670">https://nvd.nist.gov/vuln/detail/CVE-2023-25670</a>

CVE身份证	严重性	CVE详情 URL
CVE-2023-25669	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-25669">https://nvd.nist.gov/vuln/detail/CVE-2023-25669</a>
CVE-2023-25667	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-25667">https://nvd.nist.gov/vuln/detail/CVE-2023-25667</a>
CVE-2023-25666	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-25666">https://nvd.nist.gov/vuln/detail/CVE-2023-25666</a>
CVE-2023-25665	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-25665">https://nvd.nist.gov/vuln/detail/CVE-2023-25665</a>
CVE-2023-25663	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-25663">https://nvd.nist.gov/vuln/detail/CVE-2023-25663</a>
CVE-2023-25662	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-25662">https://nvd.nist.gov/vuln/detail/CVE-2023-25662</a>
CVE-2023-25660	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-25660">https://nvd.nist.gov/vuln/detail/CVE-2023-25660</a>
CVE-2023-25659	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-25659">https://nvd.nist.gov/vuln/detail/CVE-2023-25659</a>
CVE-2023-25658	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-25658">https://nvd.nist.gov/vuln/detail/CVE-2023-25658</a>
CVE-2023-25194	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-25194">https://nvd.nist.gov/vuln/detail/CVE-2023-25194</a>
CVE-2023-24816	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-24816">https://nvd.nist.gov/vuln/detail/CVE-2023-24816</a>
CVE-2023-24580	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-24580">https://nvd.nist.gov/vuln/detail/CVE-2023-24580</a>
CVE-2023-23969	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-23969">https://nvd.nist.gov/vuln/detail/CVE-2023-23969</a>

CVE身份证	严重性	CVE详情 URL
CVE-2023-20883	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-20883">https://nvd.nist.gov/vuln/detail/CVE-2023-20883</a>
CVE-2023-1436	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-1436">https://nvd.nist.gov/vuln/detail/CVE-2023-1436</a>
CVE-2023-1370	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2023-1370">https://nvd.nist.gov/vuln/detail/CVE-2023-1370</a>
CVE-2022-46751	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2022-46751">https://nvd.nist.gov/vuln/detail/CVE-2022-46751</a>
CVE-2022-45693	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2022-45693">https://nvd.nist.gov/vuln/detail/CVE-2022-45693</a>
CVE-2022-45685	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2022-45685">https://nvd.nist.gov/vuln/detail/CVE-2022-45685</a>
CVE-2022-44729	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2022-44729">https://nvd.nist.gov/vuln/detail/CVE-2022-44729</a>
CVE-2022-42969	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2022-42969">https://nvd.nist.gov/vuln/detail/CVE-2022-42969</a>
CVE-2022-42890	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2022-42890">https://nvd.nist.gov/vuln/detail/CVE-2022-42890</a>
CVE-2022-42004	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2022-42004">https://nvd.nist.gov/vuln/detail/CVE-2022-42004</a>
CVE-2022-42003	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2022-42003">https://nvd.nist.gov/vuln/detail/CVE-2022-42003</a>
CVE-2022-41704	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2022-41704">https://nvd.nist.gov/vuln/detail/CVE-2022-41704</a>
CVE-2022-40899	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2022-40899">https://nvd.nist.gov/vuln/detail/CVE-2022-40899</a>

CVE身份证	严重性	CVE详情 URL
CVE-2022-40664	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2022-40664">https://nvd.nist.gov/vuln/detail/CVE-2022-40664</a>
CVE-2022-40152	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2022-40152">https://nvd.nist.gov/vuln/detail/CVE-2022-40152</a>
CVE-2022-40150	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2022-40150">https://nvd.nist.gov/vuln/detail/CVE-2022-40150</a>
CVE-2022-40149	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2022-40149">https://nvd.nist.gov/vuln/detail/CVE-2022-40149</a>
CVE-2022-40146	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2022-40146">https://nvd.nist.gov/vuln/detail/CVE-2022-40146</a>
CVE-2022-37866	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2022-37866">https://nvd.nist.gov/vuln/detail/CVE-2022-37866</a>
CVE-2022-36364	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2022-36364">https://nvd.nist.gov/vuln/detail/CVE-2022-36364</a>
CVE-2022-3510	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2022-3510">https://nvd.nist.gov/vuln/detail/CVE-2022-3510</a>
CVE-2022-3509	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2022-3509">https://nvd.nist.gov/vuln/detail/CVE-2022-3509</a>
CVE-2022-34169	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2022-34169">https://nvd.nist.gov/vuln/detail/CVE-2022-34169</a>
CVE-2022-32532	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2022-32532">https://nvd.nist.gov/vuln/detail/CVE-2022-32532</a>
CVE-2022-3171	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2022-3171">https://nvd.nist.gov/vuln/detail/CVE-2022-3171</a>
CVE-2022-25647	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2022-25647">https://nvd.nist.gov/vuln/detail/CVE-2022-25647</a>

CVE身份证	严重性	CVE详情 URL
CVE-2022-24758	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2022-24758">https://nvd.nist.gov/vuln/detail/CVE-2022-24758</a>
CVE-2021-41303	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2021-41303">https://nvd.nist.gov/vuln/detail/CVE-2021-41303</a>
CVE-2021-37137	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2021-37137">https://nvd.nist.gov/vuln/detail/CVE-2021-37137</a>
CVE-2021-37136	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2021-37136">https://nvd.nist.gov/vuln/detail/CVE-2021-37136</a>
CVE-2021-34538	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2021-34538">https://nvd.nist.gov/vuln/detail/CVE-2021-34538</a>
CVE-2021-33813	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2021-33813">https://nvd.nist.gov/vuln/detail/CVE-2021-33813</a>
CVE-2021-33036	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2021-33036">https://nvd.nist.gov/vuln/detail/CVE-2021-33036</a>
CVE-2021-31684	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2021-31684">https://nvd.nist.gov/vuln/detail/CVE-2021-31684</a>
CVE-2021-28165	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2021-28165">https://nvd.nist.gov/vuln/detail/CVE-2021-28165</a>
CVE-2021-26919	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2021-26919">https://nvd.nist.gov/vuln/detail/CVE-2021-26919</a>
CVE-2021-25646	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2021-25646">https://nvd.nist.gov/vuln/detail/CVE-2021-25646</a>
CVE-2020-9492	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-9492">https://nvd.nist.gov/vuln/detail/CVE-2020-9492</a>
CVE-2020-9480	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-9480">https://nvd.nist.gov/vuln/detail/CVE-2020-9480</a>

CVE身份证	严重性	CVE详情 URL
CVE-2020-7692	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-7692">https://nvd.nist.gov/vuln/detail/CVE-2020-7692</a>
CVE-2020-36518	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-36518">https://nvd.nist.gov/vuln/detail/CVE-2020-36518</a>
CVE-2020-36189	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-36189">https://nvd.nist.gov/vuln/detail/CVE-2020-36189</a>
CVE-2020-36188	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-36188">https://nvd.nist.gov/vuln/detail/CVE-2020-36188</a>
CVE-2020-36187	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-36187">https://nvd.nist.gov/vuln/detail/CVE-2020-36187</a>
CVE-2020-36186	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-36186">https://nvd.nist.gov/vuln/detail/CVE-2020-36186</a>
CVE-2020-36185	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-36185">https://nvd.nist.gov/vuln/detail/CVE-2020-36185</a>
CVE-2020-36184	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-36184">https://nvd.nist.gov/vuln/detail/CVE-2020-36184</a>
CVE-2020-36183	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-36183">https://nvd.nist.gov/vuln/detail/CVE-2020-36183</a>
CVE-2020-36182	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-36182">https://nvd.nist.gov/vuln/detail/CVE-2020-36182</a>
CVE-2020-36181	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-36181">https://nvd.nist.gov/vuln/detail/CVE-2020-36181</a>
CVE-2020-36180	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-36180">https://nvd.nist.gov/vuln/detail/CVE-2020-36180</a>
CVE-2020-36179	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-36179">https://nvd.nist.gov/vuln/detail/CVE-2020-36179</a>



CVE身份证	严重性	CVE详情 URL
CVE-2020-35728	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-35728">https://nvd.nist.gov/vuln/detail/CVE -2020-35728</a>
CVE-2020-35491	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-35491">https://nvd.nist.gov/vuln/detail/CVE -2020-35491</a>
CVE-2020-35490	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-35490">https://nvd.nist.gov/vuln/detail/CVE -2020-35490</a>
CVE-2020-35214	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-35214">https://nvd.nist.gov/vuln/detail/CVE -2020-35214</a>
CVE-2020-35213	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-35213">https://nvd.nist.gov/vuln/detail/CVE -2020-35213</a>
CVE-2020-25649	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-25649">https://nvd.nist.gov/vuln/detail/CVE -2020-25649</a>
CVE-2020-24750	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-24750">https://nvd.nist.gov/vuln/detail/CVE -2020-24750</a>
CVE-2020-24616	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-24616">https://nvd.nist.gov/vuln/detail/CVE -2020-24616</a>
CVE-2020-17533	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-17533">https://nvd.nist.gov/vuln/detail/CVE -2020-17533</a>
CVE-2020-17523	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-17523">https://nvd.nist.gov/vuln/detail/CVE -2020-17523</a>
CVE-2020-14195	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-14195">https://nvd.nist.gov/vuln/detail/CVE -2020-14195</a>
CVE-2020-14062	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-14062">https://nvd.nist.gov/vuln/detail/CVE -2020-14062</a>
CVE-2020-14061	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-14061">https://nvd.nist.gov/vuln/detail/CVE -2020-14061</a>

CVE身份证	严重性	CVE详情 URL
CVE-2020-14060	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-14060">https://nvd.nist.gov/vuln/detail/CVE-2020-14060</a>
CVE-2020-13949	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-13949">https://nvd.nist.gov/vuln/detail/CVE-2020-13949</a>
CVE-2020-13936	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-13936">https://nvd.nist.gov/vuln/detail/CVE-2020-13936</a>
CVE-2020-11988	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-11988">https://nvd.nist.gov/vuln/detail/CVE-2020-11988</a>
CVE-2020-11620	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-11620">https://nvd.nist.gov/vuln/detail/CVE-2020-11620</a>
CVE-2020-11619	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-11619">https://nvd.nist.gov/vuln/detail/CVE-2020-11619</a>
CVE-2020-11113	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-11113">https://nvd.nist.gov/vuln/detail/CVE-2020-11113</a>
CVE-2020-11112	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-11112">https://nvd.nist.gov/vuln/detail/CVE-2020-11112</a>
CVE-2020-11111	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-11111">https://nvd.nist.gov/vuln/detail/CVE-2020-11111</a>
CVE-2020-10969	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-10969">https://nvd.nist.gov/vuln/detail/CVE-2020-10969</a>
CVE-2020-10968	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-10968">https://nvd.nist.gov/vuln/detail/CVE-2020-10968</a>
CVE-2020-10673	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-10673">https://nvd.nist.gov/vuln/detail/CVE-2020-10673</a>
CVE-2020-10672	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-10672">https://nvd.nist.gov/vuln/detail/CVE-2020-10672</a>

CVE身份证	严重性	CVE详情 URL
CVE-2020-10650	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2020-10650">https://nvd.nist.gov/vuln/detail/CVE-2020-10650</a>
CVE-2019-17566	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2019-17566">https://nvd.nist.gov/vuln/detail/CVE-2019-17566</a>
CVE-2019-16869	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2019-16869">https://nvd.nist.gov/vuln/detail/CVE-2019-16869</a>
CVE-2019-14439	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2019-14439">https://nvd.nist.gov/vuln/detail/CVE-2019-14439</a>
CVE-2019-12086	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2019-12086">https://nvd.nist.gov/vuln/detail/CVE-2019-12086</a>
CVE-2019-10172	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2019-10172">https://nvd.nist.gov/vuln/detail/CVE-2019-10172</a>
CVE-2019-10099	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2019-10099">https://nvd.nist.gov/vuln/detail/CVE-2019-10099</a>
CVE-2019-0205	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2019-0205">https://nvd.nist.gov/vuln/detail/CVE-2019-0205</a>
CVE-2019-0204	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2019-0204">https://nvd.nist.gov/vuln/detail/CVE-2019-0204</a>
CVE-2018-8768	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2018-8768">https://nvd.nist.gov/vuln/detail/CVE-2018-8768</a>
CVE-2018-8012	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2018-8012">https://nvd.nist.gov/vuln/detail/CVE-2018-8012</a>
CVE-2018-5968	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2018-5968">https://nvd.nist.gov/vuln/detail/CVE-2018-5968</a>
CVE-2018-3827	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2018-3827">https://nvd.nist.gov/vuln/detail/CVE-2018-3827</a>

CVE身份证	严重性	CVE详情 URL
CVE-2018-3258	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2018-3258">https://nvd.nist.gov/vuln/detail/CVE-2018-3258</a>
CVE-2018-17190	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2018-17190">https://nvd.nist.gov/vuln/detail/CVE-2018-17190</a>
CVE-2018-12023	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2018-12023">https://nvd.nist.gov/vuln/detail/CVE-2018-12023</a>
CVE-2018-12022	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2018-12022">https://nvd.nist.gov/vuln/detail/CVE-2018-12022</a>
CVE-2018-11804	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2018-11804">https://nvd.nist.gov/vuln/detail/CVE-2018-11804</a>
CVE-2018-11793	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2018-11793">https://nvd.nist.gov/vuln/detail/CVE-2018-11793</a>
CVE-2018-10936	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2018-10936">https://nvd.nist.gov/vuln/detail/CVE-2018-10936</a>
CVE-2018-10054	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2018-10054">https://nvd.nist.gov/vuln/detail/CVE-2018-10054</a>
CVE-2017-9790	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2017-9790">https://nvd.nist.gov/vuln/detail/CVE-2017-9790</a>
CVE-2017-9735	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2017-9735">https://nvd.nist.gov/vuln/detail/CVE-2017-9735</a>
CVE-2017-7687	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2017-7687">https://nvd.nist.gov/vuln/detail/CVE-2017-7687</a>
CVE-2017-7656	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2017-7656">https://nvd.nist.gov/vuln/detail/CVE-2017-7656</a>
CVE-2017-5637	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2017-5637">https://nvd.nist.gov/vuln/detail/CVE-2017-5637</a>

CVE身份证	严重性	CVE详情 URL
CVE-2017-18640	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2017-18640">https://nvd.nist.gov/vuln/detail/CVE -2017-18640</a>
CVE-2017-15288	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2017-15288">https://nvd.nist.gov/vuln/detail/CVE -2017-15288</a>
CVE-2017-12612	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2017-12612">https://nvd.nist.gov/vuln/detail/CVE -2017-12612</a>
CVE-2017-1000034	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2017-1000034">https://nvd.nist.gov/vuln/detail/CVE -2017-1000034</a>
CVE-2016-6796	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2016-6796">https://nvd.nist.gov/vuln/detail/CVE -2016-6796</a>
CVE-2016-1000352	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2016-1000352">https://nvd.nist.gov/vuln/detail/CVE -2016-1000352</a>
CVE-2016-1000344	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2016-1000344">https://nvd.nist.gov/vuln/detail/CVE -2016-1000344</a>
CVE-2016-1000343	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2016-1000343">https://nvd.nist.gov/vuln/detail/CVE -2016-1000343</a>
CVE-2016-1000342	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2016-1000342">https://nvd.nist.gov/vuln/detail/CVE -2016-1000342</a>
CVE-2016-1000340	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2016-1000340">https://nvd.nist.gov/vuln/detail/CVE -2016-1000340</a>
CVE-2016-1000338	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2016-1000338">https://nvd.nist.gov/vuln/detail/CVE -2016-1000338</a>
CVE-2015-7501	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2015-7501">https://nvd.nist.gov/vuln/detail/CVE -2015-7501</a>
CVE-2015-6420	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2015-6420">https://nvd.nist.gov/vuln/detail/CVE -2015-6420</a>

CVE身份证	严重性	CVE详情 URL
CVE-2015-5237	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2015-5237">https://nvd.nist.gov/vuln/detail/CVE-2015-5237</a>
CVE-2015-4852	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2015-4852">https://nvd.nist.gov/vuln/detail/CVE-2015-4852</a>
CVE-2015-2156	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2015-2156">https://nvd.nist.gov/vuln/detail/CVE-2015-2156</a>
CVE-2015-2080	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2015-2080">https://nvd.nist.gov/vuln/detail/CVE-2015-2080</a>
CVE-2014-0114	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2014-0114">https://nvd.nist.gov/vuln/detail/CVE-2014-0114</a>
CVE-2013-4002	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2013-4002">https://nvd.nist.gov/vuln/detail/CVE-2013-4002</a>
CVE-2012-0881	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2012-0881">https://nvd.nist.gov/vuln/detail/CVE-2012-0881</a>

## 7.2.0 修复了核心引擎的常见漏洞和漏洞

下表列出了亚马逊 EMR 7.2.0 CVEs 中修复的所有问题。

CVE身份证	严重性	CVE详情 URL
CVE-2019-12402	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2019-12402">https://nvd.nist.gov/vuln/detail/CVE-2019-12402</a>
CVE-2021-41561	HIGH	<a href="https://nvd.nist.gov/vuln/detail/CVE-2021-41561">https://nvd.nist.gov/vuln/detail/CVE-2021-41561</a>

## 7.2.0 版本组件

亚马逊 EMR 7.2.0 支持以下应用程序：

## 核心引擎：

AmazonCloudWatchAgent, 三角洲。Flink、Hadoop、、、Hive、Hudi  
HBase、Livy HCatalog、Iceberg、Phoenix、Spark、Tez、Trino、Zookeeper  
JupyterEnterpriseGateway、Zookeeper

## 额外内容：

Hue、、 JupyterHubMXNet、Oozie、Pig、Presto、Sqoop、Zeppelin TensorFlow

有关支持的软件包的更多详细信息，请参阅 [7.2.0 版本组件详细信息](#)。有关核心引擎和其他功能的更多信息，请参阅 [Amazon EMR Standard Support](#)。

## 7.2.0 版本组件详情

有关亚马逊 EMR 7.2.0 中额外套餐的更多信息，请参阅下表。

应用程序	应用程序类型	软件包	软件包类型	Package 版本
AmazonCloudWatchAgent	核心实例	hadoop-hdfs-journalnode	额外内容	3.3.6.amzn.4-1.amzn.2023
AmazonCloudWatchAgent	核心实例	护林员-kms	额外内容	2.0.0-1.amzn.2023
AmazonCloudWatchAgent	核心实例	s3-dist-cp	额外内容	2.32.0-1.amzn.2023
AmazonCloudWatchAgent	核心实例	动物园管理员	额外内容	3.9.1.amzn.1-1.amzn.2023
AmazonCloudWatchAgent	核心实例	zookeeper-server	额外内容	3.9.1.amzn.1-1.amzn.2023
Flink	核心实例	hadoop-client	额外内容	3.3.6.amzn.4-1.amzn.2023
Flink	核心实例	hadoop-hdfs	额外内容	3.3.6.amzn.4-1.amzn.2023

应用程序	应用程序类型	软件包	软件包类型	Package 版本
Flink	核心实例	hadoop-hdfs-datanode	额外内容	3.3.6.amz n.4-1.amzn.2023
Flink	核心实例	hadoop-hdfs-journalnode	额外内容	3.3.6.amz n.4-1.amzn.2023
Flink	核心实例	hadoop-hdfs-namenode	额外内容	3.3.6.amz n.4-1.amzn.2023
Flink	核心实例	hadoop-httfs	额外内容	3.3.6.amz n.4-1.amzn.2023
Flink	核心实例	hadoop-kms	额外内容	3.3.6.amz n.4-1.amzn.2023
Flink	核心实例	hadoop-Ma preduce	额外内容	3.3.6.amz n.4-1.amzn.2023
Flink	核心实例	hadoop-ma preduce-h istoryserver	额外内容	3.3.6.amz n.4-1.amzn.2023
Flink	核心实例	hadoop-yarn	额外内容	3.3.6.amz n.4-1.amzn.2023
Flink	核心实例	hadoop-yarn- nodemanager	额外内容	3.3.6.amz n.4-1.amzn.2023
Flink	核心实例	hadoop-yarn- proxyserver	额外内容	3.3.6.amz n.4-1.amzn.2023
Flink	核心实例	hadoop-ya rn-resour cemanager	额外内容	3.3.6.amz n.4-1.amzn.2023
Flink	核心实例	hadoop-yarn- timelineserver	额外内容	3.3.6.amz n.4-1.amzn.2023



应用程序	应用程序类型	软件包	软件包类型	Package 版本
Flink	核心实例	hudi	额外内容	0.14.1.amzn.1-1.amzn.2023
Flink	核心实例	护林员-kms	额外内容	2.0.0-1.amzn2023
Flink	核心实例	hadoop	额外内容	3.3.6.amzn.4-1.amzn.2023
HBase	核心实例	s3-dist-cp	额外内容	2.32.0-1.amzn2023
HCatalog	核心实例	hadoop-yarn-proxyserver	额外内容	3.3.6.amzn.4-1.amzn.2023
HCatalog	核心实例	mariadb-connector-java	额外内容	2.7.2-1
Hive	核心实例	hadoop-httfs	额外内容	3.3.6.amzn.4-1.amzn.2023
Hive	核心实例	hadoop-yarn-proxyserver	额外内容	3.3.6.amzn.4-1.amzn.2023
Hive	核心实例	hive-hbase	额外内容	3.1.3.amzn.11-1.amzn.2023
Hive	核心实例	mariadb-connector-java	额外内容	2.7.2-1
Hive	核心实例	s3-dist-cp	额外内容	2.32.0-1.amzn2023
Phoenix	核心实例	s3-dist-cp	额外内容	2.32.0-1.amzn2023

应用程序	应用程序类型	软件包	软件包类型	Package 版本
Spark	核心实例	spark-dat anucleus	额外内容	3.5.1.amz n.0-1.amz n.0-1.amzn2023
Tez	核心实例	hadoop-Ma preduce	额外内容	3.3.6.amz n.4-1.amzn.2023
Tez	核心实例	hadoop-ma preduce-h istoryserver	额外内容	3.3.6.amz n.4-1.amzn.2023
Tez	核心实例	hadoop-yarn- proxyserver	额外内容	3.3.6.amz n.4-1.amzn.2023
Trino	核心实例	bigtop-utils	额外内容	1.2.0-1.a mzn2023
Trino	核心实例	hadoop-client	额外内容	3.3.6.amz n.4-1.amzn.2023
Trino	核心实例	hadoop-hdfs	额外内容	3.3.6.amz n.4-1.amzn.2023
Trino	核心实例	hadoop-hdfs- datanode	额外内容	3.3.6.amz n.4-1.amzn.2023
Trino	核心实例	hadoop-hdfs- journalnode	额外内容	3.3.6.amz n.4-1.amzn.2023
Trino	核心实例	hadoop-hdfs- namenode	额外内容	3.3.6.amz n.4-1.amzn.2023
Trino	核心实例	hadoop-kms	额外内容	3.3.6.amz n.4-1.amzn.2023
Trino	核心实例	hadoop-Ma preduce	额外内容	3.3.6.amz n.4-1.amzn.2023

应用程序	应用程序类型	软件包	软件包类型	Package 版本
Trino	核心实例	hadoop-mapreduce-historyserver	额外内容	3.3.6.amzn.4-1.amzn.2023
Trino	核心实例	hadoop-yarn	额外内容	3.3.6.amzn.4-1.amzn.2023
Trino	核心实例	hadoop-yarn-nodemanager	额外内容	3.3.6.amzn.4-1.amzn.2023
Trino	核心实例	hadoop-yarn-proxyserver	额外内容	3.3.6.amzn.4-1.amzn.2023
Trino	核心实例	hadoop-yarn-resourcemanager	额外内容	3.3.6.amzn.4-1.amzn.2023
Trino	核心实例	hadoop-yarn-timelineserver	额外内容	3.3.6.amzn.4-1.amzn.2023
Trino	核心实例	hive	额外内容	3.1.3.amzn.11-1.amzn.2023
Trino	核心实例	hive-hcatalog-server	额外内容	3.1.3.amzn.11-1.amzn.2023
Trino	核心实例	mariadb-connector-java	额外内容	2.7.2-1
Trino	核心实例	护林员-kms	额外内容	2.0.0-1.amzn2023
Trino	核心实例	动物园管理员	额外内容	3.9.1.amzn.1-1.amzn.2023

应用程序	应用程序类型	软件包	软件包类型	Package 版本
Trino	核心实例	zookeeper-server	额外内容	3.9.1.amzn.1-1.amzn.2023
Trino	核心实例	bigtop-groovy	额外内容	2.5.4-1.amzn.2023
Trino	核心实例	bigtop-jsvc	额外内容	1.2.4-1.amzn.2023
Trino	核心实例	hadoop	额外内容	3.3.6.amzn.4-1.amzn.2023
Trino	核心实例	hadoop-lzo	额外内容	0.4.19-1.amzn.2023
Trino	核心实例	hive-h 目录	额外内容	3.1.3.amzn.11-1.amzn.2023
Trino	核心实例	hive-jdbc	额外内容	3.1.3.amzn.11-1.amzn.2023
Zookeeper	核心实例	emrfs	额外内容	2.63.0-1.amzn.2023
Zookeeper	核心实例	hadoop-client	额外内容	3.3.6.amzn.4-1.amzn.2023
Zookeeper	核心实例	hadoop-hdfs	额外内容	3.3.6.amzn.4-1.amzn.2023
Zookeeper	核心实例	hadoop-hdfs-datanode	额外内容	3.3.6.amzn.4-1.amzn.2023
Zookeeper	核心实例	hadoop-hdfs-journalnode	额外内容	3.3.6.amzn.4-1.amzn.2023

应用程序	应用程序类型	软件包	软件包类型	Package 版本
Zookeeper	核心实例	hadoop-hdfs-namenode	额外内容	3.3.6.amz n.4-1.amzn.2023
Zookeeper	核心实例	hadoop-httpfs	额外内容	3.3.6.amz n.4-1.amzn.2023
Zookeeper	核心实例	hadoop-kms	额外内容	3.3.6.amz n.4-1.amzn.2023
Zookeeper	核心实例	hadoop-Ma preduce	额外内容	3.3.6.amz n.4-1.amzn.2023
Zookeeper	核心实例	hadoop-ma preduce-h istoryserver	额外内容	3.3.6.amz n.4-1.amzn.2023
Zookeeper	核心实例	hadoop-yarn	额外内容	3.3.6.amz n.4-1.amzn.2023
Zookeeper	核心实例	hadoop-yarn- nodemanager	额外内容	3.3.6.amz n.4-1.amzn.2023
Zookeeper	核心实例	hadoop-yarn- proxyserver	额外内容	3.3.6.amz n.4-1.amzn.2023
Zookeeper	核心实例	hadoop-ya rn-resour cemanager	额外内容	3.3.6.amz n.4-1.amzn.2023
Zookeeper	核心实例	hadoop-yarn- timelineserver	额外内容	3.3.6.amz n.4-1.amzn.2023
Zookeeper	核心实例	护林员-kms	额外内容	2.0.0-1.a mzn2023

## 7.2.0 应用程序版本

此发行版支持以下应用程序

序：[AmazonCloudWatchAgent](#)、[Delta](#)、[Flink](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[Iceberg](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-7.2.0	emr-7.1.0	emr-7.0.0
AWS SDK适用于Java	2.23.18、1.12.705	2.23.18、1.12.656	2.20.160-amzn-0, 1.12.569
Python	3.9	3.9	3.9
Scala	2.12.17	2.12.17	2.12.17
AmazonCloudWatchAgent	1.300032.2	1.300032.2	1.300031.1
Delta	3.1.0	3.0.0	3.0.0
Flink	1.18.1	1.18.1	1.18.0
HBase	2.4.17	2.4.17	2.4.17
HCatalog	3.1.3	3.1.3	3.1.3
Hadoop	3.3.6	3.3.6	3.3.6

	emr-7.2.0	emr-7.1.0	emr-7.0.0
Hive	3.1.3	3.1.3	3.1.3
Hudi	0.14.1-amzn-1	0.14.1-amzn-0	0.14.0-amzn-1
Hue	4.11.0	4.11.0	4.11.0
Iceberg	1.5.0-amzn-0	1.4.3-amzn-0	1.4.2-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0
JupyterHub	1.5.0	1.5.0	1.5.0
Livy	0.8.0	0.8.0	0.7.1
MXNet	1.9.1	1.9.1	1.9.1
Oozie	5.2.1	5.2.1	5.2.1
Phoenix	5.1.3	5.1.3	5.1.3
Pig	0.17.0	0.17.0	0.17.0
Presto	0.285	0.284	0.283
Spark	3.5.1	3.5.0	3.5.0
Sqoop	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.11.0
Tez	0.10.2	0.10.2	0.10.2
Trino (PrestoSQL)	436	435	426
Zeppelin	0.10.1	0.10.1	0.10.1
ZooKeeper	3.9.1	3.9.1	3.5.10

## 7.2.0 版本说明

以下发行说明包含有关亚马逊 7.2.0 EMR 版本的信息。更改与 7.2.0 有关。

### 新功能

- 应用程序升级 – Amazon EMR 7.2.0 application upgrades include Iceberg 1.5.0-amzn-0 and Delta 3.1.0。
- 亚马逊EMR增加了支持，因此您可以将其他应用程序（例如 Flink 和 Hive）与 HBase Amazon S3 Express One Zone 存储类一起使用。
- 此版本增加了读取已恢复对象的功能，因此您可以使用S3A协议从 S3 位置读取 Glacier 对象。此功能适用于 Spark、Flink 和 Hive。
- 具有托管扩展功能的节点标签-根据市场类型或节点类型在实例上使用节点标签，以改进 Amazon 的自动扩展EMR。有关更多信息，请参阅[在 Amazon 中使用托管扩展EMR](#)。

### 已知问题

- EMRStudio 不支持 Python 3.11。

### 更改、增强功能和解决的问题

- 此版本修复了内部步骤清理操作期间可能出现的死锁问题。当步骤在EMR集群上完成时，此操作会管理这些步骤的生命周期。此问题会影响关键的 Amazon EMR 操作，例如分步操作和扩展。
- 此版本解决了具有某些预先存在的日志文件的自定义AMIs集群可能导致 Amazon EMR 日志管理守护程序失败的问题。
- Amazon EMR 7.2.0 将负责集群管理和监控活动的亚马逊EMR守护程序从 AWS SDK v1 升级到 v2。
- 当您启动带有最新补丁版本（亚马逊 EMR 5.36 或更高版本、6.6 或更高版本、7.0 或更高版本）的集群时，亚马逊将EMR使用最新的亚马逊 Linux 2023 或亚马逊 Linux 2 版本作为默认亚马逊 Linux 2 版本。EMR AMI有关更多信息，请参阅[使用亚马逊的默认 Amazon Linux AMI EMR](#)。



OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2023.5.2 240708.0	6.1.96-102.177.amzn2023	2024年7月8日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、加拿大西部 (卡尔加里)、AWS GovCloud (美国西部)、AWS

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			GovCloud (美国东部)、中国(北京)、中国(宁夏)

## 7.2.0 默认 Java 版本

对于支持 Corretto 17 (17) 的应用程序，亚马逊 7.2 及更高 EMR 版本默认附带亚马逊 Corretto 17 (在 OpenJDK n 基础上构建)，但 Apache L JDK ivy 除外。

下表显示了 Amazon EMR 7.2.0 中应用程序的默认 Java 版本。如果要更改集群的默认设置 JVM，请按照集群上运行 [配置应用程序来使用特定 Java 虚拟机](#) 的每个应用程序中的说明进行操作。一个集群只能使用一个 Java 运行时系统版本。Amazon EMR 不支持在同一集群的不同运行时版本上运行不同的节点或应用程序。

应用程序	Java/Amazon Corretto 版本 (默认为粗体)
Delta	17、11、8
Flink	17、11、8
HBase	11、8
HCatalog	17、11、8
Hadoop	17、11、8
Hive	17、11、8
Hudi	17、11、8
Iceberg	17、11、8

应用程序	Java/Amazon Corretto 版本 (默认为粗体)
Livy	17、11、8
Oozie	17、11、8
Phoenix	8
PrestoDB	8
Spark	17、11、8
火花 RAPIDS	17、11、8
Sqoop	8
Tez	17、11、8
Trino	17
Zeppelin	8
Pig	8
Zookeeper	8

## 7.2.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
adot-java-agent	1.31.0	从应用程序进程守护程序收集指标的 Java 代理。
delta	3.1.0	Delta lake 是一种适用于超大型分析数据集的开放表格式。
emr-amazon-cloudwatch-agent	1.300032.2-amzn-0	一种从 Amazon EC2 实例收集内部系统级指标和自定义应用程序指标的应用程序。
emr-ddb	5.4.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	3.11.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.15.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-notebook-env	1.16.0	适用于 EMR Notebooks ( 可提供 jupyter 企业网关 ) 的 Conda env
emr-s3-dist-cp	2.32.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	2.11.0	EMRS3Select 连接器
emr-wal-cli	1.2.2	用于 emrwal 列表/删除的 cli。
emrfs	2.63.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.18.1-amzn-1	Apache Flink 命令行客户端脚本和应用程序。

组件	版本	描述
flink-jobmanager-config	1.18.1-amzn-1	管理 Apache Flink EMR JobManager 节点上的资源。
hadoop-client	3.3.6-amzn-4	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	3.3.6-amzn-4	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	3.3.6-amzn-4	HDFS命令行客户端和库
hadoop-hdfs-namenode	3.3.6-amzn-4	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	3.3.6-amzn-4	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	3.3.6-amzn-4	HTTPHDFS操作端点。
hadoop-kms-server	3.3.6-amzn-4	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	3.3.6-amzn-4	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	3.3.6-amzn-4	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	3.3.6-amzn-4	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	3.3.6-amzn-4	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	2.4.17-amzn-6	为负责协调区域和执行管理命令的HBase集群提供服务。

组件	版本	描述
hbase-region-server	2.4.17-amzn-6	为一个或多个HBase地区提供服务的服务。
hbase-client	2.4.17-amzn-6	HBase命令行客户端。
hbase-rest-server	2.4.17-amzn-6	为其提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	2.4.17-amzn-6	为其提供 Thrift 端点的HBase 服务。
hbase-operator-tools	2.4.17-amzn-6	适用于 Apache HBase 集群的修复工具。
hcatalog-client	3.1.3-amzn-11	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	3.1.3-amzn-11	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	3.1.3-amzn-11	HTTP端点为其提供REST接口HCatalog。
hive-client	3.1.3-amzn-11	Hive 命令行客户端。
hive-hbase	3.1.3-amzn-11	Hive-hbase 客户端。
hive-metastore-server	3.1.3-amzn-11	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	3.1.3-amzn-11	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.14.1-amzn-1	增量处理框架，以支持低延迟和高效率的数据管道。

组件	版本	描述
hudi-presto	0.14.1-amzn-1	用于运行 Presto 以及 Hudi 的捆绑库。
hudi-trino	0.14.1-amzn-1	用于运行 Trino 以及 Hudi 的捆绑库。
hudi-spark	0.14.1-amzn-1	用于运行 Spark 以及 Hudi 的捆绑库。
hue-server	4.11.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
iceberg	1.5.0-amzn-0	Apache Iceberg 是一种适用于超大型分析数据集的开放表格式。
jupyterhub	1.5.0	Jupyter notebook 的多用户服务器
livy-server	0.8.0-正在孵化	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mxnet	1.9.1	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.68+	MariaDB 数据库服务器。
nvidia-cuda	11.8.0	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.1	Oozie 命令行客户端。
oozie-server	5.2.1	用于接受 Oozie 工作流请求的服务。

组件	版本	描述
opencv	4.7.0	开源计算机视觉库。
phoenix-library	5.1.3	服务器和客户端的 phoenix 库
phoenix-connectors	5.1.3	Apache Phoenix-Connectors for Spark-3
phoenix-query-server	6.0.0	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.285-amzn-0	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.285-amzn-0	用于执行查询的各个部分的服务。
presto-client	0.285-amzn-0	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
trino-coordinator	436-amzn-0	用于在 trino-worker 之中接受查询并管理查询的服务。
trino-worker	436-amzn-0	用于执行查询的各个部分的服务。
trino-client	436-amzn-0	Trino 命令行客户端，安装在 HA 集群的备用主服务器（未启动 Trino 服务器）上。
pig-client	0.17.0	Pig 命令行客户端。
r	4.3.2	用于统计计算的 R 项目
ranger-kms-server	2.0.0	Apache Ranger 密钥管理系统



组件	版本	描述
spark-client	3.5.1-amzn-0	Spark 命令行客户端。
spark-history-server	3.5.1-amzn-0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	3.5.1-amzn-0	的内存中执行引擎。YARN
spark-yarn-slave	3.5.1-amzn-0	YARN从属服务器需要的 Apache Spark 库。
spark-rapids	24.02.0-amzn-0	Nvidia Spark RAPIDS 插件可以加速 Apache Spark。GPUs
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	2.11.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.10.2-amzn-9	tez YARN 应用程序和库。
tez-on-worker	0.10.2-amzn-9	工作节点的 tez YARN 应用程序和库。
zeppelin-server	0.10.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.9.1-amzn-1	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.9.1-amzn-1	ZooKeeper 命令行客户端。

## 7.2.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

为处于运行状态的集群中的实例组指定配置时，将发生重新配置操作。Amazon EMR 仅对您修改的分类启动重新配置操作。有关更多信息，请参阅 [在正在运行的集群中重新配置实例组](#)。

### emr-7.2.0 分类

分类	描述	重新配置操作
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。	Restarts the Resource Manager service.
container-executor	更改 Hadoop YARN 的 container-executor.cfg 文件中的值。	Not available.
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。	Not available.
core-site	更改 Hadoop 的 core-site.xml 文件中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	更改 docker 相关设置。	Not available.

分类	描述	重新配置操作
emrfs-site	更改EMRFS设置。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	更改 flink-conf.yaml 设置。	Restarts Flink history server.
flink-log4j	更改 Flink log4j.properties 设置。	Restarts Flink history server.
flink-log4j-session	为 Kubernetes/Yarn 会话更改 Flink log4j-session.properties 设置。	Restarts Flink history server.
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。	Restarts Flink history server.

分类	描述	重新配置操作
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	更改 hadoop ssl 服务器配置	Not available.
hadoop-ssl-client	更改 hadoop ssl 客户端配置	Not available.
hbase	亚马逊EMR精心策划的 Apache 设置。HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.

分类	描述	重新配置操作
hbase-env	更改环境HBase中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	更改 hadoop-metrics2-hbase base.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。	Not available.
hbase-site	更改HBase的 hbase-site.xml 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	配置HDFS加密区域。	This classification should not be reconfigured.
hdfs-env	更改HDFS环境中的值。	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.
hdfs-site	更改 hdfs-site.xml HDFS 中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	更改环境HCatalog中的值。	Restarts Hive HCatalog Server.

分类	描述	重新配置操作
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。	Restarts Hive WebHCat server.
hive	亚马逊EMR精心策划的 Apache Hive 设置。	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。	Not available.
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。	Not available.
hive-env	更改 Hive 环境中的值。	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	更改 Hive 的 hive-exec-log4j2.properties 文件中的值。	Not available.
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log4j2.properties 文件中的值。	Not available.

分类	描述	重新配置操作
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。	Not available.
hive-site	更改 Hive 的 hive-site.xml 文件中的值	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值	Not available.
hue-ini	更改 Hue 的 ini 文件中的值	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	更改HTTPFS环境中的值。	Restarts Hadoop Httpfs service.
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。	Restarts Hadoop Httpfs service.
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。	Not available.
hadoop-kms-env	在 Hadoop KMS 环境中更改值。	Restarts Hadoop-KMS service.
hadoop-kms-java-home	更改 Hadoop 的 KMS java 主页	Not available.
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。	Not available.

分类	描述	重新配置操作
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	更改 Hudi 环境中的值。	Not available.
hudi-defaults	更改 Hudi 的 hudi-defaults.conf 文件中的值。	Not available.
iceberg-defaults	更改 Iceberg 的 iceberg-defaults.conf 文件中的值。	Not available.
delta-defaults	更改 Delta 的 delta-defaults.conf 文件中的值。	Not available.
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。	Not available.
jupyter-hub-conf	更改 JupyterHubs 的 jupyterhub_config.py 文件中的值。	Not available.
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。	Not available.
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。	Not available.
livy-conf	更改 Livy 的 livy.conf 文件中的值。	Restarts Livy Server.
livy-env	更改 Livy 环境中的值。	Restarts Livy Server.
livy-log4j2	更改 Livy log4j2.properties 设置。	Restarts Livy Server.
mapred-env	更改 MapReduce 应用程序环境中的值。	Restarts Hadoop MapReduce-HistoryServer.



分类	描述	重新配置操作
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	更改 Oozie 的环境中的值。	Restarts Oozie.
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。	Restarts Oozie.
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。	Restarts Oozie.
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。	Not available.
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。	Not available.
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。	Restarts Phoenix-QueryServer.
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。	Not available.
pig-env	更改 Pig 环境中的值。	Not available.
pig-properties	更改 Pig 的 pig.properties 文件中的值。	Restarts Oozie.
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。	Not available.
presto-log	更改 Presto 的 log.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-config	更改 Presto 的 config.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)

分类	描述	重新配置操作
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。	Not available.
presto-env	更改 Presto 的 presto-env.sh 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-node	更改 Presto 的 node.properties 文件中的值。	Not available.
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。	Not available.
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。	Not available.
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。	Not available.
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。	Not available.
presto-connector-lakeformation	更改 Presto 的 lakeformation.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。	Not available.
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。	Not available.
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。	Not available.
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。	Not available.
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。	Not available.
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。	Not available.
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。	Not available.
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。	Not available.
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。	Not available.
trino-log	更改 Trino 的 log.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-config	更改 Trino 的 config.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-password-authenticator	更改 Trino 的 password-authenticator.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-env	更改 Trino 的 trino-env.sh 文件中的值。	Restarts Trino-Server (for Trino)
trino-node	更改 Trino 的 node.properties 文件中的值。	Not available.

分类	描述	重新配置操作
trino-connector-blackhole	更改 Trino 的 blackhole.properties 文件中的值。	Not available.
trino-connector-cassandra	更改 Trino 的 cassandra.properties 文件中的值。	Not available.
trino-connector-delta	更改 Trino 的 delta.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-hive	更改 Trino 的 hive.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-exchange-manager	更改 Trino 的 exchange-manager.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-iceberg	更改 Trino 的 iceberg.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-hudi	更改 Trino 的 hudi.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-jmx	更改 Trino 的 jmx.properties 文件中的值。	Not available.
trino-connector-kafka	更改 Trino 的 kafka.properties 文件中的值。	Not available.
trino-connector-localfile	更改 Trino 的 localfile.properties 文件中的值。	Not available.
trino-connector-memory	更改 Trino 的 memory.properties 文件中的值。	Not available.
trino-connector-mongodb	更改 Trino 的 mongodb.properties 文件中的值。	Not available.

分类	描述	重新配置操作
trino-connector-mysql	更改 Trino 的 mysql.properties 文件中的值。	Not available.
trino-connector-postgresql	更改 Trino 的 postgresql.properties 文件中的值。	Not available.
trino-connector-raptor	更改 Trino 的 raptor.properties 文件中的值。	Not available.
trino-connector-redis	更改 Trino 的 redis.properties 文件中的值。	Not available.
trino-connector-redshift	更改 Trino 的 redshift.properties 文件中的值。	Not available.
trino-connector-tpch	更改 Trino 的 tpch.properties 文件中的值。	Not available.
trino-connector-tpcds	更改 Trino 的 tpcds.properties 文件中的值。	Not available.
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-site	更改 Ranger KMS 的 ranger-kms-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-env	在 Ranger KMS 环境中更改值。	Restarts Ranger KMS Server.
ranger-kms-logback	更改 Ranger KMS 的 kms-logback.xml 文件中的值。	Not available.
ranger-kms-db-ca	更改 S3 上我与 Ranger SQL SSL KMS 连接的 CA 文件值。	Not available.
spark	亚马逊EMR精心策划的 Apache Spark 设置。	This property modifies spark-defaults. See actions there.

分类	描述	重新配置操作
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-env	更改 Spark 环境中的值。	Restarts Spark history server and Spark thrift server.
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值	Not available.
spark-log4j2	更改 Spark 的 log4j2.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
sqoop-env	更改 Sqoop 的环境中的值。	Not available.
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。	Not available.
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。	Not available.
tez-site	更改 Tez 的 tez-site.xml 文件中的值。	Restart Oozie and HiveServer2.
yarn-env	更改YARN环境中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.

分类	描述	重新配置操作
yarn-site	更改YARN的 yarn-site.xml 文件中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	更改 Zeppelin 环境中的值。	Restarts Zeppelin.
zeppelin-site	更改 zeppelin-site.xml 中的配置设置。	Restarts Zeppelin.
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper 中的值。	Restarts Zookeeper server.
zookeeper-logback	更改 ZooKeeper的 logback.xml 文件中的值。	Restarts Zookeeper server.
emr-metrics	更改此节点的 emr 指标设置。	Restarts the CloudWatchAgent service.

## 7.2.0 更改日志

### 7.2.0 版本和发行说明的更改日志

Date	事件	描述
2024-07-25	文档发布	亚马逊 EMR 7.2.0 发行说明首次发布
2024-07-16	初始版本	亚马逊 EMR 7.2.0 首次部署到初始商业区域

## 亚马逊EMR版本 7.1.0

### 7.1.0 应用程序版本

此发行版支持以下应用程序

序：[AmazonCloudWatchAgent](#)、[Delta](#)、[Flink](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[Iceberg](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-7.2.0	emr-7.1.0	emr-7.0.0
AWS SDK适用于 Java	2.23.18、1.12.705	2.23.18、1.12.656	2.20.160-amzn-0, 1.12.569
Python	3.9	3.9	3.9
Scala	2.12.17	2.12.17	2.12.17
AmazonCloudWatchAgent	1.300032.2	1.300032.2	1.300031.1
Delta	3.1.0	3.0.0	3.0.0
Flink	1.18.1	1.18.1	1.18.0
HBase	2.4.17	2.4.17	2.4.17
HCatalog	3.1.3	3.1.3	3.1.3



	emr-7.2.0	emr-7.1.0	emr-7.0.0
Hadoop	3.3.6	3.3.6	3.3.6
Hive	3.1.3	3.1.3	3.1.3
Hudi	0.14.1-amzn-1	0.14.1-amzn-0	0.14.0-amzn-1
Hue	4.11.0	4.11.0	4.11.0
Iceberg	1.5.0-amzn-0	1.4.3-amzn-0	1.4.2-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0
JupyterHub	1.5.0	1.5.0	1.5.0
Livy	0.8.0	0.8.0	0.7.1
MXNet	1.9.1	1.9.1	1.9.1
Mahout	-	-	-
Oozie	5.2.1	5.2.1	5.2.1
Phoenix	5.1.3	5.1.3	5.1.3
Pig	0.17.0	0.17.0	0.17.0
Presto	0.285	0.284	0.283
Spark	3.5.1	3.5.0	3.5.0
Sqoop	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.11.0
Tez	0.10.2	0.10.2	0.10.2
Trino (PrestoSQL)	436	435	426
Zeppelin	0.10.1	0.10.1	0.10.1

	emr-7.2.0	emr-7.1.0	emr-7.0.0
ZooKeeper	3.9.1	3.9.1	3.5.10

## 7.1.0 版本说明

以下发行说明包含有关亚马逊 7.1.0 EMR 版本的信息。更改与 7.1.0 有关。

### 新功能

- 应用程序升级 – Amazon EMR 7.1.0 application upgrades include Livy 0.8.0, Trino 435, and ZooKeeper 3.9.1。
- [不健康的节点替换](#) — 在 Amazon EMR 7.1.0 及更高版本中，默认启用不健康的节点替换，因此 Amazon EMR 将优雅地替换您的不健康节点。为避免影响您在 Amazon 7.0.0 及更低EMR版本上的现有工作流程，如果您在集群中启用了终止保护，则会禁用不健康的节点替换。
- [CloudWatch 代理](#) — 使用 Amazon EMR 配置 CloudWatch 代理以使用其他系统指标、添加应用程序指标和更改指标目标API。

### 已知问题

- EMRStudio 不支持 Python 3.11。

### 更改、增强功能和解决的问题

- 虽然亚马逊 EMR 7.1.0 默认支持 Python 3.9，但亚马逊 7.1.0 中的 Livy 0.8.0 和 Spark 支持 Python 3.11 EMR。
- 此版本修复了 PySpark 与 Python 版本 3.11 一起使用时需要逐行运行的问题。
- 齐柏林飞艇升级 — 亚马逊 EMR 7.1.0 包括齐柏林飞艇升级到 Java v2 版。AWS SDK此次升级使齐柏林飞艇 S3 笔记本电脑能够接受定制加密材料提供商。AWS SDK适用于 Java v2 的将删除该 EncryptionMaterialsProvider 接口。升级到 Amazon EMR 7.1.0 时，如果要使用自定义加密，则必须实现[密钥环接口](#)。有关如何实现密钥环接口的示例，请参阅 [KmsKeyring.java](#)。
- 升级到 Amazon 7.1.0 EMR 版本时，请将本地磁盘加密的自定义密钥提供程序更改为使用AES/GCM/AESNoPadding算法生成密钥。如果您不更新算法，集群创建可能会失败并显示错误Local disk encryption failed on master instance (i-123456789) due to internal error.。有关创建自定义密钥提供程序的更多信息，请参阅[创建自定义密钥提供程序](#)。

- Amazon EMR 7.1.0 通过改进具有打开文件句柄的文件的日志截断逻辑，提高了节点在磁盘空间不足条件下的弹性。
- 此版本增强了编码和解码逻辑，在重启节点时，Amazon EMR 守护程序会读取和写入文件，从而最大限度地降低数据损坏和节点故障的风险。
- 当您启动具有最新补丁版本（亚马逊 EMR 5.36 或更高版本、6.6 或更高版本、7.0 或更高版本）的集群时，亚马逊将EMR使用最新的亚马逊 Linux 2023 或亚马逊 Linux 2 版本作为默认亚马逊 Linux 2 版本。EMR AMI有关更多信息，请参阅[使用亚马逊的默认 Amazon Linux AMI EMR](#)。

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2023.5.2 240708.0	6.1.96-102.177.amzn2023	2024 年 7 月 23 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			尔)、亚太地区(大阪)、亚太地区(悉尼)、亚太地区(雅加达)、非洲(开普敦)、南美洲(圣保罗)、中东(巴林)、加拿大(中部)、AWS GovCloud(美国西部)、AWS GovCloud(美国东部)、(美国东部)、中国(北京)、中国(宁夏)、亚洲太平洋(海得拉巴)、中东(UAE)、欧洲(西班牙)、欧洲(苏黎世)、亚太地区(墨尔本)、以色列(特拉维夫)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2023.3.240219.0	6.1.77-99.164.amzn2023	2024年5月8日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、加拿大西部 ( 卡尔加里 )、AWS GovCloud ( 美国西部 )、AWS GovCloud ( 美国东部 )、

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			中国 (北京)、中国 (宁夏)

### 7.1.0 默认 Java 版本

对于支持 Corretto 17 (17) 的应用程序，亚马逊7.1及更高EMR版本默认附带Amazon Corretto 17JDK (基于JDK开放版本)，但Apache Livy除外。

下表显示了 Amazon EMR 7.1.0 中应用程序的默认 Java 版本。如果要更改集群的默认设置JVM，请按照集群上运行[配置应用程序来使用特定 Java 虚拟机](#)的每个应用程序中的说明进行操作。一个集群只能使用一个 Java 运行时系统版本。Amazon EMR 不支持在同一集群的不同运行时版本上运行不同的节点或应用程序。

应用程序	Java/Amazon Corretto 版本 (默认为粗体)
Delta	17、11、8
Flink	17、11、8
HBase	11、8
HCatalog	17、11、8
Hadoop	17、11、8
Hive	17、11、8
Hudi	17、11、8
Iceberg	17、11、8
Livy	17、11、8

应用程序	Java/Amazon Corretto 版本 (默认为粗体)
Oozie	17、11、8
Phoenix	8
PrestoDB	8
Spark	17、11、8
火花 RAPIDS	17、11、8
Sqoop	8
Tez	17、11、8
Trino	17
Zeppelin	8
Pig	8
Zookeeper	8

### 7.1.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
adot-java-agent	1.31.0	从应用程序进程守护程序收集指标的 Java 代理。
delta	3.0.0	Delta lake 是一种适用于超大型分析数据集的开放表格式。
emr-amazon-cloudwatch-agent	1.300032.2-amzn-0	一种从 Amazon EC2 实例收集内部系统级指标和自定义应用程序指标的应用程序。
emr-ddb	5.3.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	3.10.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.14.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-notebook-env	1.7.0	适用于 EMR Notebooks ( 可提供 jupyter 企业网关 ) 的 Conda env
emr-s3-dist-cp	2.31.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	2.10.0	EMRS3Select 连接器
emr-wal-cli	1.2.2	用于 emrwal 列表/删除的 cli。
emrfs	2.62.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.18.1-amzn-0	Apache Flink 命令行客户端脚本和应用程序。



组件	版本	描述
flink-jobmanager-config	1.18.1-amzn-0	管理 Apache Flink EMR JobManager 节点上的资源。
hadoop-client	3.3.6-amzn-3	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	3.3.6-amzn-3	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	3.3.6-amzn-3	HDFS命令行客户端和库
hadoop-hdfs-namenode	3.3.6-amzn-3	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	3.3.6-amzn-3	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	3.3.6-amzn-3	HTTPHDFS操作端点。
hadoop-kms-server	3.3.6-amzn-3	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	3.3.6-amzn-3	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	3.3.6-amzn-3	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	3.3.6-amzn-3	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	3.3.6-amzn-3	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	2.4.17-amzn-5	为负责协调区域和执行管理命令的HBase集群提供服务。

组件	版本	描述
hbase-region-server	2.4.17-amzn-5	为一个或多个HBase地区提供服务的服务。
hbase-client	2.4.17-amzn-5	HBase命令行客户端。
hbase-rest-server	2.4.17-amzn-5	为其提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	2.4.17-amzn-5	为其提供 Thrift 端点的HBase 服务。
hbase-operator-tools	2.4.17-amzn-5	适用于 Apache HBase 集群的修复工具。
hcatalog-client	3.1.3-amzn-10	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	3.1.3-amzn-10	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	3.1.3-amzn-10	HTTP端点为其提供REST接口HCatalog。
hive-client	3.1.3-amzn-10	Hive 命令行客户端。
hive-hbase	3.1.3-amzn-10	Hive-hbase 客户端。
hive-metastore-server	3.1.3-amzn-10	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	3.1.3-amzn-10	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.14.1-amzn-0	增量处理框架，以支持低延迟和高效率的数据管道。

组件	版本	描述
hudi-presto	0.14.1-amzn-0	用于运行 Presto 以及 Hudi 的捆绑库。
hudi-trino	0.14.1-amzn-0	用于运行 Trino 以及 Hudi 的捆绑库。
hudi-spark	0.14.1-amzn-0	用于运行 Spark 以及 Hudi 的捆绑库。
hue-server	4.11.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
iceberg	1.4.3-amzn-0	Apache Iceberg 是一种适用于超大型分析数据集的开放表格式。
jupyterhub	1.5.0	Jupyter notebook 的多用户服务器
livy-server	0.8.0-正在孵化	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mxnet	1.9.1	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.68+	MariaDB 数据库服务器。
nvidia-cuda	11.8.0	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.1	Oozie 命令行客户端。
oozie-server	5.2.1	用于接受 Oozie 工作流请求的服务。

组件	版本	描述
opencv	4.7.0	开源计算机视觉库。
phoenix-library	5.1.3	服务器和客户端的 phoenix 库
phoenix-connectors	5.1.3	Apache Phoenix-Connectors for Spark-3
phoenix-query-server	5.1.3	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.284-amzn-0	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.284-amzn-0	用于执行查询的各个部分的服务。
presto-client	0.284-amzn-0	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
trino-coordinator	435-amzn-0	用于在 trino-worker 之中接受查询并管理查询的服务。
trino-worker	435-amzn-0	用于执行查询的各个部分的服务。
trino-client	435-amzn-0	Trino 命令行客户端，安装在 HA 集群的备用主服务器（未启动 Trino 服务器）上。
pig-client	0.17.0	Pig 命令行客户端。
r	4.3.2	用于统计计算的 R 项目
ranger-kms-server	2.0.0	Apache Ranger 密钥管理系统

组件	版本	描述
spark-client	3.5.0-amzn-1	Spark 命令行客户端。
spark-history-server	3.5.0-amzn-1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	3.5.0-amzn-1	的内存中执行引擎。YARN
spark-yarn-slave	3.5.0-amzn-1	YARN从属服务器需要的 Apache Spark 库。
spark-rapids	23.10.0-amzn-1	Nvidia Spark RAPIDS 插件可以加速 Apache Spark。GPUs
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	2.11.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.10.2-amzn-8	tez YARN 应用程序和库。
tez-on-worker	0.10.2-amzn-8	工作节点的 tez YARN 应用程序和库。
zeppelin-server	0.10.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.9.1-amzn-0	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.9.1-amzn-0	ZooKeeper 命令行客户端。

## 7.1.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

### emr-7.1.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-executor	更改 Hadoop YARN 的 container-executor.cfg 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
docker-conf	更改 docker 相关设置。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-session	为 Kubernetes/Yarn 会话更改 Flink log4j-session.properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置

分类	描述
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-env	更改HDFS环境中的值。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive	亚马逊EMR精心策划的 Apache Hive 设置。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。

分类	描述
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-java-home	更改 Hadoop 的 KMS java 主页
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
hudi-env	更改 Hudi 环境中的值。
hudi-defaults	更改 Hudi 的 hudi-defaults.conf 文件中的值。



分类	描述
iceberg-defaults	更改 Iceberg 的 iceberg-defaults.conf 文件中的值。
delta-defaults	更改 Delta 的 delta-defaults.conf 文件中的值。
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 JupyterHubs的 jupyterhub_config.py 文件中的值。
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j2	更改 Livy log4j2.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。

分类	描述
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-lakeformation	更改 Presto 的 lakeformation.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。

分类	描述
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。
trino-log	更改 Trino 的 log.properties 文件中的值。
trino-config	更改 Trino 的 config.properties 文件中的值。
trino-password-authenticator	更改 Trino 的 password-authenticator.properties 文件中的值。
trino-env	更改 Trino 的 trino-env.sh 文件中的值。
trino-node	更改 Trino 的 node.properties 文件中的值。
trino-connector-blackhole	更改 Trino 的 blackhole.properties 文件中的值。
trino-connector-cassandra	更改 Trino 的 cassandra.properties 文件中的值。
trino-connector-delta	更改 Trino 的 delta.properties 文件中的值。
trino-connector-hive	更改 Trino 的 hive.properties 文件中的值。

分类	描述
trino-exchange-manager	更改 Trino 的 exchange-manager.properties 文件中的值。
trino-connector-iceberg	更改 Trino 的 iceberg.properties 文件中的值。
trino-connector-hudi	更改 Trino 的 hudi.properties 文件中的值。
trino-connector-jmx	更改 Trino 的 jmx.properties 文件中的值。
trino-connector-kafka	更改 Trino 的 kafka.properties 文件中的值。
trino-connector-localfile	更改 Trino 的 localfile.properties 文件中的值。
trino-connector-memory	更改 Trino 的 memory.properties 文件中的值。
trino-connector-mongodb	更改 Trino 的 mongodb.properties 文件中的值。
trino-connector-mysql	更改 Trino 的 mysql.properties 文件中的值。
trino-connector-postgresql	更改 Trino 的 postgresql.properties 文件中的值。
trino-connector-raptor	更改 Trino 的 raptor.properties 文件中的值。
trino-connector-redis	更改 Trino 的 redis.properties 文件中的值。
trino-connector-redshift	更改 Trino 的 redshift.properties 文件中的值。
trino-connector-tpch	更改 Trino 的 tpch.properties 文件中的值。
trino-connector-tpcds	更改 Trino 的 tpcds.properties 文件中的值。
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。
ranger-kms-site	更改 Ranger KMS 的 ranger-kms-site KMS ger .xml 文件中的值。

分类	描述
ranger-kms-env	在游侠KMS环境中更改值。
ranger-kms-logback	更改 Ranger KMS 的 kms-logback.xml 文件中的值。
ranger-kms-db-ca	更改 S3 上我与 Ranger SQL SSL KMS 连接的 CA 文件值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j2	更改 Spark 的 log4j2.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zeppelin-site	更改 zeppelin-site.xml 中的配置设置。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-logback	更改 ZooKeeper的 logback.xml 文件中的值。

分类	描述
emr-metrics	更改此节点的 emr 指标设置。

## 7.1.0 更改日志

### 7.1.0 版本和发行说明的更改日志

Date	事件	描述
2024-05-10	文档发布	亚马逊 EMR 7.1.0 发行说明首次发布
2024-05-08	部署完成	Amazon EMR 7.1.0 已全面部署到所有 <a href="#">受支持的区域</a>
2024-04-23	初始版本	亚马逊 EMR 7.1.0 首次部署到初始商业区域

## 亚马逊EMR版本 7.0.0

### 7.0.0 应用程序版本

此发行版支持以下应用程序

序：[AmazonCloudWatchAgent](#)、[Delta](#)、[Flink](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[Iceberg](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

## 应用程序版本信息

	emr-7.2.0	emr-7.1.0	emr-7.0.0
AWS SDK适用于Java	2.23.18、1.12.705	2.23.18、1.12.656	2.20.160-amzn-0, 1.12.569
Python	3.9	3.9	3.9
Scala	2.12.17	2.12.17	2.12.17
AmazonCloudWatchAgent	1.300032.2	1.300032.2	1.300031.1
Delta	3.1.0	3.0.0	3.0.0
Flink	1.18.1	1.18.1	1.18.0
HBase	2.4.17	2.4.17	2.4.17
HCatalog	3.1.3	3.1.3	3.1.3
Hadoop	3.3.6	3.3.6	3.3.6
Hive	3.1.3	3.1.3	3.1.3
Hudi	0.14.1-amzn-1	0.14.1-amzn-0	0.14.0-amzn-1
Hue	4.11.0	4.11.0	4.11.0
Iceberg	1.5.0-amzn-0	1.4.3-amzn-0	1.4.2-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0
JupyterHub	1.5.0	1.5.0	1.5.0
Livy	0.8.0	0.8.0	0.7.1
MXNet	1.9.1	1.9.1	1.9.1
Mahout	-	-	-

	emr-7.2.0	emr-7.1.0	emr-7.0.0
Oozie	5.2.1	5.2.1	5.2.1
Phoenix	5.1.3	5.1.3	5.1.3
Pig	0.17.0	0.17.0	0.17.0
Presto	0.285	0.284	0.283
Spark	3.5.1	3.5.0	3.5.0
Sqoop	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.11.0
Tez	0.10.2	0.10.2	0.10.2
Trino (PrestoSQL)	436	435	426
Zeppelin	0.10.1	0.10.1	0.10.1
ZooKeeper	3.9.1	3.9.1	3.5.10

## 7.0.0 发布说明

以下发行说明包含 Amazon 7.0.0 EMR 版本的信息。更改与 6.15.0 有关。

### 新功能

- 应用程序升级 – Amazon EMR 7.0.0 application upgrades include Python 3.9, Spark 3.5, Flink 1.18, and Delta 3.0。此版本还增加了对 [亚马逊 CloudWatch 代理](#) 应用程序的支持，并取消了对 Ganglia 的支持。
- Amazon Corretto 17 — EMR 亚马逊发布的7.0及更高版本默认为支持Corretto 17 (17) 的应用程序附带亚马逊Corretto 17 (基于开放版本JDK)，但Apache Livy除外。JDK有关此版本中应用程序支持的JDK版本的更多信息，请参阅[7.0.0 默认 Java 版本](#)。
- 亚马逊 Linux 2023 — 在 7.0 版本中，亚马逊EMR集群现在默认在 AL2 023 上运行。有关这对默认AMI版本的影响的信息，请参阅《Amazon EMR 管理指南》中的[软件更新注意事项](#)。另请注意，AL2023 删除了 Python 2.7，因此现在任何需要 Python 的组件都应该使用 Python 3 编写。



- Outposts 上的 S3 使用 — **s3a** 亚马逊EMR现在支持带有文件系统的 Outposts 存储桶上的 Amazon S3。s3a有关 S3 on Outposts 的更多信息，请参阅《Amazon Simple Storage Service 用户指南》中的[什么是 S3 on Outposts ?](#)。

## 已知问题

- 请注意，如果您使用的是 Python 版本 3.10 或更高版本 PySpark，则一次只能运行一行。必须逐行运行每行。

## 更改、增强功能和解决的问题

- 为了维护集群中所有实例的状态，Amazon EMR 与 Apache、Apache 和 K YARN erber HDFS os 集成。在 7.0 中，我们增强了这些集成，以便可靠地移除因扩展或其他操作而终止的实例的状态。这对于启用了托管扩展功能的长使用寿命集群尤其重要，因为它们经常扩展，并在生命周期内累积成千上万个终止的实例。
- 此版本改进了 Kerberos 配置，使其仅包括对AES基于密码的支持。在 Amazon KDC 7.0.0 及更高 EMR版本上运行的EMR集群不再支持使用非AES基于密码的 Kerberos。AES基于密码的密码可为您的集群提供最强的安全性。
- 作为 AWS SDK 2.x 迁移的一部分，亚马逊 EMR 7.0 包括对 Spark Kinesis 连接器的更新以实现兼容性。此更新在 Apache Spark 社区版本中不可用。如果您使用低于 7.0 的亚马逊EMR版本中的 Spark Kinesis 连接器，则必须先将应用程序代码迁移到在 SDK 2.x 上运行，然后才能将工作负载迁移到亚马逊 7.0。EMR有关更多信息，请参阅[将适用于亚马逊 7.0 的 Spark Kinesis 连接器迁移到 SDK 2.x EMR](#)。
- 当您启动具有最新补丁版本（亚马逊 EMR 5.36 或更高版本、6.6 或更高版本、7.0 或更高版本）的集群时，亚马逊将EMR使用最新的亚马逊 Linux 2023 或亚马逊 Linux 2 版本作为默认亚马逊 Linux 2 版本。EMR AMI有关更多信息，请参阅[使用亚马逊的默认 Amazon Linux AMI EMR](#)。

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2023.5.2 240708.1	6.1.96-102.177.amzn2023	2024 年 7 月 23 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			俄州)、美国西部(加利福尼亚北部)、美国西部(俄勒冈)、欧洲(斯德哥尔摩)、欧洲(米兰)、欧洲(西班牙)、欧洲(法兰克福)、欧洲(苏黎世)、欧洲(伦敦)、欧洲(巴黎)、亚太地区(香港)、亚太地区(孟买)、亚太地区(海得拉巴)、亚太地区(东京)、亚太地区(东京)、亚太地区(首尔)、亚太地区(大阪)、亚太地区(新加坡)、亚太地区(悉尼)、亚太地区(雅加达)、亚太地区(墨尔本)、非洲(开普敦)、南美洲(圣保罗)、中东(巴林)、中东( )、中东(UAE)、加拿大(中部)、以色列(特拉维夫)、加拿大西部(卡尔加里)、AWS GovCloud(美国西部)、AWS GovCloud(美国东部)、中国(北京)、中国(宁夏)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2023.3.240304.0	6.1.79-99.164.amzn2023	2024 年 3 月 12 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、加拿大西部 ( 卡尔加里 )、AWS GovCloud ( 美国西部 )、AWS GovCloud ( 美国东部 )、

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2023.3.240219.0	6.1.77-99.164.amzn2023	2024 年 3 月 1 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、加拿大西部 (卡尔加里)、AWS GovCloud (美国西部)、AWS GovCloud (美国东部)、

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2023.3.240205.0	6.1.75-99.163.amzn2023	2024年2月19日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、加拿大西部 ( 卡尔加里 )、AWS GovCloud ( 美国西部 )、AWS GovCloud ( 美国东部 )、

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			中国 ( 北京 )、中国 ( 宁夏 )



OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2023.3.240122.0	6.1.72-96.166.amzn2023	2024年2月5日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、加拿大西部 (卡尔加里)、AWS GovCloud (美国西部)、AWS GovCloud (美国东部)、

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2023.3.240108.0	6.1.72-96.166.amzn2023	2024 年 1 月 24 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、加拿大西部 (卡尔加里)、AWS GovCloud (美国西部)、AWS GovCloud (美国东部)、

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2023.3.2 231211.4	6.1.66-91.160.amzn2023	2023 年 12 月 19 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			(美国东部)、中国(北京)、中国(宁夏)

## 7.0.0 默认 Java 版本

对于支持 Corretto 17 (17) 的应用程序，亚马逊7.0及更高EMR版本默认附带Amazon Corretto 17 (基于JDK开放版本)，但Apache Livy除JDK外。

下表显示了 Amazon EMR 7.0.0 中应用程序的默认 Java 版本。如果要更改集群的默认设置JVM，请按照集群上运行[配置应用程序来使用特定 Java 虚拟机](#)的每个应用程序中的说明进行操作。一个集群只能使用一个 Java 运行时系统版本。Amazon EMR 不支持在同一集群的不同运行时版本上运行不同的节点或应用程序。

应用程序	Java/Amazon Corretto 版本 (默认为粗体)
Delta	17、11、8
Flink	11、8
HBase	11、8
HCatalog	17、11、8
Hadoop	17、11、8
Hive	17、11、8
Hudi	17、11、8
Iceberg	17、11、8
Livy	17、11、8

应用程序	Java/Amazon Corretto 版本 (默认为粗体)
Oozie	17、11、8
Phoenix	8
PrestoDB	8
Spark	17、11、8
火花 RAPIDS	17、11、8
Sqoop	8
Tez	17、11、8
Trino	17
Zeppelin	8
Pig	8
Zookeeper	8

## 7.0.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
adot-java-agent	1.31.0	从应用程序进程守护程序收集指标的 Java 代理。
delta	3.0.0	Delta lake 是一种适用于超大型分析数据集的开放表格式。
delta-standalone-connectors	0.6.0	Delta Connectors 提供不同的运行时，将 Delta Lake 与 Flink、Hive 和 Presto 等引擎集成。
emr-amazon-cloudwatch-agent	1.300031.1-amzn-0	一种从 Amazon EC2 实例收集内部系统级指标和自定义应用程序指标的应用程序。
emr-ddb	5.2.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	3.9.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.13.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-notebook-env	1.7.0	适用于 EMR Notebooks ( 可提供 jupyter 企业网关 ) 的 Conda env
emr-s3-dist-cp	2.30.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	2.9.0	EMRS3Select 连接器
emr-wal-cli	1.2.1	用于 emrwal 列表/删除的 cli。



组件	版本	描述
emrfs	2.61.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.18.0-amzn-0	Apache Flink 命令行客户端脚本和应用程序。
flink-jobmanager-config	1.18.0-amzn-0	管理 Apache Flink EMR JobManager 节点上的资源。
hadoop-client	3.3.6-amzn-2	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	3.3.6-amzn-2	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	3.3.6-amzn-2	HDFS命令行客户端和库
hadoop-hdfs-namenode	3.3.6-amzn-2	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	3.3.6-amzn-2	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	3.3.6-amzn-2	HTTPHDFS操作端点。
hadoop-kms-server	3.3.6-amzn-2	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	3.3.6-amzn-2	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	3.3.6-amzn-2	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	3.3.6-amzn-2	YARN用于分配和管理群集资源和分布式应用程序的服务。

组件	版本	描述
hadoop-yarn-timeline-server	3.3.6-amzn-2	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	2.4.17-amzn-4	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	2.4.17-amzn-4	为一个或多个HBase地区提供服务的服务。
hbase-client	2.4.17-amzn-4	HBase命令行客户端。
hbase-rest-server	2.4.17-amzn-4	为其提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	2.4.17-amzn-4	为其提供 Thrift 端点的HBase服务。
hbase-operator-tools	2.4.17-amzn-4	适用于 Apache HBase 集群的修复工具。
hcatalog-client	3.1.3-amzn-9	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	3.1.3-amzn-9	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	3.1.3-amzn-9	HTTP端点为其提供REST接口HCatalog。
hive-client	3.1.3-amzn-9	Hive 命令行客户端。
hive-hbase	3.1.3-amzn-9	Hive-hbase 客户端。
hive-metastore-server	3.1.3-amzn-9	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。

组件	版本	描述
hive-server2	3.1.3-amzn-9	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.14.0-amzn-1	增量处理框架，以支持低延迟和高效率的数据管道。
hudi-presto	0.14.0-amzn-1	用于运行 Presto 以及 Hudi 的捆绑库。
hudi-trino	0.14.0-amzn-1	用于运行 Trino 以及 Hudi 的捆绑库。
hudi-spark	0.14.0-amzn-1	用于运行 Spark 以及 Hudi 的捆绑库。
hue-server	4.11.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
iceberg	1.4.2-amzn-0	Apache Iceberg 是一种适用于超大型分析数据集的开放表格式。
jupyterhub	1.5.0	Jupyter notebook 的多用户服务器
livy-server	0.7.1-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mxnet	1.9.1	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.68+	MariaDB 数据库服务器。

组件	版本	描述
nvidia-cuda	11.8.0	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.1	Oozie 命令行客户端。
oozie-server	5.2.1	用于接受 Oozie 工作流请求的服务。
opencv	4.7.0	开源计算机视觉库。
phoenix-library	5.1.3	服务器和客户端的 phoenix 库
phoenix-connectors	5.1.3	Apache Phoenix-Connectors for Spark-3
phoenix-query-server	5.1.3	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.283-amzn-1	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.283-amzn-1	用于执行查询的各个部分的服务。
presto-client	0.283-amzn-1	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
trino-coordinator	426-amzn-1	用于在 trino-worker 之中接受查询并管理查询的服务。
trino-worker	426-amzn-1	用于执行查询的各个部分的服务。

组件	版本	描述
trino-client	426-amzn-1	Trino 命令行客户端，安装在 HA 集群的备用主服务器（未启动 Trino 服务器）上。
pig-client	0.17.0	Pig 命令行客户端。
r	4.1.3	用于统计计算的 R 项目
ranger-kms-server	2.0.0	Apache Ranger 密钥管理系统
spark-client	3.5.0-amzn-0	Spark 命令行客户端。
spark-history-server	3.5.0-amzn-0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	3.5.0-amzn-0	的内存中执行引擎。YARN
spark-yarn-slave	3.5.0-amzn-0	YARN从属服务器需要的 Apache Spark 库。
spark-rapids	23.10.0-amzn-0	Nvidia Spark RAPIDS 插件可以加速 Apache Spark。GPUs
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	2.11.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.10.2-amzn-7	tez YARN 应用程序和库。
tez-on-worker	0.10.2-amzn-7	工作节点的 tez YARN 应用程序和库。
webserver	2.4.58	Apache HTTP 服务器。

组件	版本	描述
zeppelin-server	0.10.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.5.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.5.10	ZooKeeper 命令行客户端。

## 7.0.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

### emr-7.0.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-executor	更改 Hadoop YARN 的 container-executor.cfg 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
docker-conf	更改 docker 相关设置。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。

分类	描述
flink-log4j-session	为 Kubernetes/Yarn 会话更改 Flink log4j-session.properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-env	更改HDFS环境中的值。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。

分类	描述
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive	亚马逊EMR精心策划的 Apache Hive 设置。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。



分类	描述
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-java-home	更改 Hadoop 的 KMS java 主页
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
hudi-env	更改 Hudi 环境中的值。
hudi-defaults	更改 Hudi 的 hudi-defaults.conf 文件中的值。
iceberg-defaults	更改 Iceberg 的 iceberg-defaults.conf 文件中的值。
delta-defaults	更改 Delta 的 delta-defaults.conf 文件中的值。
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 JupyterHubs的 jupyterhub_config.py 文件中的值。
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j2	更改 Livy log4j2.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。

分类	描述
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。

分类	描述
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-lakeformation	更改 Presto 的 lakeformation.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。
trino-log	更改 Trino 的 log.properties 文件中的值。
trino-config	更改 Trino 的 config.properties 文件中的值。
trino-password-authenticator	更改 Trino 的 password-authenticator.properties 文件中的值。

分类	描述
trino-env	更改 Trino 的 trino-env.sh 文件中的值。
trino-node	更改 Trino 的 node.properties 文件中的值。
trino-connector-blackhole	更改 Trino 的 blackhole.properties 文件中的值。
trino-connector-cassandra	更改 Trino 的 cassandra.properties 文件中的值。
trino-connector-delta	更改 Trino 的 delta.properties 文件中的值。
trino-connector-hive	更改 Trino 的 hive.properties 文件中的值。
trino-exchange-manager	更改 Trino 的 exchange-manager.properties 文件中的值。
trino-connector-iceberg	更改 Trino 的 iceberg.properties 文件中的值。
trino-connector-hudi	更改 Trino 的 hudi.properties 文件中的值。
trino-connector-jmx	更改 Trino 的 jmx.properties 文件中的值。
trino-connector-kafka	更改 Trino 的 kafka.properties 文件中的值。
trino-connector-localfile	更改 Trino 的 localfile.properties 文件中的值。
trino-connector-memory	更改 Trino 的 memory.properties 文件中的值。
trino-connector-mongodb	更改 Trino 的 mongodb.properties 文件中的值。
trino-connector-mysql	更改 Trino 的 mysql.properties 文件中的值。
trino-connector-postgresql	更改 Trino 的 postgresql.properties 文件中的值。
trino-connector-raptor	更改 Trino 的 raptor.properties 文件中的值。

分类	描述
trino-connector-redis	更改 Trino 的 redis.properties 文件中的值。
trino-connector-redshift	更改 Trino 的 redshift.properties 文件中的值。
trino-connector-tpch	更改 Trino 的 tpch.properties 文件中的值。
trino-connector-tpcds	更改 Trino 的 tpcds.properties 文件中的值。
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。
ranger-kms-site	更改 Ran ranger-kms-site KMS ger .xml 文件中的值。
ranger-kms-env	在游侠KMS环境中更改值。
ranger-kms-logback	更改 Ranger KMS 的 kms-logback.xml 文件中的值。
ranger-kms-db-ca	更改 S3 上我与 Ranger SQL SSL KMS 连接的 CA 文件值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j2	更改 Spark 的 log4j2.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。

分类	描述
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zeppelin-site	更改 zeppelin-site.xml 中的配置设置。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 7.0.0 更改日志

### 7.0.0 发行版的更改日志和发布说明

Date	事件	描述
2023-12-29	部署完成	<a href="#">Amazon EMR 7.0.0 已完全部署到所有支持的区域</a>
2023-12-28	文档发布	亚马逊 EMR 7.0.0 发行说明首次发布
2023-12-19	初始版本	亚马逊 EMR 7.0.0 首次部署到初始商业区域

## 亚马逊 EMR 6.x 发布版本

本部分包含每个 Amazon EMR 6.x 发行版本中可用的应用程序版本、发行说明、组件版本和配置分类。

启动集群时，您可以从 Amazon 的多个版本中进行选择EMR。这允许您测试和使用满足您解决方案兼容性需求的应用程序版本。您可以使用发行版标注指定版本号。版本标签的格式是 `emr-x.x.x`。例如，`emr-7.2.0`。

Amazon 的新EMR版本将在几天内在不同地区上市，从最初发布日期的第一个地区开始。在此期间，您所在区域可能无法提供最新发行版。

有关每个 Amazon EMR 6.x 版本中应用程序版本的完整表，请参阅[Amazon EMR 6.x 发行版中的应用程序版本](#)。

## 主题

- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [亚马逊EMR版本 6.15.0](#)
- [亚马逊EMR版本 6.14.0](#)
- [亚马逊EMR版本 6.13.0](#)
- [亚马逊EMR版本 6.12.0](#)
- [亚马逊EMR版本 6.11.1](#)
- [亚马逊EMR版本 6.11.0](#)
- [亚马逊EMR版本 6.10.1](#)
- [亚马逊EMR版本 6.10.0](#)
- [亚马逊EMR版本 6.9.1](#)
- [亚马逊EMR版本 6.9.0](#)
- [亚马逊EMR版本 6.8.1](#)
- [亚马逊EMR版本 6.8.0](#)
- [亚马逊EMR版本 6.7.0](#)
- [亚马逊EMR版本 6.6.0](#)
- [亚马逊EMR版本 6.5.0](#)
- [亚马逊EMR版本 6.4.0](#)
- [亚马逊EMR版本 6.3.1](#)
- [亚马逊EMR版本 6.3.0](#)
- [亚马逊EMR版本 6.2.1](#)
- [亚马逊EMR版本 6.2.0](#)
- [亚马逊EMR版本 6.1.1](#)

- [亚马逊EMR版本 6.1.0](#)
- [亚马逊EMR版本 6.0.1](#)
- [亚马逊EMR版本 6.0.0](#)

## Amazon EMR 6.x 发行版中的应用程序版本

有关列出每个 Amazon EMR 6.x 发行版中可用的应用程序版本的综合表格，请在浏览器打开 [Amazon EMR 6.x 发行版中的应用程序版本](#)。

### 亚马逊EMR版本 6.15.0

#### 6.15.0 应用程序版本

此发行版支持以下应用程序

序：[Delta](#)、[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[Iceberg](#)、[JupyterEnterpriseC](#)和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-6.15.0	emr-6.14.0	emr-6.13.0	emr-6.12.0
适用于 Java 的 AWS SDK	2.20.160-amzn-0, 1.12.569	1.12.543	1.12.513	1.12.490
Python	2.7、3.7	2.7、3.7	2.7、3.7	2.7、3.7
Scala	2.12.17	2.12.15	2.12.15	2.12.15



	emr-6.15.0	emr-6.14.0	emr-6.13.0	emr-6.12.0
AmazonCloudWatchAgent	-	-	-	-
Delta	2.4.0	2.4.0	2.4.0	2.4.0
Flink	1.17.1	1.17.1	1.17.0	1.17.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.17	2.4.17	2.4.17	2.4.17
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.6	3.3.3	3.3.3	3.3.3
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.14.0-amzn-0	0.13.1-amzn-2	0.13.1-amzn-1	0.13.1-amzn-0
Hue	4.11.0	4.11.0	4.11.0	4.11.0
Iceberg	1.4.0-amzn-0	1.3.1-amzn-0	1.3.0-amzn-1	1.3.0-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0	2.6.0
JupyterHub	1.5.0	1.5.0	1.5.0	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.3	5.1.3	5.1.3	5.1.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0

	emr-6.15.0	emr-6.14.0	emr-6.13.0	emr-6.12.0
Presto	0.283	0.281	0.281	0.281
Spark	3.4.1	3.4.1	3.4.1	3.4.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.11.0	2.11.0
Tez	0.10.2	0.10.2	0.10.2	0.10.2
Trino (PrestoSQL)	426	422	414	414
Zeppelin	0.10.1	0.10.1	0.10.1	0.10.1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

## 6.15.0 发布说明

以下发行说明包含 Amazon 6.15. EMR 0 版本的信息。更改与 6.14.0 有关。有关发布时间表的更多信息，请参阅 [6.15.0 更改日志](#)。

### 新功能

- 应用程序升级 – Amazon EMR 6.15.0 application upgrades include Apache Hadoop 3.3.6, Apache Hudi 0.14.0-amzn-0, Iceberg 1.4.0-amzn-0, and Trino 426.
- 在@@ [EMR集群上运行的启动速度更快 EC2](#) — 现在，在集EC2群上启动 Amaz EMR on 的速度最多可提高 35%。通过这一改进，大多数客户可以在 5 分钟或更短时间内启动集群。
- [CodeWhisperer 适用于 EMR Studio](#) — 现在，你可以将亚马逊 CodeWhisperer 与 Amazon EMR Studio 配合使用，在编写代码时获得实时推荐 JupyterLab。CodeWhisperer 可以完成您的注释、完成单行代码、提出 line-by-line 建议并生成完整格式的函数。
- [使用 Flink 可以缩短任务重启时间](#) — 在 Amazon EMR 6.15.0 及更高版本中，Apache Flink 可以使用多种新机制来缩短任务恢复或扩展操作期间的任务重启时间。这样可以优化执行图的恢复和重启速度，从而提高作业稳定性。

- [开放表格的表级和细粒度访问控制](#) — 在 Amazon EMR 6.15.0 及更高版本中，当您在访问 Glue 数据目录中数据的集群上在 Amazon EMR 上 EC2 运行 Spark 作业时，您可以使用 AWS Lake Formation 对基于 Hudi、Iceberg 或 Delta Lake 的表应用表、行、列和单元格级别的权限。
- Hadoop 升级 — 亚马逊 EMR 6.15.0 包括 Apache Hadoop 升级到 3.3.6 版。Hadoop 3.3.6 是部署亚马逊 EMR 6.15 时的最新版本，由 Apache 于 2023 年 6 月发布。亚马逊之前发布的版本 EMR ( 6.9.0 至 6.14.x ) 使用了 Hadoop 3.3.3。

此次升级包括数百项改进和修复，其功能包括可重新配置的数据节点参数、在所有实时数据节点上启动批量重新配置操作的 DFSAdmin 选项，以及允许大量搜索的读者指定多个读取范围 API 的向量。Hadoop 3.3.6 还增加了对其预写日志 (WAL) 的支持 HDFS APIs 和语义，因此 HBase 可以在其他存储系统实现上运行。有关更多信息，请参阅 Apache Hadoop 文档中的 [3.3.4](#)、[3.3.5](#) 和 [3.3.6](#) 版本的变更日志。

- 支持 Java，版本 2 —— 如果应用程序支持 v2，亚马逊 EMR 6.15.0 应用程序可以用 AWS SDK 于 Java 版本 [1.12.569](#) 或 [2.20.160](#)。AWS SDK 适用于 Java 2.x 的版本是对 1.x 版本代码库的重大改写。它基于 Java 8+ 构建，并增加了几个请求次数较多的功能。其中包括对非阻塞 I/O 的支持，以及在运行时插入其他 HTTP 实现的能力。有关更多信息，包括从 Java v1 到 v2 SDK 的迁移指南，请参阅 [AWS SDK 适用于 Java 的版本 2](#) 指南。

## 更改、增强功能和解决的问题

- 为了提高 EMR 集群的高可用性，此版本允许连接到使用 IPv6 终端节点的本地主机上的 Amazon EMR 守护程序。
- 此版本启用 TLS 1.2，以便在您的高可用性 ZooKeeper 集群的所有主节点上与预配置进行通信。
- 此版本改进了对主节点上维护的 ZooKeeper 事务日志文件的管理，以最大限度地减少日志文件越界增长并中断集群操作的情况。
- 此版本使高可用 EMR 性集群的节点内通信更具弹性。此改进减小了引导操作失败或集群启动失败的几率。
- Amazon EMR 6.15.0 中的 Tez 引入了配置，您可以指定这些配置来异步打开 Tez 分组拆分中的输入拆分。当单个 Tez 分组拆分中有大量输入拆分时，这样可以提高读取查询的性能。有关更多信息，请参阅 [Tez 异步拆分打开](#)。
- 当您启动带有最新补丁版本 ( 亚马逊 EMR 5.36 或更高版本、6.6 或更高版本、7.0 或更高版本 ) 的集群时，亚马逊将 EMR 使用最新的亚马逊 Linux 2023 或亚马逊 Linux 2 版本作为默认亚马逊 Linux 2 版本。EMR AMI 有关更多信息，请参阅 [使用亚马逊的默认 Amazon Linux AMI EMR](#)。

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 709.1	4.14.348	2024年7月23日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、AWS GovCloud ( 美国

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			东部 )、( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )、亚洲太平洋 ( 海得拉巴 )、中东 ( UAE )、欧洲 ( 西班牙 )、欧洲 ( 苏黎世 )、亚太地区 ( 墨尔本 )、以色列 ( 特拉维夫 )、加拿大西部 ( 卡尔加里 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 223.0	4.14.336	2024 年 3 月 8 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、中国 ( 北

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 131.0	4.14.336	2024年2月14日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、中国 ( 北



OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024124.0	4.14.336	2024 年 2 月 7 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、中国 ( 北

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 109.0	4.14.334	2024 年 1 月 24 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、中国 ( 北

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 218.0	4.14.330	2024 年 1 月 2 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 206.0	4.14.330	2023 年 12 月 22 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023116.0	4.14.328	2023 年 12 月 11 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 101.0	4.14.327	2023 年 11 月 13 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			(美国东部)、中国(北京)、中国(宁夏)

## 6.15.0 默认 Java 版本

亚马逊EMR发布的6.12.0及更高版本默认支持所有带有Amazon Corretto 8的应用程序，但Trino除外。对于 Trino，从亚马逊 EMR 6.9.0 版本开始，亚马逊默认支持 Amazon Corretto 17。EMR亚马逊EMR还支持某些带有 Amazon Corretto 11 和 17 的应用程序。下表列出了这些应用程序。如果要更改集群的默认设置JVM，请按照集群上运行[配置应用程序来使用特定 Java 虚拟机](#)的每个应用程序中的说明进行操作。一个集群只能使用一个 Java 运行时系统版本。Amazon EMR 不支持在同一集群的不同运行时版本上运行不同的节点或应用程序。

虽然亚马逊在 Apache Spark、Apache Hadoop 和 Apache Hive 上同时EMR支持 Amazon Corretto 11 和 17，但当你使用这些版本的 Corretto 时，某些工作负载的性能可能会下降。我们建议您在更改默认值之前先测试工作负载。

下表显示了 Amazon EMR 6.15.0 中应用程序的默认 Java 版本：

应用程序	Java/Amazon Corretto 版本 (默认为粗体)
Delta	17、11、8
Flink	11、8
Ganglia	8
HBase	11、8
HCatalog	17、11、8
Hadoop	17、11、8

应用程序	Java/Amazon Corretto 版本 (默认为粗体)
Hive	17、11、8
Hudi	17、11、8
Iceberg	17、11、8
Livy	17、11、8
Oozie	17、11、8
Phoenix	8
PrestoDB	8
Spark	17、11、8
火花 RAPIDS	17、11、8
Sqoop	8
Tez	17、11、8
Trino	17
Zeppelin	8
Pig	8
Zookeeper	8

## 6.15.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 `CommunityVersion-amzn-EmrVersion` 的发行版标注。`EmrVersion` 从 0 开始。例如，如果 `myapp-component` 以 2.2

版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.4.2	亚马逊 SageMaker Spark SDK
delta	2.4.0	Delta lake 是一种适用于超大型分析数据集的开放表格式。
delta-standalone-connectors	0.6.0	Delta Connectors 提供不同的运行时，将 Delta Lake 与 Flink、Hive 和 Presto 等引擎集成。
emr-ddb	5.2.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	3.8.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.12.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-notebook-env	1.7.0	适用于 EMR Notebooks ( 可提供 jupyter 企业网关 ) 的 Conda env
emr-s3-dist-cp	2.29.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	2.8.0	EMRS3Select 连接器
emr-wal-cli	1.2.0	用于 emrwal 列表/删除的 cli。
emrfs	2.60.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。

组件	版本	描述
flink-client	1.17.1-amzn-1	Apache Flink 命令行客户端脚本和应用程序。
flink-jobmanager-config	1.17.1-amzn-1	管理 Apache Flink EMR JobManager 节点上的资源。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	3.3.6-amzn-1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	3.3.6-amzn-1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	3.3.6-amzn-1	HDFS命令行客户端和库
hadoop-hdfs-namenode	3.3.6-amzn-1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	3.3.6-amzn-1	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	3.3.6-amzn-1	HTTPHDFS操作端点。
hadoop-kms-server	3.3.6-amzn-1	基于 Hadoop 的加密密钥管理服务。 KeyProvider API



组件	版本	描述
hadoop-mapred	3.3.6-amzn-1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	3.3.6-amzn-1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	3.3.6-amzn-1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	3.3.6-amzn-1	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	2.4.17-amzn-3	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	2.4.17-amzn-3	为一个或多个HBase地区提供服务的服务。
hbase-client	2.4.17-amzn-3	HBase命令行客户端。
hbase-rest-server	2.4.17-amzn-3	为其提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	2.4.17-amzn-3	为其提供 Thrift 端点的HBase 服务。
hbase-operator-tools	2.4.17-amzn-3	适用于 Apache HBase 集群的修复工具。
hcatalog-client	3.1.3-amzn-8	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	3.1.3-amzn-8	提供服务HCatalog，为分布式应用程序提供表和存储管理层。

组件	版本	描述
hcatalog-webhcat-server	3.1.3-amzn-8	HTTP端点为其提供REST接口HCatalog。
hive-client	3.1.3-amzn-8	Hive 命令行客户端。
hive-hbase	3.1.3-amzn-8	Hive-hbase 客户端。
hive-metastore-server	3.1.3-amzn-8	用于访问 Hive 元数据仓库的服务，这是一个语义存储库，用于存储 Hadoop 操作SQL的元数据。
hive-server2	3.1.3-amzn-8	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.14.0-amzn-0	增量处理框架，以支持低延迟和高效率的数据管道。
hudi-presto	0.14.0-amzn-0	用于运行 Presto 以及 Hudi 的捆绑库。
hudi-trino	0.14.0-amzn-0	用于运行 Trino 以及 Hudi 的捆绑库。
hudi-spark	0.14.0-amzn-0	用于运行 Spark 以及 Hudi 的捆绑库。
hue-server	4.11.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
iceberg	1.4.0-amzn-0	Apache Iceberg 是一种适用于超大型分析数据集的开放表格式。
jupyterhub	1.5.0	Jupyter notebook 的多用户服务器

组件	版本	描述
livy-server	0.7.1-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mxnet	1.9.1	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.68+	MariaDB 数据库服务器。
nvidia-cuda	11.8.0	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.1	Oozie 命令行客户端。
oozie-server	5.2.1	用于接受 Oozie 工作流请求的服务。
opencv	4.7.0	开源计算机视觉库。
phoenix-library	5.1.3	服务器和客户端的 phoenix 库
phoenix-connectors	5.1.3	Apache Phoenix-Connectors for Spark-3
phoenix-query-server	5.1.3	轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限 API
presto-coordinator	0.283-amzn-0	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.283-amzn-0	用于执行查询的各个部分的服务。

组件	版本	描述
presto-client	0.283-amzn-0	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
trino-coordinator	426-amzn-0	用于在 trino-worker 之中接受查询并管理查询的服务。
trino-worker	426-amzn-0	用于执行查询的各个部分的服务。
trino-client	426-amzn-0	Trino 命令行客户端，安装在 HA 集群的备用主服务器（未启动 Trino 服务器）上。
pig-client	0.17.0	Pig 命令行客户端。
r	4.0.2	用于统计计算的 R 项目
ranger-kms-server	2.0.0	Apache Ranger 密钥管理系统
spark-client	3.4.1-amzn-2	Spark 命令行客户端。
spark-history-server	3.4.1-amzn-2	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	3.4.1-amzn-2	的内存中执行引擎。YARN
spark-yarn-slave	3.4.1-amzn-2	YARN从属服务器需要的 Apache Spark 库。
spark-rapids	23.08.1-amzn-0	Nvidia Spark RAPIDS 插件可以加速 Apache Spark。GPUs
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。

组件	版本	描述
tensorflow	2.11.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.10.2-amzn-6	tez YARN 应用程序和库。
tez-on-worker	0.10.2-amzn-6	工作节点的 tez YARN 应用程序和库。
webserver	2.4.41+	Apache HTTP 服务器。
zeppelin-server	0.10.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.5.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.5.10	ZooKeeper 命令行客户端。

## 6.15.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

为处于运行状态的集群中的实例组指定配置时，将发生重新配置操作。Amazon EMR 仅对您修改的分类启动重新配置操作。有关更多信息，请参阅 [在正在运行的集群中重新配置实例组](#)。

### emr-6.15.0 分类

分类	描述	重新配置操作
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。	Restarts the ResourceManager service.
container-executor	更改 Hadoop YARN 的 container-executor.cfg 文件中的值。	Not available.

分类	描述	重新配置操作
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。	Not available.
core-site	更改 Hadoop 的 core-site.xml 文件中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	更改 docker 相关设置。	Not available.

分类	描述	重新配置操作
emrfs-site	更改EMRFS设置。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	更改 flink-conf.yaml 设置。	Restarts Flink history server.
flink-log4j	更改 Flink log4j.properties 设置。	Restarts Flink history server.
flink-log4j-session	为 Kubernetes/Yarn 会话更改 Flink log4j-session.properties 设置。	Restarts Flink history server.
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。	Restarts Flink history server.

分类	描述	重新配置操作
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	更改 hadoop ssl 服务器配置	Not available.
hadoop-ssl-client	更改 hadoop ssl 客户端配置	Not available.
hbase	亚马逊EMR精心策划的 Apache 设置。HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.



分类	描述	重新配置操作
hbase-env	更改环境HBase中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	更改 hadoop-metrics2-hbase base.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。	Not available.
hbase-site	更改HBase的 hbase-site.xml 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	配置HDFS加密区域。	This classification should not be reconfigured.
hdfs-env	更改HDFS环境中的值。	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	更改环境HCatalog中的值。	Restarts Hive HCatalog Server.

分类	描述	重新配置操作
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。	Restarts Hive WebHCat server.
hive	亚马逊EMR精心策划的 Apache Hive 设置。	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。	Not available.
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。	Not available.
hive-env	更改 Hive 环境中的值。	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	更改 Hive 的 hive-exec-log4j2.properties 文件中的值。	Not available.
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log4j2.properties 文件中的值。	Not available.

分类	描述	重新配置操作
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。	Not available.
hive-site	更改 Hive 的 hive-site.xml 文件中的值	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值	Not available.
hue-ini	更改 Hue 的 ini 文件中的值	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	更改HTTPFS环境中的值。	Restarts Hadoop Httpfs service.
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。	Restarts Hadoop Httpfs service.
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。	Not available.
hadoop-kms-env	在 Hadoop KMS 环境中更改值。	Restarts Hadoop-KMS service.
hadoop-kms-java-home	更改 Hadoop 的 KMS java 主页	Not available.
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。	Not available.

分类	描述	重新配置操作
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	更改 Hudi 环境中的值。	Not available.
hudi-defaults	更改 Hudi 的 hudi-defaults.conf 文件中的值。	Not available.
iceberg-defaults	更改 Iceberg 的 iceberg-defaults.conf 文件中的值。	Not available.
delta-defaults	更改 Delta 的 delta-defaults.conf 文件中的值。	Not available.
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。	Not available.
jupyter-hub-conf	更改 JupyterHubs 的 jupyterhub_config.py 文件中的值。	Not available.
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。	Not available.
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。	Not available.
livy-conf	更改 Livy 的 livy.conf 文件中的值。	Restarts Livy Server.
livy-env	更改 Livy 环境中的值。	Restarts Livy Server.
livy-log4j2	更改 Livy log4j2.properties 设置。	Restarts Livy Server.
mapred-env	更改 MapReduce 应用程序环境中的值。	Restarts Hadoop MapReduce-HistoryServer.

分类	描述	重新配置操作
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	更改 Oozie 的环境中的值。	Restarts Oozie.
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。	Restarts Oozie.
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。	Restarts Oozie.
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。	Not available.
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。	Not available.
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。	Restarts Phoenix-QueryServer.
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。	Not available.
pig-env	更改 Pig 环境中的值。	Not available.
pig-properties	更改 Pig 的 pig.properties 文件中的值。	Restarts Oozie.
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。	Not available.
presto-log	更改 Presto 的 log.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-config	更改 Presto 的 config.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)

分类	描述	重新配置操作
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。	Not available.
presto-env	更改 Presto 的 presto-env.sh 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-node	更改 Presto 的 node.properties 文件中的值。	Not available.
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。	Not available.
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。	Not available.
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。	Not available.
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。	Not available.
presto-connector-lakeformation	更改 Presto 的 lakeformation.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。	Not available.
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。	Not available.
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。	Not available.
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。	Not available.
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。	Not available.
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。	Not available.
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。	Not available.
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。	Not available.
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。	Not available.
trino-log	更改 Trino 的 log.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-config	更改 Trino 的 config.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-password-authenticator	更改 Trino 的 password-authenticator.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-env	更改 Trino 的 trino-env.sh 文件中的值。	Restarts Trino-Server (for Trino)
trino-node	更改 Trino 的 node.properties 文件中的值。	Not available.

分类	描述	重新配置操作
trino-connector-blackhole	更改 Trino 的 blackhole.properties 文件中的值。	Not available.
trino-connector-cassandra	更改 Trino 的 cassandra.properties 文件中的值。	Not available.
trino-connector-delta	更改 Trino 的 delta.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-hive	更改 Trino 的 hive.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-exchange-manager	更改 Trino 的 exchange-manager.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-iceberg	更改 Trino 的 iceberg.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-hudi	更改 Trino 的 hudi.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-jmx	更改 Trino 的 jmx.properties 文件中的值。	Not available.
trino-connector-kafka	更改 Trino 的 kafka.properties 文件中的值。	Not available.
trino-connector-localfile	更改 Trino 的 localfile.properties 文件中的值。	Not available.
trino-connector-memory	更改 Trino 的 memory.properties 文件中的值。	Not available.
trino-connector-mongodb	更改 Trino 的 mongodb.properties 文件中的值。	Not available.



分类	描述	重新配置操作
trino-connector-mysql	更改 Trino 的 mysql.properties 文件中的值。	Not available.
trino-connector-postgresql	更改 Trino 的 postgresql.properties 文件中的值。	Not available.
trino-connector-raptor	更改 Trino 的 raptor.properties 文件中的值。	Not available.
trino-connector-redis	更改 Trino 的 redis.properties 文件中的值。	Not available.
trino-connector-redshift	更改 Trino 的 redshift.properties 文件中的值。	Not available.
trino-connector-tpch	更改 Trino 的 tpch.properties 文件中的值。	Not available.
trino-connector-tpcds	更改 Trino 的 tpcds.properties 文件中的值。	Not available.
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-site	更改 Ranger KMS 的 ranger-kms-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-env	在 Ranger KMS 环境中更改值。	Restarts Ranger KMS Server.
ranger-kms-logback	更改 Ranger KMS 的 kms-logback.xml 文件中的值。	Not available.
ranger-kms-db-ca	更改 S3 上我与 Ranger SQL SSL KMS 连接的 CA 文件值。	Not available.
spark	亚马逊EMR精心策划的 Apache Spark 设置。	This property modifies spark-defaults. See actions there.

分类	描述	重新配置操作
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-env	更改 Spark 环境中的值。	Restarts Spark history server and Spark thrift server.
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值	Not available.
spark-log4j2	更改 Spark 的 log4j2.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
sqoop-env	更改 Sqoop 的环境中的值。	Not available.
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。	Not available.
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。	Not available.
tez-site	更改 Tez 的 tez-site.xml 文件中的值。	Restart Oozie and HiveServer2.
yarn-env	更改YARN环境中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.

分类	描述	重新配置操作
yarn-site	更改 yarn-site.xml 文件YARN 中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	更改 Zeppelin 环境中的值。	Restarts Zeppelin.
zeppelin-site	更改 zeppelin-site.xml 中的配置设置。	Restarts Zeppelin.
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。	Restarts Zookeeper server.
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。	Restarts Zookeeper server.

## 6.15.0 更改日志

### 6.15.0 发行版的更改日志和发布说明

Date	事件	描述
2023-11-17	文档发布	亚马逊 EMR 6.15.0 发行说明首次发布
2023-11-17	部署完成	<a href="#">Amazon EMR 6.15.0 已完全部署到所有支持的区域</a>
2023-11-13	初始版本	亚马逊 EMR 6.15.0 首次部署到初始商业区域

## 亚马逊EMR版本 6.14.0

### 6.14.0 应用程序版本

此发行版支持以下应用程序

序：[Delta](#)、[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[Iceberg](#)、[JupyterEnterpriseC](#)和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-6.14.0	emr-6.13.0	emr-6.12.0	emr-6.11.1
适用于 Java 的 AWS SDK	1.12.543	1.12.513	1.12.490	1.12.446
Python	2.7、3.7	2.7、3.7	2.7、3.7	2.7、3.7
Scala	2.12.15	2.12.15	2.12.15	2.12.15
AmazonCloudWatchAgent	-	-	-	-
Delta	2.4.0	2.4.0	2.4.0	2.2.0
Flink	1.17.1	1.17.0	1.17.0	1.16.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.17	2.4.17	2.4.17	2.4.15

	emr-6.14.0	emr-6.13.0	emr-6.12.0	emr-6.11.1
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.3.3	3.3.3	3.3.3
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.13.1-amzn-2	0.13.1-amzn-1	0.13.1-amzn-0	0.13.0-amzn-0
Hue	4.11.0	4.11.0	4.11.0	4.11.0
Iceberg	1.3.1-amzn-0	1.3.0-amzn-1	1.3.0-amzn-0	1.2.0-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0	2.6.0
JupyterHub	1.5.0	1.5.0	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.3	5.1.3	5.1.3	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.281	0.281	0.281	0.279
Spark	3.4.1	3.4.1	3.4.0	3.3.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.11.0	2.11.0
Tez	0.10.2	0.10.2	0.10.2	0.10.2

	emr-6.14.0	emr-6.13.0	emr-6.12.0	emr-6.11.1
Trino (PrestoSQL)	422	414	414	410
Zeppelin	0.10.1	0.10.1	0.10.1	0.10.1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

## 6.14.0 发布说明

以下发行说明包含有关亚马逊 6.14.0 EMR 版本的信息。更改与 6.13.0 有关。有关发布时间表的更多信息，请参阅 [6.14.0 更改日志](#)。

### 新功能

- Amazon EMR 6.14.0 supports Apache Spark 3.4.1, Apache Spark RAPIDS 23.06.0-amzn-2, Flink 1.17.1, Iceberg 1.3.1, and Trino 422.
- 对于您使用 [Amazon EMR 6.14.0 及更高版本创建的集群](#)，[ap-southeast-3 亚太地区（雅加达）地区现已推出 Amazon EMR 托管扩展](#)。

### 更改、增强功能和解决的问题

- 6.14.0 版本优化了日志管理，亚马逊EMR在亚马逊上运行。EC2因此，您可能会看到集群日志的存储成本略有降低。
- 6.14.0 版本改进了扩展工作流程，以考虑不同的核心实例，这些实例的 Amazon EBS 卷大小差异很大。此改进仅适用于核心节点；任务节点的缩减操作不受影响。
- 6.14.0 版本改进了 Amazon 与开源应用程序EMR交互的方式，例如。Apache Hadoop YARN ResourceManager and HDFS NameNode此改进降低了集群扩展导致操作延迟的风险，并减少了由于开源应用程序连接问题导致的启动故障。
- 6.14.0 发行版优化了集群启动时的应用程序安装。这缩短了某些 Amazon EMR 应用程序组合的集群启动时间。
- 6.14.0 版本修复了一个问题，即在自定义域中运行的集群遇到核心或任务节点重启时，集群缩小操作可能会停滞的问题。VPC

- 当您启动带有最新补丁版本 ( 亚马逊 EMR 5.36 或更高版本、6.6 或更高版本、7.0 或更高版本 ) 的集群时，亚马逊将EMR使用最新的亚马逊 Linux 2023 或亚马逊 Linux 2 版本作为默认亚马逊 Linux 2 版本。EMR AMI有关更多信息，请参阅[使用亚马逊的默认 Amazon Linux AMI EMR](#)。

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 709.1	4.14.348	2024年7月23日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、AWS GovCloud ( 美国



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			东部 )、( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )、亚洲太平洋 ( 海得拉巴 )、中东 ( UAE )、欧洲 ( 西班牙 )、欧洲 ( 苏黎世 )、亚太地区 ( 墨尔本 )、以色列 ( 特拉维夫 )、加拿大西部 ( 卡尔加里 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 223.0	4.14.336	2024 年 3 月 8 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、中国 ( 北

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 131.0	4.14.336	2024年2月14日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、中国 ( 北

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024124.0	4.14.336	2024 年 2 月 7 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、中国 (北

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 109.0	4.14.334	2024 年 1 月 24 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、中国 (北



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 218.0	4.14.330	2024 年 1 月 2 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 206.0	4.14.330	2023 年 12 月 22 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023116.0	4.14.328	2023 年 12 月 11 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 101.0	4.14.327	2023 年 11 月 17 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )
2.0.2023 906.0	4.14.322	2023 年 9 月 11 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )

## 6.14.0 默认 Java 版本

亚马逊EMR发布的6.12.0及更高版本默认支持所有带有Amazon Corretto 8的应用程序，但Trino除外。对于 Trino，从亚马逊 EMR 6.9.0 版本开始，亚马逊默认支持 Amazon Corretto 17。EMR亚马逊EMR还支持某些带有 Amazon Corretto 11 和 17 的应用程序。下表列出了这些应用程序。如果要更改集群的默认设置JVM，请按照集群上运行[配置应用程序来使用特定 Java 虚拟机](#)的每个应用程序中的说明进行操作。一个集群只能使用一个 Java 运行时系统版本。Amazon EMR 不支持在同一集群的不同运行时版本上运行不同的节点或应用程序。

虽然亚马逊在 Apache Spark、Apache Hadoop 和 Apache Hive 上同时EMR支持 Amazon Corretto 11 和 17，但当你使用这些版本的 Corretto 时，某些工作负载的性能可能会下降。我们建议您在更改默认值之前先测试工作负载。

下表显示了 Amazon EMR 6.14.0 中应用程序的默认 Java 版本：

应用程序	Java/Amazon Corretto 版本 ( 默认为粗体 )
Delta	17、11、8
Flink	11、8
Ganglia	8
HBase	11、8
HCatalog	17、11、8
Hadoop	17、11、8
Hive	17、11、8
Hudi	17、11、8
Iceberg	17、11、8
Livy	17、11、8
Oozie	17、11、8
Phoenix	8

应用程序	Java/Amazon Corretto 版本 (默认为粗体)
PrestoDB	8
Spark	17、11、8
火花 RAPIDS	17、11、8
Sqoop	8
Tez	17、11、8
Trino	17
Zeppelin	8
Pig	8
Zookeeper	8

## 6.14.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>aws-sagemaker-spark-sdk</code>	1.4.2	亚马逊 SageMaker Spark SDK
<code>delta</code>	2.4.0	Delta lake 是一种适用于超大型分析数据集的开放表格式。

组件	版本	描述
delta-standalone-connectors	0.6.0	Delta Connectors 提供不同的运行时，将 Delta Lake 与 Flink、Hive 和 Presto 等引擎集成。
emr-ddb	5.1.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	3.7.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.11.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-notebook-env	1.7.0	适用于 EMR Notebooks ( 可提供 jupyter 企业网关 ) 的 Conda env
emr-s3-dist-cp	2.28.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	2.7.0	EMRS3Select 连接器
emr-wal-cli	1.1.0	用于 emrwal 列表/删除的 cli。
emrfs	2.59.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.17.1-amzn-0	Apache Flink 命令行客户端脚本和应用程序。
flink-jobmanager-config	1.17.1-amzn-0	管理 Apache Flink EMR JobManager 节点上的资源。

组件	版本	描述
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	3.3.3-amzn-6	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	3.3.3-amzn-6	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	3.3.3-amzn-6	HDFS命令行客户端和库
hadoop-hdfs-namenode	3.3.3-amzn-6	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	3.3.3-amzn-6	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	3.3.3-amzn-6	HTTPHDFS操作端点。
hadoop-kms-server	3.3.3-amzn-6	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	3.3.3-amzn-6	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。

组件	版本	描述
hadoop-yarn-nodemanager	3.3.3-amzn-6	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	3.3.3-amzn-6	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	3.3.3-amzn-6	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	2.4.17-amzn-2	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	2.4.17-amzn-2	为一个或多个HBase地区提供服务的服务。
hbase-client	2.4.17-amzn-2	HBase命令行客户端。
hbase-rest-server	2.4.17-amzn-2	为其提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	2.4.17-amzn-2	为其提供 Thrift 端点的HBase服务。
hbase-operator-tools	2.4.17-amzn-2	适用于 Apache HBase 集群的修复工具。
hcatalog-client	3.1.3-amzn-7	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	3.1.3-amzn-7	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	3.1.3-amzn-7	HTTP端点为其提供REST接口HCatalog。
hive-client	3.1.3-amzn-7	Hive 命令行客户端。

组件	版本	描述
hive-hbase	3.1.3-amzn-7	Hive-hbase 客户端。
hive-metastore-server	3.1.3-amzn-7	用于访问 Hive 元数据仓库的服务，这是一个语义存储库，用于存储 Hadoop 操作SQL的元数据。
hive-server2	3.1.3-amzn-7	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.13.1-amzn-2	增量处理框架，以支持低延迟和高效率的数据管道。
hudi-presto	0.13.1-amzn-2	用于运行 Presto 以及 Hudi 的捆绑库。
hudi-trino	0.13.1-amzn-2	用于运行 Trino 以及 Hudi 的捆绑库。
hudi-spark	0.13.1-amzn-2	用于运行 Spark 以及 Hudi 的捆绑库。
hue-server	4.11.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
iceberg	1.3.1-amzn-0	Apache Iceberg 是一种适用于超大型分析数据集的开放表格式。
jupyterhub	1.5.0	Jupyter notebook 的多用户服务器
livy-server	0.7.1-incubating	REST用于与 Apache Spark 交互的接口

组件	版本	描述
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mxnet	1.9.1	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.68+	MariaDB 数据库服务器。
nvidia-cuda	11.8.0	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.1	Oozie 命令行客户端。
oozie-server	5.2.1	用于接受 Oozie 工作流请求的服务。
opencv	4.7.0	开源计算机视觉库。
phoenix-library	5.1.3	服务器和客户端的 phoenix 库
phoenix-connectors	5.1.3	Apache Phoenix-Connectors for Spark-3
phoenix-query-server	5.1.3	轻量级服务器，提供JDBC对Avatica 的访问权限以及协议缓冲区和JSON格式访问权限 API
presto-coordinator	0.281-amzn-2	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.281-amzn-2	用于执行查询的各个部分的服务。
presto-client	0.281-amzn-2	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。



组件	版本	描述
trino-coordinator	422-amzn-0	用于在 trino-worker 之中接受查询并管理查询的服务。
trino-worker	422-amzn-0	用于执行查询的各个部分的服务。
trino-client	422-amzn-0	Trino 命令行客户端，安装在 HA 集群的备用主服务器（未启动 Trino 服务器）上。
pig-client	0.17.0	Pig 命令行客户端。
r	4.0.2	用于统计计算的 R 项目
ranger-kms-server	2.0.0	Apache Ranger 密钥管理系统
spark-client	3.4.1-amzn-1	Spark 命令行客户端。
spark-history-server	3.4.1-amzn-1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	3.4.1-amzn-1	的内存中执行引擎。YARN
spark-yarn-slave	3.4.1-amzn-1	YARN从属服务器需要的 Apache Spark 库。
spark-rapids	23.06.0-amzn-2	Nvidia Spark RAPIDS 插件可以加速 Apache Spark。GPUs
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	2.11.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.10.2-amzn-5	tez YARN 应用程序和库。

组件	版本	描述
tez-on-worker	0.10.2-amzn-5	工作节点的 tez YARN 应用程序和库。
webserver	2.4.41+	Apache HTTP 服务器。
zeppelin-server	0.10.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.5.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.5.10	ZooKeeper 命令行客户端。

## 6.14.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

为处于运行状态的集群中的实例组指定配置时，将发生重新配置操作。Amazon EMR 仅对您修改的分类启动重新配置操作。有关更多信息，请参阅 [在正在运行的集群中重新配置实例组](#)。

### emr-6.14.0 分类

分类	描述	重新配置操作
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。	Restarts the ResourceManager service.
container-executor	更改 Hadoop YARN 的 container-executor.cfg 文件中的值。	Not available.
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。	Not available.

分类	描述	重新配置操作
core-site	更改 Hadoop 的 core-site.xml 文件中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	更改 docker 相关设置。	Not available.
emrfs-site	更改EMRFS设置。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	更改 flink-conf.yaml 设置。	Restarts Flink history server.

分类	描述	重新配置操作
flink-log4j	更改 Flink log4j.properties 设置。	Restarts Flink history server.
flink-log4j-session	为 Kubernetes/Yarn 会话更改 Flink log4j-session.properties 设置。	Restarts Flink history server.
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。	Restarts Flink history server.
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.

分类	描述	重新配置操作
hadoop-ssl-server	更改 hadoop ssl 服务器配置	Not available.
hadoop-ssl-client	更改 hadoop ssl 客户端配置	Not available.
hbase	亚马逊EMR精心策划的 Apache 设置。HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	更改环境HBase中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	更改 hadoop-metrics2-hbase base.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。	Not available.
hbase-site	更改HBase的 hbase-site.xml 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	配置HDFS加密区域。	This classification should not be reconfigured.
hdfs-env	更改HDFS环境中的值。	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

分类	描述	重新配置操作
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	更改环境HCatalog中的值。	Restarts Hive HCatalog Server.
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。	Restarts Hive WebHCat server.
hive	亚马逊EMR精心策划的 Apache Hive 设置。	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。	Not available.
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。	Not available.

分类	描述	重新配置操作
hive-env	更改 Hive 环境中的值。	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	更改 Hive 的 hive-exec-log4j2.properties 文件中的值。	Not available.
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log4j2.properties 文件中的值。	Not available.
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。	Not available.
hive-site	更改 Hive 的 hive-site.xml 文件中的值	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值	Not available.
hue-ini	更改 Hue 的 ini 文件中的值	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	更改HTTPFS环境中的值。	Restarts Hadoop Httpfs service.
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。	Restarts Hadoop Httpfs service.

分类	描述	重新配置操作
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。	Not available.
hadoop-kms-env	在 Hadoop KMS 环境中更改值。	Restarts Hadoop-KMS service.
hadoop-kms-java-home	更改 Hadoop 的 KMS java 主页	Not available.
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。	Not available.
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	更改 Hudi 环境中的值。	Not available.
hudi-defaults	更改 Hudi 的 hudi-defaults.conf 文件中的值。	Not available.
iceberg-defaults	更改 Iceberg 的 iceberg-defaults.conf 文件中的值。	Not available.
delta-defaults	更改 Delta 的 delta-defaults.conf 文件中的值。	Not available.
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。	Not available.
jupyter-hub-conf	更改 JupyterHubs 的 jupyterhub_config.py 文件中的值。	Not available.
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。	Not available.
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。	Not available.



分类	描述	重新配置操作
livy-conf	更改 Livy 的 livy.conf 文件中的值。	Restarts Livy Server.
livy-env	更改 Livy 环境中的值。	Restarts Livy Server.
livy-log4j2	更改 Livy log4j2.properties 设置。	Restarts Livy Server.
mapred-env	更改 MapReduce 应用程序环境中的值。	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	更改 Oozie 的环境中的值。	Restarts Oozie.
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。	Restarts Oozie.
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。	Restarts Oozie.
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。	Not available.
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。	Not available.
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。	Restarts Phoenix-QueryServer.
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。	Not available.
pig-env	更改 Pig 环境中的值。	Not available.

分类	描述	重新配置操作
pig-properties	更改 Pig 的 pig.properties 文件中的值。	Restarts Oozie.
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。	Not available.
presto-log	更改 Presto 的 log.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-config	更改 Presto 的 config.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。	Not available.
presto-env	更改 Presto 的 presto-env.sh 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-node	更改 Presto 的 node.properties 文件中的值。	Not available.
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。	Not available.
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。	Not available.
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。	Not available.
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-lakeformation	更改 Presto 的 lakeformation.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。	Not available.
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。	Not available.
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。	Not available.
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。	Not available.
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。	Not available.
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。	Not available.
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。	Not available.
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。	Not available.
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。	Not available.
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。	Not available.
trino-log	更改 Trino 的 log.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-config	更改 Trino 的 config.properties 文件中的值。	Restarts Trino-Server (for Trino)

分类	描述	重新配置操作
trino-password-authenticator	更改 Trino 的 password-authenticator.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-env	更改 Trino 的 trino-env.sh 文件中的值。	Restarts Trino-Server (for Trino)
trino-node	更改 Trino 的 node.properties 文件中的值。	Not available.
trino-connector-blackhole	更改 Trino 的 blackhole.properties 文件中的值。	Not available.
trino-connector-cassandra	更改 Trino 的 cassandra.properties 文件中的值。	Not available.
trino-connector-delta	更改 Trino 的 delta.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-hive	更改 Trino 的 hive.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-exchange-manager	更改 Trino 的 exchange-manager.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-iceberg	更改 Trino 的 iceberg.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-hudi	更改 Trino 的 hudi.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-jmx	更改 Trino 的 jmx.properties 文件中的值。	Not available.
trino-connector-kafka	更改 Trino 的 kafka.properties 文件中的值。	Not available.

分类	描述	重新配置操作
trino-connector-localfile	更改 Trino 的 localfile.properties 文件中的值。	Not available.
trino-connector-memory	更改 Trino 的 memory.properties 文件中的值。	Not available.
trino-connector-mongodb	更改 Trino 的 mongodb.properties 文件中的值。	Not available.
trino-connector-mysql	更改 Trino 的 mysql.properties 文件中的值。	Not available.
trino-connector-postgresql	更改 Trino 的 postgresql.properties 文件中的值。	Not available.
trino-connector-raptor	更改 Trino 的 raptor.properties 文件中的值。	Not available.
trino-connector-redis	更改 Trino 的 redis.properties 文件中的值。	Not available.
trino-connector-redshift	更改 Trino 的 redshift.properties 文件中的值。	Not available.
trino-connector-tpch	更改 Trino 的 tpch.properties 文件中的值。	Not available.
trino-connector-tpcds	更改 Trino 的 tpcds.properties 文件中的值。	Not available.
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-site	更改 Ranger KMS 的 ranger-kms-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-env	在游侠KMS环境中更改值。	Restarts Ranger KMS Server.

分类	描述	重新配置操作
ranger-kms-logback	更改 Ranger KMS 的 kms-logback.xml 文件中的值。	Not available.
ranger-kms-db-ca	更改 S3 上我与 Ranger SQL SSL KMS 连接的 CA 文件值。	Not available.
spark	亚马逊EMR精心策划的 Apache Spark 设置。	This property modifies spark-defaults. See actions there.
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-env	更改 Spark 环境中的值。	Restarts Spark history server and Spark thrift server.
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值	Not available.
spark-log4j2	更改 Spark 的 log4j2.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
sqoop-env	更改 Sqoop 的环境中的值。	Not available.
sqoop-orahome-site	更改 Sqoop OraOop 的 orahome-site.xml 文件中的值。	Not available.
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。	Not available.
tez-site	更改 Tez 的 tez-site.xml 文件中的值。	Restart Oozie and HiveServer2.

分类	描述	重新配置操作
yarn-env	更改YARN环境中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	更改 yarn-site.xml 文件YARN 中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	更改 Zeppelin 环境中的值。	Restarts Zeppelin.
zeppelin-site	更改 zeppelin-site.xml 中的配置设置。	Restarts Zeppelin.
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。	Restarts Zookeeper server.
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。	Restarts Zookeeper server.

## 6.14.0 更改日志

### 6.14.0 发行版的更改日志和发布说明

Date	事件	描述
2023-11-02	部署完成	<a href="#">Amazon EMR 6.14.0 已完全部署到所有支持的区域</a>

Date	事件	描述
2023-10-10	文档发布	亚马逊 EMR 6.14.0 发行说明首次发布
2023-10-04	初始版本	亚马逊 EMR 6.14.0 首次部署到初始商业区域

## 亚马逊EMR版本 6.13.0

### 6.13.0 应用程序版本

此发行版支持以下应用程序

序：[Delta](#)、[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[Iceberg](#)、[JupyterEnterpriseC](#)和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-6.13.0	emr-6.12.0	emr-6.11.1	emr-6.11.0
适用于 Java 的 AWS SDK	1.12.513	1.12.490	1.12.446	1.12.446
Python	2.7、3.7	2.7、3.7	2.7、3.7	2.7、3.7
Scala	2.12.15	2.12.15	2.12.15	2.12.15



	emr-6.13.0	emr-6.12.0	emr-6.11.1	emr-6.11.0
AmazonCloudWatchAgent	-	-	-	-
Delta	2.4.0	2.4.0	2.2.0	2.2.0
Flink	1.17.0	1.17.0	1.16.0	1.16.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.17	2.4.17	2.4.15	2.4.15
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.3.3	3.3.3	3.3.3
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.13.1-amzn-1	0.13.1-amzn-0	0.13.0-amzn-0	0.13.0-amzn-0
Hue	4.11.0	4.11.0	4.11.0	4.11.0
Iceberg	1.3.0-amzn-1	1.3.0-amzn-0	1.2.0-amzn-0	1.2.0-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0	2.6.0
JupyterHub	1.5.0	1.4.1	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.3	5.1.3	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0

	emr-6.13.0	emr-6.12.0	emr-6.11.1	emr-6.11.0
Presto	0.281	0.281	0.279	0.279
Spark	3.4.1	3.4.0	3.3.2	3.3.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.11.0	2.11.0
Tez	0.10.2	0.10.2	0.10.2	0.10.2
Trino (PrestoSQL)	414	414	410	410
Zeppelin	0.10.1	0.10.1	0.10.1	0.10.1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

## 6.13.0 发布说明

以下发行说明包含 Amazon 6.13.0 EMR 版本的信息。更改与 6.12.0 有关。有关发布时间表的更多信息，请参阅 [6.13.0 更改日志](#)。

### 新功能

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

### 更改、增强功能和解决的问题

- 6.13.0 版本改进了 Amazon EMR 日志管理守护程序，确保在发出集群终止命令时，所有日志均按固定节奏上传到 Amazon S3。这有助于更快地终止集群。
- 6.13.0 版本增强了 Amazon EMR 日志管理功能，可确保将所有日志文件一致且及时地上传到 Amazon S3。这尤其有利于长时间运行的 EMR 集群。
- 当您启动带有最新补丁版本（亚马逊 EMR 5.36 或更高版本、6.6 或更高版本、7.0 或更高版本）的集群时，亚马逊将 EMR 使用最新的亚马逊 Linux 2023 或亚马逊 Linux 2 版本作为默认亚马逊 Linux 2 版本。EMR AMI 有关更多信息，请参阅 [使用亚马逊的默认 Amazon Linux AMI EMR](#)。

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 709.1	4.14.348	2024年7月23日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、AWS GovCloud (美国西部)、AWS GovCloud (美国

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			东部 )、( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )、亚洲太平洋 ( 海得拉巴 )、中东 ( UAE )、欧洲 ( 西班牙 )、欧洲 ( 苏黎世 )、亚太地区 ( 墨尔本 )、以色列 ( 特拉维夫 )、加拿大西部 ( 卡尔加里 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 223.0	4.14.336	2024 年 3 月 8 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、中国 (北

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 131.0	4.14.336	2024年2月14日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、中国 ( 北

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)



OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024124.0	4.14.336	2024 年 2 月 7 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、中国 (北

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 109.0	4.14.334	2024 年 1 月 24 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、中国 ( 北

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 218.0	4.14.330	2024 年 1 月 2 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 206.0	4.14.330	2023 年 12 月 22 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023116.0	4.14.328	2023 年 12 月 11 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 101.0	4.14.327	2023 年 11 月 16 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023020.1	4.14.326	2023 年 11 月 7 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023012.1	4.14.326	2023 年 10 月 26 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )



OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 926.0	4.14.322	2023 年 10 月 19 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 906.0	4.14.322	2023 年 10 月 4 日	(美国东部)、中国(北京)、中国(宁夏)  美国东部(弗吉尼亚北部)、美国东部(俄亥俄州)、美国西部(加利福尼亚北部)、美国西部(俄勒冈)、欧洲(斯德哥尔摩)、欧洲(米兰)、欧洲(西班牙)、欧洲(法兰克福)、欧洲(苏黎世)、欧洲(伦敦)、欧洲(巴黎)、亚太地区(香港)、亚太地区(孟买)、亚太地区(海得拉巴)、亚太地区(东京)、亚太地区(东京)、亚太地区(首尔)、亚太地区(大阪)、亚太地区(新加坡)、亚太地区(悉尼)、亚太地区(雅加达)、亚太地区(墨尔本)、非洲(开普敦)、南美洲(圣保罗)、中东(巴林)、中东( )、中东(UAE)、加拿大(中部)、以色列(特拉维夫)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.202308080.0	4.14.320	2023 年 8 月 24 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)

## 6.13.0 默认 Java 版本

亚马逊EMR发布的6.12.0及更高版本默认支持所有带有Amazon Corretto 8的应用程序，但Trino除外。对于 Trino，从亚马逊 EMR 6.9.0 版本开始，亚马逊默认支持 Amazon Corretto 17。EMR亚马逊EMR还支持某些带有 Amazon Corretto 11 和 17 的应用程序。下表列出了这些应用程序。如果要更改集群的默认设置JVM，请按照集群上运行[配置应用程序来使用特定 Java 虚拟机](#)的每个应用程序中的说明进行操作。一个集群只能使用一个 Java 运行时系统版本。Amazon EMR 不支持在同一集群的不同运行时版本上运行不同的节点或应用程序。

虽然亚马逊在 Apache Spark、Apache Hadoop 和 Apache Hive 上同时EMR支持 Amazon Corretto 11 和 17，但当你使用这些版本的 Corretto 时，某些工作负载的性能可能会下降。我们建议您在更改默认值之前先测试工作负载。

下表显示了 Amazon EMR 6.13.0 中应用程序的默认 Java 版本：

应用程序	Java/Amazon Corretto 版本 ( 默认为粗体 )
Delta	17、11、8
Flink	11、8
Ganglia	8
HBase	11、8
HCatalog	17、11、8
Hadoop	17、11、8
Hive	17、11、8
Hudi	17、11、8
Iceberg	17、11、8
Livy	17、11、8
Oozie	17、11、8
Phoenix	8

应用程序	Java/Amazon Corretto 版本 (默认为粗体)
PrestoDB	8
Spark	17、11、8
火花 RAPIDS	17、11、8
Sqoop	8
Tez	17、11、8
Trino	17
Zeppelin	8
Pig	8
Zookeeper	8

### 6.13.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>aws-sagemaker-spark-sdk</code>	1.4.2	亚马逊 SageMaker Spark SDK
<code>delta</code>	2.4.0	Delta lake 是一种适用于超大型分析数据集的开放表格式。

组件	版本	描述
delta-standalone-connectors	0.6.0	Delta Connectors 提供不同的运行时，将 Delta Lake 与 Flink、Hive 和 Presto 等引擎集成。
emr-ddb	5.1.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	3.6.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.10.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-notebook-env	1.7.0	适用于 EMR Notebooks ( 可提供 jupyter 企业网关 ) 的 Conda env
emr-s3-dist-cp	2.27.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	2.6.0	EMRS3Select 连接器
emr-wal-cli	1.1.0	用于 emrwal 列表/删除的 cli。
emrfs	2.58.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.17.0	Apache Flink 命令行客户端脚本和应用程序。
flink-jobmanager-config	1.17.0	管理 Apache Flink EMR JobManager 节点上的资源。

组件	版本	描述
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	3.3.3-amzn-5	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	3.3.3-amzn-5	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	3.3.3-amzn-5	HDFS命令行客户端和库
hadoop-hdfs-namenode	3.3.3-amzn-5	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	3.3.3-amzn-5	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	3.3.3-amzn-5	HTTPHDFS操作端点。
hadoop-kms-server	3.3.3-amzn-5	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	3.3.3-amzn-5	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。

组件	版本	描述
hadoop-yarn-nodemanager	3.3.3-amzn-5	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	3.3.3-amzn-5	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	3.3.3-amzn-5	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	2.4.17-amzn-1	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	2.4.17-amzn-1	为一个或多个HBase地区提供服务的服务。
hbase-client	2.4.17-amzn-1	HBase命令行客户端。
hbase-rest-server	2.4.17-amzn-1	为其提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	2.4.17-amzn-1	为其提供 Thrift 端点的HBase服务。
hbase-operator-tools	2.4.17-amzn-1	适用于 Apache HBase 集群的修复工具。
hcatalog-client	3.1.3-amzn-6	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	3.1.3-amzn-6	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	3.1.3-amzn-6	HTTP端点为其提供REST接口HCatalog。
hive-client	3.1.3-amzn-6	Hive 命令行客户端。



组件	版本	描述
hive-hbase	3.1.3-amzn-6	Hive-hbase 客户端。
hive-metastore-server	3.1.3-amzn-6	用于访问 Hive 元数据仓库的服务，这是一个语义存储库，用于存储 Hadoop 操作SQL的元数据。
hive-server2	3.1.3-amzn-6	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.13.1-amzn-1	增量处理框架，以支持低延迟和高效率的数据管道。
hudi-presto	0.13.1-amzn-1	用于运行 Presto 以及 Hudi 的捆绑库。
hudi-trino	0.13.1-amzn-1	用于运行 Trino 以及 Hudi 的捆绑库。
hudi-spark	0.13.1-amzn-1	用于运行 Spark 以及 Hudi 的捆绑库。
hue-server	4.11.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
iceberg	1.3.0-amzn-1	Apache Iceberg 是一种适用于超大型分析数据集的开放表格式。
jupyterhub	1.5.0	Jupyter notebook 的多用户服务器
livy-server	0.7.1-incubating	REST用于与 Apache Spark 交互的接口

组件	版本	描述
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mxnet	1.9.1	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.68+	MariaDB 数据库服务器。
nvidia-cuda	11.8.0	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.1	Oozie 命令行客户端。
oozie-server	5.2.1	用于接受 Oozie 工作流请求的服务。
opencv	4.7.0	开源计算机视觉库。
phoenix-library	5.1.3	服务器和客户端的 phoenix 库
phoenix-connectors	5.1.3	Apache Phoenix-Connectors for Spark-3
phoenix-query-server	5.1.3	轻量级服务器，提供JDBC对Avatica 的访问权限以及协议缓冲区和JSON格式访问权限 API
presto-coordinator	0.281-amzn-1	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.281-amzn-1	用于执行查询的各个部分的服务。
presto-client	0.281-amzn-1	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。

组件	版本	描述
trino-coordinator	414-amzn-1	用于在 trino-worker 之中接受查询并管理查询的服务。
trino-worker	414-amzn-1	用于执行查询的各个部分的服务。
trino-client	414-amzn-1	Trino 命令行客户端，安装在 HA 集群的备用主服务器（未启动 Trino 服务器）上。
pig-client	0.17.0	Pig 命令行客户端。
r	4.0.2	用于统计计算的 R 项目
ranger-kms-server	2.0.0	Apache Ranger 密钥管理系统
spark-client	3.4.1-amzn-0	Spark 命令行客户端。
spark-history-server	3.4.1-amzn-0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	3.4.1-amzn-0	的内存中执行引擎。YARN
spark-yarn-slave	3.4.1-amzn-0	YARN从属服务器需要的 Apache Spark 库。
spark-rapids	23.06.0-amzn-1	Nvidia Spark RAPIDS 插件可以加速 Apache Spark。GPUs
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	2.11.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.10.2-amzn-4	tez YARN 应用程序和库。

组件	版本	描述
tez-on-worker	0.10.2-amzn-4	工作节点的 tez YARN 应用程序和库。
webserver	2.4.41+	Apache HTTP 服务器。
zeppelin-server	0.10.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.5.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.5.10	ZooKeeper 命令行客户端。

### 6.13.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

为处于运行状态的集群中的实例组指定配置时，将发生重新配置操作。Amazon EMR 仅对您修改的分类启动重新配置操作。有关更多信息，请参阅 [在正在运行的集群中重新配置实例组](#)。

#### emr-6.13.0 分类

分类	描述	重新配置操作
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。	Restarts the ResourceManager service.
container-executor	更改 Hadoop YARN 的 container-executor.cfg 文件中的值。	Not available.
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。	Not available.

分类	描述	重新配置操作
core-site	更改 Hadoop 的 core-site.xml 文件中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	更改 docker 相关设置。	Not available.
emrfs-site	更改EMRFS设置。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	更改 flink-conf.yaml 设置。	Restarts Flink history server.

分类	描述	重新配置操作
flink-log4j	更改 Flink log4j.properties 设置。	Restarts Flink history server.
flink-log4j-session	为 Kubernetes/Yarn 会话更改 Flink log4j-session.properties 设置。	Restarts Flink history server.
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。	Restarts Flink history server.
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.

分类	描述	重新配置操作
hadoop-ssl-server	更改 hadoop ssl 服务器配置	Not available.
hadoop-ssl-client	更改 hadoop ssl 客户端配置	Not available.
hbase	亚马逊EMR精心策划的 Apache 设置。HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	更改环境HBase中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	更改 hadoop-metrics2-hbase base.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。	Not available.
hbase-site	更改HBase的 hbase-site.xml 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	配置HDFS加密区域。	This classification should not be reconfigured.
hdfs-env	更改HDFS环境中的值。	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

分类	描述	重新配置操作
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	更改环境HCatalog中的值。	Restarts Hive HCatalog Server.
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。	Restarts Hive WebHCat server.
hive	亚马逊EMR精心策划的 Apache Hive 设置。	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。	Not available.
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。	Not available.



分类	描述	重新配置操作
hive-env	更改 Hive 环境中的值。	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	更改 Hive 的 hive-exec-log4j2.properties 文件中的值。	Not available.
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log4j2.properties 文件中的值。	Not available.
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。	Not available.
hive-site	更改 Hive 的 hive-site.xml 文件中的值	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值	Not available.
hue-ini	更改 Hue 的 ini 文件中的值	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	更改 HTTPFS 环境中的值。	Restarts Hadoop Httpfs service.
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。	Restarts Hadoop Httpfs service.

分类	描述	重新配置操作
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。	Not available.
hadoop-kms-env	在 Hadoop KMS 环境中更改值。	Restarts Hadoop-KMS service.
hadoop-kms-java-home	更改 Hadoop 的 KMS java 主页	Not available.
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。	Not available.
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	更改 Hudi 环境中的值。	Not available.
hudi-defaults	更改 Hudi 的 hudi-defaults.conf 文件中的值。	Not available.
iceberg-defaults	更改 Iceberg 的 iceberg-defaults.conf 文件中的值。	Not available.
delta-defaults	更改 Delta 的 delta-defaults.conf 文件中的值。	Not available.
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。	Not available.
jupyter-hub-conf	更改 JupyterHubs 的 jupyterhub_config.py 文件中的值。	Not available.
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。	Not available.
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。	Not available.

分类	描述	重新配置操作
livy-conf	更改 Livy 的 livy.conf 文件中的值。	Restarts Livy Server.
livy-env	更改 Livy 环境中的值。	Restarts Livy Server.
livy-log4j2	更改 Livy log4j2.properties 设置。	Restarts Livy Server.
mapred-env	更改 MapReduce 应用程序环境中的值。	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	更改 Oozie 的环境中的值。	Restarts Oozie.
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。	Restarts Oozie.
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。	Restarts Oozie.
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。	Not available.
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。	Not available.
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。	Restarts Phoenix-QueryServer.
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。	Not available.
pig-env	更改 Pig 环境中的值。	Not available.

分类	描述	重新配置操作
pig-properties	更改 Pig 的 pig.properties 文件中的值。	Restarts Oozie.
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。	Not available.
presto-log	更改 Presto 的 log.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-config	更改 Presto 的 config.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。	Not available.
presto-env	更改 Presto 的 presto-env.sh 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-node	更改 Presto 的 node.properties 文件中的值。	Not available.
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。	Not available.
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。	Not available.
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。	Not available.
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-lakeformat ion	更改 Presto 的 lakeforma tion.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	更改 Presto 的 localfile .properties 文件中的值。	Not available.
presto-connector-memory	更改 Presto 的 memory.pr operties 文件中的值。	Not available.
presto-connector-mongodb	更改 Presto 的 mongodb.p roperties 文件中的值。	Not available.
presto-connector-mysql	更改 Presto 的 mysql.pro perties 文件中的值。	Not available.
presto-connector-postgresql	更改 Presto 的 postgresq l.properties 文件中的值。	Not available.
presto-connector-raptor	更改 Presto 的 raptor.pr operties 文件中的值。	Not available.
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。	Not available.
presto-connector-redshift	更改 Presto 的 redshift. properties 文件中的值。	Not available.
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。	Not available.
presto-connector-tpcds	更改 Presto 的 tpcds.pro perties 文件中的值。	Not available.
trino-log	更改 Trino 的 log.properties 文 件中的值。	Restarts Trino-Server (for Trino)
trino-config	更改 Trino 的 config.properties 文件中的值。	Restarts Trino-Server (for Trino)

分类	描述	重新配置操作
trino-password-authenticator	更改 Trino 的 password-authenticator.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-env	更改 Trino 的 trino-env.sh 文件中的值。	Restarts Trino-Server (for Trino)
trino-node	更改 Trino 的 node.properties 文件中的值。	Not available.
trino-connector-blackhole	更改 Trino 的 blackhole.properties 文件中的值。	Not available.
trino-connector-cassandra	更改 Trino 的 cassandra.properties 文件中的值。	Not available.
trino-connector-delta	更改 Trino 的 delta.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-hive	更改 Trino 的 hive.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-exchange-manager	更改 Trino 的 exchange-manager.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-iceberg	更改 Trino 的 iceberg.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-hudi	更改 Trino 的 hudi.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-jmx	更改 Trino 的 jmx.properties 文件中的值。	Not available.
trino-connector-kafka	更改 Trino 的 kafka.properties 文件中的值。	Not available.

分类	描述	重新配置操作
trino-connector-localfile	更改 Trino 的 localfile.properties 文件中的值。	Not available.
trino-connector-memory	更改 Trino 的 memory.properties 文件中的值。	Not available.
trino-connector-mongodb	更改 Trino 的 mongodb.properties 文件中的值。	Not available.
trino-connector-mysql	更改 Trino 的 mysql.properties 文件中的值。	Not available.
trino-connector-postgresql	更改 Trino 的 postgresql.properties 文件中的值。	Not available.
trino-connector-raptor	更改 Trino 的 raptor.properties 文件中的值。	Not available.
trino-connector-redis	更改 Trino 的 redis.properties 文件中的值。	Not available.
trino-connector-redshift	更改 Trino 的 redshift.properties 文件中的值。	Not available.
trino-connector-tpch	更改 Trino 的 tpch.properties 文件中的值。	Not available.
trino-connector-tpcds	更改 Trino 的 tpcds.properties 文件中的值。	Not available.
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-site	更改 Ranger KMS 的 ranger-kms-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-env	在游侠KMS环境中更改值。	Restarts Ranger KMS Server.

分类	描述	重新配置操作
ranger-kms-logback	更改 Ranger KMS 的 kms-logback.xml 文件中的值。	Not available.
ranger-kms-db-ca	更改 S3 上我与 Ranger SQL SSL KMS 连接的 CA 文件值。	Not available.
spark	亚马逊EMR精心策划的 Apache Spark 设置。	This property modifies spark-defaults. See actions there.
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-env	更改 Spark 环境中的值。	Restarts Spark history server and Spark thrift server.
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值	Not available.
spark-log4j2	更改 Spark 的 log4j2.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
sqoop-env	更改 Sqoop 的环境中的值。	Not available.
sqoop-orahome-site	更改 Sqoop OraOop 的 orahome-site.xml 文件中的值。	Not available.
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。	Not available.
tez-site	更改 Tez 的 tez-site.xml 文件中的值。	Restart Oozie and HiveServer2.



分类	描述	重新配置操作
yarn-env	更改YARN环境中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	更改 yarn-site.xml 文件YARN 中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	更改 Zeppelin 环境中的值。	Restarts Zeppelin.
zeppelin-site	更改 zeppelin-site.xml 中的配置设置。	Restarts Zeppelin.
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。	Restarts Zookeeper server.
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。	Restarts Zookeeper server.

## 6.13.0 更改日志

发行版 6.13.0 的更改日志和发布说明

Date	事件	描述
2023-09-23	部署完成	<a href="#">Amazon EMR 6.13.0 已完全部署到所有支持的区域</a>

Date	事件	描述
2023-09-12	文档发布	亚马逊 EMR 6.13.0 发行说明首次发布
2023-09-01	初始版本	亚马逊 EMR 6.13.0 首次部署到初始商业区域

## 亚马逊EMR版本 6.12.0

### 6.12.0 应用程序版本

此发行版支持以下应用程序

序：[Delta](#)、[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[Iceberg](#)、[JupyterEnterpriseC](#)和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-6.12.0	emr-6.11.1	emr-6.11.0	emr-6.10.1
适用于 Java 的 AWS SDK	1.12.490	1.12.446	1.12.446	1.12.397
Python	2.7、3.7	2.7、3.7	2.7、3.7	2.7、3.7
Scala	2.12.15	2.12.15	2.12.15	2.12.15

	emr-6.12.0	emr-6.11.1	emr-6.11.0	emr-6.10.1
AmazonCloudWatchAgent	-	-	-	-
Delta	2.4.0	2.2.0	2.2.0	2.2.0
Flink	1.17.0	1.16.0	1.16.0	1.16.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.17	2.4.15	2.4.15	2.4.15
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.3.3	3.3.3	3.3.3
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.13.1-amzn-0	0.13.0-amzn-0	0.13.0-amzn-0	0.12.2-amzn-0
Hue	4.11.0	4.11.0	4.11.0	4.10.0
Iceberg	1.3.0-amzn-0	1.2.0-amzn-0	1.2.0-amzn-0	1.1.0-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0	2.6.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.5.0
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.3	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0

	emr-6.12.0	emr-6.11.1	emr-6.11.0	emr-6.10.1
Presto	0.281	0.279	0.279	0.278
Spark	3.4.0	3.3.2	3.3.2	3.3.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.11.0	2.11.0
Tez	0.10.2	0.10.2	0.10.2	0.10.2
Trino (PrestoSQL)	414	410	410	403
Zeppelin	0.10.1	0.10.1	0.10.1	0.10.1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

## 6.12.0 发布说明

以下发行说明包含有关亚马逊 6.12. EMR 0 版本的信息。更改与 6.11.0 有关。有关发布时间表的更多信息，请参阅 [6.12.0 更改日志](#)。

### 新功能

- Amazon EMR 6.12.0 supports Apache Spark 3.4.0, Apache Spark RAPIDS 23.06.0-amzn-0, CUDA 11.8.0, Apache Hudi 0.13.1-amzn-0, Apache Iceberg 1.3.0-amzn-0, Trino 414, and PrestoDB 0.281.
- 亚马逊EMR发布了 6.12.0 及更高版本的支持，与 Apache Livy、Apache Hive 到 HiveServer 2 (HS2)、Trino、Presto 和 Hue LDAP 集成。你还可以在使用 6.12.0 或更高版本的EMR集群上安装 Apache Spark 和 Apache Hadoop，然后对其进行配置以供使用。LDAP有关更多信息，请参阅[使用 Active Directory 或LDAP服务器进行亚马逊身份验证EMR](#)。

### 更改、增强功能和解决的问题

- 亚马逊EMR版本 6.12.0 及更高版本为 Flink 提供了 Java 11 运行时支持。有关更多信息，请参阅 [将 Flink 配置为使用 Java 11 运行](#)。

- 6.12.0 版本为运行 Presto 或 Trino 的集群的集群扩展工作流程添加了新的重试机制。EMR这一改进降低了由于单个调整大小操作失败而导致集群大小调整无限期停滞的风险。它还可以提高集群利用率，因为您的集群可以更快地向上和向下扩展。
- 6.12.0 版本修复了一个问题，即当处于正常停用状态的核心节点在完全停用之前出于任何原因变得运行不正常时，集群的缩减操作可能会停滞不前。
- 6.12.0 版本改进了集群缩小逻辑，因此您的集群不会尝试将核心节点缩小到低于集群HDFS重复因子设置的范围。这符合您的数据冗余要求，并减少了扩展操作可能停滞的机会。
- 6.12.0 版本EMR通过提高记录实例状态变化的速度，提高了 Amazon 运行状况监控服务的性能和效率。这一改进降低了运行多个自定义客户端工具或第三方应用程序的集群节点性能下降的机会。
- 6.12.0 版本提高了 Amazon 集群日志管理守护程序的性能。EMR因此，对于以高并发度运行步骤的 EMR集群，性能下降的可能性较小。
- 在 Amazon 6.12.0 EMR 版本中，日志管理守护程序已升级，可以识别所有正在使用本地实例存储上的打开文件句柄的日志，以及相关的进程。此次升级可确保 Amazon 在日志存档到 Amazon S3 后 EMR正确删除文件并回收存储空间。
- 6.12.0 版本包括日志管理进程守护程序增强功能，可删除本地集群文件系统中空的、未使用的步骤目录。过多的空目录会降低 Amazon EMR 守护程序的性能并导致磁盘过度使用。
- 6.12.0 版本允许对YARN时间轴服务器日志进行日志轮换。这样可以最大限度地减少磁盘过度使用情况，特别是对于长时间运行的集群。
- 在 Amazon EMR 6.10.0 及更高版本中，默认根卷大小已增加到 15 GB。早期版本的默认根卷大小为 10 GB。
- 当您启动带有最新补丁版本（亚马逊 EMR 5.36 或更高版本、6.6 或更高版本、7.0 或更高版本）的集群时，亚马逊将EMR使用最新的亚马逊 Linux 2023 或亚马逊 Linux 2 版本作为默认亚马逊 Linux 2 版本。EMR AMI有关更多信息，请参阅[使用亚马逊的默认 Amazon Linux AMI EMR](#)。

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 709.1	4.14.348	2024年7月23日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、AWS GovCloud ( 美国东部 )、( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )、亚洲太平洋 ( 海得拉巴 )、中东 ( UAE )、欧洲 ( 西班牙 )、欧洲 ( 苏黎世 )、亚太地区

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 墨尔本 )、以色列 ( 特拉维夫 )、加拿大西部 ( 卡尔加里 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 223.0	4.14.336	2024 年 3 月 8 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、中国 ( 北



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 131.0	4.14.336	2024年2月14日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、中国 ( 北

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024124.0	4.14.336	2024 年 2 月 7 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、中国 ( 北

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 109.0	4.14.334	2024 年 1 月 24 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、中国 ( 北

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 218.0	4.14.330	2024 年 1 月 2 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 206.0	4.14.330	2023 年 12 月 22 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023116.0	4.14.328	2023 年 12 月 11 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 101.0	4.14.327	2023 年 11 月 16 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023020.1	4.14.326	2023 年 11 月 7 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023012.1	4.14.326	2023 年 10 月 26 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 926.0	4.14.322	2023 年 10 月 19 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 906.0	4.14.322	2023 年 10 月 4 日	(美国东部)、中国(北京)、中国(宁夏)  美国东部(弗吉尼亚北部)、美国东部(俄亥俄州)、美国西部(加利福尼亚北部)、美国西部(俄勒冈)、欧洲(斯德哥尔摩)、欧洲(米兰)、欧洲(西班牙)、欧洲(法兰克福)、欧洲(苏黎世)、欧洲(伦敦)、欧洲(巴黎)、亚太地区(香港)、亚太地区(孟买)、亚太地区(海得拉巴)、亚太地区(东京)、亚太地区(东京)、亚太地区(首尔)、亚太地区(大阪)、亚太地区(新加坡)、亚太地区(悉尼)、亚太地区(雅加达)、亚太地区(墨尔本)、非洲(开普敦)、南美洲(圣保罗)、中东(巴林)、中东( )、中东(UAE)、加拿大(中部)、以色列(特拉维夫)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.20230822.0	4.14.322	2023 年 8 月 30 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 808.0	4.14.320	2023 年 8 月 24 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 727.0	4.14.320	2023 年 8 月 14 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )



OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 719.0	4.14.320	2023 年 8 月 2 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 628.0	4.14.318	2023 年 7 月 12 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 (巴林)、中东 (UAE)、加拿大 (中部)

## 6.12.0 默认 Java 版本

亚马逊EMR发布的6.12.0及更高版本默认支持所有带有Amazon Corretto 8的应用程序，但Trino除外。对于 Trino，从亚马逊 EMR 6.9.0 版本开始，亚马逊默认支持 Amazon Corretto 17。EMR亚马逊EMR

还支持某些带有 Amazon Corretto 11 和 17 的应用程序。下表列出了这些应用程序。如果要更改集群的默认设置JVM，请按照集群上运行[配置应用程序来使用特定 Java 虚拟机](#)的每个应用程序中的说明进行操作。一个集群只能使用一个 Java 运行时系统版本。Amazon EMR 不支持在同一集群的不同运行时版本上运行不同的节点或应用程序。

虽然亚马逊在 Apache Spark、Apache Hadoop 和 Apache Hive 上同时EMR支持 Amazon Corretto 11 和 17，但当你使用这些版本的 Corretto 时，某些工作负载的性能可能会下降。我们建议您在更改默认值之前先测试工作负载。

下表显示了 Amazon EMR 6.12.0 中应用程序的默认 Java 版本：

应用程序	Java/Amazon Corretto 版本 ( 默认为粗体 )
Delta	17、11、8
Flink	11、8
Ganglia	8
HBase	11、8
HCatalog	17、11、8
Hadoop	17、11、8
Hive	17、11、8
Hudi	17、11、8
Iceberg	17、11、8
Livy	17、11、8
Oozie	17、11、8
Phoenix	8
PrestoDB	8
Spark	17、11、8

应用程序	Java/Amazon Corretto 版本 (默认为粗体)
火花 RAPIDS	17、11、8
Sqoop	8
Tez	17、11、8
Trino	17
Zeppelin	8
Pig	8
Zookeeper	8

## 6.12.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.4.2	亚马逊 SageMaker Spark SDK
delta	2.4.0	Delta lake 是一种适用于超大型分析数据集的开放表格式。
delta-standalone-connectors	0.6.0	Delta Connectors 提供不同的运行时，将 Delta Lake 与 Flink、Hive 和 Presto 等引擎集成。

组件	版本	描述
emr-ddb	5.1.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	3.5.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.9.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-notebook-env	1.7.0	适用于 EMR Notebooks ( 可提供 jupyter 企业网关 ) 的 Conda env
emr-s3-dist-cp	2.26.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	2.5.0	EMRS3Select 连接器
emr-wal-cli	1.1.0	用于 emrwal 列表/删除的 cli。
emrfs	2.57.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.17.0	Apache Flink 命令行客户端脚本和应用程序。
flink-jobmanager-config	1.17.0	管理 Apache Flink EMR JobManager 节点上的资源。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。

组件	版本	描述
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	3.3.3-amzn-4	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	3.3.3-amzn-4	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	3.3.3-amzn-4	HDFS命令行客户端和库
hadoop-hdfs-namenode	3.3.3-amzn-4	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	3.3.3-amzn-4	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	3.3.3-amzn-4	HTTPHDFS操作端点。
hadoop-kms-server	3.3.3-amzn-4	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	3.3.3-amzn-4	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	3.3.3-amzn-4	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	3.3.3-amzn-4	YARN用于分配和管理群集资源和分布式应用程序的服务。

组件	版本	描述
hadoop-yarn-timeline-server	3.3.3-amzn-4	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	2.4.17-amzn-0	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	2.4.17-amzn-0	为一个或多个HBase地区提供服务的服务。
hbase-client	2.4.17-amzn-0	HBase命令行客户端。
hbase-rest-server	2.4.17-amzn-0	为其提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	2.4.17-amzn-0	为其提供 Thrift 端点的HBase服务。
hbase-operator-tools	2.4.17-amzn-0	适用于 Apache HBase 集群的修复工具。
hcatalog-client	3.1.3-amzn-5	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	3.1.3-amzn-5	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	3.1.3-amzn-5	HTTP端点为其提供REST接口HCatalog。
hive-client	3.1.3-amzn-5	Hive 命令行客户端。
hive-hbase	3.1.3-amzn-5	Hive-hbase 客户端。
hive-metastore-server	3.1.3-amzn-5	用于访问 Hive 元数据仓库的服务，这是一个语义存储库，用于存储 Hadoop 操作SQL的元数据。

组件	版本	描述
hive-server2	3.1.3-amzn-5	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.13.1-amzn-0	增量处理框架，以支持低延迟和高效率的数据管道。
hudi-presto	0.13.1-amzn-0	用于运行 Presto 以及 Hudi 的捆绑库。
hudi-trino	0.13.1-amzn-0	用于运行 Trino 以及 Hudi 的捆绑库。
hudi-spark	0.13.1-amzn-0	用于运行 Spark 以及 Hudi 的捆绑库。
hue-server	4.11.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
iceberg	1.3.0-amzn-0	Apache Iceberg 是一种适用于超大型分析数据集的开放表格式。
jupyterhub	1.4.1	Jupyter notebook 的多用户服务器
livy-server	0.7.1-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mxnet	1.9.1	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.68+	MariaDB 数据库服务器。



组件	版本	描述
nvidia-cuda	11.8.0	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.1	Oozie 命令行客户端。
oozie-server	5.2.1	用于接受 Oozie 工作流请求的服务。
opencv	4.7.0	开源计算机视觉库。
phoenix-library	5.1.3	服务器和客户端的 phoenix 库
phoenix-connectors	5.1.3	Apache Phoenix-Connectors for Spark-3
phoenix-query-server	5.1.3	轻量级服务器，提供JDBC对Avatica 的访问权限以及协议缓冲区和JSON格式访问权限 API
presto-coordinator	0.281-amzn-0	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.281-amzn-0	用于执行查询的各个部分的服务。
presto-client	0.281-amzn-0	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
trino-coordinator	414-amzn-0	用于在 trino-worker 之中接受查询并管理查询的服务。
trino-worker	414-amzn-0	用于执行查询的各个部分的服务。
trino-client	414-amzn-0	Trino 命令行客户端，安装在 HA 集群的备用主服务器（未启动 Trino 服务器）上。

组件	版本	描述
pig-client	0.17.0	Pig 命令行客户端。
r	4.0.2	用于统计计算的 R 项目
ranger-kms-server	2.0.0	Apache Ranger 密钥管理系统
spark-client	3.4.0-amzn-0	Spark 命令行客户端。
spark-history-server	3.4.0-amzn-0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	3.4.0-amzn-0	的内存中执行引擎。YARN
spark-yarn-slave	3.4.0-amzn-0	YARN从属服务器需要的 Apache Spark 库。
spark-rapids	23.06.0-amzn-0	Nvidia Spark RAPIDS 插件可以加速 Apache Spark。GPUs
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	2.11.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.10.2-amzn-3	tez YARN 应用程序和库。
tez-on-worker	0.10.2-amzn-3	工作节点的 tez YARN 应用程序和库。
webserver	2.4.41+	Apache HTTP 服务器。
zeppelin-server	0.10.1	支持交互式数据分析的基于 Web 的笔记本电脑。

组件	版本	描述
zookeeper-server	3.5.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.5.10	ZooKeeper 命令行客户端。

## 6.12.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

为处于运行状态的集群中的实例组指定配置时，将发生重新配置操作。Amazon EMR 仅对您修改的分类启动重新配置操作。有关更多信息，请参阅 [在正在运行的集群中重新配置实例组](#)。

### emr-6.12.0 分类

分类	描述	重新配置操作
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。	Restarts the ResourceManager service.
container-executor	更改 Hadoop YARN 的 container-executor.cfg 文件中的值。	Not available.
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。	Not available.
core-site	更改 Hadoop 的 core-site.xml 文件中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServ

分类	描述	重新配置操作
		er, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	更改 docker 相关设置。	Not available.
emrfs-site	更改EMRFS设置。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	更改 flink-conf.yaml 设置。	Restarts Flink history server.
flink-log4j	更改 Flink log4j.properties 设置。	Restarts Flink history server.
flink-log4j-session	为 Kubernetes/Yarn 会话更改 Flink log4j-session.properties 设置。	Restarts Flink history server.
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。	Restarts Flink history server.

分类	描述	重新配置操作
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	更改 hadoop ssl 服务器配置	Not available.
hadoop-ssl-client	更改 hadoop ssl 客户端配置	Not available.
hbase	亚马逊EMR精心策划的 Apache 设置。HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.

分类	描述	重新配置操作
hbase-env	更改环境HBase中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	更改 hadoop-metrics2-hbase base.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。	Not available.
hbase-site	更改HBase的 hbase-site.xml 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	配置HDFS加密区域。	This classification should not be reconfigured.
hdfs-env	更改HDFS环境中的值。	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	更改环境HCatalog中的值。	Restarts Hive HCatalog Server.

分类	描述	重新配置操作
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。	Restarts Hive WebHCat server.
hive	亚马逊EMR精心策划的 Apache Hive 设置。	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。	Not available.
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。	Not available.
hive-env	更改 Hive 环境中的值。	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	更改 Hive 的 hive-exec-log4j2.properties 文件中的值。	Not available.
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log4j2.properties 文件中的值。	Not available.

分类	描述	重新配置操作
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。	Not available.
hive-site	更改 Hive 的 hive-site.xml 文件中的值	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值	Not available.
hue-ini	更改 Hue 的 ini 文件中的值	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	更改HTTPFS环境中的值。	Restarts Hadoop Httpfs service.
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。	Restarts Hadoop Httpfs service.
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。	Not available.
hadoop-kms-env	在 Hadoop KMS 环境中更改值。	Restarts Hadoop-KMS service.
hadoop-kms-java-home	更改 Hadoop 的 KMS java 主页	Not available.
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。	Not available.



分类	描述	重新配置操作
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	更改 Hudi 环境中的值。	Not available.
hudi-defaults	更改 Hudi 的 hudi-defaults.conf 文件中的值。	Not available.
iceberg-defaults	更改 Iceberg 的 iceberg-defaults.conf 文件中的值。	Not available.
delta-defaults	更改 Delta 的 delta-defaults.conf 文件中的值。	Not available.
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。	Not available.
jupyter-hub-conf	更改 JupyterHubs 的 jupyterhub_config.py 文件中的值。	Not available.
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。	Not available.
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。	Not available.
livy-conf	更改 Livy 的 livy.conf 文件中的值。	Restarts Livy Server.
livy-env	更改 Livy 环境中的值。	Restarts Livy Server.
livy-log4j2	更改 Livy log4j2.properties 设置。	Restarts Livy Server.
mapred-env	更改 MapReduce 应用程序环境中的值。	Restarts Hadoop MapReduce-HistoryServer.

分类	描述	重新配置操作
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	更改 Oozie 的环境中的值。	Restarts Oozie.
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。	Restarts Oozie.
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。	Restarts Oozie.
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。	Not available.
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。	Not available.
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。	Restarts Phoenix-QueryServer.
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。	Not available.
pig-env	更改 Pig 环境中的值。	Not available.
pig-properties	更改 Pig 的 pig.properties 文件中的值。	Restarts Oozie.
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。	Not available.
presto-log	更改 Presto 的 log.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-config	更改 Presto 的 config.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)

分类	描述	重新配置操作
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。	Not available.
presto-env	更改 Presto 的 presto-env.sh 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-node	更改 Presto 的 node.properties 文件中的值。	Not available.
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。	Not available.
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。	Not available.
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。	Not available.
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。	Not available.
presto-connector-lakeformation	更改 Presto 的 lakeformation.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。	Not available.
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。	Not available.
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。	Not available.
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。	Not available.
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。	Not available.
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。	Not available.
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。	Not available.
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。	Not available.
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。	Not available.
trino-log	更改 Trino 的 log.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-config	更改 Trino 的 config.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-password-authenticator	更改 Trino 的 password-authenticator.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-env	更改 Trino 的 trino-env.sh 文件中的值。	Restarts Trino-Server (for Trino)
trino-node	更改 Trino 的 node.properties 文件中的值。	Not available.

分类	描述	重新配置操作
trino-connector-blackhole	更改 Trino 的 blackhole.properties 文件中的值。	Not available.
trino-connector-cassandra	更改 Trino 的 cassandra.properties 文件中的值。	Not available.
trino-connector-delta	更改 Trino 的 delta.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-hive	更改 Trino 的 hive.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-exchange-manager	更改 Trino 的 exchange-manager.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-iceberg	更改 Trino 的 iceberg.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-hudi	更改 Trino 的 hudi.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-jmx	更改 Trino 的 jmx.properties 文件中的值。	Not available.
trino-connector-kafka	更改 Trino 的 kafka.properties 文件中的值。	Not available.
trino-connector-localfile	更改 Trino 的 localfile.properties 文件中的值。	Not available.
trino-connector-memory	更改 Trino 的 memory.properties 文件中的值。	Not available.
trino-connector-mongodb	更改 Trino 的 mongodb.properties 文件中的值。	Not available.

分类	描述	重新配置操作
trino-connector-mysql	更改 Trino 的 mysql.properties 文件中的值。	Not available.
trino-connector-postgresql	更改 Trino 的 postgresql.properties 文件中的值。	Not available.
trino-connector-raptor	更改 Trino 的 raptor.properties 文件中的值。	Not available.
trino-connector-redis	更改 Trino 的 redis.properties 文件中的值。	Not available.
trino-connector-redshift	更改 Trino 的 redshift.properties 文件中的值。	Not available.
trino-connector-tpch	更改 Trino 的 tpch.properties 文件中的值。	Not available.
trino-connector-tpcds	更改 Trino 的 tpcds.properties 文件中的值。	Not available.
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-site	更改 Ranger KMS 的 ranger-kms-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-env	在 Ranger KMS 环境中更改值。	Restarts Ranger KMS Server.
ranger-kms-logback	更改 Ranger KMS 的 kms-logback.xml 文件中的值。	Not available.
ranger-kms-db-ca	更改 S3 上我与 Ranger SQL SSL KMS 连接的 CA 文件值。	Not available.
spark	亚马逊EMR精心策划的 Apache Spark 设置。	This property modifies spark-defaults. See actions there.

分类	描述	重新配置操作
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-env	更改 Spark 环境中的值。	Restarts Spark history server and Spark thrift server.
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值	Not available.
spark-log4j2	更改 Spark 的 log4j2.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
sqoop-env	更改 Sqoop 的环境中的值。	Not available.
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。	Not available.
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。	Not available.
tez-site	更改 Tez 的 tez-site.xml 文件中的值。	Restart Oozie and HiveServer2.
yarn-env	更改YARN环境中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduceHistoryServer.

分类	描述	重新配置操作
yarn-site	更改 yarn-site.xml 文件YARN 中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	更改 Zeppelin 环境中的值。	Restarts Zeppelin.
zeppelin-site	更改 zeppelin-site.xml 中的配置设置。	Restarts Zeppelin.
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。	Restarts Zookeeper server.
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。	Restarts Zookeeper server.

## 6.12.0 更改日志

发行版 6.12.0 的更改日志和发布说明

Date	事件	描述
2023-07-27	更新文档	更新 6.12 的 Java 选项并添加 Oozie 教程进行更新 JVM
2023-07-21	部署完成	<a href="#">Amazon EMR 6.12.0 已完全部署到所有支持的区域</a>
2023-07-21	文档发布	亚马逊 EMR 6.12.0 发行说明首次发布
2023-07-12	初始版本	亚马逊 EMR 6.12.0 首次部署到初始商业区域



## 亚马逊EMR版本 6.11.1

### 6.11.1 应用程序版本

此发行版支持以下应用程序

序：[Delta](#)、[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[Iceberg](#)、[JupyterEnterpriseC](#)和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-6.11.1	emr-6.11.0	emr-6.10.1	emr-6.10.0
适用于 Java 的 AWS SDK	1.12.446	1.12.446	1.12.397	1.12.397
Python	2.7、3.7	2.7、3.7	2.7、3.7	2.7、3.7
Scala	2.12.15	2.12.15	2.12.15	2.12.15
AmazonCloudWatchAgent	-	-	-	-
Delta	2.2.0	2.2.0	2.2.0	2.2.0
Flink	1.16.0	1.16.0	1.16.0	1.16.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.15	2.4.15	2.4.15	2.4.15

	emr-6.11.1	emr-6.11.0	emr-6.10.1	emr-6.10.0
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.3.3	3.3.3	3.3.3
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.13.0-amzn-0	0.13.0-amzn-0	0.12.2-amzn-0	0.12.2-amzn-0
Hue	4.11.0	4.11.0	4.10.0	4.10.0
Iceberg	1.2.0-amzn-0	1.2.0-amzn-0	1.1.0-amzn-0	1.1.0-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0	2.6.0
JupyterHub	1.4.1	1.4.1	1.5.0	1.5.0
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.279	0.279	0.278	0.278
Spark	3.3.2	3.3.2	3.3.1	3.3.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.11.0	2.11.0
Tez	0.10.2	0.10.2	0.10.2	0.10.2

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

## 6.11.1 发布说明

以下发行说明包含有关亚马逊 6.11.1 EMR 版本的信息。更改与 6.11.0 有关。有关发布时间表的更多信息，请参阅 [6.11.1 更改日志](#)。

### 更改、增强功能和解决的问题

- 由于锁争用，如果在尝试停用节点的同时添加或移除节点，则该节点可能会陷入死锁。结果，Hadoop 资源管理器 (YARN) 变得无响应，并会影响所有传入和当前正在运行的容器。
- 此版本包括一项更改，允许高可用性集群在重启后从故障状态中恢复。
- 此版本包含针对 Hue 和的安全补丁HBase。
- 此版本修复了在 Spark with Amazon 上运行工作负载的集群EMR可能会在静默中收到使用contains、startsWithendsWith、和like的错误结果的问题。当您对在 Amazon EMR Hive3 Metastore 服务器中包含元数据的分区字段使用表达式时，就会出现此问题 ()。HMS
- 此版本修复了当没有用户定义的函数时，Glue 端的限制问题 () UDF。
- 此版本修复了一个问题，即节点日志聚合服务在日志推送器可以将容器日志推送到 S3 之前删除容器日志，以防YARN停用。
- 此版本修复了 Hadoop FairShare p 启用节点标签时调度器指标的问题。
- 此版本修复了您在 spark-defaults.conf 中为 spark.yarn.heterogeneousExecutors.enabled 配置设置默认 true 值时影响 Spark 性能的问题。
- 此版本修复了 Reduce Task 无法读取随机数据的问题。该问题因内存损坏错误导致 Hive 查询失败。
- 此版本为运行 Presto 或 Trino 的集群的集群扩展工作流程添加了新的重试机制。EMR这一改进降低了由于单个调整大小操作失败而导致集群大小调整无限期停滞的风险。它还可以提高集群利用率，因为您的集群可以更快地向上和向下扩展。

- 此版本改进了集群缩容逻辑，因此您的集群不会尝试将核心节点缩小到低于集群HDFS重复因子设置的范围。这符合您的数据冗余要求，并减少了扩展操作可能停滞的机会。
- 日志管理进程守护程序已升级，可以识别本地实例存储中所有包含打开文件句柄的使用中的日志，以及相关的进程。此次升级可确保 Amazon 在日志存档到 Amazon S3 后EMR正确删除文件并回收存储空间。
- 此版本包括日志管理进程守护程序增强功能，可删除本地集群文件系统中空的、未使用的步骤目录。过多的空目录会降低 Amazon EMR 守护程序的性能并导致磁盘过度使用。
- 当您启动带有最新补丁版本（亚马逊 EMR 5.36 或更高版本、6.6 或更高版本、7.0 或更高版本）的集群时，亚马逊将EMR使用最新的亚马逊 Linux 2023 或亚马逊 Linux 2 版本作为默认亚马逊 Linux 2 版本。EMR AMI有关更多信息，请参阅[使用亚马逊的默认 Amazon Linux AMI EMR](#)。

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 709.1	4.14.348	2024年7月23日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、AWS GovCloud ( 美国东部 )、( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )、亚洲太平洋 ( 海得拉巴 )、中东 ( UAE )、欧洲 ( 西班牙 )、欧洲 ( 苏黎世 )、亚太地区 ( 墨尔本 )、以色列 ( 特拉维夫 )、加拿大西部 ( 卡尔加里 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 223.0	4.14.336	2024 年 3 月 8 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、中国 ( 北

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 131.0	4.14.336	2024年2月14日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、中国 ( 北



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024124.0	4.14.336	2024 年 2 月 7 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、中国 (北

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 109.0	4.14.334	2024 年 1 月 24 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、中国 (北

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 218.0	4.14.330	2024 年 1 月 2 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 206.0	4.14.330	2023 年 12 月 22 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023116.0	4.14.328	2023 年 12 月 11 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 101.0	4.14.327	2023 年 11 月 16 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023020.1	4.14.326	2023 年 11 月 7 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023012.1	4.14.326	2023 年 10 月 26 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 926.0	4.14.322	2023 年 10 月 19 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 906.0	4.14.322	2023 年 10 月 4 日	(美国东部)、中国(北京)、中国(宁夏)  美国东部(弗吉尼亚北部)、美国东部(俄亥俄州)、美国西部(加利福尼亚北部)、美国西部(俄勒冈)、欧洲(斯德哥尔摩)、欧洲(米兰)、欧洲(西班牙)、欧洲(法兰克福)、欧洲(苏黎世)、欧洲(伦敦)、欧洲(巴黎)、亚太地区(香港)、亚太地区(孟买)、亚太地区(海得拉巴)、亚太地区(东京)、亚太地区(东京)、亚太地区(首尔)、亚太地区(大阪)、亚太地区(新加坡)、亚太地区(悉尼)、亚太地区(雅加达)、亚太地区(墨尔本)、非洲(开普敦)、南美洲(圣保罗)、中东(巴林)、中东( )、中东(UAE)、加拿大(中部)、以色列(特拉维夫)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.20230822.0	4.14.322	2023 年 8 月 30 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 808.0	4.14.320	2023 年 8 月 24 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 727.0	4.14.320	2023 年 8 月 14 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、欧洲地区 (斯德哥尔摩)、欧洲地区 (米兰)、欧洲地区 (法兰克福)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)

### 6.11.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.4.2	亚马逊 SageMaker Spark SDK
delta	2.2.0	Delta lake 是一种适用于超大型分析数据集的开放表格式。
delta-standalone-connectors	0.6.0	Delta Connectors 提供不同的运行时，将 Delta Lake 与 Flink、Hive 和 Presto 等引擎集成。
emr-ddb	5.1.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	3.4.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.8.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-notebook-env	1.7.0	适用于 EMR Notebooks ( 可提供 jupyter 企业网关 ) 的 Conda env
emr-s3-dist-cp	2.25.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	2.4.0	EMRS3Select 连接器
emr-wal-cli	1.1.0	用于 emrwal 列表/删除的 cli。

组件	版本	描述
emrfs	2.56.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.16.0	Apache Flink 命令行客户端脚本和应用程序。
flink-jobmanager-config	1.16.0	管理 Apache Flink EMR JobManager 节点上的资源。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	3.3.3-amzn-3.1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	3.3.3-amzn-3.1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	3.3.3-amzn-3.1	HDFS命令行客户端和库
hadoop-hdfs-namenode	3.3.3-amzn-3.1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	3.3.3-amzn-3.1	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	3.3.3-amzn-3.1	HTTPHDFS操作端点。



组件	版本	描述
hadoop-kms-server	3.3.3-amzn-3.1	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	3.3.3-amzn-3.1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	3.3.3-amzn-3.1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	3.3.3-amzn-3.1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	3.3.3-amzn-3.1	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	2.4.15-amzn-1.1	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	2.4.15-amzn-1.1	为一个或多个HBase地区提供服务的服务。
hbase-client	2.4.15-amzn-1.1	HBase命令行客户端。
hbase-rest-server	2.4.15-amzn-1.1	为其提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	2.4.15-amzn-1.1	为其提供 Thrift 端点的HBase 服务。
hbase-operator-tools	2.4.15-amzn-1.1	适用于 Apache HBase 集群的修复工具。
hcatalog-client	3.1.3-amzn-4.1	用于操作 hcatalog-server 的“hcat”命令行客户端。

组件	版本	描述
hcatalog-server	3.1.3-amzn-4.1	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	3.1.3-amzn-4.1	HTTP端点为其提供REST接口HCatalog。
hive-client	3.1.3-amzn-4.1	Hive 命令行客户端。
hive-hbase	3.1.3-amzn-4.1	Hive-hbase 客户端。
hive-metastore-server	3.1.3-amzn-4.1	用于访问 Hive 元数据仓库的服务，这是一个语义存储库，用于存储 Hadoop 操作SQL的元数据。
hive-server2	3.1.3-amzn-4.1	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.13.0-amzn-0	增量处理框架，以支持低延迟和高效率的数据管道。
hudi-presto	0.13.0-amzn-0	用于运行 Presto 以及 Hudi 的捆绑库。
hudi-trino	0.13.0-amzn-0	用于运行 Trino 以及 Hudi 的捆绑库。
hudi-spark	0.13.0-amzn-0	用于运行 Spark 以及 Hudi 的捆绑库。
hue-server	4.11.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序

组件	版本	描述
iceberg	1.2.0-amzn-0	Apache Iceberg 是一种适用于超大型分析数据集的开放表格式。
jupyterhub	1.4.1	Jupyter notebook 的多用户服务器
livy-server	0.7.1-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mxnet	1.9.1	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.68+	MariaDB 数据库服务器。
nvidia-cuda	11.8.0	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.1	Oozie 命令行客户端。
oozie-server	5.2.1	用于接受 Oozie 工作流请求的服务。
opencv	4.5.0	开源计算机视觉库。
phoenix-library	5.1.2	服务器和客户端的 phoenix 库
phoenix-connectors	5.1.2	Apache Phoenix-Connectors for Spark-3
phoenix-query-server	5.1.2	轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限 API

组件	版本	描述
presto-coordinator	0.279-amzn-0	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.279-amzn-0	用于执行查询的各个部分的服务。
presto-client	0.279-amzn-0	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
trino-coordinator	410-amzn-0	用于在 trino-worker 之中接受查询并管理查询的服务。
trino-worker	410-amzn-0	用于执行查询的各个部分的服务。
trino-client	410-amzn-0	Trino 命令行客户端，安装在 HA 集群的备用主服务器（未启动 Trino 服务器）上。
pig-client	0.17.0	Pig 命令行客户端。
r	4.0.2	用于统计计算的 R 项目
ranger-kms-server	2.0.0	Apache Ranger 密钥管理系统
spark-client	3.3.2-amzn-0.1	Spark 命令行客户端。
spark-history-server	3.3.2-amzn-0.1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	3.3.2-amzn-0.1	的内存中执行引擎。YARN
spark-yarn-slave	3.3.2-amzn-0.1	YARN从属服务器需要的 Apache Spark 库。

组件	版本	描述
spark-rapids	23.02.0-amzn-0	Nvidia Spark RAPIDS 插件可以加速 Apache Spark。GPUs
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	2.11.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.10.2-amzn-2.1	tez YARN 应用程序和库。
tez-on-worker	0.10.2-amzn-2.1	工作节点的 tez YARN 应用程序和库。
webserver	2.4.41+	Apache HTTP 服务器。
zeppelin-server	0.10.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.5.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.5.10	ZooKeeper 命令行客户端。

### 6.11.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

为处于运行状态的集群中的实例组指定配置时，将发生重新配置操作。Amazon EMR 仅对您修改的分类启动重新配置操作。有关更多信息，请参阅 [在正在运行的集群中重新配置实例组](#)。

## emr-6.11.1 分类

分类	描述	重新配置操作
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。	Restarts the Resource Manager service.
container-executor	更改 Hadoop YARN 的 container-executor.cfg 文件中的值。	Not available.
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。	Not available.
core-site	更改 Hadoop 的 core-site.xml 文件中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	更改 docker 相关设置。	Not available.
emrfs-site	更改EMRFS设置。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager,

分类	描述	重新配置操作
		NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	更改 flink-conf.yaml 设置。	Restarts Flink history server.
flink-log4j	更改 Flink log4j.properties 设置。	Restarts Flink history server.
flink-log4j-session	为 Kubernetes/Yarn 会话更改 Flink log4j-session.properties 设置。	Restarts Flink history server.
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。	Restarts Flink history server.
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.

分类	描述	重新配置操作
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	更改 hadoop ssl 服务器配置	Not available.
hadoop-ssl-client	更改 hadoop ssl 客户端配置	Not available.
hbase	亚马逊EMR精心策划的 Apache 设置。HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	更改环境HBase中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。	Not available.



分类	描述	重新配置操作
hbase-site	更改HBase的 hbase-site.xml 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	配置HDFS加密区域。	This classification should not be reconfigured.
hdfs-env	更改HDFS环境中的值。	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	更改环境HCatalog中的值。	Restarts Hive HCatalog Server.
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。	Restarts Hive WebHCat server.

分类	描述	重新配置操作
hcatalog-webhcat-site	更改 HCatalog WebHCat 的 webhcat-site.xml 文件中的值。	Restarts Hive WebHCat server.
hive	亚马逊EMR精心策划的 Apache Hive 设置。	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。	Not available.
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。	Not available.
hive-env	更改 Hive 环境中的值。	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	更改 Hive 的 hive-exec-log4j2.properties 文件中的值。	Not available.
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log4j2.properties 文件中的值。	Not available.
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。	Not available.
hive-site	更改 Hive 的 hive-site.xml 文件中的值	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.

分类	描述	重新配置操作
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值	Not available.
hue-ini	更改 Hue 的 ini 文件中的值	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	更改HTTPFS环境中的值。	Restarts Hadoop Httpfs service.
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。	Restarts Hadoop Httpfs service.
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。	Not available.
hadoop-kms-env	在 Hadoop KMS 环境中更改值。	Restarts Hadoop-KMS service.
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。	Not available.
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	更改 Hudi 环境中的值。	Not available.
hudi-defaults	更改 Hudi 的 hudi-defaults.conf 文件中的值。	Not available.
iceberg-defaults	更改 Iceberg 的 iceberg-defaults.conf 文件中的值。	Not available.
delta-defaults	更改 Delta 的 delta-defaults.conf 文件中的值。	Not available.

分类	描述	重新配置操作
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。	Not available.
jupyter-hub-conf	更改 JupyterHubs的 jupyterhub_config.py 文件中的值。	Not available.
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。	Not available.
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。	Not available.
livy-conf	更改 Livy 的 livy.conf 文件中的值。	Restarts Livy Server.
livy-env	更改 Livy 环境中的值。	Restarts Livy Server.
livy-log4j2	更改 Livy log4j2.properties 设置。	Restarts Livy Server.
mapred-env	更改 MapReduce 应用程序环境中的值。	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	更改 Oozie 的环境中的值。	Restarts Oozie.
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。	Restarts Oozie.
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。	Restarts Oozie.
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。	Not available.

分类	描述	重新配置操作
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。	Not available.
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。	Restarts Phoenix-QueryServer.
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。	Not available.
pig-env	更改 Pig 环境中的值。	Not available.
pig-properties	更改 Pig 的 pig.properties 文件中的值。	Restarts Oozie.
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。	Not available.
presto-log	更改 Presto 的 log.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-config	更改 Presto 的 config.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。	Not available.
presto-env	更改 Presto 的 presto-env.sh 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-node	更改 Presto 的 node.properties 文件中的值。	Not available.
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。	Not available.
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。	Not available.
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。	Not available.
presto-connector-lakeformation	更改 Presto 的 lakeformation.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。	Not available.
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。	Not available.
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。	Not available.
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。	Not available.
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。	Not available.
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。	Not available.
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。	Not available.
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。	Not available.
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。	Not available.
trino-log	更改 Trino 的 log.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-config	更改 Trino 的 config.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-password-authenticator	更改 Trino 的 password-authenticator.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-env	更改 Trino 的 trino-env.sh 文件中的值。	Restarts Trino-Server (for Trino)
trino-node	更改 Trino 的 node.properties 文件中的值。	Not available.
trino-connector-blackhole	更改 Trino 的 blackhole.properties 文件中的值。	Not available.
trino-connector-cassandra	更改 Trino 的 cassandra.properties 文件中的值。	Not available.
trino-connector-delta	更改 Trino 的 delta.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-hive	更改 Trino 的 hive.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-exchange-manager	更改 Trino 的 exchange-manager.properties 文件中的值。	Restarts Trino-Server (for Trino)

分类	描述	重新配置操作
trino-connector-iceberg	更改 Trino 的 iceberg.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-hudi	更改 Trino 的 hudi.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-jmx	更改 Trino 的 jmx.properties 文件中的值。	Not available.
trino-connector-kafka	更改 Trino 的 kafka.properties 文件中的值。	Not available.
trino-connector-localfile	更改 Trino 的 localfile.properties 文件中的值。	Not available.
trino-connector-memory	更改 Trino 的 memory.properties 文件中的值。	Not available.
trino-connector-mongodb	更改 Trino 的 mongodb.properties 文件中的值。	Not available.
trino-connector-mysql	更改 Trino 的 mysql.properties 文件中的值。	Not available.
trino-connector-postgresql	更改 Trino 的 postgresql.properties 文件中的值。	Not available.
trino-connector-raptor	更改 Trino 的 raptor.properties 文件中的值。	Not available.
trino-connector-redis	更改 Trino 的 redis.properties 文件中的值。	Not available.
trino-connector-redshift	更改 Trino 的 redshift.properties 文件中的值。	Not available.
trino-connector-tpch	更改 Trino 的 tpch.properties 文件中的值。	Not available.



分类	描述	重新配置操作
trino-connector-tpcds	更改 Trino 的 tpcds.properties 文件中的值。	Not available.
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-site	更改 Ranger KMS 的 ranger-kms-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-env	在 Ranger KMS 环境中更改值。	Restarts Ranger KMS Server.
ranger-kms-logback	更改 Ranger KMS 的 kms-logback.xml 文件中的值。	Not available.
ranger-kms-db-ca	更改 S3 上我与 Ranger SQL SSL KMS 连接的 CA 文件值。	Not available.
spark	亚马逊EMR精心策划的 Apache Spark 设置。	This property modifies spark-defaults. See actions there.
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-env	更改 Spark 环境中的值。	Restarts Spark history server and Spark thrift server.
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值	Not available.
spark-log4j2	更改 Spark 的 log4j2.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
sqoop-env	更改 Sqoop 的环境中的值。	Not available.

分类	描述	重新配置操作
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。	Not available.
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。	Not available.
tez-site	更改 Tez 的 tez-site.xml 文件中的值。	Restart Oozie and HiveServer2.
yarn-env	更改YARN环境中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	更改 yarn-site.xml 文件YARN 中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	更改 Zeppelin 环境中的值。	Restarts Zeppelin.
zeppelin-site	更改 zeppelin-site.xml 中的配置设置。	Restarts Zeppelin.
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。	Restarts Zookeeper server.
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。	Restarts Zookeeper server.

## 6.11.1 更改日志

发行版 6.11.1 的更改日志和发布说明

Date	事件	描述
2023-08-30	更新发行说明	在发行说明中添加了几个与控制面板相关的修复
2023-08-21	文档发布	亚马逊 EMR 6.11.1 发行说明首次发布
2023-08-16	部署完成	<a href="#">亚马逊 EMR 6.11.1 已全面部署到所有受支持的区域</a>
2023-08-04	初始版本	亚马逊 EMR 6.11.1 首次部署到有限的商业区域

## 亚马逊EMR版本 6.11.0

### 6.11.0 应用程序版本

此发行版支持以下应用程序

序：[Delta](#)、[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[Iceberg](#)、[JupyterEnterpriseC](#)和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

## 应用程序版本信息

	emr-6.11.0	emr-6.10.1	emr-6.10.0	emr-6.9.1
适用于 Java 的 AWS SDK	1.12.446	1.12.397	1.12.397	1.12.170
Python	2.7、3.7	2.7、3.7	2.7、3.7	2.7、3.7
Scala	2.12.15	2.12.15	2.12.15	2.12.15
AmazonCloudWatchAgent	-	-	-	-
Delta	2.2.0	2.2.0	2.2.0	2.1.0
Flink	1.16.0	1.16.0	1.16.0	1.15.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.15	2.4.15	2.4.15	2.4.13
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.3.3	3.3.3	3.3.3
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.13.0-amzn-0	0.12.2-amzn-0	0.12.2-amzn-0	0.12.1-amzn-0
Hue	4.11.0	4.10.0	4.10.0	4.10.0
Iceberg	1.2.0-amzn-0	1.1.0-amzn-0	1.1.0-amzn-0	0.14.1-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0	2.6.0
JupyterHub	1.4.1	1.5.0	1.5.0	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1

	emr-6.11.0	emr-6.10.1	emr-6.10.0	emr-6.9.1
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.279	0.278	0.278	0.276
Spark	3.3.2	3.3.1	3.3.1	3.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.11.0	2.10.0
Tez	0.10.2	0.10.2	0.10.2	0.10.2
Trino (PrestoSQL)	410	403	403	398
Zeppelin	0.10.1	0.10.1	0.10.1	0.10.1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

## 6.11.0 发布说明

以下发行说明包含有关亚马逊 6.11.0 EMR 版本的信息。更改与 6.10.0 有关。有关发布时间表的信息，请参阅[更改日志](#)。

### 新功能

- 亚马逊 EMR 6.11.0 支持 Apache Spark 3.3.2-amzn-0、Apache Spark 23.02.0-amzn-0、11 CUDA .8.0、Apache Hudi 0.13.0-amzn-0、Apache Iceberg RAPIDS 1.2.0-amzn-0、Trino 410-amzn-0 和 PrestoDB 0.279-amzn-0。

## 更改、增强功能和解决的问题

- 在亚马逊 EMR 6.11.0 中，DynamoDB 连接器已升级到 5.0.0 版。5.0.0 版本使用。AWS SDK for Java 2.x 之前的版本使用的是 AWS SDK for Java 1.x。由于此次升级，我们强烈建议您在将 DynamoDB 连接器与 Amazon 6.11 配合使用之前，先测试您的代码。EMR
- EMR 亚马逊 6.11.0 的 DynamoDB 连接器调用 DynamoDB 服务时，它将使用您为该属性提供的区域值。dynamodb.endpoint 我们建议您在 `dynamodb.endpoint` 时也配置 `dynamodb.region`，并且两个属性都以相同的 AWS 区域为目标。如果您使用 `dynamodb.endpoint` 但未进行配置 `dynamodb.region`，则适用于 Amazon 6.11.0 的 DynamoDB 连接器将返回无效的区域异常，并尝试协调来自 EC2 亚马逊实例元 AWS 区域数据服务的信息 (IMDS)。如果连接器无法从中检索区域 IMDS，则默认为美国东部 (弗吉尼亚北部) (`us-east-1`)。以下错误是您未正确配置该 `dynamodb.region` 属性时可能会遇到的无效区域异常的示例：

```
software.amazon.awssdk.services.dynamodb.model.DynamoDbException: Credential should be scoped to a valid region.
```

有关受 AWS SDK for Java 升级到 2.x 影响的类的更多信息，请参阅 Amazon-EMR DynamoDB 连接器存储库中的 [AWS SDK for Java 从 1.x 升级到 2.x \(#175\)](#) 提交。GitHub
- 此版本修复了在执行列重命名操作后使用 Delta Lake 在 Amazon S3 中存储 Delta 表数据时列数据变为 NULL 的问题。有关 Delta Lake 中此实验性功能的更多信息，请参阅《Delta Lake User Guide》中的 [Column rename operation](#)。
- 6.11.0 版本修复了通过从具有多个主节点的集群中复制一个主节点来创建边缘节点时可能出现的问题。复制的边缘节点可能会导致缩减操作的延迟，或者导致主节点的内存使用率过高。有关如何创建边缘节点以及与 EMR 集群通信的更多信息，请参阅存储库中的 [Edge Node Create](#) `aws-samples` or. GitHub
- 6.11.0 版本改进了亚马逊用于在重启后 EMR 将 Amazon 卷 EBS 重新安装到实例的自动化流程。
- 6.11.0 版本修复了导致亚马逊向亚马逊 EMR 发布的 Hadoop 指标间歇性出现差距的问题。CloudWatch
- 6.11.0 版本修复了 EMR 群集中的一个问题，即由于磁盘过度使用而导致对包含群集节点排除列表的 YARN 配置文件的更新中断。不完整的更新阻碍了未来对集群的缩减操作。此版本可确保您的集群保持正常运行，并确保扩展操作按预期进行。
- 在 Amazon EMR 6.10.0 及更高版本中，默认根卷大小已增加到 15 GB。早期版本的默认根卷大小为 10 GB。
- Hadoop 3.3.3 在 YARN ([YARN-9608](#)) 中引入了一项更改，即在应用程序完成之前，容器运行的节点一直处于停用状态。此更改可确保如随机数据等本地数据不会丢失，并且您无需重新运行作业。无论是否启用托管扩展，这种方法还可能导致集群的资源利用不足。

在亚马逊EMR版本 6.11.0 及更高版本以及 6.8.1、6.9.1 和 6.10.1 中，将的值设置 `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` 为 `in` 以 `false` 解决此问题。 `yarn-site.xml`

虽然该修复程序解决了 YARN -9608 引入的问题，但由于启用了托管扩展的集群上的洗牌数据丢失，它可能会导致 Hive 作业失败。在此版本中，我们还通过设置 `Hive yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-shuffle-data` 工作负载来降低这种风险。此配置仅在 Amazon 6.11.0 及更高EMR版本中可用。

- 当您启动带有最新补丁版本（亚马逊 EMR 5.36 或更高版本、6.6 或更高版本、7.0 或更高版本）的集群时，亚马逊将EMR使用最新的亚马逊 Linux 2023 或亚马逊 Linux 2 版本作为默认亚马逊 Linux 2 版本。EMR AMI有关更多信息，请参阅[使用亚马逊的默认 Amazon Linux AMI EMR](#)。

### Note

此版本不再获得自动AMI更新，因为又有 1 个补丁版本接替了该版本。补丁版本以第二位小数点后的数字 (6.8.1) 表示。要查看您是否使用的是最新补丁版本，请查看[发布指南](#)中的可用版本，或者在控制台中创建集群时查看 Amazon EMR 版本下拉列表，或者使用[ListReleaseLabels](#)API或[list-release-labels](#)CLI操作。要获取有关新版本的更新，请订阅“[新增内容?](#)”RSS 页面。

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 709.1	4.14.348	2024年7月23日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			兰)、欧洲(伦敦)、欧洲(巴黎)、亚太地区(香港)、亚太地区(孟买)、亚太地区(东京)、亚太地区(首尔)、亚太地区(大阪)、亚太地区(新加坡)、亚太地区(新加坡)、亚太地区(新加坡)、亚太地区(孟买)、亚太地区(东京)、亚太地区(首尔)、亚太地区(大阪)、亚太地区(悉尼)、亚太地区(雅加达)、非洲(开普敦)、南美洲(圣保罗)、中东(巴林)、加拿大(中部)、AWS GovCloud(美国西部)、AWS GovCloud(美国东部)、(美国东部)、中国(北京)、中国(宁夏)、亚洲太平洋(海得拉巴)、中东(UAE)、欧洲(西班牙)、欧洲(苏黎世)、亚太地区(墨尔本)、以色列(特拉维夫)、加拿大西部(卡尔加里)



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.20230808.0	4.14.320	2023 年 8 月 24 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 米兰 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 727.0	4.14.320	2023 年 8 月 14 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 719.0	4.14.320	2023 年 8 月 2 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 628.0	4.14.318	2023 年 7 月 12 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、加拿大 ( 中部 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( 巴林 )、中东 ( UAE )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 612.0	4.14.314	2023 年 6 月 23 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、加拿大 (中部)、欧洲 (斯德哥尔摩)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (米兰)、欧洲 (西班牙)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 (巴林)、中东 (UAE)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 504.1	4.14.313	2023 年 5 月 16 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 (巴林)、中东 (UAE)、加拿大 (中部)

### 6.11.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon

最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.4.2	亚马逊 SageMaker Spark SDK
delta	2.2.0	Delta lake 是一种适用于超大型分析数据集的开放表格式。
delta-standalone-connectors	0.6.0	Delta Connectors 提供不同的运行时，将 Delta Lake 与 Flink、Hive 和 Presto 等引擎集成。
emr-ddb	5.1.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	3.4.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.8.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-notebook-env	1.7.0	适用于 EMR Notebooks ( 可提供 jupyter 企业网关 ) 的 Conda env
emr-s3-dist-cp	2.25.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	2.4.0	EMRS3Select 连接器

组件	版本	描述
emr-wal-cli	1.1.0	用于 emrwal 列表/删除的 cli。
emrfs	2.56.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.16.0	Apache Flink 命令行客户端脚本和应用程序。
flink-jobmanager-config	1.16.0	管理 Apache Flink EMR JobManager 节点上的资源。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	3.3.3-amzn-3	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	3.3.3-amzn-3	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	3.3.3-amzn-3	HDFS命令行客户端和库
hadoop-hdfs-namenode	3.3.3-amzn-3	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	3.3.3-amzn-3	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。



组件	版本	描述
hadoop-httfs-server	3.3.3-amzn-3	HTTPHDFS操作端点。
hadoop-kms-server	3.3.3-amzn-3	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	3.3.3-amzn-3	MapReduce 用于运行 MapReduce 应用程序的执行引擎。
hadoop-yarn-nodemanager	3.3.3-amzn-3	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	3.3.3-amzn-3	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	3.3.3-amzn-3	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	2.4.15-amzn-1	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	2.4.15-amzn-1	为一个或多个HBase地区提供服务的服务。
hbase-client	2.4.15-amzn-1	HBase命令行客户端。
hbase-rest-server	2.4.15-amzn-1	为其提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	2.4.15-amzn-1	为其提供 Thrift 端点的HBase 服务。
hbase-operator-tools	2.4.15-amzn-1	适用于 Apache HBase 集群的修复工具。
hcatalog-client	3.1.3-amzn-4	用于操作 hcatalog-server 的“hcat”命令行客户端。

组件	版本	描述
hcatalog-server	3.1.3-amzn-4	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	3.1.3-amzn-4	HTTP端点为其提供REST接口HCatalog。
hive-client	3.1.3-amzn-4	Hive 命令行客户端。
hive-hbase	3.1.3-amzn-4	Hive-hbase 客户端。
hive-metastore-server	3.1.3-amzn-4	用于访问 Hive 元数据仓库的服务，这是一个语义存储库，用于存储 Hadoop 操作SQL的元数据。
hive-server2	3.1.3-amzn-4	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.13.0-amzn-0	增量处理框架，以支持低延迟和高效率的数据管道。
hudi-presto	0.13.0-amzn-0	用于运行 Presto 以及 Hudi 的捆绑库。
hudi-trino	0.13.0-amzn-0	用于运行 Trino 以及 Hudi 的捆绑库。
hudi-spark	0.13.0-amzn-0	用于运行 Spark 以及 Hudi 的捆绑库。
hue-server	4.11.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序

组件	版本	描述
iceberg	1.2.0-amzn-0	Apache Iceberg 是一种适用于超大型分析数据集的开放表格式。
jupyterhub	1.4.1	Jupyter notebook 的多用户服务器
livy-server	0.7.1-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mxnet	1.9.1	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.68+	MariaDB 数据库服务器。
nvidia-cuda	11.8.0	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.1	Oozie 命令行客户端。
oozie-server	5.2.1	用于接受 Oozie 工作流请求的服务。
opencv	4.5.0	开源计算机视觉库。
phoenix-library	5.1.2	服务器和客户端的 phoenix 库
phoenix-connectors	5.1.2	Apache Phoenix-Connectors for Spark-3
phoenix-query-server	5.1.2	轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限 API

组件	版本	描述
presto-coordinator	0.279-amzn-0	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.279-amzn-0	用于执行查询的各个部分的服务。
presto-client	0.279-amzn-0	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
trino-coordinator	410-amzn-0	用于在 trino-worker 之中接受查询并管理查询的服务。
trino-worker	410-amzn-0	用于执行查询的各个部分的服务。
trino-client	410-amzn-0	Trino 命令行客户端，安装在 HA 集群的备用主服务器（未启动 Trino 服务器）上。
pig-client	0.17.0	Pig 命令行客户端。
r	4.0.2	用于统计计算的 R 项目
ranger-kms-server	2.0.0	Apache Ranger 密钥管理系统
spark-client	3.3.2-amzn-0	Spark 命令行客户端。
spark-history-server	3.3.2-amzn-0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	3.3.2-amzn-0	的内存中执行引擎。YARN
spark-yarn-slave	3.3.2-amzn-0	YARN从属服务器需要的 Apache Spark 库。

组件	版本	描述
spark-rapids	23.02.0-amzn-0	Nvidia Spark RAPIDS 插件可以加速 Apache Spark。GPUs
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	2.11.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.10.2-amzn-2	tez YARN 应用程序和库。
tez-on-worker	0.10.2-amzn-2	工作节点的 tez YARN 应用程序和库。
webserver	2.4.41+	Apache HTTP 服务器。
zeppelin-server	0.10.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.5.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.5.10	ZooKeeper 命令行客户端。

### 6.11.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

为处于运行状态的集群中的实例组指定配置时，将发生重新配置操作。Amazon EMR 仅对您修改的分类启动重新配置操作。有关更多信息，请参阅 [在正在运行的集群中重新配置实例组](#)。

## emr-6.11.0 分类

分类	描述	重新配置操作
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。	Restarts the Resource Manager service.
container-executor	更改 Hadoop YARN 的 container-executor.cfg 文件中的值。	Not available.
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。	Not available.
core-site	更改 Hadoop 的 core-site.xml 文件中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	更改 docker 相关设置。	Not available.
emrfs-site	更改EMRFS设置。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager,

分类	描述	重新配置操作
		NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	更改 flink-conf.yaml 设置。	Restarts Flink history server.
flink-log4j	更改 Flink log4j.properties 设置。	Restarts Flink history server.
flink-log4j-session	为 Kubernetes/Yarn 会话更改 Flink log4j-session.properties 设置。	Restarts Flink history server.
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。	Restarts Flink history server.
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.

分类	描述	重新配置操作
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	更改 hadoop ssl 服务器配置	Not available.
hadoop-ssl-client	更改 hadoop ssl 客户端配置	Not available.
hbase	亚马逊EMR精心策划的 Apache 设置。HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	更改环境HBase中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	更改 hadoop-metrics2-hbase base.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。	Not available.



分类	描述	重新配置操作
hbase-site	更改HBase的 hbase-site.xml 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	配置HDFS加密区域。	This classification should not be reconfigured.
hdfs-env	更改HDFS环境中的值。	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	更改环境HCatalog中的值。	Restarts Hive HCatalog Server.
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。	Restarts Hive WebHCat server.

分类	描述	重新配置操作
hcatalog-webhcat-site	更改 HCatalog WebHCat 的 webhcat-site.xml 文件中的值。	Restarts Hive WebHCat server.
hive	亚马逊EMR精心策划的 Apache Hive 设置。	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。	Not available.
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。	Not available.
hive-env	更改 Hive 环境中的值。	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	更改 Hive 的 hive-exec-log4j2.properties 文件中的值。	Not available.
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log4j2.properties 文件中的值。	Not available.
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。	Not available.
hive-site	更改 Hive 的 hive-site.xml 文件中的值	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.

分类	描述	重新配置操作
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值	Not available.
hue-ini	更改 Hue 的 ini 文件中的值	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	更改HTTPFS环境中的值。	Restarts Hadoop Httpfs service.
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。	Restarts Hadoop Httpfs service.
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。	Not available.
hadoop-kms-env	在 Hadoop KMS 环境中更改值。	Restarts Hadoop-KMS service.
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。	Not available.
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	更改 Hudi 环境中的值。	Not available.
hudi-defaults	更改 Hudi 的 hudi-defaults.conf 文件中的值。	Not available.
iceberg-defaults	更改 Iceberg 的 iceberg-defaults.conf 文件中的值。	Not available.
delta-defaults	更改 Delta 的 delta-defaults.conf 文件中的值。	Not available.

分类	描述	重新配置操作
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。	Not available.
jupyter-hub-conf	更改 JupyterHubs的 jupyterhub_config.py 文件中的值。	Not available.
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。	Not available.
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。	Not available.
livy-conf	更改 Livy 的 livy.conf 文件中的值。	Restarts Livy Server.
livy-env	更改 Livy 环境中的值。	Restarts Livy Server.
livy-log4j2	更改 Livy log4j2.properties 设置。	Restarts Livy Server.
mapred-env	更改 MapReduce 应用程序环境中的值。	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	更改 Oozie 的环境中的值。	Restarts Oozie.
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。	Restarts Oozie.
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。	Restarts Oozie.
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。	Not available.

分类	描述	重新配置操作
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。	Not available.
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。	Restarts Phoenix-QueryServer.
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。	Not available.
pig-env	更改 Pig 环境中的值。	Not available.
pig-properties	更改 Pig 的 pig.properties 文件中的值。	Restarts Oozie.
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。	Not available.
presto-log	更改 Presto 的 log.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-config	更改 Presto 的 config.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。	Not available.
presto-env	更改 Presto 的 presto-env.sh 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-node	更改 Presto 的 node.properties 文件中的值。	Not available.
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。	Not available.
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。	Not available.
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。	Not available.
presto-connector-lakeformation	更改 Presto 的 lakeformation.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。	Not available.
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。	Not available.
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。	Not available.
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。	Not available.
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。	Not available.
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。	Not available.
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。	Not available.
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。	Not available.
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。	Not available.
trino-log	更改 Trino 的 log.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-config	更改 Trino 的 config.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-password-authenticator	更改 Trino 的 password-authenticator.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-env	更改 Trino 的 trino-env.sh 文件中的值。	Restarts Trino-Server (for Trino)
trino-node	更改 Trino 的 node.properties 文件中的值。	Not available.
trino-connector-blackhole	更改 Trino 的 blackhole.properties 文件中的值。	Not available.
trino-connector-cassandra	更改 Trino 的 cassandra.properties 文件中的值。	Not available.
trino-connector-delta	更改 Trino 的 delta.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-hive	更改 Trino 的 hive.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-exchange-manager	更改 Trino 的 exchange-manager.properties 文件中的值。	Restarts Trino-Server (for Trino)

分类	描述	重新配置操作
trino-connector-iceberg	更改 Trino 的 iceberg.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-hudi	更改 Trino 的 hudi.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-jmx	更改 Trino 的 jmx.properties 文件中的值。	Not available.
trino-connector-kafka	更改 Trino 的 kafka.properties 文件中的值。	Not available.
trino-connector-localfile	更改 Trino 的 localfile.properties 文件中的值。	Not available.
trino-connector-memory	更改 Trino 的 memory.properties 文件中的值。	Not available.
trino-connector-mongodb	更改 Trino 的 mongodb.properties 文件中的值。	Not available.
trino-connector-mysql	更改 Trino 的 mysql.properties 文件中的值。	Not available.
trino-connector-postgresql	更改 Trino 的 postgresql.properties 文件中的值。	Not available.
trino-connector-raptor	更改 Trino 的 raptor.properties 文件中的值。	Not available.
trino-connector-redis	更改 Trino 的 redis.properties 文件中的值。	Not available.
trino-connector-redshift	更改 Trino 的 redshift.properties 文件中的值。	Not available.
trino-connector-tpch	更改 Trino 的 tpch.properties 文件中的值。	Not available.



分类	描述	重新配置操作
trino-connector-tpcds	更改 Trino 的 tpcds.properties 文件中的值。	Not available.
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-site	更改 Ranger KMS 的 ranger-kms-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-env	在 Ranger KMS 环境中更改值。	Restarts Ranger KMS Server.
ranger-kms-logback	更改 Ranger KMS 的 kms-logback.xml 文件中的值。	Not available.
ranger-kms-db-ca	更改 S3 上我与 Ranger SQL SSL KMS 连接的 CA 文件值。	Not available.
spark	亚马逊EMR精心策划的 Apache Spark 设置。	This property modifies spark-defaults. See actions there.
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-env	更改 Spark 环境中的值。	Restarts Spark history server and Spark thrift server.
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值	Not available.
spark-log4j2	更改 Spark 的 log4j2.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
sqoop-env	更改 Sqoop 的环境中的值。	Not available.

分类	描述	重新配置操作
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。	Not available.
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。	Not available.
tez-site	更改 Tez 的 tez-site.xml 文件中的值。	Restart Oozie and HiveServer2.
yarn-env	更改YARN环境中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	更改 yarn-site.xml 文件YARN 中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	更改 Zeppelin 环境中的值。	Restarts Zeppelin.
zeppelin-site	更改 zeppelin-site.xml 中的配置设置。	Restarts Zeppelin.
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。	Restarts Zookeeper server.
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。	Restarts Zookeeper server.

## 6.11.0 更改日志

发行版 6.11.0 的更改日志和发布说明

Date	事件	描述
2023-08-21	更新	修复了 Hadoop 3.3.3 引入的问题。
2023-07-26	更新	新的操作系统版本标签 2.0.20230612.0 和 2.0.20230628.0 。
2023-06-09	部署完成	<a href="#">Amazon EMR 6.11.0 已完全部署到所有支持的区域</a>
2023-06-09	文档发布	亚马逊 EMR 6.11.0 发行说明首次发布
2023-06-08	初始版本	亚马逊 EMR 6.11.0 首次部署到初始商业区域

## 亚马逊EMR版本 6.10.1

### 6.10.1 应用程序版本

此发行版支持以下应用程序

序：[Delta](#)、[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[Iceberg](#)、[JupyterEnterpriseC](#)和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

## 应用程序版本信息

	emr-6.10.1	emr-6.10.0	emr-6.9.1	emr-6.9.0
适用于 Java 的 AWS SDK	1.12.397	1.12.397	1.12.170	1.12.170
Python	2.7、3.7	2.7、3.7	2.7、3.7	2.7、3.7
Scala	2.12.15	2.12.15	2.12.15	2.12.15
AmazonCloudWatchAgent	-	-	-	-
Delta	2.2.0	2.2.0	2.1.0	2.1.0
Flink	1.16.0	1.16.0	1.15.2	1.15.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.15	2.4.15	2.4.13	2.4.13
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.3.3	3.3.3	3.3.3
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.12.2-amzn-0	0.12.2-amzn-0	0.12.1-amzn-0	0.12.1-amzn-0
Hue	4.10.0	4.10.0	4.10.0	4.10.0
Iceberg	1.1.0-amzn-0	1.1.0-amzn-0	0.14.1-amzn-0	0.14.1-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0	2.6.0
JupyterHub	1.5.0	1.5.0	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1

	emr-6.10.1	emr-6.10.0	emr-6.9.1	emr-6.9.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.278	0.278	0.276	0.276
Spark	3.3.1	3.3.1	3.3.0	3.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.10.0	2.10.0
Tez	0.10.2	0.10.2	0.10.2	0.10.2
Trino (PrestoSQL)	403	403	398	398
Zeppelin	0.10.1	0.10.1	0.10.1	0.10.1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

## 6.10.1 发布说明

以下发行说明包含有关亚马逊 6.10.1 EMR 版本的信息。更改与 6.10.0 有关。有关发布时间表的更多信息，请参阅 [6.10.1 更改日志](#)。

### 更改、增强功能和解决的问题

- 由于锁争用，如果在尝试停用节点的同时添加或移除节点，则该节点可能会陷入死锁。结果，Hadoop 资源管理器 (YARN) 变得无响应，并会影响所有传入和当前正在运行的容器。
- Hadoop 3.3.3 在 YARN ([YARN-9608](#)) 中引入了一项更改，即在应用程序完成之前，容器运行的节点一直处于停用状态。此更改可确保如随机数据等本地数据不会丢失，并且您无需重新运行作业。无论是是否启用托管扩展，这种方法还可能导致集群的资源利用不足。

在亚马逊EMR版本 6.11.0 及更高版本以及 6.8.1、6.9.1 和 6.10.1 中，将的值设置 `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` 为 `in` 以 `false` 解决此问题。 `yarn-site.xml`

虽然该修复程序解决了 YARN -9608 引入的问题，但由于启用了托管扩展的集群上的洗牌数据丢失，它可能会导致 Hive 作业失败。在此版本中，我们还通过设置 `Hive yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-shuffle-data` 工作负载来降低这种风险。此配置仅在 Amazon 6.11.0 及更高EMR版本中可用。

- 使用实例组配置的集群中的主节点失效转移后，指标收集器不会向控制面板发送任何指标。
- 此版本包括一项更改，允许高可用性集群在重启后从故障状态中恢复。
- 此版本包含针对 Hue 和的安全补丁HBase。
- 此版本修复了在 Spark with Amazon 上运行工作负载的集群EMR可能会在静默中收到使用 `contains`、`startsWith`、`endsWith`、和 `like` 的错误结果的问题。当您对在 Amazon EMR Hive3 Metastore 服务器中包含元数据的分区字段使用表达式时，就会出现此问题 ()。HMS
- 此版本修复了当没有用户定义的函数时，Glue 端的限制问题 () UDF。
- 此版本修复了一个问题，即节点日志聚合服务在日志推送器可以将容器日志推送到 S3 之前删除容器日志，以防YARN停用。
- 此版本修复了 Hadoop FairShare `p` 启用节点标签时调度器指标的问题。
- 此版本修复了您在 `spark-defaults.conf` 中为 `spark.yarn.heterogeneousExecutors.enabled` 配置设置默认 `true` 值时影响 Spark 性能的问题。
- 此版本修复了 Reduce Task 无法读取随机数据的问题。该问题因内存损坏错误导致 Hive 查询失败。
- 此版本为运行 Presto 或 Trino 的集群的集群扩展工作流程添加了新的重试机制。EMR这一改进降低了由于单个调整大小操作失败而导致集群大小调整无限期停滞的风险。它还可以提高集群利用率，因为您的集群可以更快地向上和向下扩展。
- 此版本改进了集群扩容逻辑，因此您的集群不会尝试将核心节点缩小到低于集群HDFS重复因子设置的范围。这符合您的数据冗余要求，并减少了扩展操作可能停滞的机会。
- 日志管理进程守护程序已升级，可以识别本地实例存储中所有包含打开文件句柄的使用中的日志，以及相关的进程。此次升级可确保 Amazon 在日志存档到 Amazon S3 后EMR正确删除文件并回收存储空间。
- 此版本包括日志管理进程守护程序增强功能，可删除本地集群文件系统中空的、未使用的步骤目录。过多的空目录会降低 Amazon EMR 守护程序的性能并导致磁盘过度使用。

- 此版本修复了通过从具有多个主节点的集群中复制一个主节点来创建边缘节点时可能出现的问题。复制的边缘节点可能会导致缩减操作的延迟，或者导致主节点的内存使用率过高。有关如何创建边缘节点以及与EMR集群通信的更多信息，请参阅存储库中的 [Edge Node Create aws-samples](#) 或 [GitHub](#)。
- 此版本改进了 Amazon 用于在重启后EMR将 Amazon EBS 卷重新挂载到实例的自动化流程。
- 此版本修复了导致亚马逊向亚马逊EMR发布的 Hadoop 指标间歇性出现差距的问题。CloudWatch
- 此版本修复了EMR群集的一个问题，即由于磁盘过度使用而中断了对包含群集节点排除列表的YARN 配置文件的更新。不完整的更新阻碍了未来对集群的缩减操作。此版本可确保您的集群保持正常运行，并确保扩展操作按预期进行。
- 当您启动带有最新补丁版本（亚马逊 EMR 5.36 或更高版本、6.6 或更高版本、7.0 或更高版本）的集群时，亚马逊将EMR使用最新的亚马逊 Linux 2023 或亚马逊 Linux 2 版本作为默认亚马逊 Linux 2 版本。EMR AMI有关更多信息，请参阅[使用亚马逊的默认 Amazon Linux AMI EMR](#)。

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 709.1	4.14.348	2024年7月23日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、AWS GovCloud ( 美国东部 )、( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )、亚洲太平洋 ( 海得拉巴 )、中东 ( UAE )、欧洲 ( 西班牙 )、欧洲 ( 苏黎世 )、亚太地区 ( 墨尔本 )、以色列 ( 特拉维夫 )、加拿大西部 ( 卡尔加里 )



OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 223.0	4.14.336	2024 年 3 月 8 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、中国 ( 北

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 131.0	4.14.336	2024年2月14日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、中国 ( 北

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024124.0	4.14.336	2024 年 2 月 7 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、中国 (北

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 109.0	4.14.334	2024 年 1 月 24 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、中国 ( 北

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)



OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 218.0	4.14.330	2024 年 1 月 2 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 206.0	4.14.330	2023 年 12 月 22 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023116.0	4.14.328	2023 年 12 月 11 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 101.0	4.14.327	2023 年 11 月 16 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023020.1	4.14.326	2023 年 11 月 7 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023012.1	4.14.326	2023 年 10 月 26 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 926.0	4.14.322	2023 年 10 月 19 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 906.0	4.14.322	2023 年 10 月 4 日	(美国东部)、中国(北京)、中国(宁夏) 美国东部(弗吉尼亚北部)、美国东部(俄亥俄州)、美国西部(加利福尼亚北部)、美国西部(俄勒冈)、欧洲(斯德哥尔摩)、欧洲(米兰)、欧洲(西班牙)、欧洲(法兰克福)、欧洲(苏黎世)、欧洲(伦敦)、欧洲(巴黎)、亚太地区(香港)、亚太地区(孟买)、亚太地区(海得拉巴)、亚太地区(东京)、亚太地区(东京)、亚太地区(首尔)、亚太地区(大阪)、亚太地区(新加坡)、亚太地区(悉尼)、亚太地区(雅加达)、亚太地区(墨尔本)、非洲(开普敦)、南美洲(圣保罗)、中东(巴林)、中东( )、中东(UAE)、加拿大(中部)、以色列(特拉维夫)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.20230822.0	4.14.322	2023 年 8 月 30 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 808.0	4.14.320	2023 年 8 月 24 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)



OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 727.0	4.14.320	2023 年 8 月 14 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、欧洲地区 (斯德哥尔摩)、欧洲地区 (米兰)、欧洲地区 (法兰克福)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)

### 6.10.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.4.2	亚马逊 SageMaker Spark SDK
delta	2.2.0	Delta lake 是一种适用于超大型分析数据集的开放表格式。
emr-ddb	4.16.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	3.3.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.7.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-notebook-env	1.7.0	适用于 EMR Notebooks ( 可提供 jupyter 企业网关 ) 的 Conda env
emr-s3-dist-cp	2.24.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	2.3.0	EMRS3Select 连接器
emr-wal-cli	1.0.0	用于 emrwal 列表/删除的 cli。
emrfs	2.55.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.16.0	Apache Flink 命令行客户端脚本和应用程序。

组件	版本	描述
flink-jobmanager-config	1.16.0	管理 Apache Flink EMR JobManager 节点上的资源。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	3.3.3-amzn-2.1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	3.3.3-amzn-2.1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	3.3.3-amzn-2.1	HDFS命令行客户端和库
hadoop-hdfs-namenode	3.3.3-amzn-2.1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	3.3.3-amzn-2.1	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	3.3.3-amzn-2.1	HTTPHDFS操作端点。
hadoop-kms-server	3.3.3-amzn-2.1	基于 Hadoop 的加密密钥管理服务。 KeyProvider API

组件	版本	描述
hadoop-mapred	3.3.3-amzn-2.1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	3.3.3-amzn-2.1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	3.3.3-amzn-2.1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	3.3.3-amzn-2.1	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	2.4.15-amzn-0.1	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	2.4.15-amzn-0.1	为一个或多个HBase地区提供服务的服务。
hbase-client	2.4.15-amzn-0.1	HBase命令行客户端。
hbase-rest-server	2.4.15-amzn-0.1	为其提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	2.4.15-amzn-0.1	为其提供 Thrift 端点的HBase 服务。
hbase-operator-tools	2.4.15-amzn-0.1	适用于 Apache HBase 集群的修复工具。
hcatalog-client	3.1.3-amzn-3.1	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	3.1.3-amzn-3.1	提供服务HCatalog，为分布式应用程序提供表和存储管理层。

组件	版本	描述
hcatalog-webhcat-server	3.1.3-amzn-3.1	HTTP端点为其提供REST接口HCatalog。
hive-client	3.1.3-amzn-3.1	Hive 命令行客户端。
hive-hbase	3.1.3-amzn-3.1	Hive-hbase 客户端。
hive-metastore-server	3.1.3-amzn-3.1	用于访问 Hive 元数据仓库的服务，这是一个语义存储库，用于存储 Hadoop 操作SQL的元数据。
hive-server2	3.1.3-amzn-3.1	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.12.2-amzn-0	增量处理框架，以支持低延迟和高效率的数据管道。
hudi-presto	0.12.2-amzn-0	用于运行 Presto 以及 Hudi 的捆绑库。
hudi-trino	0.12.2-amzn-0	用于运行 Trino 以及 Hudi 的捆绑库。
hudi-spark	0.12.2-amzn-0	用于运行 Spark 以及 Hudi 的捆绑库。
hue-server	4.10.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
iceberg	1.1.0-amzn-0	Apache Iceberg 是一种适用于超大型分析数据集的开放表格式。
jupyterhub	1.5.0	Jupyter notebook 的多用户服务器

组件	版本	描述
livy-server	0.7.1-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mxnet	1.9.1	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.68+	MariaDB 数据库服务器。
nvidia-cuda	11.8.0	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.1	Oozie 命令行客户端。
oozie-server	5.2.1	用于接受 Oozie  workflow 请求的服务。
opencv	4.5.0	开源计算机视觉库。
phoenix-library	5.1.2	服务器和客户端的 phoenix 库
phoenix-connectors	6.0.0-SNAPSHOT	Apache Phoenix-Connectors for Spark-3
phoenix-query-server	6.0.0	轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限 API
presto-coordinator	0.278.1-amzn-0	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.278.1-amzn-0	用于执行查询的各个部分的服务。

组件	版本	描述
presto-client	0.278.1-amzn-0	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
trino-coordinator	403-amzn-0	用于在 trino-worker 之中接受查询并管理查询的服务。
trino-worker	403-amzn-0	用于执行查询的各个部分的服务。
trino-client	403-amzn-0	Trino 命令行客户端，安装在 HA 集群的备用主服务器（未启动 Trino 服务器）上。
pig-client	0.17.0	Pig 命令行客户端。
r	4.0.2	用于统计计算的 R 项目
ranger-kms-server	2.0.0	Apache Ranger 密钥管理系统
spark-client	3.3.1-amzn-0.1	Spark 命令行客户端。
spark-history-server	3.3.1-amzn-0.1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	3.3.1-amzn-0.1	的内存中执行引擎。YARN
spark-yarn-slave	3.3.1-amzn-0.1	YARN从属服务器需要的 Apache Spark 库。
spark-rapids	22.12.0-amzn-0	Nvidia Spark RAPIDS 插件可以加速 Apache Spark。GPUs
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。

组件	版本	描述
tensorflow	2.11.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.10.2-amzn-1.1	tez YARN 应用程序和库。
tez-on-worker	0.10.2-amzn-1.1	工作节点的 tez YARN 应用程序和库。
webserver	2.4.41+	Apache HTTP 服务器。
zeppelin-server	0.10.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.5.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.5.10	ZooKeeper 命令行客户端。

## 6.10.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

为处于运行状态的集群中的实例组指定配置时，将发生重新配置操作。Amazon EMR 仅对您修改的分类启动重新配置操作。有关更多信息，请参阅 [在正在运行的集群中重新配置实例组](#)。

### emr-6.10.1 分类

分类	描述	重新配置操作
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。	Restarts the ResourceManager service.
container-executor	更改 Hadoop YARN 的 container-executor.cfg 文件中的值。	Not available.



分类	描述	重新配置操作
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。	Not available.
core-site	更改 Hadoop 的 core-site.xml 文件中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	更改 docker 相关设置。	Not available.

分类	描述	重新配置操作
emrfs-site	更改EMRFS设置。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	更改 flink-conf.yaml 设置。	Restarts Flink history server.
flink-log4j	更改 Flink log4j.properties 设置。	Restarts Flink history server.
flink-log4j-session	为 Kubernetes/Yarn 会话更改 Flink log4j-session.properties 设置。	Restarts Flink history server.
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。	Restarts Flink history server.

分类	描述	重新配置操作
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	更改 hadoop ssl 服务器配置	Not available.
hadoop-ssl-client	更改 hadoop ssl 客户端配置	Not available.
hbase	亚马逊EMR精心策划的 Apache 设置。HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.

分类	描述	重新配置操作
hbase-env	更改环境HBase中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	更改 hadoop-metrics2-hbase base.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。	Not available.
hbase-site	更改HBase的 hbase-site.xml 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	配置HDFS加密区域。	This classification should not be reconfigured.
hdfs-env	更改HDFS环境中的值。	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	更改环境HCatalog中的值。	Restarts Hive HCatalog Server.

分类	描述	重新配置操作
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。	Restarts Hive WebHCat server.
hive	亚马逊EMR精心策划的 Apache Hive 设置。	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。	Not available.
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。	Not available.
hive-env	更改 Hive 环境中的值。	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	更改 Hive 的 hive-exec-log4j2.properties 文件中的值。	Not available.
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log4j2.properties 文件中的值。	Not available.

分类	描述	重新配置操作
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。	Not available.
hive-site	更改 Hive 的 hive-site.xml 文件中的值	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值	Not available.
hue-ini	更改 Hue 的 ini 文件中的值	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	更改HTTPFS环境中的值。	Restarts Hadoop Httpfs service.
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。	Restarts Hadoop Httpfs service.
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。	Not available.
hadoop-kms-env	在 Hadoop KMS 环境中更改值。	Restarts Hadoop-KMS service.
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。	Not available.
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。	Restarts Hadoop-KMS and Ranger-KMS service.

分类	描述	重新配置操作
hudi-env	更改 Hudi 环境中的值。	Not available.
hudi-defaults	更改 Hudi 的 hudi-defaults.conf 文件中的值。	Not available.
iceberg-defaults	更改 Iceberg 的 iceberg-defaults.conf 文件中的值。	Not available.
delta-defaults	更改 Delta 的 delta-defaults.conf 文件中的值。	Not available.
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。	Not available.
jupyter-hub-conf	更改 JupyterHubs 的 jupyterhub_config.py 文件中的值。	Not available.
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。	Not available.
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。	Not available.
livy-conf	更改 Livy 的 livy.conf 文件中的值。	Restarts Livy Server.
livy-env	更改 Livy 环境中的值。	Restarts Livy Server.
livy-log4j2	更改 Livy log4j2.properties 设置。	Restarts Livy Server.
mapred-env	更改 MapReduce 应用程序环境中的值。	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。	Restarts Hadoop MapReduce-HistoryServer.

分类	描述	重新配置操作
oozie-env	更改 Oozie 的环境中的值。	Restarts Oozie.
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。	Restarts Oozie.
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。	Restarts Oozie.
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。	Not available.
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。	Not available.
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。	Restarts Phoenix-QueryServer.
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。	Not available.
pig-env	更改 Pig 环境中的值。	Not available.
pig-properties	更改 Pig 的 pig.properties 文件中的值。	Restarts Oozie.
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。	Not available.
presto-log	更改 Presto 的 log.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-config	更改 Presto 的 config.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)



分类	描述	重新配置操作
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。	Not available.
presto-env	更改 Presto 的 presto-env.sh 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-node	更改 Presto 的 node.properties 文件中的值。	Not available.
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。	Not available.
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。	Not available.
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。	Not available.
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。	Not available.
presto-connector-lakeformation	更改 Presto 的 lakeformation.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。	Not available.
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。	Not available.
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。	Not available.
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。	Not available.
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。	Not available.
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。	Not available.
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。	Not available.
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。	Not available.
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。	Not available.
trino-log	更改 Trino 的 log.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-config	更改 Trino 的 config.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-password-authenticator	更改 Trino 的 password-authenticator.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-env	更改 Trino 的 trino-env.sh 文件中的值。	Restarts Trino-Server (for Trino)
trino-node	更改 Trino 的 node.properties 文件中的值。	Not available.

分类	描述	重新配置操作
trino-connector-blackhole	更改 Trino 的 blackhole.properties 文件中的值。	Not available.
trino-connector-cassandra	更改 Trino 的 cassandra.properties 文件中的值。	Not available.
trino-connector-delta	更改 Trino 的 delta.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-hive	更改 Trino 的 hive.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-exchange-manager	更改 Trino 的 exchange-manager.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-iceberg	更改 Trino 的 iceberg.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-hudi	更改 Trino 的 hudi.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-jmx	更改 Trino 的 jmx.properties 文件中的值。	Not available.
trino-connector-kafka	更改 Trino 的 kafka.properties 文件中的值。	Not available.
trino-connector-localfile	更改 Trino 的 localfile.properties 文件中的值。	Not available.
trino-connector-memory	更改 Trino 的 memory.properties 文件中的值。	Not available.
trino-connector-mongodb	更改 Trino 的 mongodb.properties 文件中的值。	Not available.

分类	描述	重新配置操作
trino-connector-mysql	更改 Trino 的 mysql.properties 文件中的值。	Not available.
trino-connector-postgresql	更改 Trino 的 postgresql.properties 文件中的值。	Not available.
trino-connector-raptor	更改 Trino 的 raptor.properties 文件中的值。	Not available.
trino-connector-redis	更改 Trino 的 redis.properties 文件中的值。	Not available.
trino-connector-redshift	更改 Trino 的 redshift.properties 文件中的值。	Not available.
trino-connector-tpch	更改 Trino 的 tpch.properties 文件中的值。	Not available.
trino-connector-tpcds	更改 Trino 的 tpcds.properties 文件中的值。	Not available.
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-site	更改 Ranger KMS 的 ranger-kms-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-env	在 Ranger KMS 环境中更改值。	Restarts Ranger KMS Server.
ranger-kms-logback	更改 Ranger KMS 的 kms-logback.xml 文件中的值。	Not available.
ranger-kms-db-ca	更改 S3 上我与 Ranger SQL SSL KMS 连接的 CA 文件值。	Not available.
spark	亚马逊EMR精心策划的 Apache Spark 设置。	This property modifies spark-defaults. See actions there.

分类	描述	重新配置操作
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-env	更改 Spark 环境中的值。	Restarts Spark history server and Spark thrift server.
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值	Not available.
spark-log4j2	更改 Spark 的 log4j2.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
sqoop-env	更改 Sqoop 的环境中的值。	Not available.
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。	Not available.
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。	Not available.
tez-site	更改 Tez 的 tez-site.xml 文件中的值。	Restart Oozie and HiveServer2.
yarn-env	更改YARN环境中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.

分类	描述	重新配置操作
yarn-site	更改 yarn-site.xml 文件YARN 中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	更改 Zeppelin 环境中的值。	Restarts Zeppelin.
zeppelin-site	更改 zeppelin-site.xml 中的配置设置。	Restarts Zeppelin.
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。	Restarts Zookeeper server.
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。	Restarts Zookeeper server.

## 6.10.1 更改日志

发行版 6.10.1 的更改日志和发布说明

Date	事件	描述
2023-08-30	更新发行说明	在发行说明中添加了几个与控制面板相关的修复
2023-08-21	文档发布	亚马逊 EMR 6.10.1 发行说明首次发布
2023-08-16	部署完成	<a href="#">Amazon EMR 6.10.1 已全面部署到所有支持的区域</a>
2023-08-04	初始版本	亚马逊 EMR 6.10.1 首次部署到有限的商业区域

## 亚马逊EMR版本 6.10.0

### 6.10.0 应用程序版本

此发行版支持以下应用程序

序：[Delta](#)、[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[Iceberg](#)、[JupyterEnterpriseC](#)和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-6.10.0	emr-6.9.1	emr-6.9.0	emr-6.8.1
适用于 Java 的 AWS SDK	1.12.397	1.12.170	1.12.170	1.12.170
Python	2.7、3.7	2.7、3.7	2.7、3.7	2.7、3.7
Scala	2.12.15	2.12.15	2.12.15	2.12.15
AmazonCloudWatchAgent	-	-	-	-
Delta	2.2.0	2.1.0	2.1.0	-
Flink	1.16.0	1.15.2	1.15.2	1.15.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.15	2.4.13	2.4.13	2.4.12

	emr-6.10.0	emr-6.9.1	emr-6.9.0	emr-6.8.1
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.3.3	3.3.3	3.2.1
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.12.2-amzn-0	0.12.1-amzn-0	0.12.1-amzn-0	0.11.1-amzn-0
Hue	4.10.0	4.10.0	4.10.0	4.10.0
Iceberg	1.1.0-amzn-0	0.14.1-amzn-0	0.14.1-amzn-0	0.14.0-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0	2.1.0
JupyterHub	1.5.0	1.4.1	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.278	0.276	0.276	0.273
Spark	3.3.1	3.3.0	3.3.0	3.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.10.0	2.10.0	2.9.1
Tez	0.10.2	0.10.2	0.10.2	0.9.2



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

## 6.10.0 发布说明

以下发行说明包含有关亚马逊 6.10. EMR 0 版本的信息。更改与 6.9.0 有关。有关发布时间表的信息，请参阅[更改日志](#)。

### 新功能

- 亚马逊 EMR 6.10.0 支持 Apache Spark 3.1、Apache Spark RAPIDS 22.12.0、CUDA 11.8.0、Apache Hudi 0.12.2-amzn-0、Apache Iceberg 1.1.0-amzn-0、Trino 403 和 PrestoDB 0.278.1。
- 亚马逊 EMR 6.10.0 包含原生 Trino-Hudi 连接器，可提供对 Hudi 表中数据的读取权限。您可以使用 `trino-cli --catalog hudi` 激活连接器，并使用 `trino-connector-hudi` 配置连接器以满足您的要求。与 Amazon 的原生集成 EMR 意味着您不再需要使用 `trino-connector-hive` 来查询 Hudi 表。有关新连接器支持的配置列表，请参阅 Trino 文档的 [Hudi connector](#) 页面。
- 亚马逊 EMR 发布的 6.10.0 及更高版本支持 Apache Zeppelin 与 Apache Flink 集成。请参阅在 [Amazon EMR 中通过 Zeppelin 使用 Flink 作业](#) 了解更多信息。

### 已知问题

- Hadoop 3.3.3 在 YARN ([YARN-9608](#)) 中引入了一项更改，即在应用程序完成之前，容器运行的节点一直处于停用状态。此更改可确保如随机数据等本地数据不会丢失，并且您无需重新运行作业。无论是否启用托管扩展，这种方法还可能导致集群的资源利用不足。

要在 Amazon EMR 6.10.0 中解决此问题，您可以将的值设置为 `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications`。false `yarn-site.xml` 在亚马逊 EMR 版本 6.11.0 及更高版本以及 6.8.1、6.9.1 和 6.10.1 中，配置 false 默认设置为以解决此问题。

## 更改、增强功能和解决的问题

- 亚马逊 EMR 6.10.0 消除了对 [Apache Spark Amazon Redshift 集成的](#) 依赖，并自动将所需的 Spark-Redshift 相关的 jar 添加到 Spark:、和的 执行器类路径中。minimal-json.jar spark-redshift.jar spark-avro.jar RedshiftJDBC.jar
- 6.10.0 版本改进了集群上的日志管理守护程序，可以监视集群中的其他日志文件夹。EMR 这一改进最大限度地减少了磁盘过度使用情况。
- 6.10.0 版本在集群上日志管理进程守护程序停止后会自动重启该守护程序。这一改进降低了由于磁盘过度使用而导致节点出现运行状况不佳的风险。
- Amazon EMR 6.10.0 支持用于 EMRFS 用户映射的区域终端节点。
- 在 Amazon EMR 6.10.0 及更高版本中，默认根卷大小已增加到 15 GB。早期版本的默认根卷大小为 10 GB。
- 6.10.0 版本修复了当所有剩余的 Spark 执行器都在使用资源管理器的停用主机上时，导致 Spark 作业停滞的问题。YARN
- 在 Amazon EMR 6.0 到 6.9.x 中，具有动态分区和 BY 或 BY 子句 ORDER 的 INSERT 查询将始终有 SORT 两个缩减器。此问题是由 OSS 更改 [HIVE-20703](#) 引起的，该更改将动态排序分区优化置于基于成本的决策之下。如果您的工作负载不需要对动态分区进行排序，建议将 hive.optimize.sort.dynamic.partition.threshold 属性设置为 -1 以禁用新功能并获得计算正确的 Reducer 数量。此问题已在 OSS Hive 中作为 [HIVE-22269](#) 的一部分修复，并已在亚马逊 6.10.0 中修复。EMR
- 当您启动带有最新补丁版本（亚马逊 EMR 5.36 或更高版本、6.6 或更高版本、7.0 或更高版本）的集群时，亚马逊将 EMR 使用最新的亚马逊 Linux 2023 或亚马逊 Linux 2 版本作为默认亚马逊 Linux 2 版本。EMR AMI 有关更多信息，请参阅 [使用亚马逊的默认 Amazon Linux AMI EMR](#)。

### Note

此版本不再获得自动 AMI 更新，因为又有 1 个补丁版本接替了该版本。补丁版本以第二位小数点后的数字 (6.8.1) 表示。要查看您是否使用的是最新补丁版本，请查看 [发布指南](#) 中的可用版本，或者在控制台中创建集群时查看 Amazon EMR 版本下拉列表，或者使用 [ListReleaseLabels API](#) 或 [list-release-labels CLI](#) 操作。要获取有关新版本的更新，请订阅 [“新增内容？”](#) RSS 页面。

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 709.1	4.14.348	2024年7月23日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、AWS GovCloud ( 美国

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			东部 )、( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )、亚洲太平洋 ( 海得拉巴 )、中东 ( UAE )、欧洲 ( 西班牙 )、欧洲 ( 苏黎世 )、亚太地区 ( 墨尔本 )、以色列 ( 特拉维夫 )、加拿大西部 ( 卡尔加里 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.20230808.0	4.14.320	2023 年 8 月 24 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 米兰 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 727.0	4.14.320	2023 年 8 月 14 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 719.0	4.14.320	2023 年 8 月 2 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 628.0	4.14.318	2023 年 7 月 12 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、加拿大 (中部)、欧洲 (斯德哥尔摩)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (米兰)、欧洲 (西班牙)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 (巴林)、中东 (UAE)



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 612.0	4.14.314	2023 年 6 月 23 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、加拿大 ( 中部 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( 巴林 )、中东 ( UAE )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 504.1	4.14.313	2023 年 5 月 16 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、加拿大 (中部)、欧洲 (斯德哥尔摩)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (米兰)、欧洲 (西班牙)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 (巴林)、中东 (UAE)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 418.0	4.14.311	2023 年 5 月 3 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、加拿大 ( 中部 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( 巴林 )、中东 ( UAE )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 404.1	4.14.311	2023 年 4 月 18 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、加拿大 (中部)、欧洲 (斯德哥尔摩)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、欧洲 (法兰克福)、欧洲 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (首尔)、亚太地区 (首尔)、亚太地区 (首尔)、亚太地区 (首尔)、亚太地区 (首尔)、亚太地区 (香港)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 (UAE)
2.0.2023 404.0	4.14.311	2023 年 4 月 10 日	美国东部 (弗吉尼亚州北部)、欧洲地区 (巴黎)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 320.0	4.14.309	2023 年 3 月 30 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、加拿大 (中部)、欧洲 (斯德哥尔摩)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、欧洲 (法兰克福)、欧洲 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (首尔)、亚太地区 (首尔)、亚太地区 (首尔)、亚太地区 (首尔)、亚太地区 (首尔)、亚太地区 (香港)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 (UAE)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.202307.0	4.14.304	2023 年 2 月 22 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、加拿大 (中部)、欧洲 (斯德哥尔摩)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、欧洲 (法兰克福)、欧洲 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (首尔)、亚太地区 (首尔)、亚太地区 (首尔)、亚太地区 (首尔)、亚太地区 (香港)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 (UAE)

### 6.10.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon

最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.4.2	亚马逊 SageMaker Spark SDK
delta	2.2.0	Delta lake 是一种适用于超大型分析数据集的开放表格式。
emr-ddb	4.16.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	3.3.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.7.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-notebook-env	1.7.0	适用于 EMR Notebooks ( 可提供 jupyter 企业网关 ) 的 Conda env
emr-s3-dist-cp	2.24.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	2.3.0	EMRS3Select 连接器
emr-wal-cli	1.0.0	用于 emrwal 列表/删除的 cli。
emrfs	2.55.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。

组件	版本	描述
flink-client	1.16.0	Apache Flink 命令行客户端脚本和应用程序。
flink-jobmanager-config	1.16.0	管理 Apache Flink EMR JobManager 节点上的资源。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	3.3.3-amzn-2	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	3.3.3-amzn-2	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	3.3.3-amzn-2	HDFS命令行客户端和库
hadoop-hdfs-namenode	3.3.3-amzn-2	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	3.3.3-amzn-2	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	3.3.3-amzn-2	HTTPHDFS操作端点。
hadoop-kms-server	3.3.3-amzn-2	基于 Hadoop 的加密密钥管理服务。 KeyProvider API



组件	版本	描述
hadoop-mapred	3.3.3-amzn-2	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	3.3.3-amzn-2	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	3.3.3-amzn-2	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	3.3.3-amzn-2	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	2.4.15-amzn-0	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	2.4.15-amzn-0	为一个或多个HBase地区提供服务的服务。
hbase-client	2.4.15-amzn-0	HBase命令行客户端。
hbase-rest-server	2.4.15-amzn-0	为其提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	2.4.15-amzn-0	为其提供 Thrift 端点的HBase 服务。
hbase-operator-tools	2.4.15-amzn-0	适用于 Apache HBase 集群的修复工具。
hcatalog-client	3.1.3-amzn-3	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	3.1.3-amzn-3	提供服务HCatalog，为分布式应用程序提供表和存储管理层。

组件	版本	描述
hcatalog-webhcat-server	3.1.3-amzn-3	HTTP端点为其提供REST接口HCatalog。
hive-client	3.1.3-amzn-3	Hive 命令行客户端。
hive-hbase	3.1.3-amzn-3	Hive-hbase 客户端。
hive-metastore-server	3.1.3-amzn-3	用于访问 Hive 元数据仓库的服务，这是一个语义存储库，用于存储 Hadoop 操作SQL的元数据。
hive-server2	3.1.3-amzn-3	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.12.2-amzn-0	增量处理框架，以支持低延迟和高效率的数据管道。
hudi-presto	0.12.2-amzn-0	用于运行 Presto 以及 Hudi 的捆绑库。
hudi-trino	0.12.2-amzn-0	用于运行 Trino 以及 Hudi 的捆绑库。
hudi-spark	0.12.2-amzn-0	用于运行 Spark 以及 Hudi 的捆绑库。
hue-server	4.10.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
iceberg	1.1.0-amzn-0	Apache Iceberg 是一种适用于超大型分析数据集的开放表格式。
jupyterhub	1.5.0	Jupyter notebook 的多用户服务器

组件	版本	描述
livy-server	0.7.1-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mxnet	1.9.1	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.68+	MariaDB 数据库服务器。
nvidia-cuda	11.8.0	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.1	Oozie 命令行客户端。
oozie-server	5.2.1	用于接受 Oozie 工作流请求的服务。
opencv	4.5.0	开源计算机视觉库。
phoenix-library	5.1.2	服务器和客户端的 phoenix 库
phoenix-connectors	6.0.0-SNAPSHOT	Apache Phoenix-Connectors for Spark-3
phoenix-query-server	6.0.0	轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限 API
presto-coordinator	0.278.1-amzn-0	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.278.1-amzn-0	用于执行查询的各个部分的服务。

组件	版本	描述
presto-client	0.278.1-amzn-0	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
trino-coordinator	403-amzn-0	用于在 trino-worker 之中接受查询并管理查询的服务。
trino-worker	403-amzn-0	用于执行查询的各个部分的服务。
trino-client	403-amzn-0	Trino 命令行客户端，安装在 HA 集群的备用主服务器（未启动 Trino 服务器）上。
pig-client	0.17.0	Pig 命令行客户端。
r	4.0.2	用于统计计算的 R 项目
ranger-kms-server	2.0.0	Apache Ranger 密钥管理系统
spark-client	3.3.1-amzn-0	Spark 命令行客户端。
spark-history-server	3.3.1-amzn-0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	3.3.1-amzn-0	的内存中执行引擎。YARN
spark-yarn-slave	3.3.1-amzn-0	YARN从属服务器需要的 Apache Spark 库。
spark-rapids	22.12.0-amzn-0	Nvidia Spark RAPIDS 插件可以加速 Apache Spark。GPUs
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。

组件	版本	描述
tensorflow	2.11.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.10.2-amzn-1	tez YARN 应用程序和库。
tez-on-worker	0.10.2-amzn-1	工作节点的 tez YARN 应用程序和库。
webserver	2.4.41+	Apache HTTP 服务器。
zeppelin-server	0.10.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.5.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.5.10	ZooKeeper 命令行客户端。

## 6.10.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

为处于运行状态的集群中的实例组指定配置时，将发生重新配置操作。Amazon EMR 仅对您修改的分类启动重新配置操作。有关更多信息，请参阅 [在正在运行的集群中重新配置实例组](#)。

### emr-6.10.0 分类

分类	描述	重新配置操作
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。	Restarts the ResourceManager service.
container-executor	更改 Hadoop YARN 的 container-executor.cfg 文件中的值。	Not available.

分类	描述	重新配置操作
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。	Not available.
core-site	更改 Hadoop 的 core-site.xml 文件中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	更改 docker 相关设置。	Not available.

分类	描述	重新配置操作
emrfs-site	更改EMRFS设置。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	更改 flink-conf.yaml 设置。	Restarts Flink history server.
flink-log4j	更改 Flink log4j.properties 设置。	Restarts Flink history server.
flink-log4j-session	为 Kubernetes/Yarn 会话更改 Flink log4j-session.properties 设置。	Restarts Flink history server.
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。	Restarts Flink history server.

分类	描述	重新配置操作
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	更改 hadoop ssl 服务器配置	Not available.
hadoop-ssl-client	更改 hadoop ssl 客户端配置	Not available.
hbase	亚马逊EMR精心策划的 Apache 设置。HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.



分类	描述	重新配置操作
hbase-env	更改环境HBase中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	更改 hadoop-metrics2-hbase base.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。	Not available.
hbase-site	更改HBase的 hbase-site.xml 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	配置HDFS加密区域。	This classification should not be reconfigured.
hdfs-env	更改HDFS环境中的值。	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	更改环境HCatalog中的值。	Restarts Hive HCatalog Server.

分类	描述	重新配置操作
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。	Restarts Hive WebHCat server.
hive	亚马逊EMR精心策划的 Apache Hive 设置。	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。	Not available.
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。	Not available.
hive-env	更改 Hive 环境中的值。	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	更改 Hive 的 hive-exec-log4j2.properties 文件中的值。	Not available.
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log4j2.properties 文件中的值。	Not available.

分类	描述	重新配置操作
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。	Not available.
hive-site	更改 Hive 的 hive-site.xml 文件中的值	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值	Not available.
hue-ini	更改 Hue 的 ini 文件中的值	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	更改HTTPFS环境中的值。	Restarts Hadoop Httpfs service.
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。	Restarts Hadoop Httpfs service.
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。	Not available.
hadoop-kms-env	在 Hadoop KMS 环境中更改值。	Restarts Hadoop-KMS service.
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。	Not available.
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。	Restarts Hadoop-KMS and Ranger-KMS service.

分类	描述	重新配置操作
hudi-env	更改 Hudi 环境中的值。	Not available.
hudi-defaults	更改 Hudi 的 hudi-defaults.conf 文件中的值。	Not available.
iceberg-defaults	更改 Iceberg 的 iceberg-defaults.conf 文件中的值。	Not available.
delta-defaults	更改 Delta 的 delta-defaults.conf 文件中的值。	Not available.
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。	Not available.
jupyter-hub-conf	更改 JupyterHubs 的 jupyterhub_config.py 文件中的值。	Not available.
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。	Not available.
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。	Not available.
livy-conf	更改 Livy 的 livy.conf 文件中的值。	Restarts Livy Server.
livy-env	更改 Livy 环境中的值。	Restarts Livy Server.
livy-log4j2	更改 Livy log4j2.properties 设置。	Restarts Livy Server.
mapred-env	更改 MapReduce 应用程序环境中的值。	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。	Restarts Hadoop MapReduce-HistoryServer.

分类	描述	重新配置操作
oozie-env	更改 Oozie 的环境中的值。	Restarts Oozie.
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。	Restarts Oozie.
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。	Restarts Oozie.
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。	Not available.
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。	Not available.
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。	Restarts Phoenix-QueryServer.
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。	Not available.
pig-env	更改 Pig 环境中的值。	Not available.
pig-properties	更改 Pig 的 pig.properties 文件中的值。	Restarts Oozie.
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。	Not available.
presto-log	更改 Presto 的 log.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-config	更改 Presto 的 config.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)

分类	描述	重新配置操作
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。	Not available.
presto-env	更改 Presto 的 presto-env.sh 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-node	更改 Presto 的 node.properties 文件中的值。	Not available.
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。	Not available.
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。	Not available.
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。	Not available.
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。	Not available.
presto-connector-lakeformation	更改 Presto 的 lakeformation.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。	Not available.
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。	Not available.
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。	Not available.
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。	Not available.
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。	Not available.
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。	Not available.
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。	Not available.
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。	Not available.
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。	Not available.
trino-log	更改 Trino 的 log.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-config	更改 Trino 的 config.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-password-authenticator	更改 Trino 的 password-authenticator.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-env	更改 Trino 的 trino-env.sh 文件中的值。	Restarts Trino-Server (for Trino)
trino-node	更改 Trino 的 node.properties 文件中的值。	Not available.

分类	描述	重新配置操作
trino-connector-blackhole	更改 Trino 的 blackhole.properties 文件中的值。	Not available.
trino-connector-cassandra	更改 Trino 的 cassandra.properties 文件中的值。	Not available.
trino-connector-delta	更改 Trino 的 delta.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-hive	更改 Trino 的 hive.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-exchange-manager	更改 Trino 的 exchange-manager.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-iceberg	更改 Trino 的 iceberg.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-hudi	更改 Trino 的 hudi.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-jmx	更改 Trino 的 jmx.properties 文件中的值。	Not available.
trino-connector-kafka	更改 Trino 的 kafka.properties 文件中的值。	Not available.
trino-connector-localfile	更改 Trino 的 localfile.properties 文件中的值。	Not available.
trino-connector-memory	更改 Trino 的 memory.properties 文件中的值。	Not available.
trino-connector-mongodb	更改 Trino 的 mongodb.properties 文件中的值。	Not available.



分类	描述	重新配置操作
trino-connector-mysql	更改 Trino 的 mysql.properties 文件中的值。	Not available.
trino-connector-postgresql	更改 Trino 的 postgresql.properties 文件中的值。	Not available.
trino-connector-raptor	更改 Trino 的 raptor.properties 文件中的值。	Not available.
trino-connector-redis	更改 Trino 的 redis.properties 文件中的值。	Not available.
trino-connector-redshift	更改 Trino 的 redshift.properties 文件中的值。	Not available.
trino-connector-tpch	更改 Trino 的 tpch.properties 文件中的值。	Not available.
trino-connector-tpcds	更改 Trino 的 tpcds.properties 文件中的值。	Not available.
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-site	更改 Ranger KMS 的 ranger-kms-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-env	在 Ranger KMS 环境中更改值。	Restarts Ranger KMS Server.
ranger-kms-logback	更改 Ranger KMS 的 kms-logback.xml 文件中的值。	Not available.
ranger-kms-db-ca	更改 S3 上我与 Ranger SQL SSL KMS 连接的 CA 文件值。	Not available.
spark	亚马逊EMR精心策划的 Apache Spark 设置。	This property modifies spark-defaults. See actions there.

分类	描述	重新配置操作
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-env	更改 Spark 环境中的值。	Restarts Spark history server and Spark thrift server.
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值	Not available.
spark-log4j2	更改 Spark 的 log4j2.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
sqoop-env	更改 Sqoop 的环境中的值。	Not available.
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。	Not available.
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。	Not available.
tez-site	更改 Tez 的 tez-site.xml 文件中的值。	Restart Oozie and HiveServer2.
yarn-env	更改YARN环境中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.

分类	描述	重新配置操作
yarn-site	更改 yarn-site.xml 文件YARN 中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	更改 Zeppelin 环境中的值。	Restarts Zeppelin.
zeppelin-site	更改 zeppelin-site.xml 中的配置设置。	Restarts Zeppelin.
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。	Restarts Zookeeper server.
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。	Restarts Zookeeper server.

## 6.10.0 更改日志

发行版 6.10.0 的更改日志和发布说明

Date	事件	描述
2023-08-21	更新	添加了 Hadoop 3.3.3 引入的一个已知问题。
2023-07-26	更新	新的操作系统版本标签 2.0.20230612.0 和 2.0.20230628.0。
2023-03-02	部署完成	Amazon EMR 6.10 已全面部署到所有 <a href="#">受支持的区域</a>
2023-03-02	文档发布	亚马逊 EMR 6.10 发行说明首次发布

Date	事件	描述
2023-02-27	初始版本	亚马逊 EMR 6.10 已部署到有限的商业区域

## 亚马逊EMR版本 6.9.1

### 6.9.1 应用程序版本

此发行版支持以下应用程序

序：[Delta](#)、[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[Iceberg](#)、[JupyterEnterpriseC](#)和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-6.9.1	emr-6.9.0	emr-6.8.1	emr-6.8.0
适用于 Java 的 AWS SDK	1.12.170	1.12.170	1.12.170	1.12.170
Python	2.7、3.7	2.7、3.7	2.7、3.7	2.7、3.7
Scala	2.12.15	2.12.15	2.12.15	2.12.15
AmazonCloudWatchAgent	-	-	-	-
Delta	2.1.0	2.1.0	-	-

	emr-6.9.1	emr-6.9.0	emr-6.8.1	emr-6.8.0
Flink	1.15.2	1.15.2	1.15.1	1.15.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.13	2.4.13	2.4.12	2.4.12
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.3.3	3.2.1	3.2.1
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.12.1-amzn-0	0.12.1-amzn-0	0.11.1-amzn-0	0.11.1-amzn-0
Hue	4.10.0	4.10.0	4.10.0	4.10.0
Iceberg	0.14.1-amzn-0	0.14.1-amzn-0	0.14.0-amzn-0	0.14.0-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.276	0.276	0.273	0.273
Spark	3.3.0	3.3.0	3.3.0	3.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7

	emr-6.9.1	emr-6.9.0	emr-6.8.1	emr-6.8.0
TensorFlow	2.10.0	2.10.0	2.9.1	2.9.1
Tez	0.10.2	0.10.2	0.9.2	0.9.2
Trino (PrestoSQL)	398	398	388	388
Zeppelin	0.10.1	0.10.1	0.10.1	0.10.1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

## 6.9.1 发布说明

以下发行说明包含 Amazon 6.9.1 EMR 版本的信息。更改与 6.9.0 有关。有关发布时间表的更多信息，请参阅 [6.9.1 更改日志](#)。

### 更改、增强功能和解决的问题

- Hadoop 3.3.3 在 YARN ([YARN-9608](#)) 中引入了一项更改，即在应用程序完成之前，容器运行的节点一直处于停用状态。此更改可确保如随机数据等本地数据不会丢失，并且您无需重新运行作业。无论是否启用托管扩展，这种方法还可能导致集群的资源利用不足。

在亚马逊EMR版本 6.11.0 及更高版本以及 6.8.1、6.9.1 和 6.10.1 中，将的值设置 `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` 为 `in` 以 `false` 解决此问题。 `yarn-site.xml`

虽然该修复程序解决了 YARN -9608 引入的问题，但由于启用了托管扩展的集群上的洗牌数据丢失，它可能会导致 Hive 作业失败。在此版本中，我们还通过设置 Hive `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-shuffle-data` 工作负载来降低这种风险。此配置仅在 Amazon 6.11.0 及更高EMR版本中可用。

- 使用实例组配置的集群中的主节点失效转移后，指标收集器不会向控制面板发送任何指标。
- 此版本消除了对指标收集器端点的HTTP请求失败时进行重试。
- 此版本包括一项更改，允许高可用性集群在重启后从故障状态中恢复。
- 此版本修复了用户创建的大量内容UIDs导致溢出异常的问题。
- 此版本修复了 Amazon EMR 重新配置过程中的超时问题。
- 此版本包含安全修复。

- 此版本修复了在 Spark with Amazon 上运行工作负载的集群EMR可能会在静默中收到使用contains、startsWithendsWith、和like的错误结果的问题。当您对在 Amazon EMR Hive3 Metastore 服务器中包含元数据的分区字段使用表达式时，就会出现此问题 ()。HMS
- 在 Amazon EMR 6.0 到 6.9.x 中，具有动态分区和 BY 或 BY 子句ORDER的INSERT查询将始终有SORT两个缩减器。此问题是由OSS更改 [HIVE-20703](#) 引起的，该更改将动态排序分区优化置于基于成本的决策之下。如果您的工作负载不需要对动态分区进行排序，建议将hive.optimize.sort.dynamic.partition.threshold 属性设置为 -1 以禁用新功能并获得计算正确的 Reducer 数量。此问题已在 OSS Hive 中作为 [HIVE-22269](#) 的一部分修复，并已在亚马逊 6.10.0 中修复。EMR
- 当您HDFS用作临时目录并启用合并小文件并且该表包含静态分区路径时，Hive 可能会丢失数据。
- 此版本修复了ETL任务结束时启用合并小文件 ( 默认禁用 ) 时 Hive 的性能问题。
- 此版本修复了当没有用户定义的函数时，Glue 端的限制问题 () UDF。
- 此版本修复了一个问题，即节点日志聚合服务在日志推送器可以将容器日志推送到 S3 之前删除容器日志，以防YARN停用。
- 此版本修复了使用永久存储文件跟踪功能对压缩/存档文件的处理。HBase
- 此版本修复了您在 spark-defaults.conf 中为 spark.yarn.heterogeneousExecutors.enabled 配置设置默认 true 值时影响 Spark 性能的问题。
- 此版本修复了 Reduce Task 无法读取随机数据的问题。该问题因内存损坏错误导致 Hive 查询失败。
- 此版本修复了在节点更换期间 HDFS NameNode (NN) 服务停留在安全模式下会导致节点配置器失败的问题。
- 此版本为运行 Presto 或 Trino 的集群的集群扩展工作流程添加了新的重试机制。EMR这一改进降低了由于单个调整大小操作失败而导致集群大小调整无限期停滞的风险。它还可以提高集群利用率，因为您的集群可以更快地向上和向下扩展。
- 此版本改进了集群扩容逻辑，因此您的集群不会尝试将核心节点缩小到低于集群HDFS重复因子设置的范围。这符合您的数据冗余要求，并减少了扩展操作可能停滞的机会。
- 日志管理进程守护程序已升级，可以识别本地实例存储中所有包含打开文件句柄的使用中的日志，以及相关的进程。此次升级可确保 Amazon 在日志存档到 Amazon S3 后EMR正确删除文件并回收存储空间。
- 此版本包括日志管理进程守护程序增强功能，可删除本地集群文件系统中空的、未使用的步骤目录。过多的空目录会降低 Amazon EMR 守护程序的性能并导致磁盘过度使用。
- 此版本修复了通过从具有多个主节点的集群中复制一个主节点来创建边缘节点时可能出现的问题。复制的边缘节点可能会导致缩减操作的延迟，或者导致主节点的内存使用率过高。有关如何创建边缘节点以及与EMR集群通信的更多信息，请参阅存储库中的 [Edge Node Creat](#) aws-samples or。GitHub

- 此版本改进了 Amazon 用于在重启后EMR将 Amazon EBS 卷重新挂载到实例的自动化流程。
- 此版本修复了导致亚马逊向亚马逊EMR发布的 Hadoop 指标间歇性出现差距的问题。 CloudWatch
- 此版本修复了EMR群集的一个问题，即由于磁盘过度使用而中断了对包含群集节点排除列表的YARN 配置文件的更新。不完整的更新阻碍了未来对集群的缩减操作。此版本可确保您的集群保持正常运行，并确保扩展操作按预期进行。
- 此版本改进了集群上的日志管理守护程序，可以监视集群中的其他日志文件夹。EMR这一改进最大限度地减少了磁盘过度使用情况。
- 此版本在集群上日志管理进程守护程序停止后会自动重启该守护程序。这一改进降低了由于磁盘过度使用而导致节点出现运行状况不佳的风险。
- 当您启动带有最新补丁版本（亚马逊 EMR 5.36 或更高版本、6.6 或更高版本、7.0 或更高版本）的集群时，亚马逊将EMR使用最新的亚马逊 Linux 2023 或亚马逊 Linux 2 版本作为默认亚马逊 Linux 2 版本。EMR AMI有关更多信息，请参阅[使用亚马逊的默认 Amazon Linux AMI EMR](#)。

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 709.1	4.14.348	2024年7月23日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、AWS GovCloud ( 美国东部 )、( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )、亚洲太平洋 ( 海得拉巴 )、中东 ( UAE )、欧洲 ( 西班牙 )、欧洲 ( 苏黎世 )、亚太地区 ( 墨尔本 )、以色列 ( 特拉维夫 )、加拿大西部 ( 卡尔加里 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024223.0	4.14.336	2024 年 3 月 8 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			GovCloud (美国东部)、中国(北京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 131.0	4.14.336	2024年2月14日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			GovCloud (美国东部)、中国(北京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024124.0	4.14.336	2024 年 2 月 7 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			GovCloud (美国东部)、中国(北京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 109.0	4.14.334	2024 年 1 月 24 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS



OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			GovCloud (美国东部)、中国(北京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 218.0	4.14.330	2024 年 1 月 2 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			GovCloud ( 美国东部 )、 中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 206.0	4.14.330	2023 年 12 月 22 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			GovCloud ( 美国东部 )、 中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 116.0	4.14.328	2023 年 12 月 11 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			GovCloud ( 美国东部 )、 中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 101.0	4.14.327	2023 年 11 月 16 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS



OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			GovCloud (美国东部)、中国(北京)、中国(宁夏)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023020.1	4.14.326	2023 年 11 月 7 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			GovCloud ( 美国东部 )、 中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023012.1	4.14.326	2023 年 10 月 26 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			GovCloud (美国东部)、中国(北京)、中国(宁夏)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 926.0	4.14.322	2023 年 10 月 19 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 4.14.322 906.0		2023 年 10 月 4 日	GovCloud (美国东部)、中国 (北京)、中国 (宁夏)  美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、欧洲地区 (斯德哥尔摩)、欧洲地区 (米兰)、欧洲地区 (法兰克福)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、加拿大 (中部)、以色列 (特拉维夫)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.20230822.0	4.14.322	2023 年 8 月 30 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、欧洲地区 (斯德哥尔摩)、欧洲地区 (米兰)、欧洲地区 (法兰克福)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、加拿大 (中部)、以色列 (特拉维夫)



OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 808.0	4.14.320	2023 年 8 月 24 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、欧洲地区 (斯德哥尔摩)、欧洲地区 (米兰)、欧洲地区 (法兰克福)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、加拿大 (中部)、以色列 (特拉维夫)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 727.0	4.14.320	2023 年 8 月 14 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、欧洲地区 (斯德哥尔摩)、欧洲地区 (米兰)、欧洲地区 (法兰克福)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)

### 6.9.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.4.2	亚马逊 SageMaker Spark SDK
delta	2.1.0	Delta lake 是一种适用于超大型分析数据集的开放表格式。
emr-ddb	4.16.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	3.3.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.6.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-notebook-env	1.7.0	适用于 EMR Notebooks ( 可提供 jupyter 企业网关 ) 的 Conda env
emr-s3-dist-cp	2.23.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	2.2.0	EMRS3Select 连接器
emrfs	2.54.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.15.2	Apache Flink 命令行客户端脚本和应用程序。

组件	版本	描述
flink-jobmanager-config	1.15.2	管理 Apache Flink EMR JobManager 节点上的资源。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	3.3.3-amzn-1.1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	3.3.3-amzn-1.1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	3.3.3-amzn-1.1	HDFS命令行客户端和库
hadoop-hdfs-namenode	3.3.3-amzn-1.1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	3.3.3-amzn-1.1	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	3.3.3-amzn-1.1	HTTPHDFS操作端点。
hadoop-kms-server	3.3.3-amzn-1.1	基于 Hadoop 的加密密钥管理服务。 KeyProvider API

组件	版本	描述
hadoop-mapred	3.3.3-amzn-1.1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	3.3.3-amzn-1.1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	3.3.3-amzn-1.1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	3.3.3-amzn-1.1	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	2.4.13-amzn-0.1	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	2.4.13-amzn-0.1	为一个或多个HBase地区提供服务的服务。
hbase-client	2.4.13-amzn-0.1	HBase命令行客户端。
hbase-rest-server	2.4.13-amzn-0.1	为其提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	2.4.13-amzn-0.1	为其提供 Thrift 端点的HBase 服务。
hbase-operator-tools	2.4.13-amzn-0.1	适用于 Apache HBase 集群的修复工具。
hcatalog-client	3.1.3-amzn-2.1	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	3.1.3-amzn-2.1	提供服务HCatalog，为分布式应用程序提供表和存储管理层。

组件	版本	描述
hcatalog-webhcat-server	3.1.3-amzn-2.1	HTTP端点为其提供REST接口HCatalog。
hive-client	3.1.3-amzn-2.1	Hive 命令行客户端。
hive-hbase	3.1.3-amzn-2.1	Hive-hbase 客户端。
hive-metastore-server	3.1.3-amzn-2.1	用于访问 Hive 元数据仓库的服务，这是一个语义存储库，用于存储 Hadoop 操作SQL的元数据。
hive-server2	3.1.3-amzn-2.1	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.12.1-amzn-0	增量处理框架，以支持低延迟和高效率的数据管道。
hudi-presto	0.12.1-amzn-0	用于运行 Presto 以及 Hudi 的捆绑库。
hudi-trino	0.12.1-amzn-0	用于运行 Trino 以及 Hudi 的捆绑库。
hudi-spark	0.12.1-amzn-0	用于运行 Spark 以及 Hudi 的捆绑库。
hue-server	4.10.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
iceberg	0.14.1-amzn-0	Apache Iceberg 是一种适用于超大型分析数据集的开放表格式。
jupyterhub	1.4.1	Jupyter notebook 的多用户服务器

组件	版本	描述
livy-server	0.7.1-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mxnet	1.9.1	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.68+	MariaDB 数据库服务器。
nvidia-cuda	11.7.0	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.1	Oozie 命令行客户端。
oozie-server	5.2.1	用于接受 Oozie  workflow 请求的服务。
opencv	4.5.0	开源计算机视觉库。
phoenix-library	5.1.2	服务器和客户端的 phoenix 库
phoenix-connectors	6.0.0-SNAPSHOT	Apache Phoenix-Connectors for Spark-3
phoenix-query-server	6.0.0	轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限 API
presto-coordinator	0.276-amzn-0	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.276-amzn-0	用于执行查询的各个部分的服务。

组件	版本	描述
presto-client	0.276-amzn-0	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
trino-coordinator	398-amzn-0	用于在 trino-worker 之中接受查询并管理查询的服务。
trino-worker	398-amzn-0	用于执行查询的各个部分的服务。
trino-client	398-amzn-0	Trino 命令行客户端，安装在 HA 集群的备用主服务器（未启动 Trino 服务器）上。
pig-client	0.17.0	Pig 命令行客户端。
r	4.0.2	用于统计计算的 R 项目
ranger-kms-server	2.0.0	Apache Ranger 密钥管理系统
spark-client	3.3.0-amzn-1.1	Spark 命令行客户端。
spark-history-server	3.3.0-amzn-1.1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	3.3.0-amzn-1.1	的内存中执行引擎。YARN
spark-yarn-slave	3.3.0-amzn-1.1	YARN从属服务器需要的 Apache Spark 库。
spark-rapids	22.08.0-amzn-0	Nvidia Spark RAPIDS 插件可以加速 Apache Spark。GPUs
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。



组件	版本	描述
tensorflow	2.10.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.10.2-amzn-0.1	tez YARN 应用程序和库。
webserver	2.4.41+	Apache HTTP 服务器。
zeppelin-server	0.10.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.5.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.5.10	ZooKeeper 命令行客户端。

### 6.9.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

为处于运行状态的集群中的实例组指定配置时，将发生重新配置操作。Amazon EMR 仅对您修改的分类启动重新配置操作。有关更多信息，请参阅 [在正在运行的集群中重新配置实例组](#)。

#### emr-6.9.1 分类

分类	描述	重新配置操作
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。	Restarts the ResourceManager service.
container-executor	更改 Hadoop YARN 的 container-executor.cfg 文件中的值。	Not available.

分类	描述	重新配置操作
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。	Not available.
core-site	更改 Hadoop 的 core-site.xml 文件中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	更改 docker 相关设置。	Not available.

分类	描述	重新配置操作
emrfs-site	更改EMRFS设置。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	更改 flink-conf.yaml 设置。	Restarts Flink history server.
flink-log4j	更改 Flink log4j.properties 设置。	Restarts Flink history server.
flink-log4j-session	为 Kubernetes/Yarn 会话更改 Flink log4j-session.properties 设置。	Restarts Flink history server.
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。	Restarts Flink history server.

分类	描述	重新配置操作
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	更改 hadoop ssl 服务器配置	Not available.
hadoop-ssl-client	更改 hadoop ssl 客户端配置	Not available.
hbase	亚马逊EMR精心策划的 Apache 设置。HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.

分类	描述	重新配置操作
hbase-env	更改环境HBase中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	更改 hadoop-metrics2-hbase base.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。	Not available.
hbase-site	更改HBase的 hbase-site.xml 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	配置HDFS加密区域。	This classification should not be reconfigured.
hdfs-env	更改HDFS环境中的值。	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	更改环境HCatalog中的值。	Restarts Hive HCatalog Server.

分类	描述	重新配置操作
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。	Restarts Hive WebHCat server.
hive	亚马逊EMR精心策划的 Apache Hive 设置。	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。	Not available.
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。	Not available.
hive-env	更改 Hive 环境中的值。	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	更改 Hive 的 hive-exec-log4j2.properties 文件中的值。	Not available.
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log4j2.properties 文件中的值。	Not available.

分类	描述	重新配置操作
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。	Not available.
hive-site	更改 Hive 的 hive-site.xml 文件中的值	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值	Not available.
hue-ini	更改 Hue 的 ini 文件中的值	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	更改HTTPFS环境中的值。	Restarts Hadoop Httpfs service.
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。	Restarts Hadoop Httpfs service.
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。	Not available.
hadoop-kms-env	在 Hadoop KMS 环境中更改值。	Restarts Hadoop-KMS service.
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。	Not available.
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。	Restarts Hadoop-KMS and Ranger-KMS service.

分类	描述	重新配置操作
hudi-env	更改 Hudi 环境中的值。	Not available.
hudi-defaults	更改 Hudi 的 hudi-defaults.conf 文件中的值。	Not available.
iceberg-defaults	更改 Iceberg 的 iceberg-defaults.conf 文件中的值。	Not available.
delta-defaults	更改 Delta 的 delta-defaults.conf 文件中的值。	Not available.
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。	Not available.
jupyter-hub-conf	更改 JupyterHubs 的 jupyterhub_config.py 文件中的值。	Not available.
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。	Not available.
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。	Not available.
livy-conf	更改 Livy 的 livy.conf 文件中的值。	Restarts Livy Server.
livy-env	更改 Livy 环境中的值。	Restarts Livy Server.
livy-log4j2	更改 Livy log4j2.properties 设置。	Restarts Livy Server.
mapred-env	更改 MapReduce 应用程序环境中的值。	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。	Restarts Hadoop MapReduce-HistoryServer.



分类	描述	重新配置操作
oozie-env	更改 Oozie 的环境中的值。	Restarts Oozie.
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。	Restarts Oozie.
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。	Restarts Oozie.
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。	Not available.
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。	Not available.
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。	Restarts Phoenix-QueryServer.
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。	Not available.
pig-env	更改 Pig 环境中的值。	Not available.
pig-properties	更改 Pig 的 pig.properties 文件中的值。	Restarts Oozie.
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。	Not available.
presto-log	更改 Presto 的 log.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-config	更改 Presto 的 config.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)

分类	描述	重新配置操作
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。	Not available.
presto-env	更改 Presto 的 presto-env.sh 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-node	更改 Presto 的 node.properties 文件中的值。	Not available.
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。	Not available.
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。	Not available.
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。	Not available.
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。	Not available.
presto-connector-lakeformation	更改 Presto 的 lakeformation.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。	Not available.
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。	Not available.
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。	Not available.
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。	Not available.
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。	Not available.
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。	Not available.
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。	Not available.
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。	Not available.
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。	Not available.
trino-log	更改 Trino 的 log.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-config	更改 Trino 的 config.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-password-authenticator	更改 Trino 的 password-authenticator.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-env	更改 Trino 的 trino-env.sh 文件中的值。	Restarts Trino-Server (for Trino)
trino-node	更改 Trino 的 node.properties 文件中的值。	Not available.

分类	描述	重新配置操作
trino-connector-blackhole	更改 Trino 的 blackhole.properties 文件中的值。	Not available.
trino-connector-cassandra	更改 Trino 的 cassandra.properties 文件中的值。	Not available.
trino-connector-delta	更改 Trino 的 delta.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-hive	更改 Trino 的 hive.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-exchange-manager	更改 Trino 的 exchange-manager.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-iceberg	更改 Trino 的 iceberg.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-jmx	更改 Trino 的 jmx.properties 文件中的值。	Not available.
trino-connector-kafka	更改 Trino 的 kafka.properties 文件中的值。	Not available.
trino-connector-localfile	更改 Trino 的 localfile.properties 文件中的值。	Not available.
trino-connector-memory	更改 Trino 的 memory.properties 文件中的值。	Not available.
trino-connector-mongodb	更改 Trino 的 mongodb.properties 文件中的值。	Not available.
trino-connector-mysql	更改 Trino 的 mysql.properties 文件中的值。	Not available.

分类	描述	重新配置操作
trino-connector-postgresql	更改 Trino 的 postgresql.properties 文件中的值。	Not available.
trino-connector-raptor	更改 Trino 的 raptor.properties 文件中的值。	Not available.
trino-connector-redis	更改 Trino 的 redis.properties 文件中的值。	Not available.
trino-connector-redshift	更改 Trino 的 redshift.properties 文件中的值。	Not available.
trino-connector-tpch	更改 Trino 的 tpch.properties 文件中的值。	Not available.
trino-connector-tpcds	更改 Trino 的 tpcds.properties 文件中的值。	Not available.
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-site	更改 Ranger KMS 的 ranger-kms-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-env	在 Ranger KMS 环境中更改值。	Restarts Ranger KMS Server.
ranger-kms-log4j	更改 Ranger 的 kms-log4j.properties 文件中的值。KMS	Not available.
ranger-kms-db-ca	更改 S3 上我与 Ranger SQL SSL KMS 连接的 CA 文件值。	Not available.
spark	亚马逊EMR精心策划的 Apache Spark 设置。	This property modifies spark-defaults. See actions there.
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。	Restarts Spark history server and Spark thrift server.

分类	描述	重新配置操作
spark-env	更改 Spark 环境中的值。	Restarts Spark history server and Spark thrift server.
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值	Not available.
spark-log4j2	更改 Spark 的 log4j2.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
sqoop-env	更改 Sqoop 的环境中的值。	Not available.
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。	Not available.
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。	Not available.
tez-site	更改 Tez 的 tez-site.xml 文件中的值。	Restart Oozie and HiveServer2.
yarn-env	更改YARN环境中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	更改 yarn-site.xml 文件YARN 中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.

分类	描述	重新配置操作
zeppelin-env	更改 Zeppelin 环境中的值。	Restarts Zeppelin.
zeppelin-site	更改 zeppelin-site.xml 中的配置设置。	Restarts Zeppelin.
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。	Restarts Zookeeper server.
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。	Restarts Zookeeper server.

## 6.9.1 更改日志

发行版 6.9.1 的更改日志和发布说明

Date	事件	描述
2023-08-30	更新发行说明	在发行说明中添加了几个与控制面板相关的修复
2023-08-21	文档发布	亚马逊 EMR 6.9.1 发行说明首次发布
2023-08-16	部署完成	Amazon EMR 6.9.1 已完全部署到所有 <a href="#">支持的区域</a>
2023-08-04	初始版本	亚马逊 EMR 6.9.1 首次部署到有限的商业区域

## 亚马逊EMR版本 6.9.0

### 6.9.0 应用程序版本

此发行版支持以下应用程序：

[Delta](#)、[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[Iceberg](#)、[JupyterEnterpriseGateway](#) 和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本 ( 如果适用 )。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-6.9.0	emr-6.8.1	emr-6.8.0	emr-6.7.0
适用于 Java 的 AWS SDK	1.12.170	1.12.170	1.12.170	1.12.170
Python	2.7、3.7	2.7、3.7	2.7、3.7	2.7、3.7
Scala	2.12.15	2.12.15	2.12.15	2.12.15
AmazonCloudWatchAgent	-	-	-	-
Delta	2.1.0	-	-	-
Flink	1.15.2	1.15.1	1.15.1	1.14.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.13	2.4.12	2.4.12	2.4.4
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.2.1	3.2.1	3.2.1
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.12.1-amzn-0	0.11.1-amzn-0	0.11.1-amzn-0	0.11.0-amzn-0



	emr-6.9.0	emr-6.8.1	emr-6.8.0	emr-6.7.0
Hue	4.10.0	4.10.0	4.10.0	4.10.0
Iceberg	0.14.1-amzn-0	0.14.0-amzn-0	0.14.0-amzn-0	0.13.1-amzn-0
JupyterEnterpriseGateway	2.6.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.8.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.276	0.273	0.273	0.272
Spark	3.3.0	3.3.0	3.3.0	3.2.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.10.0	2.9.1	2.9.1	2.4.1
Tez	0.10.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	398	388	388	378
Zeppelin	0.10.1	0.10.1	0.10.1	0.10.0
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.7

## 6.9.0 发布说明

以下发行说明包含 Amazon 6.9. EMR 0 版本的信息。更改与亚马逊 6.8.0 EMR 版本有关。有关发布时间的信息，请参阅[更改日志](#)。

### 新功能

- 亚马逊EMR版本 6.9.0 支持 Apache Spark RAPIDS 22.08.0、Apache Hudi 0.12.1、Apache Iceberg 0.14.1、Trino 398 和 Tez 0.10.2。
- 亚马逊 6.9.0 EMR 版本包括一个新的开源应用程序 2.1.0。 [Delta Lake](#)
- 适用于 Apache Spark 的 Amazon Redshift 集成包含在亚马逊 6.9.0 及更高EMR版本中。本地集成之前是一种开源工具，现在是 Spark 连接器，您可以将其用于构建 Apache Spark 应用程序，这些应用程序可在 Amazon Redshift 和 Amazon Redshift Serverless 中读取和写入数据。有关更多信息，请参阅 [使用亚马逊 Redshift 与 Apache Spark 的集成与亚马逊 EMR](#)。
- Amazon EMR 版本 6.9.0 增加了对集群缩小期间将日志存档到 Amazon S3 的支持。之前，您只能在集群终止期间将日志文件存档到 Amazon S3。这项新功能可确保即使在节点终止后，集群上生成的日志文件仍保留在 Amazon S3 上。有关更多信息，请参阅[配置集群日志记录和调试](#)。
- 为了支持长时间运行的查询，Trino 现在包括容错执行机制。容错执行通过重试失败的查询或其组件任务来减少查询失败。有关更多信息，请参阅 [Trino 中的容错执行](#)。
- 您可以使用亚马逊EMR上的 Apache Flink 来统一 BATCH和STREAM处理 Apache Hive 表或任何 Flink 表源的元数据，例如 Iceberg、Kinesis 或 Kafka。你可以使用、或 Amazon 将 AWS Glue 数据目录指定为 Flink 的 AWS Management Console元数据库。AWS CLI EMR API有关更多信息，请参阅 [在亚马逊中配置 Flink EMR](#)。
- 现在，你可以在装有 Amazon Studio 的EMR集群上为亚马逊上的 Apache Spark、Apache Hive 和 Presto 查询指定 AWS Identity and Access Management (IAM) 运行时角色和 AWS Lake Formation 基于访问控制EC2的权限。 SageMaker 有关更多信息，请参阅[为 Amazon EMR 步骤配置运行时角色](#)。

### 已知问题

- 对于亚马逊EMR版本 6.9.0，Trino 不适用于启用 Apache Ranger 的集群。如果您需要将 Trino 与 Ranger 结合使用，请联系 [AWS Support](#)。
- 如果您使用适用于 Apache Spark 的 Amazon Redshift 集成，并且具有 Parquet 格式的时间、timetz、时间戳或 timestampz (精度为微秒)，连接器会将时间值舍入为最接近的毫秒值。解决方法是使用文本卸载格式 unload\_s3\_format 参数。

- 当你使用带有 Hive 分区位置格式的 Spark 来读取 Amazon S3 中的数据，并在亚马逊EMR版本 5.30.0 到 5.36.0 和 6.2.0 到 6.9.0 上运行 Spark 时，你可能会遇到一个问题，导致集群无法正确读取数据。如果您的分区具有以下所有特征，会发生这种情况：
  - 从同一个表扫描两个或多个分区。
  - 至少有一个分区目录路径是至少一个其他分区目录路径的前缀，例如，s3://bucket/table/p=a 是 s3://bucket/table/p=a b 的前缀。
  - 另一个分区目录中前缀之后的第一个字符的值为 UTF -8，小于该/字符 (U+002F)。例如，在 s3://bucket/table/p=a b 中，a 和 b 之间出现的空格字符 (U+0020) 就属于此类。请注意，还有其他 14 个非控制字符：!"#\$%&'()\*+,-。有关更多信息，请参阅 [UTF-8 编码表和 Unicode 字符](#)。

解决方法是在 spark-defaults 分类中将

spark.sql.sources.fastS3PartitionDiscovery.enabled 配置设置为 false。

- 从 Amazon SageMaker Studio 到亚马逊EMR集群的连接可能会间歇性失败，并显示 403 禁止的响应代码。当在集群上设置IAM角色的时间超过 60 秒时，就会发生此错误。解决方法是，您可以安装 Amazon EMR 补丁以启用重试并将超时时间延长到最少 300 秒。启动集群时，按照以下步骤应用引导操作。

1. 从以下 Amazon S3 URIs 下载引导脚本和RPM文件。

```
s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/gcsc/replace-rpms.sh
s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/gcsc/emr-secret-agent-1.18.0-SNAPSHOT20221121212949.noarch.rpm
```

2. 将上一步中的文件上传到您自己的 Amazon S3 存储桶中。存储桶必须与您计划启动集群的 AWS 区域 位置相同。
3. 在启动EMR集群时包括以下引导操作。Replace ( 替换 ) *bootstrap\_URI* 以及 *RPM\_URI* 以及 URIs来自亚马逊 S3 的相应内容。

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

- 在 Amazon 5.36.0 和 6.6.0 到 6.9.0 EMR 版本中，由于 Log4 SecretAgent j2 属性中的文件名模式配置不正确，RecordServer服务组件可能会丢失日志数据。错误的配置导致组件每天只生成一个日志文件。当应用轮换策略时，它会重写现有文件，而不是按预期生成新的日志文件。应变方法是使用引导操作每小时生成一次日志文件，并在文件名中附加一个自动增量的整数来处理轮换。

对于 Amazon EMR 6.0 到 6.9.0 版本，请在启动集群时使用以下引导操作。

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

对于 Amazon EMR 5.36.0，请在启动集群时使用以下引导操作。

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

- Apache Flink 提供原生 S3 FileSystem 和 Hadoop FileSystem 连接器，允许应用程序创建 FileSink 并将数据写入 Amazon S3。这 FileSink 会失败，但有以下两个例外情况之一。

```
java.lang.UnsupportedOperationException: Recoverable writers on Hadoop are only supported for HDFS
```

```
Caused by: java.lang.NoSuchMethodError:
  org.apache.hadoop.io.retry.RetryPolicies.retryOtherThanRemoteAndSaslException(Lorg/apache/hadoop/io/retry/RetryPolicy;Ljava/util/Map;)Lorg/apache/hadoop/io/retry/RetryPolicy;
                                     at
  org.apache.hadoop.yarn.client.RMProxy.createRetryPolicy(RMProxy.java:302) ~[hadoop-yarn-common-3.3.3-amzn-0.jar:?]
```

要解决这个问题，你可以安装一个 Amazon EMR 补丁，该补丁修复了 Flink 中的上述问题。要在启动集群时应用引导操作，请完成以下步骤。

1. 将 flink-rpm 下载到 Amazon S3 存储桶中。你的RPM路是s3://*DOC-EXAMPLE-BUCKET*/rpms/flink/。
2. 使用以下URI命令从 Amazon S3 下载引导脚本和RPM文件。*regionName*替换为您计划启动集群 AWS 区域 的位置。

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

3. Hadoop 3.3.3 在 YARN ([YARN-9608](#)) 中引入了一项更改，即在应用程序完成之前，容器运行的节点一直处于停用状态。此更改可确保如随机数据等本地数据不会丢失，并且您无需重新运行作业。在 Amazon EMR 6.8.0 和 6.9.0 中，这种方法还可能导致启用或未启用托管扩展的集群上的资源利用不足。

在 [Amazon EMR 6.10.0](#) 中，有一种解决方法可以将 `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` 值设置为 `in`。在 `false` `yarn-site.xml` 在亚马逊EMR版本 6.11.0 及更高版本以及 6.8.1、6.9.1 和 6.10.1 中，配置 `false` 默认设置为以解决此问题。

## 更改、增强和解决的问题

- 对于亚马逊 6.9.0 及更高EMR版本，亚马逊EMR安装的所有使用 Log4j 库的组件都使用 Log4j 版本 2.17.1 或更高版本。
- 当您在EMR亚马逊 6.0、6.7.0 和 6.8.0 版本上将 DynamoDB 连接器与 Spark 配合使用时，即使输入拆分引用非空数据，所有从表中读取的内容都会返回空结果。亚马逊 6.9.0 EMR 版本修复了这个问题。
- 当使用 Spark 读取数据时，亚马逊 EMR 6.9.0 增加了对基于 Lake Formation 的 Apache Hudi 访问控制的有限支持。SQL支持使用 Spark 的SELECT查询，SQL并且仅限于列级访问控制。有关更多信息，请参阅 [Hudi 和 Lake Formation](#)。
- 当您使用 Amazon EMR 6.9.0 创建启用了 [节点标签](#) 的 Hadoop 集群时，[YARN指标](#) 会API返回所有分区的汇总信息，而不是默认分区。有关更多信息，请参阅 [YARN-11414](#)。
- 在亚马逊6.9.0 EMR 版本中，我们已将Trino更新到使用Java 17的398版本。之前支持的亚马逊 EMR 6.8.0 版 Trino 版本是在 Java 11 上运行的 Trino 388。有关此变更的更多信息，请参阅 Trino 博客上的 [Trino updates to Java 17](#)。
- 此版本修复了 Apache BigTop 和 Amaz EMR on EC2 集群启动序列之间的时间序列不匹配问题。当系统尝试同时执行两个或多个操作而不是按正确的顺序执行它们时，就会发生这种计时序列不匹配。因此，某些集群配置会遇到实例启动超时和较慢的集群启动时间。
- 当您启动带有最新补丁版本（亚马逊 EMR 5.36 或更高版本、6.6 或更高版本、7.0 或更高版本）的集群时，亚马逊将EMR使用最新的亚马逊 Linux 2023 或亚马逊 Linux 2 版本作为默认亚马逊 Linux 2 版本。EMR AMI有关更多信息，请参阅[使用亚马逊的默认 Amazon Linux AMI EMR](#)。

### Note

此版本不再获得自动AMI更新，因为又有 1 个补丁版本接替了该版本。补丁版本以第二位小数点后的数字 (6.8.1) 表示。要查看您是否使用的是最新补丁版本，请查看[发布指南](#)中的可用版本，或者在控制台中创建集群时查看 Amazon EMR 版本下拉列表，或者使用[ListReleaseLabels](#)API或[list-release-labels](#)CLI操作。要获取有关新版本的更新，请订阅“[新增内容?](#)”RSS 页面。

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 709.1	4.14.348	2024年7月23日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、AWS GovCloud ( 美国

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			东部 )、( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )、亚洲太平洋 ( 海得拉巴 )、中东 ( UAE )、欧洲 ( 西班牙 )、欧洲 ( 苏黎世 )、亚太地区 ( 墨尔本 )、以色列 ( 特拉维夫 )、加拿大西部 ( 卡尔加里 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.20230808.0	4.14.320	2023 年 8 月 24 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、欧洲地区 (斯德哥尔摩)、欧洲地区 (米兰)、欧洲地区 (法兰克福)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、以色列 (特拉维夫)



OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 727.0	4.14.320	2023 年 8 月 14 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 719.0	4.14.320	2023 年 8 月 2 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 628.0	4.14.318	2023 年 7 月 12 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 612.0	4.14.314	2023 年 6 月 23 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 504.1	4.14.313	2023 年 5 月 16 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 418.0	4.14.311	2023 年 5 月 3 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 404.1	4.14.311	2023 年 4 月 18 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)
2.0.2023 404.0	4.14.311	2023 年 4 月 10 日	美国东部 (弗吉尼亚州北部)、欧洲地区 (巴黎)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 320.0	4.14.309	2023 年 3 月 30 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )



OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.202307.0	4.14.305	2023 年 3 月 15 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 207.0	4.14.304	2023 年 2 月 22 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022 210.1	4.14.301	2023 年 1 月 12 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022 103.3	4.14.296	2022 年 12 月 5 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

### 6.9.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 emr 或 aws 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.4.2	亚马逊 SageMaker Spark SDK
delta	2.1.0	Delta lake 是一种适用于超大型分析数据集的开放表格式。
emr-ddb	4.16.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	3.3.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.6.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-notebook-env	1.7.0	适用于 EMR Notebooks ( 可提供 jupyter 企业网关 ) 的 Conda env
emr-s3-dist-cp	2.23.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	2.2.0	EMRS3Select 连接器
emrfs	2.54.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.15.2	Apache Flink 命令行客户端脚本和应用程序。

组件	版本	描述
flink-jobmanager-config	1.15.2	管理 Apache Flink EMR JobManager 节点上的资源。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	3.3.3-amzn-1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	3.3.3-amzn-1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	3.3.3-amzn-1	HDFS命令行客户端和库
hadoop-hdfs-namenode	3.3.3-amzn-1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	3.3.3-amzn-1	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	3.3.3-amzn-1	HTTPHDFS操作端点。
hadoop-kms-server	3.3.3-amzn-1	基于 Hadoop 的加密密钥管理服务。 KeyProvider API

组件	版本	描述
hadoop-mapred	3.3.3-amzn-1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	3.3.3-amzn-1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	3.3.3-amzn-1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	3.3.3-amzn-1	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	2.4.13-amzn-0	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	2.4.13-amzn-0	为一个或多个HBase地区提供服务的服务。
hbase-client	2.4.13-amzn-0	HBase命令行客户端。
hbase-rest-server	2.4.13-amzn-0	为其提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	2.4.13-amzn-0	为其提供 Thrift 端点的HBase 服务。
hbase-operator-tools	2.4.13-amzn-0	适用于 Apache HBase 集群的修复工具。
hcatalog-client	3.1.3-amzn-2	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	3.1.3-amzn-2	提供服务HCatalog，为分布式应用程序提供表和存储管理层。

组件	版本	描述
hcatalog-webhcat-server	3.1.3-amzn-2	HTTP端点为其提供REST接口HCatalog。
hive-client	3.1.3-amzn-2	Hive 命令行客户端。
hive-hbase	3.1.3-amzn-2	Hive-hbase 客户端。
hive-metastore-server	3.1.3-amzn-2	用于访问 Hive 元数据仓库的服务，这是一个语义存储库，用于存储 Hadoop 操作SQL的元数据。
hive-server2	3.1.3-amzn-2	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.12.1-amzn-0	增量处理框架，以支持低延迟和高效率的数据管道。
hudi-presto	0.12.1-amzn-0	用于运行 Presto 以及 Hudi 的捆绑库。
hudi-trino	0.12.1-amzn-0	用于运行 Trino 以及 Hudi 的捆绑库。
hudi-spark	0.12.1-amzn-0	用于运行 Spark 以及 Hudi 的捆绑库。
hue-server	4.10.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
iceberg	0.14.1-amzn-0	Apache Iceberg 是一种适用于超大型分析数据集的开放表格式。
jupyterhub	1.4.1	Jupyter notebook 的多用户服务器



组件	版本	描述
livy-server	0.7.1-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mxnet	1.9.1	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.68+	MariaDB 数据库服务器。
nvidia-cuda	11.7.0	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.1	Oozie 命令行客户端。
oozie-server	5.2.1	用于接受 Oozie 工作流请求的服务。
opencv	4.5.0	开源计算机视觉库。
phoenix-library	5.1.2	服务器和客户端的 phoenix 库
phoenix-connectors	6.0.0-SNAPSHOT	Apache Phoenix-Connectors for Spark-3
phoenix-query-server	6.0.0	轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限 API
presto-coordinator	0.276-amzn-0	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.276-amzn-0	用于执行查询的各个部分的服务。

组件	版本	描述
presto-client	0.276-amzn-0	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
trino-coordinator	398-amzn-0	用于在 trino-worker 之中接受查询并管理查询的服务。
trino-worker	398-amzn-0	用于执行查询的各个部分的服务。
trino-client	398-amzn-0	Trino 命令行客户端，安装在 HA 集群的备用主服务器（未启动 Trino 服务器）上。
pig-client	0.17.0	Pig 命令行客户端。
r	4.0.2	用于统计计算的 R 项目
ranger-kms-server	2.0.0	Apache Ranger 密钥管理系统
spark-client	3.3.0-amzn-1	Spark 命令行客户端。
spark-history-server	3.3.0-amzn-1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	3.3.0-amzn-1	的内存中执行引擎。YARN
spark-yarn-slave	3.3.0-amzn-1	YARN从属服务器需要的 Apache Spark 库。
spark-rapids	22.08.0-amzn-0	Nvidia Spark RAPIDS 插件可以加速 Apache Spark。GPUs
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。

组件	版本	描述
tensorflow	2.10.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.10.2-amzn-0	tez YARN 应用程序和库。
webserver	2.4.41+	Apache HTTP 服务器。
zeppelin-server	0.10.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.5.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.5.10	ZooKeeper 命令行客户端。

## 6.9.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

为处于运行状态的集群中的实例组指定配置时，将发生重新配置操作。Amazon EMR 仅对您修改的分类启动重新配置操作。有关更多信息，请参阅 [在正在运行的集群中重新配置实例组](#)。

### emr-6.9.0 分类

分类	描述	重新配置操作
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。	Restarts the ResourceManager service.
container-executor	更改 Hadoop YARN 的 container-executor.cfg 文件中的值。	Not available.

分类	描述	重新配置操作
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。	Not available.
core-site	更改 Hadoop 的 core-site.xml 文件中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	更改 docker 相关设置。	Not available.

分类	描述	重新配置操作
emrfs-site	更改EMRFS设置。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	更改 flink-conf.yaml 设置。	Restarts Flink history server.
flink-log4j	更改 Flink log4j.properties 设置。	Restarts Flink history server.
flink-log4j-session	为 Kubernetes/Yarn 会话更改 Flink log4j-session.properties 设置。	Restarts Flink history server.
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。	Restarts Flink history server.

分类	描述	重新配置操作
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	更改 hadoop ssl 服务器配置	Not available.
hadoop-ssl-client	更改 hadoop ssl 客户端配置	Not available.
hbase	亚马逊EMR精心策划的 Apache 设置。HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.

分类	描述	重新配置操作
hbase-env	更改环境HBase中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	更改 hadoop-metrics2-hbase base.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。	Not available.
hbase-site	更改HBase的 hbase-site.xml 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	配置HDFS加密区域。	This classification should not be reconfigured.
hdfs-env	更改HDFS环境中的值。	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	更改环境HCatalog中的值。	Restarts Hive HCatalog Server.

分类	描述	重新配置操作
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。	Restarts Hive WebHCat server.
hive	亚马逊EMR精心策划的 Apache Hive 设置。	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。	Not available.
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。	Not available.
hive-env	更改 Hive 环境中的值。	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	更改 Hive 的 hive-exec-log4j2.properties 文件中的值。	Not available.
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log4j2.properties 文件中的值。	Not available.



分类	描述	重新配置操作
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。	Not available.
hive-site	更改 Hive 的 hive-site.xml 文件中的值	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值	Not available.
hue-ini	更改 Hue 的 ini 文件中的值	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	更改HTTPFS环境中的值。	Restarts Hadoop Httpfs service.
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。	Restarts Hadoop Httpfs service.
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。	Not available.
hadoop-kms-env	在 Hadoop KMS 环境中更改值。	Restarts Hadoop-KMS service.
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。	Not available.
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。	Restarts Hadoop-KMS and Ranger-KMS service.

分类	描述	重新配置操作
hudi-env	更改 Hudi 环境中的值。	Not available.
hudi-defaults	更改 Hudi 的 hudi-defaults.conf 文件中的值。	Not available.
iceberg-defaults	更改 Iceberg 的 iceberg-defaults.conf 文件中的值。	Not available.
delta-defaults	更改 Delta 的 delta-defaults.conf 文件中的值。	Not available.
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。	Not available.
jupyter-hub-conf	更改 JupyterHubs 的 jupyterhub_config.py 文件中的值。	Not available.
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。	Not available.
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。	Not available.
livy-conf	更改 Livy 的 livy.conf 文件中的值。	Restarts Livy Server.
livy-env	更改 Livy 环境中的值。	Restarts Livy Server.
livy-log4j2	更改 Livy log4j2.properties 设置。	Restarts Livy Server.
mapred-env	更改 MapReduce 应用程序环境中的值。	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。	Restarts Hadoop MapReduce-HistoryServer.

分类	描述	重新配置操作
oozie-env	更改 Oozie 的环境中的值。	Restarts Oozie.
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。	Restarts Oozie.
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。	Restarts Oozie.
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。	Not available.
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。	Not available.
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。	Restarts Phoenix-QueryServer.
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。	Not available.
pig-env	更改 Pig 环境中的值。	Not available.
pig-properties	更改 Pig 的 pig.properties 文件中的值。	Restarts Oozie.
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。	Not available.
presto-log	更改 Presto 的 log.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-config	更改 Presto 的 config.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)

分类	描述	重新配置操作
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。	Not available.
presto-env	更改 Presto 的 presto-env.sh 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-node	更改 Presto 的 node.properties 文件中的值。	Not available.
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。	Not available.
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。	Not available.
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。	Not available.
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。	Not available.
presto-connector-lakeformation	更改 Presto 的 lakeformation.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。	Not available.
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。	Not available.
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。	Not available.
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。	Not available.
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。	Not available.
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。	Not available.
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。	Not available.
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。	Not available.
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。	Not available.
trino-log	更改 Trino 的 log.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-config	更改 Trino 的 config.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-password-authenticator	更改 Trino 的 password-authenticator.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-env	更改 Trino 的 trino-env.sh 文件中的值。	Restarts Trino-Server (for Trino)
trino-node	更改 Trino 的 node.properties 文件中的值。	Not available.

分类	描述	重新配置操作
trino-connector-blackhole	更改 Trino 的 blackhole.properties 文件中的值。	Not available.
trino-connector-cassandra	更改 Trino 的 cassandra.properties 文件中的值。	Not available.
trino-connector-delta	更改 Trino 的 delta.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-hive	更改 Trino 的 hive.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-exchange-manager	更改 Trino 的 exchange-manager.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-iceberg	更改 Trino 的 iceberg.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-jmx	更改 Trino 的 jmx.properties 文件中的值。	Not available.
trino-connector-kafka	更改 Trino 的 kafka.properties 文件中的值。	Not available.
trino-connector-localfile	更改 Trino 的 localfile.properties 文件中的值。	Not available.
trino-connector-memory	更改 Trino 的 memory.properties 文件中的值。	Not available.
trino-connector-mongodb	更改 Trino 的 mongodb.properties 文件中的值。	Not available.
trino-connector-mysql	更改 Trino 的 mysql.properties 文件中的值。	Not available.

分类	描述	重新配置操作
trino-connector-postgresql	更改 Trino 的 postgresql.properties 文件中的值。	Not available.
trino-connector-raptor	更改 Trino 的 raptor.properties 文件中的值。	Not available.
trino-connector-redis	更改 Trino 的 redis.properties 文件中的值。	Not available.
trino-connector-redshift	更改 Trino 的 redshift.properties 文件中的值。	Not available.
trino-connector-tpch	更改 Trino 的 tpch.properties 文件中的值。	Not available.
trino-connector-tpcds	更改 Trino 的 tpcds.properties 文件中的值。	Not available.
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-site	更改 Ranger KMS 的 ranger-kms-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-env	在 Ranger KMS 环境中更改值。	Restarts Ranger KMS Server.
ranger-kms-log4j	更改 Ranger 的 kms-log4j.properties 文件中的值。KMS	Not available.
ranger-kms-db-ca	更改 S3 上我与 Ranger SQL SSL KMS 连接的 CA 文件值。	Not available.
spark	亚马逊EMR精心策划的 Apache Spark 设置。	This property modifies spark-defaults. See actions there.
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。	Restarts Spark history server and Spark thrift server.

分类	描述	重新配置操作
spark-env	更改 Spark 环境中的值。	Restarts Spark history server and Spark thrift server.
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值	Not available.
spark-log4j2	更改 Spark 的 log4j2.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
sqoop-env	更改 Sqoop 的环境中的值。	Not available.
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。	Not available.
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。	Not available.
tez-site	更改 Tez 的 tez-site.xml 文件中的值。	Restart Oozie and HiveServer2.
yarn-env	更改YARN环境中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	更改 yarn-site.xml 文件YARN 中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.



分类	描述	重新配置操作
zeppelin-env	更改 Zeppelin 环境中的值。	Restarts Zeppelin.
zeppelin-site	更改 zeppelin-site.xml 中的配置设置。	Restarts Zeppelin.
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。	Restarts Zookeeper server.
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。	Restarts Zookeeper server.

## 6.9.0 更改日志

### 发行版 6.9.0 的更改日志和发布说明

Date	事件	描述
2023-08-30	更新发行说明	添加了对定时间序列不匹配问题的修复
2023-08-21	更新发行说明	在 Hadoop 3.3.3 中添加了一个已知问题。
2023-07-26	更新	新的操作系统版本标签 2.0.20230612.0 和 2.0.20230628.0 。
2022-12-13	发布说明已更新	增加了运行时的功能和已知问题 SageMaker
2022-11-29	发布说明和文档已更新	添加了适用于 Apache Spark 的 Amazon Redshift 集成功能
2022-11-23	发布说明已更新	Log4j 条目已删除
2022-11-18	部署完成	Amazon EMR 6.9 已完全部署到所有 <a href="#">受支持的区域</a>

Date	事件	描述
2022-11-18	文档发布	亚马逊 EMR 6.9 发行说明首次发布
2022-11-14	初始版本	亚马逊 EMR 6.9 已部署到有限的商业区域

## 亚马逊EMR版本 6.8.1

### 6.8.1 应用程序版本

此发行版支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[Iceberg](#)、[JupyterEnterpriseGateway](#) 和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-6.8.1	emr-6.8.0	emr-6.7.0	emr-6.6.0
适用于 Java 的 AWS SDK	1.12.170	1.12.170	1.12.170	1.12.170
Python	2.7、3.7	2.7、3.7	2.7、3.7	2.7、3.7
Scala	2.12.15	2.12.15	2.12.15	2.12.10

	emr-6.8.1	emr-6.8.0	emr-6.7.0	emr-6.6.0
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.15.1	1.15.1	1.14.2	1.14.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.12	2.4.12	2.4.4	2.4.4
HCatalog	3.1.3	3.1.3	3.1.3	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.3	3.1.3	3.1.3	3.1.2
Hudi	0.11.1-amzn-0	0.11.1-amzn-0	0.11.0-amzn-0	0.10.1-amzn-0
Hue	4.10.0	4.10.0	4.10.0	4.10.0
Iceberg	0.14.0-amzn-0	0.14.0-amzn-0	0.13.1-amzn-0	0.13.1
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.8.0	1.8.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0

	emr-6.8.1	emr-6.8.0	emr-6.7.0	emr-6.6.0
Presto	0.273	0.273	0.272	0.267
Spark	3.3.0	3.3.0	3.2.1	3.2.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.9.1	2.9.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	388	388	378	367
Zeppelin	0.10.1	0.10.1	0.10.0	0.10.0
ZooKeeper	3.5.10	3.5.10	3.5.7	3.5.7

## 6.8.1 发布说明

以下发行说明包含有关亚马逊 6.8.1 EMR 版本的信息。更改与 6.8.0 有关。有关发布时间表的更多信息，请参阅 [6.8.1 更改日志](#)。

### 更改、增强功能和解决的问题

- Hadoop 3.3.3 在 YARN ([YARN-9608](#)) 中引入了一项更改，即在应用程序完成之前，容器运行的节点一直处于停用状态。此更改可确保如随机数据等本地数据不会丢失，并且您无需重新运行作业。无论是否启用托管扩展，这种方法还可能导致集群的资源利用不足。

在亚马逊EMR版本 6.11.0 及更高版本以及 6.8.1、6.9.1 和 6.10.1 中，将的值设置 `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` 为 `in` 以 `false` 解决此问题。 `yarn-site.xml`

虽然该修复程序解决了 YARN -9608 引入的问题，但由于启用了托管扩展的集群上的洗牌数据丢失，它可能会导致 Hive 作业失败。在此版本中，我们还通过设置 Hive `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-shuffle-data` 工作负载来降低这种风险。此配置仅在 Amazon 6.11.0 及更高EMR版本中可用。

- 使用实例组配置的集群中的主节点失效转移后，指标收集器不会向控制面板发送任何指标。

- 此版本消除了对指标收集器端点的HTTP请求失败时进行重试。
- 此版本包括一项更改，允许高可用性集群在重启后从故障状态中恢复。
- 此版本修复了用户创建的大量内容UIDs导致溢出异常的问题。
- 此版本修复了 Amazon EMR 重新配置过程中的超时问题。
- 此版本可防止出现重新配置失败可能会中断其他不相关的进程的问题。
- 此版本包含安全修复。
- 此版本修复了在 Spark with Amazon 上运行工作负载的集群EMR可能会在静默中收到使用contains、startsWithendsWith、和like的错误结果的问题。当您对在 Amazon EMR Hive3 Metastore 服务器中包含元数据的分区字段使用表达式时，就会出现此问题 ()。HMS
- 在 Amazon EMR 6.0 到 6.9.x 中，具有动态分区和 BY 或 BY 子句ORDER的INSERT查询将始终有SORT两个缩减器。此问题是由OSS更改 [HIVE-20703](#) 引起的，该更改将动态排序分区优化置于基于成本的决策之下。如果您的工作负载不需要对动态分区进行排序，建议将hive.optimize.sort.dynamic.partition.threshold 属性设置为 -1 以禁用新功能并获得计算正确的 Reducer 数量。此问题已在 OSS Hive 中作为 [HIVE-22269](#) 的一部分修复，并已在亚马逊 6.10.0 中修复。EMR
- 当您HDFS用作临时目录并启用合并小文件并且该表包含静态分区路径时，Hive 可能会丢失数据。
- 此版本修复了ETL任务结束时启用合并小文件（默认禁用）时 Hive 的性能问题。
- 此版本修复了当没有用户定义的函数时，Glue 端的限制问题 () UDF。
- 此版本修复了一个问题，即节点日志聚合服务在日志推送器可以将容器日志推送到 S3 之前删除容器日志，以防YARN停用。
- 此版本修复了使用永久存储文件跟踪功能对压缩/存档文件的处理。HBase
- 此版本修复了您在 spark-defaults.conf 中为 spark.yarn.heterogeneousExecutors.enabled 配置设置默认 true 值时影响 Spark 性能的问题。
- 此版本修复了 Reduce Task 无法读取随机数据的问题。该问题因内存损坏错误导致 Hive 查询失败。
- 此版本修复了在节点更换期间 HDFS NameNode (NN) 服务停留在安全模式下会导致节点配置器失败的问题。
- 此版本为运行 Presto 或 Trino 的集群的集群扩展工作流程添加了新的重试机制。EMR这一改进降低了由于单个调整大小操作失败而导致集群大小调整无限期停滞的风险。它还可以提高集群利用率，因为您的集群可以更快地向上和向下扩展。
- 此版本改进了集群缩容逻辑，因此您的集群不会尝试将核心节点缩小到低于集群HDFS重复因子设置的范围。这符合您的数据冗余要求，并减少了扩展操作可能停滞的机会。

- 日志管理进程守护程序已升级，可以识别本地实例存储中所有包含打开文件句柄的使用中的日志，以及相关的进程。此次升级可确保 Amazon 在日志存档到 Amazon S3 后EMR正确删除文件并回收存储空间。
- 此版本包括日志管理进程守护程序增强功能，可删除本地集群文件系统中空的、未使用的步骤目录。过多的空目录会降低 Amazon EMR 守护程序的性能并导致磁盘过度使用。
- 此版本修复了通过从具有多个主节点的集群中复制一个主节点来创建边缘节点时可能出现的问题。复制的边缘节点可能会导致缩减操作的延迟，或者导致主节点的内存使用率过高。有关如何创建边缘节点以及与EMR集群通信的更多信息，请参阅存储库中的 [Edge Node Create](#) `aws-samples` or。GitHub
- 此版本改进了 Amazon 用于在重启后EMR将 Amazon EBS 卷重新挂载到实例的自动化流程。
- 此版本修复了导致亚马逊向亚马逊EMR发布的 Hadoop 指标间歇性出现差距的问题。CloudWatch
- 此版本修复了EMR群集的一个问题，即由于磁盘过度使用而中断了对包含群集节点排除列表的YARN 配置文件的更新。不完整的更新阻碍了未来对集群的缩减操作。此版本可确保您的集群保持正常运行，并确保扩展操作按预期进行。
- 此版本改进了集群上的日志管理守护程序，可以监视集群中的其他日志文件夹。EMR这一改进最大限度地减少了磁盘过度使用情况。
- 此版本在集群上日志管理进程守护程序停止后会自动重启该守护程序。这一改进降低了由于磁盘过度使用而导致节点出现运行状况不佳的风险。
- 此版本增加了对在集群缩减期间将日志存档到 Amazon S3 的支持。之前，您只能在集群终止期间将日志文件存档到 Amazon S3。这项新功能可确保即使在节点终止后，集群上生成的日志文件仍保留在 Amazon S3 上。有关更多信息，请参阅[配置集群日志记录和调试](#)。
- 此版本修复了URI用于引导操作的 Amazon S3 以端口号结尾时出现的问题，例如：`a.b.c.d:4345`。Amazon 错误地解析EMR了这些URIs内容，因此任何相关的引导操作都将失败。
- 此版本修复了 Apache BigTop 和 Amaz EMR on EC2 集群启动序列之间的时间序列不匹配问题。当系统尝试同时执行两个或多个操作而不是按正确的顺序执行它们时，就会发生这种计时序列不匹配。因此，某些集群配置会遇到实例启动超时和较慢的集群启动时间。
- 当您启动带有最新补丁版本（亚马逊 EMR 5.36 或更高版本、6.6 或更高版本、7.0 或更高版本）的集群时，亚马逊将EMR使用最新的亚马逊 Linux 2023 或亚马逊 Linux 2 版本作为默认亚马逊 Linux 2 版本。EMR AMI有关更多信息，请参阅[使用亚马逊的默认 Amazon Linux AMI EMR](#)。

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 709.1	4.14.348	2024年7月23日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、AWS GovCloud ( 美国

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			东部)、(美国东部)、中国(北京)、中国(宁夏)、亚洲太平洋(海得拉巴)、中东(UAE)、欧洲(西班牙)、欧洲(苏黎世)、亚太地区(墨尔本)、以色列(特拉维夫)、



OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024223.0	4.14.336	2024 年 3 月 8 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			GovCloud ( 美国东部 )、 中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 131.0	4.14.336	2024年2月14日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			GovCloud (美国东部)、中国(北京)、中国(宁夏)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 124.0	4.14.336	2024 年 2 月 7 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			GovCloud (美国东部)、中国(北京)、中国(宁夏)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 109.0	4.14.334	2024 年 1 月 24 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			GovCloud (美国东部)、中国(北京)、中国(宁夏)



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 218.0	4.14.330	2024 年 1 月 2 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			GovCloud (美国东部)、中国(北京)、中国(宁夏)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 206.0	4.14.330	2023 年 12 月 22 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			GovCloud (美国东部)、中国(北京)、中国(宁夏)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023116.0	4.14.328	2023 年 12 月 11 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			GovCloud ( 美国东部 )、 中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 101.0	4.14.327	2023 年 11 月 16 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			GovCloud (美国东部)、中国(北京)、中国(宁夏)



OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023020.1	4.14.326	2023 年 11 月 7 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			GovCloud ( 美国东部 )、 中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023012.1	4.14.326	2023 年 10 月 26 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			GovCloud ( 美国东部 )、 中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 926.0	4.14.322	2023 年 10 月 19 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 4.14.322 906.0		2023 年 10 月 4 日	GovCloud (美国东部)、中国(北京)、中国(宁夏)  美国东部(弗吉尼亚州北部)、美国东部(俄亥俄州)、美国西部(北加利福尼亚)、美国西部(俄勒冈州)、欧洲地区(斯德哥尔摩)、欧洲地区(米兰)、欧洲地区(法兰克福)、欧洲地区(爱尔兰)、欧洲地区(伦敦)、欧洲地区(巴黎)、亚太地区(香港)、亚太地区(孟买)、亚太地区(东京)、亚太地区(首尔)、亚太地区(大阪)、亚太地区(新加坡)、亚太地区(悉尼)、亚太地区(雅加达)、亚太地区(墨尔本)、非洲(开普敦)、南美洲(圣保罗)、加拿大(中部)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 822.0	4.14.322	2023 年 8 月 30 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 米兰 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、加拿大 ( 中部 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 808.0	4.14.320	2023 年 8 月 24 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、欧洲地区 (斯德哥尔摩)、欧洲地区 (米兰)、欧洲地区 (法兰克福)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、加拿大 (中部)



OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 727.0	4.14.320	2023 年 8 月 14 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、欧洲地区 (斯德哥尔摩)、欧洲地区 (米兰)、欧洲地区 (法兰克福)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)

### 6.8.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.4.2	亚马逊 SageMaker Spark SDK
emr-ddb	4.16.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	3.2.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.5.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-notebook-env	1.7.0	适用于 EMR Notebooks ( 可提供 jupyter 企业网关 ) 的 Conda env
emr-s3-dist-cp	2.22.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	2.1.0	EMRS3Select 连接器
emrfs	2.53.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.15.1	Apache Flink 命令行客户端脚本和应用程序。
flink-jobmanager-config	1.15.1	管理 Apache Flink EMR JobManager 节点上的资源。

组件	版本	描述
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	3.2.1-amzn-8.1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	3.2.1-amzn-8.1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	3.2.1-amzn-8.1	HDFS命令行客户端和库
hadoop-hdfs-namenode	3.2.1-amzn-8.1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	3.2.1-amzn-8.1	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	3.2.1-amzn-8.1	HTTPHDFS操作端点。
hadoop-kms-server	3.2.1-amzn-8.1	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	3.2.1-amzn-8.1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。

组件	版本	描述
hadoop-yarn-nodemanager	3.2.1-amzn-8.1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	3.2.1-amzn-8.1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	3.2.1-amzn-8.1	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	2.4.12-amzn-0.1	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	2.4.12-amzn-0.1	为一个或多个HBase地区提供服务的服务。
hbase-client	2.4.12-amzn-0.1	HBase命令行客户端。
hbase-rest-server	2.4.12-amzn-0.1	为其提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	2.4.12-amzn-0.1	为其提供 Thrift 端点的HBase服务。
hbase-operator-tools	2.4.12-amzn-0.1	适用于 Apache HBase 集群的修复工具。
hcatalog-client	3.1.3-amzn-1.1	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	3.1.3-amzn-1.1	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	3.1.3-amzn-1.1	HTTP端点为其提供REST接口HCatalog。
hive-client	3.1.3-amzn-1.1	Hive 命令行客户端。

组件	版本	描述
hive-hbase	3.1.3-amzn-1.1	Hive-hbase 客户端。
hive-metastore-server	3.1.3-amzn-1.1	用于访问 Hive 元数据仓库的服务，这是一个语义存储库，用于存储 Hadoop 操作SQL的元数据。
hive-server2	3.1.3-amzn-1.1	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.11.1-amzn-0	增量处理框架，以支持低延迟和高效率的数据管道。
hudi-presto	0.11.1-amzn-0	用于运行 Presto 以及 Hudi 的捆绑库。
hudi-trino	0.11.1-amzn-0	用于运行 Trino 以及 Hudi 的捆绑库。
hudi-spark	0.11.1-amzn-0	用于运行 Spark 以及 Hudi 的捆绑库。
hue-server	4.10.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
iceberg	0.14.0-amzn-0	Apache Iceberg 是一种适用于超大型分析数据集的开放表格式。
jupyterhub	1.4.1	Jupyter notebook 的多用户服务器
livy-server	0.7.1-incubating	REST用于与 Apache Spark 交互的接口

组件	版本	描述
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mxnet	1.9.1	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.68+	MariaDB 数据库服务器。
nvidia-cuda	11.7.0	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.1	Oozie 命令行客户端。
oozie-server	5.2.1	用于接受 Oozie 工作流请求的服务。
opencv	4.5.0	开源计算机视觉库。
phoenix-library	5.1.2	服务器和客户端的 phoenix 库
phoenix-connectors	5.1.2	Apache Phoenix-Connectors for Spark-3
phoenix-query-server	5.1.2	轻量级服务器，提供JDBC对Avatica 的访问权限以及协议缓冲区和JSON格式访问权限 API
presto-coordinator	0.273.3-amzn-0	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.273.3-amzn-0	用于执行查询的各个部分的服务。
presto-client	0.273.3-amzn-0	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。

组件	版本	描述
trino-coordinator	388-amzn-0	用于在 trino-worker 之中接受查询并管理查询的服务。
trino-worker	388-amzn-0	用于执行查询的各个部分的服务。
trino-client	388-amzn-0	Trino 命令行客户端，安装在 HA 集群的备用主服务器（未启动 Trino 服务器）上。
pig-client	0.17.0	Pig 命令行客户端。
r	4.0.2	用于统计计算的 R 项目
ranger-kms-server	2.0.0	Apache Ranger 密钥管理系统
spark-client	3.3.0-amzn-0.1	Spark 命令行客户端。
spark-history-server	3.3.0-amzn-0.1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	3.3.0-amzn-0.1	的内存中执行引擎。YARN
spark-yarn-slave	3.3.0-amzn-0.1	YARN从属服务器需要的 Apache Spark 库。
spark-rapids	22.06.0-amzn-0	Nvidia Spark RAPIDS 插件可以加速 Apache Spark。GPUs
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	2.9.1	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.2	tez YARN 应用程序和库。

组件	版本	描述
webserver	2.4.41+	Apache HTTP 服务器。
zeppelin-server	0.10.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.5.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.5.10	ZooKeeper 命令行客户端。

### 6.8.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

为处于运行状态的集群中的实例组指定配置时，将发生重新配置操作。Amazon EMR 仅对您修改的分类启动重新配置操作。有关更多信息，请参阅 [在正在运行的集群中重新配置实例组](#)。

#### emr-6.8.1 分类

分类	描述	重新配置操作
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。	Restarts the ResourceManager service.
container-executor	更改 Hadoop YARN 的 container-executor.cfg 文件中的值。	Not available.
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。	Not available.
core-site	更改 Hadoop 的 core-site.xml 文件中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode,



分类	描述	重新配置操作
		Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	更改 docker 相关设置。	Not available.
emrfs-site	更改EMRFS设置。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	更改 flink-conf.yaml 设置。	Restarts Flink history server.
flink-log4j	更改 Flink log4j.properties 设置。	Restarts Flink history server.

分类	描述	重新配置操作
flink-log4j-session	为 Kubernetes/Yarn 会话更改 Flink log4j-session.properties 设置。	Restarts Flink history server.
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。	Restarts Flink history server.
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	更改 hadoop ssl 服务器配置	Not available.
hadoop-ssl-client	更改 hadoop ssl 客户端配置	Not available.

分类	描述	重新配置操作
hbase	亚马逊EMR精心策划的Apache 设置。HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	更改环境HBase中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	更改 hadoop-metrics2-hbase base.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。	Not available.
hbase-site	更改HBase的 hbase-site.xml 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	配置HDFS加密区域。	This classification should not be reconfigured.
hdfs-env	更改HDFS环境中的值。	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

分类	描述	重新配置操作
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	更改环境HCatalog中的值。	Restarts Hive HCatalog Server.
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。	Restarts Hive WebHCat server.
hive	亚马逊EMR精心策划的 Apache Hive 设置。	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。	Not available.
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。	Not available.

分类	描述	重新配置操作
hive-env	更改 Hive 环境中的值。	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	更改 Hive 的 hive-exec-log4j2.properties 文件中的值。	Not available.
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log4j2.properties 文件中的值。	Not available.
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。	Not available.
hive-site	更改 Hive 的 hive-site.xml 文件中的值	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值	Not available.
hue-ini	更改 Hue 的 ini 文件中的值	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	更改HTTPFS环境中的值。	Restarts Hadoop Httpfs service.
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。	Restarts Hadoop Httpfs service.

分类	描述	重新配置操作
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。	Not available.
hadoop-kms-env	在 Hadoop KMS 环境中更改值。	Restarts Hadoop-KMS service.
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。	Not available.
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	更改 Hudi 环境中的值。	Not available.
hudi-defaults	更改 Hudi 的 hudi-defaults.conf 文件中的值。	Not available.
iceberg-defaults	更改 Iceberg 的 iceberg-defaults.conf 文件中的值。	Not available.
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。	Not available.
jupyter-hub-conf	更改 JupyterHubs 的 jupyterhub_config.py 文件中的值。	Not available.
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。	Not available.
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。	Not available.
livy-conf	更改 Livy 的 livy.conf 文件中的值。	Restarts Livy Server.
livy-env	更改 Livy 环境中的值。	Restarts Livy Server.

分类	描述	重新配置操作
livy-log4j	更改 Livy log4j.properties 设置。	Restarts Livy Server.
mapred-env	更改 MapReduce 应用程序环境中的值。	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	更改 Oozie 的环境中的值。	Restarts Oozie.
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。	Restarts Oozie.
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。	Restarts Oozie.
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。	Not available.
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。	Not available.
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。	Restarts Phoenix-QueryServer.
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。	Not available.
pig-env	更改 Pig 环境中的值。	Not available.
pig-properties	更改 Pig 的 pig.properties 文件中的值。	Restarts Oozie.
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-log	更改 Presto 的 log.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-config	更改 Presto 的 config.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。	Not available.
presto-env	更改 Presto 的 presto-env.sh 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-node	更改 Presto 的 node.properties 文件中的值。	Not available.
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。	Not available.
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。	Not available.
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。	Not available.
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。	Not available.
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。	Not available.
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。	Not available.



分类	描述	重新配置操作
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。	Not available.
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。	Not available.
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。	Not available.
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。	Not available.
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。	Not available.
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。	Not available.
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。	Not available.
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。	Not available.
trino-log	更改 Trino 的 log.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-config	更改 Trino 的 config.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-password-authenticator	更改 Trino 的 password-authenticator.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-env	更改 Trino 的 trino-env.sh 文件中的值。	Restarts Trino-Server (for Trino)

分类	描述	重新配置操作
trino-node	更改 Trino 的 node.properties 文件中的值。	Not available.
trino-connector-blackhole	更改 Trino 的 blackhole.properties 文件中的值。	Not available.
trino-connector-cassandra	更改 Trino 的 cassandra.properties 文件中的值。	Not available.
trino-connector-hive	更改 Trino 的 hive.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-iceberg	更改 Trino 的 iceberg.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-jmx	更改 Trino 的 jmx.properties 文件中的值。	Not available.
trino-connector-kafka	更改 Trino 的 kafka.properties 文件中的值。	Not available.
trino-connector-localfile	更改 Trino 的 localfile.properties 文件中的值。	Not available.
trino-connector-memory	更改 Trino 的 memory.properties 文件中的值。	Not available.
trino-connector-mongodb	更改 Trino 的 mongodb.properties 文件中的值。	Not available.
trino-connector-mysql	更改 Trino 的 mysql.properties 文件中的值。	Not available.
trino-connector-postgresql	更改 Trino 的 postgresql.properties 文件中的值。	Not available.
trino-connector-raptor	更改 Trino 的 raptor.properties 文件中的值。	Not available.

分类	描述	重新配置操作
trino-connector-redis	更改 Trino 的 redis.properties 文件中的值。	Not available.
trino-connector-redshift	更改 Trino 的 redshift.properties 文件中的值。	Not available.
trino-connector-tpch	更改 Trino 的 tpch.properties 文件中的值。	Not available.
trino-connector-tpcds	更改 Trino 的 tpcds.properties 文件中的值。	Not available.
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-site	更改 Ranger KMS 的 ranger-kms-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-env	在 Ranger KMS 环境中更改值。	Restarts Ranger KMS Server.
ranger-kms-log4j	更改 Ranger 的 kms-log4j.properties 文件中的值。KMS	Not available.
ranger-kms-db-ca	更改 S3 上我与 Ranger SQL SSL KMS 连接的 CA 文件值。	Not available.
spark	亚马逊EMR精心策划的 Apache Spark 设置。	This property modifies spark-defaults. See actions there.
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-env	更改 Spark 环境中的值。	Restarts Spark history server and Spark thrift server.
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值	Not available.

分类	描述	重新配置操作
spark-log4j2	更改 Spark 的 log4j2.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
sqoop-env	更改 Sqoop 的环境中的值。	Not available.
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。	Not available.
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。	Not available.
tez-site	更改 Tez 的 tez-site.xml 文件中的值。	Restart Oozie and HiveServer2.
yarn-env	更改YARN环境中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	更改YARN的 yarn-site.xml 文件中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	更改 Zeppelin 环境中的值。	Restarts Zeppelin.
zeppelin-site	更改 zeppelin-site.xml 中的配置设置。	Restarts Zeppelin.

分类	描述	重新配置操作
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。	Restarts Zookeeper server.
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。	Restarts Zookeeper server.

## 6.8.1 更改日志

发行版 6.8.1 的更改日志和发布说明

Date	事件	描述
2023-08-30	更新发行说明	在发行说明中添加了几个与控制面板相关的修复
2023-08-21	文档发布	亚马逊 EMR 6.8.1 发行说明首次发布
2023-08-16	部署完成	Amazon EMR 6.8.1 已完全部署到所有 <a href="#">支持的区域</a>
2023-08-04	初始版本	亚马逊 EMR 6.8.1 首次部署到有限的商业区域

## 亚马逊EMR版本 6.8.0

### 6.8.0 应用程序版本

此发行版支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[Iceberg](#)、[JupyterEnterpriseGateway](#)和[ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

## 应用程序版本信息

	emr-6.8.0	emr-6.7.0	emr-6.6.0	emr-6.5.0
适用于 Java 的 AWS SDK	1.12.170	1.12.170	1.12.170	1.12.31
Python	2.7、3.7	2.7、3.7	2.7、3.7	2.7、3.7
Scala	2.12.15	2.12.15	2.12.10	2.12.10
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.15.1	1.14.2	1.14.2	1.14.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.12	2.4.4	2.4.4	2.4.4
HCatalog	3.1.3	3.1.3	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.3	3.1.3	3.1.2	3.1.2
Hudi	0.11.1-amzn-0	0.11.0-amzn-0	0.10.1-amzn-0	0.9.0-amzn-1
Hue	4.10.0	4.10.0	4.10.0	4.9.0
Iceberg	0.14.0-amzn-0	0.13.1-amzn-0	0.13.1	0.12.0

	emr-6.8.0	emr-6.7.0	emr-6.6.0	emr-6.5.0
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.8.0	1.8.0	1.8.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.273	0.272	0.267	0.261
Spark	3.3.0	3.2.1	3.2.0	3.1.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.9.1	2.4.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	388	378	367	360
Zeppelin	0.10.1	0.10.0	0.10.0	0.10.0
ZooKeeper	3.5.10	3.5.7	3.5.7	3.5.7

## 6.8.0 发布说明

以下发行说明包含 Amazon 6.8.0 EMR 版本的信息。更改与 6.7.0 有关。

## 新功能

- Amaz EMR on steps 功能现在支持 Apache Livy 终端节点和JDBC/ODBC客户端。有关更多信息，请参阅[为 Amazon EMR 步骤配置运行时角色](#)。
- 亚马逊EMR版本6.8.0附带了Apache HBase 版本2.4.12。在此HBase版本中，您可以存档和删除HBase表。Amazon S3 存档过程会将所有表文件重命名为存档目录。这一过程成本高昂且时间较长。现在，您可以跳过存档过程，快速删除大型表。有关更多信息，请参阅[使用外HBase壳](#)。

## 已知问题

- Hadoop 3.3.3 在 YARN ([YARN-9608](#)) 中引入了一项更改，即在应用程序完成之前，容器运行的节点一直处于停用状态。此更改可确保如随机数据等本地数据不会丢失，并且您无需重新运行作业。在 Amazon EMR 6.8.0 和 6.9.0 中，这种方法还可能导致启用或未启用托管扩展的集群上的资源利用不足。

在 [Amazon EMR 6.10.0](#) 中，有一种解决方法可以将 `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` 值设置为 `in`。 `false` `yarn-site.xml` 在亚马逊EMR版本 6.11.0 及更高版本以及 6.8.1、6.9.1 和 6.10.1 中，配置 `false` 默认设置为以解决此问题。

## 更改、增强和解决的问题

- 当亚马逊EMR发布 6.5.0、6.6.0 或 6.7.0 通过 Apache Spark shell 读取 Apache Phoenix 表时，亚马逊制作了一个。EMR NoSuchMethodError 亚马逊EMR版本 6.8.0 修复了这个问题。
- 亚马逊EMR版本6.8.0附带A [pache Hudi 0.11.1](#)；但是，EMR亚马逊6.8.0集群也与Hudi 0.12.0的开源集群兼容。 `hudi-spark3.3-bundle_2.12`
- 亚马逊EMR版本 6.8.0 附带了 Apache Spark 3.0。此 Spark 发行版使用 Apache Log4j 2 和 `log4j2.properties` 文件，在 Spark 进程中配置 Log4j。如果您在集群中使用 Spark 或使用自定义配置参数创建EMR集群，并且想要升级到 Amazon EMR 版本 6.8.0，则必须迁移到 Apache Log4j 2 的新 `spark-log4j2` 配置分类和密钥格式。有关更多信息，请参阅 [从 Apache Log4j 1.x 迁移到 Log4j 2.x](#)。
- 当您启动带有最新补丁版本（亚马逊 EMR 5.36 或更高版本、6.6 或更高版本、7.0 或更高版本）的集群时，亚马逊将EMR使用最新的亚马逊 Linux 2023 或亚马逊 Linux 2 版本作为默认亚马逊 Linux 2 版本。EMR AMI有关更多信息，请参阅[使用亚马逊的默认 Amazon Linux AMI EMR](#)。



**Note**

此版本不再获得自动AMI更新，因为又有 1 个补丁版本接替了该版本。补丁版本以第二位小数点后的数字 (6.8.**1**) 表示。要查看您是否使用的是最新补丁版本，请查看[发布指南](#)中的可用版本，或者在控制台中创建集群时查看 Amazon EMR 版本下拉列表，或者使用[ListReleaseLabels](#)API或[list-release-labels](#)CLI操作。要获取有关新版本的更新，请订阅“[新增内容？](#)”RSS 页面。

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 709.1	4.14.348	2024年7月23日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、AWS GovCloud (美国西部)、AWS GovCloud (美国

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			东部)、( 美国东部)、中国( 北京)、中国( 宁夏)、亚洲太平洋( 海得拉巴)、中东( UAE)、欧洲( 西班牙)、欧洲( 苏黎世)、亚太地区( 墨尔本)、以色列( 特拉维夫)
2.0.2023 808.0	4.14.320	2023 年 8 月 24 日	美国东部( 弗吉尼亚州北部)、美国东部( 俄亥俄州)、美国西部( 北加利福尼亚)、美国西部( 俄勒冈州)、亚太地区( 香港)、亚太地区( 孟买)、亚太地区( 东京)、亚太地区( 首尔)、亚太地区( 大阪)、亚太地区( 新加坡)、亚太地区( 悉尼)、亚太地区( 雅加达)、亚太地区( 墨尔本)、非洲( 开普敦)、南美洲( 圣保罗)、中东( 巴林)、加拿大( 中部)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 727.0	4.14.320	2023 年 8 月 14 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、欧洲地区 (斯德哥尔摩)、欧洲地区 (米兰)、欧洲地区 (法兰克福)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 719.0	4.14.320	2023 年 8 月 2 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中央 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 628.0	4.14.318	2023 年 7 月 12 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 612.0	4.14.314	2023 年 6 月 23 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 504.1	4.14.313	2023 年 5 月 16 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 418.0	4.14.311	2023 年 5 月 3 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 404.1	4.14.311	2023 年 4 月 18 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)
2.0.2023 404.0	4.14.311	2023 年 4 月 10 日	美国东部 (弗吉尼亚州北部)、欧洲地区 (巴黎)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 320.0	4.14.309	2023 年 3 月 30 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.202307.0	4.14.305	2023 年 3 月 15 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 207.0	4.14.304	2023 年 2 月 22 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023119.1	4.14.301	2023 年 2 月 3 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022 210.1	4.14.301	2023 年 12 月 22 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022 103.3	4.14.296	2022 年 12 月 5 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )



OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022004.0	4.14.294	2022 年 11 月 2 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022 912.1	4.14.291	2022 年 9 月 6 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

## 已知问题

- 当您在EMR亚马逊 6.0、6.7.0 和 6.8.0 版本上将 DynamoDB 连接器与 Spark 配合使用时，即使输入拆分引用非空数据，所有从表中读取的内容都会返回空结果。这是因为 Spark 3.2.0 将 `spark.hadoopRDD.ignoreEmptySplits` 默认设置为 `true`。解决方法是将 `spark.hadoopRDD.ignoreEmptySplits` 显式设置为 `false`。亚马逊 6.9.0 EMR 版本修复了这个问题。

- 当你使用带有 Hive 分区位置格式的 Spark 来读取 Amazon S3 中的数据，并在亚马逊EMR版本 5.30.0 到 5.36.0 和 6.2.0 到 6.9.0 上运行 Spark 时，你可能会遇到一个问题，导致集群无法正确读取数据。如果您的分区具有以下所有特征，会发生这种情况：
  - 从同一个表扫描两个或多个分区。
  - 至少有一个分区目录路径是至少一个其他分区目录路径的前缀，例如，s3://bucket/table/p=a 是 s3://bucket/table/p=a b 的前缀。
  - 另一个分区目录中前缀之后的第一个字符的值为 UTF -8，小于该/字符 (U+002F)。例如，在 s3://bucket/table/p=a b 中，a 和 b 之间出现的空格字符 (U+0020) 就属于此类。请注意，还有其他 14 个非控制字符：! "\$ % & ' ( ) \* + , - 。有关更多信息，请参阅 [UTF-8 编码表和 Unicode 字符](#)。

解决方法是在 spark-defaults 分类中将

spark.sql.sources.fastS3PartitionDiscovery.enabled 配置设置为 false。

- 在 Amazon 5.36.0 和 6.6.0 到 6.9.0 EMR 版本中，由于 Log4 SecretAgent j2 属性中的文件名模式配置不正确，RecordServer 服务组件可能会丢失日志数据。错误的配置导致组件每天只生成一个日志文件。当应用轮换策略时，它会重写现有文件，而不是按预期生成新的日志文件。应变方法是使用引导操作每小时生成一次日志文件，并在文件名中附加一个自动增量的整数来处理轮换。

对于 Amazon EMR 6.0 到 6.9.0 版本，请在启动集群时使用以下引导操作。

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

对于 Amazon EMR 5.36.0，请在启动集群时使用以下引导操作。

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

有关发布时间表的更多信息，请参阅[更改日志](#)。

## 6.8.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 emr 或 aws 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.4.2	亚马逊 SageMaker Spark SDK
emr-ddb	4.16.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	3.2.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.5.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-notebook-env	1.7.0	适用于 EMR Notebooks ( 可提供 jupyter 企业网关 ) 的 Conda env
emr-s3-dist-cp	2.22.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	2.1.0	EMRS3Select 连接器
emrfs	2.53.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.15.1	Apache Flink 命令行客户端脚本和应用程序。
flink-jobmanager-config	1.15.1	管理 Apache Flink EMR JobManager 节点上的资源。

组件	版本	描述
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	3.2.1-amzn-8	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	3.2.1-amzn-8	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	3.2.1-amzn-8	HDFS命令行客户端和库
hadoop-hdfs-namenode	3.2.1-amzn-8	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	3.2.1-amzn-8	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	3.2.1-amzn-8	HTTPHDFS操作端点。
hadoop-kms-server	3.2.1-amzn-8	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	3.2.1-amzn-8	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。

组件	版本	描述
hadoop-yarn-nodemanager	3.2.1-amzn-8	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	3.2.1-amzn-8	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	3.2.1-amzn-8	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	2.4.12-amzn-0	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	2.4.12-amzn-0	为一个或多个HBase地区提供服务的服务。
hbase-client	2.4.12-amzn-0	HBase命令行客户端。
hbase-rest-server	2.4.12-amzn-0	为其提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	2.4.12-amzn-0	为其提供 Thrift 端点的HBase服务。
hbase-operator-tools	2.4.12-amzn-0	适用于 Apache HBase 集群的修复工具。
hcatalog-client	3.1.3-amzn-1	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	3.1.3-amzn-1	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	3.1.3-amzn-1	HTTP端点为其提供REST接口HCatalog。
hive-client	3.1.3-amzn-1	Hive 命令行客户端。

组件	版本	描述
hive-hbase	3.1.3-amzn-1	Hive-hbase 客户端。
hive-metastore-server	3.1.3-amzn-1	用于访问 Hive 元数据仓库的服务，这是一个语义存储库，用于存储 Hadoop 操作SQL的元数据。
hive-server2	3.1.3-amzn-1	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.11.1-amzn-0	增量处理框架，以支持低延迟和高效率的数据管道。
hudi-presto	0.11.1-amzn-0	用于运行 Presto 以及 Hudi 的捆绑库。
hudi-trino	0.11.1-amzn-0	用于运行 Trino 以及 Hudi 的捆绑库。
hudi-spark	0.11.1-amzn-0	用于运行 Spark 以及 Hudi 的捆绑库。
hue-server	4.10.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
iceberg	0.14.0-amzn-0	Apache Iceberg 是一种适用于超大型分析数据集的开放表格式。
jupyterhub	1.4.1	Jupyter notebook 的多用户服务器
livy-server	0.7.1-incubating	REST用于与 Apache Spark 交互的接口

组件	版本	描述
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mxnet	1.9.1	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.68+	MariaDB 数据库服务器。
nvidia-cuda	11.7.0	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.1	Oozie 命令行客户端。
oozie-server	5.2.1	用于接受 Oozie 工作流请求的服务。
opencv	4.5.0	开源计算机视觉库。
phoenix-library	5.1.2	服务器和客户端的 phoenix 库
phoenix-connectors	5.1.2	Apache Phoenix-Connectors for Spark-3
phoenix-query-server	5.1.2	轻量级服务器，提供JDBC对Avatica 的访问权限以及协议缓冲区和JSON格式访问权限 API
presto-coordinator	0.273.3-amzn-0	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.273.3-amzn-0	用于执行查询的各个部分的服务。
presto-client	0.273.3-amzn-0	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。



组件	版本	描述
trino-coordinator	388-amzn-0	用于在 trino-worker 之中接受查询并管理查询的服务。
trino-worker	388-amzn-0	用于执行查询的各个部分的服务。
trino-client	388-amzn-0	Trino 命令行客户端，安装在 HA 集群的备用主服务器（未启动 Trino 服务器）上。
pig-client	0.17.0	Pig 命令行客户端。
r	4.0.2	用于统计计算的 R 项目
ranger-kms-server	2.0.0	Apache Ranger 密钥管理系统
spark-client	3.3.0-amzn-0	Spark 命令行客户端。
spark-history-server	3.3.0-amzn-0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	3.3.0-amzn-0	的内存中执行引擎。YARN
spark-yarn-slave	3.3.0-amzn-0	YARN从属服务器需要的 Apache Spark 库。
spark-rapids	22.06.0-amzn-0	Nvidia Spark RAPIDS 插件可以加速 Apache Spark。GPUs
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	2.9.1	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.2	tez YARN 应用程序和库。

组件	版本	描述
webserver	2.4.41+	Apache HTTP 服务器。
zeppelin-server	0.10.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.5.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.5.10	ZooKeeper 命令行客户端。

### 6.8.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

为处于运行状态的集群中的实例组指定配置时，将发生重新配置操作。Amazon EMR 仅对您修改的分类启动重新配置操作。有关更多信息，请参阅 [在正在运行的集群中重新配置实例组](#)。

#### emr-6.8.0 分类

分类	描述	重新配置操作
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。	Restarts the ResourceManager service.
container-executor	更改 Hadoop YARN 的 container-executor.cfg 文件中的值。	Not available.
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。	Not available.
core-site	更改 Hadoop 的 core-site.xml 文件中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode,

分类	描述	重新配置操作
		Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	更改 docker 相关设置。	Not available.
emrfs-site	更改EMRFS设置。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	更改 flink-conf.yaml 设置。	Restarts Flink history server.
flink-log4j	更改 Flink log4j.properties 设置。	Restarts Flink history server.

分类	描述	重新配置操作
flink-log4j-session	为 Kubernetes/Yarn 会话更改 Flink log4j-session.properties 设置。	Restarts Flink history server.
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。	Restarts Flink history server.
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	更改 hadoop ssl 服务器配置	Not available.
hadoop-ssl-client	更改 hadoop ssl 客户端配置	Not available.

分类	描述	重新配置操作
hbase	亚马逊EMR精心策划的 Apache 设置。HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	更改环境HBase中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	更改 hadoop-metrics2-hbase base.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。	Not available.
hbase-site	更改HBase的 hbase-site.xml 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	配置HDFS加密区域。	This classification should not be reconfigured.
hdfs-env	更改HDFS环境中的值。	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

分类	描述	重新配置操作
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	更改环境HCatalog中的值。	Restarts Hive HCatalog Server.
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。	Restarts Hive WebHCat server.
hive	亚马逊EMR精心策划的 Apache Hive 设置。	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。	Not available.
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。	Not available.

分类	描述	重新配置操作
hive-env	更改 Hive 环境中的值。	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	更改 Hive 的 hive-exec-log4j2.properties 文件中的值。	Not available.
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log4j2.properties 文件中的值。	Not available.
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。	Not available.
hive-site	更改 Hive 的 hive-site.xml 文件中的值	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值	Not available.
hue-ini	更改 Hue 的 ini 文件中的值	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	更改HTTPFS环境中的值。	Restarts Hadoop Httpfs service.
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。	Restarts Hadoop Httpfs service.

分类	描述	重新配置操作
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。	Not available.
hadoop-kms-env	在 Hadoop KMS 环境中更改值。	Restarts Hadoop-KMS service.
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。	Not available.
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	更改 Hudi 环境中的值。	Not available.
hudi-defaults	更改 Hudi 的 hudi-defaults.conf 文件中的值。	Not available.
iceberg-defaults	更改 Iceberg 的 iceberg-defaults.conf 文件中的值。	Not available.
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。	Not available.
jupyter-hub-conf	更改 JupyterHubs 的 jupyterhub_config.py 文件中的值。	Not available.
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。	Not available.
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。	Not available.
livy-conf	更改 Livy 的 livy.conf 文件中的值。	Restarts Livy Server.
livy-env	更改 Livy 环境中的值。	Restarts Livy Server.



分类	描述	重新配置操作
livy-log4j	更改 Livy log4j.properties 设置。	Restarts Livy Server.
mapred-env	更改 MapReduce 应用程序环境中的值。	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	更改 Oozie 的环境中的值。	Restarts Oozie.
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。	Restarts Oozie.
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。	Restarts Oozie.
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。	Not available.
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。	Not available.
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。	Restarts Phoenix-QueryServer.
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。	Not available.
pig-env	更改 Pig 环境中的值。	Not available.
pig-properties	更改 Pig 的 pig.properties 文件中的值。	Restarts Oozie.
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-log	更改 Presto 的 log.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-config	更改 Presto 的 config.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。	Not available.
presto-env	更改 Presto 的 presto-env.sh 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-node	更改 Presto 的 node.properties 文件中的值。	Not available.
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。	Not available.
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。	Not available.
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。	Not available.
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。	Not available.
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。	Not available.
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。	Not available.
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。	Not available.
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。	Not available.
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。	Not available.
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。	Not available.
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。	Not available.
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。	Not available.
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。	Not available.
trino-log	更改 Trino 的 log.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-config	更改 Trino 的 config.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-password-authenticator	更改 Trino 的 password-authenticator.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-env	更改 Trino 的 trino-env.sh 文件中的值。	Restarts Trino-Server (for Trino)

分类	描述	重新配置操作
trino-node	更改 Trino 的 node.properties 文件中的值。	Not available.
trino-connector-blackhole	更改 Trino 的 blackhole.properties 文件中的值。	Not available.
trino-connector-cassandra	更改 Trino 的 cassandra.properties 文件中的值。	Not available.
trino-connector-hive	更改 Trino 的 hive.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-iceberg	更改 Trino 的 iceberg.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-jmx	更改 Trino 的 jmx.properties 文件中的值。	Not available.
trino-connector-kafka	更改 Trino 的 kafka.properties 文件中的值。	Not available.
trino-connector-localfile	更改 Trino 的 localfile.properties 文件中的值。	Not available.
trino-connector-memory	更改 Trino 的 memory.properties 文件中的值。	Not available.
trino-connector-mongodb	更改 Trino 的 mongodb.properties 文件中的值。	Not available.
trino-connector-mysql	更改 Trino 的 mysql.properties 文件中的值。	Not available.
trino-connector-postgresql	更改 Trino 的 postgresql.properties 文件中的值。	Not available.
trino-connector-raptor	更改 Trino 的 raptor.properties 文件中的值。	Not available.

分类	描述	重新配置操作
trino-connector-redis	更改 Trino 的 redis.properties 文件中的值。	Not available.
trino-connector-redshift	更改 Trino 的 redshift.properties 文件中的值。	Not available.
trino-connector-tpch	更改 Trino 的 tpch.properties 文件中的值。	Not available.
trino-connector-tpcds	更改 Trino 的 tpcds.properties 文件中的值。	Not available.
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-site	更改 Ranger KMS 的 ranger-kms-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-env	在 Ranger KMS 环境中更改值。	Restarts Ranger KMS Server.
ranger-kms-log4j	更改 Ranger 的 kms-log4j.properties 文件中的值。KMS	Not available.
ranger-kms-db-ca	更改 S3 上我与 Ranger SQL SSL KMS 连接的 CA 文件值。	Not available.
spark	亚马逊EMR精心策划的 Apache Spark 设置。	This property modifies spark-defaults. See actions there.
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-env	更改 Spark 环境中的值。	Restarts Spark history server and Spark thrift server.
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值	Not available.

分类	描述	重新配置操作
spark-log4j2	更改 Spark 的 log4j2.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
sqoop-env	更改 Sqoop 的环境中的值。	Not available.
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。	Not available.
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。	Not available.
tez-site	更改 Tez 的 tez-site.xml 文件中的值。	Restart Oozie and HiveServer2.
yarn-env	更改YARN环境中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	更改YARN的 yarn-site.xml 文件中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	更改 Zeppelin 环境中的值。	Restarts Zeppelin.
zeppelin-site	更改 zeppelin-site.xml 中的配置设置。	Restarts Zeppelin.

分类	描述	重新配置操作
zookeeper-config	更改 ZooKeeper 的 zoo.cfg 文件中的值。	Restarts Zookeeper server.
zookeeper-log4j	更改 ZooKeeper 的 log4j.properties 文件中的值。	Restarts Zookeeper server.

## 6.8.0 更改日志

发行版 6.8.0 的更改日志和发布说明

Date	事件	描述
2023-08-21	更新	在 Hadoop 3.3.3 中添加了一个已知问题。
2023-07-26	更新	新的操作系统版本标签 2.0.20230612.0 和 2.0.20230628.0。
2022-09-06	部署完成	Amazon EMR 6.8 已完全部署到所有 <a href="#">受支持的区域</a>
2022-09-06	初次发布	亚马逊 EMR 6.8 发行说明首次发布
2022-08-31	初始版本	亚马逊 EMR 6.8 已在有限的商业区域发布

## 亚马逊EMR版本 6.7.0

### 6.7.0 应用程序版本

此发行版支持以下应用程序：

[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[Iceberg](#)、[JupyterEnterpriseGateway](#) 和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本 ( 如果适用 )。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-6.7.0	emr-6.6.0	emr-6.5.0	emr-6.4.0
适用于 Java 的 AWS SDK	1.12.170	1.12.170	1.12.31	1.12.31
Python	2.7、3.7	2.7、3.7	2.7、3.7	2.7、3.7
Scala	2.12.15	2.12.10	2.12.10	2.12.10
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.14.2	1.14.2	1.14.0	1.13.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.4	2.4.4	2.4.4	2.4.4
HCatalog	3.1.3	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.3	3.1.2	3.1.2	3.1.2
Hudi	0.11.0-amzn-0	0.10.1-amzn-0	0.9.0-amzn-1	0.8.0-amzn-0



	emr-6.7.0	emr-6.6.0	emr-6.5.0	emr-6.4.0
Hue	4.10.0	4.10.0	4.9.0	4.9.0
Iceberg	0.13.1-amzn-0	0.13.1	0.12.0	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.8.0	1.8.0	1.8.0	1.8.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.272	0.267	0.261	0.254.1
Spark	3.2.1	3.2.0	3.1.2	3.1.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	378	367	360	359
Zeppelin	0.10.0	0.10.0	0.10.0	0.9.0
ZooKeeper	3.5.7	3.5.7	3.5.7	3.5.7

## 6.7.0 发布说明

以下发行说明包含 Amazon 6.7. EMR 0 版本的信息。更改与 6.6.0 有关。

首次发布日期：2022 年 7 月 15 日

### 新功能

- 亚马逊EMR现在支持 Apache Spark 3.2.1、Apache Hive 3.1.3、0.11 HUDI、PrestoDB 0.272 和 Trino 0.378。
- 通过集群上的 Amazon EMR 步骤 ( Spark、Hive ) 支持基于IAM角色和湖泊组成的访问控制。EC2
- 在启用 Apache Ranger 的集群上支持 Apache Spark 数据定义语句。现在，这包括支持 Trino 应用程序在启用 Apache Ranger 的集群上读取和写入 Apache Hive 元数据。有关更多信息，请参阅在[亚马逊上使用 Trino 和 Apache Ranger 启用联合治理](#)。EMR
- 当您启动带有最新补丁版本 ( 亚马逊 EMR 5.36 或更高版本、6.6 或更高版本、7.0 或更高版本 ) 的集群时，亚马逊将EMR使用最新的亚马逊 Linux 2023 或亚马逊 Linux 2 版本作为默认亚马逊 Linux 2 版本。EMR AMI有关更多信息，请参阅[使用亚马逊的默认 Amazon Linux AMI EMR](#)。

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 709.1	4.14.348	2024年7月23日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、AWS GovCloud ( 美国东部 )、( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )、亚洲太平洋 ( 海得拉巴 )、中东 ( UAE )、欧洲 ( 西班牙 )、欧洲 ( 苏黎世 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 223.0	4.14.336	2024 年 3 月 8 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 131.0	4.14.336	2024年2月14日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			中国(北京)、中国(宁夏)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024124.0	4.14.336	2024 年 2 月 7 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 109.0	4.14.334	2024 年 1 月 24 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			中国(北京)、中国(宁夏)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 218.0	4.14.330	2024 年 1 月 2 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 206.0	4.14.330	2023 年 12 月 22 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			中国(北京)、中国(宁夏)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023116.0	4.14.328	2023 年 12 月 11 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 101.0	4.14.327	2023 年 11 月 16 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023020.1	4.14.326	2023 年 11 月 7 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023012.1	4.14.326	2023 年 10 月 26 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 926.0	4.14.322	2023 年 10 月 19 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 906.0	4.14.322	2023 年 10 月 4 日	中国 ( 北京 )、中国 ( 宁夏 )  美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 米兰 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 822.0	4.14.322	2023 年 8 月 30 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 米兰 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 808.0	4.14.320	2023 年 8 月 24 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 米兰 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 727.0	4.14.320	2023 年 8 月 14 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、欧洲地区 (斯德哥尔摩)、欧洲地区 (米兰)、欧洲地区 (法兰克福)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 719.0	4.14.320	2023 年 8 月 2 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 (巴林)、中东 (UAE)、加拿大 (中部)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 628.0	4.14.318	2023 年 7 月 12 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 612.0	4.14.314	2023 年 6 月 23 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 504.1	4.14.313	2023 年 5 月 16 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )



OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 418.0	4.14.311	2023 年 5 月 3 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 404.1	4.14.311	2023 年 4 月 18 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)
2.0.2023 404.0	4.14.311	2023 年 4 月 10 日	美国东部 (弗吉尼亚州北部)、欧洲地区 (巴黎)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 320.0	4.14.309	2023 年 3 月 30 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 307.0	4.14.305	2023 年 3 月 15 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 207.0	4.14.304	2023 年 2 月 22 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 119.1	4.14.301	2023 年 2 月 3 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022 210.1	4.14.301	2023 年 12 月 22 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022 103.3	4.14.296	2022 年 12 月 5 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022004.0	4.14.294	2022 年 11 月 2 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022 912.1	4.14.291	2022 年 10 月 7 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)
2.0.2022 719.0	4.14.287	2022 年 8 月 10 日	us-west-1 , eu-west-3 , eu-north-1 , ap-south-1 , me-south-1

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022 606.1	4.14.281	2022 年 7 月 15 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

## 已知问题

- 当亚马逊EMR版本 6.5.0、6.6.0 或 6.7.0 通过 Apache Spark shell 读取 Apache Phoenix 表时，会发生这种情况NoSuchMethodError是因为亚马逊使用了错误的。EMR Hbase.compat.version亚马逊EMR版本 6.8.0 修复了这个问题。
- 当您在EMR亚马逊 6.0、6.7.0 和 6.8.0 版本上将 DynamoDB 连接器与 Spark 配合使用时，即使输入拆分引用非空数据，所有从表中读取的内容都会返回空结果。这是因为 Spark

3.2.0 将 `spark.hadoopRDD.ignoreEmptySplits` 默认设置为 `true`。解决方法是将 `spark.hadoopRDD.ignoreEmptySplits` 显式设置为 `false`。亚马逊 6.9.0 EMR 版本修复了这个问题。

- 当你使用带有 Hive 分区位置格式的 Spark 来读取 Amazon S3 中的数据，并在亚马逊EMR版本 5.30.0 到 5.36.0 和 6.2.0 到 6.9.0 上运行 Spark 时，你可能会遇到一个问题，导致集群无法正确读取数据。如果您的分区具有以下所有特征，会发生这种情况：
  - 从同一个表扫描两个或多个分区。
  - 至少有一个分区目录路径是至少一个其他分区目录路径的前缀，例如，`s3://bucket/table/p=a` 是 `s3://bucket/table/p=a b` 的前缀。
  - 另一个分区目录中前缀之后的第一个字符的值为 UTF -8，小于该/字符 (U+002F)。例如，在 `s3://bucket/table/p=a b` 中，`a` 和 `b` 之间出现的空格字符 (U+0020) 就属于此类。请注意，还有其他 14 个非控制字符：`!"#$%&'()*+,-.`。有关更多信息，请参阅 [UTF-8 编码表和 Unicode 字符](#)。

解决方法是在 `spark-defaults` 分类中将 `spark.sql.sources.fastS3PartitionDiscovery.enabled` 配置设置为 `false`。

- 在 Amazon 5.36.0 和 6.6.0 到 6.9.0 EMR 版本中，由于 Log4 SecretAgent j2 属性中的文件名模式配置不正确，RecordServer 服务组件可能会丢失日志数据。错误的配置导致组件每天只生成一个日志文件。当应用轮换策略时，它会重写现有文件，而不是按预期生成新的日志文件。应变方法是使用引导操作每小时生成一次日志文件，并在文件名中附加一个自动增量的整数来处理轮换。

对于 Amazon EMR 6.0 到 6.9.0 版本，请在启动集群时使用以下引导操作。

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

对于 Amazon EMR 5.36.0，请在启动集群时使用以下引导操作。

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

- 在 Amazon EMR 6.7 或更低版本上运行的集群 `GetClusterSessionCredentialsAPI` 不支持。

## 6.7.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon

最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.4.1	亚马逊 SageMaker Spark SDK
emr-ddb	4.16.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	3.2.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.5.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-notebook-env	1.6.0	适用于 EMR Notebooks ( 可提供 jupyter 企业网关 ) 的 Conda env
emr-s3-dist-cp	2.22.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	2.1.0	EMRS3Select 连接器
emrfs	2.52.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.14.2	Apache Flink 命令行客户端脚本和应用程序。

组件	版本	描述
flink-jobmanager-config	1.14.2	管理 Apache Flink EMR JobManager 节点上的资源。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	3.2.1-amzn-7	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	3.2.1-amzn-7	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	3.2.1-amzn-7	HDFS命令行客户端和库
hadoop-hdfs-namenode	3.2.1-amzn-7	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	3.2.1-amzn-7	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	3.2.1-amzn-7	HTTPHDFS操作端点。
hadoop-kms-server	3.2.1-amzn-7	基于 Hadoop 的加密密钥管理服务。 KeyProvider API

组件	版本	描述
hadoop-mapred	3.2.1-amzn-7	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	3.2.1-amzn-7	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	3.2.1-amzn-7	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	3.2.1-amzn-7	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	2.4.4-amzn-3	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	2.4.4-amzn-3	为一个或多个HBase地区提供服务的服务。
hbase-client	2.4.4-amzn-3	HBase命令行客户端。
hbase-rest-server	2.4.4-amzn-3	为其提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	2.4.4-amzn-3	为其提供 Thrift 端点的HBase 服务。
hbase-operator-tools	2.4.4-amzn-3	适用于 Apache HBase 集群的修复工具。
hcatalog-client	3.1.3-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	3.1.3-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。

组件	版本	描述
hcatalog-webhcat-server	3.1.3-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	3.1.3-amzn-0	Hive 命令行客户端。
hive-hbase	3.1.3-amzn-0	Hive-hbase 客户端。
hive-metastore-server	3.1.3-amzn-0	用于访问 Hive 元数据仓库的服务，这是一个语义存储库，用于存储 Hadoop 操作SQL的元数据。
hive-server2	3.1.3-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.11.0-amzn-0	增量处理框架，以支持低延迟和高效率的数据管道。
hudi-presto	0.11.0-amzn-0	用于运行 Presto 以及 Hudi 的捆绑库。
hudi-trino	0.11.0-amzn-0	用于运行 Trino 以及 Hudi 的捆绑库。
hudi-spark	0.11.0-amzn-0	用于运行 Spark 以及 Hudi 的捆绑库。
hue-server	4.10.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
iceberg	0.13.1-amzn-0	Apache Iceberg 是一种适用于超大型分析数据集的开放表格式。
jupyterhub	1.4.1	Jupyter notebook 的多用户服务器



组件	版本	描述
livy-server	0.7.1-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mxnet	1.8.0	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.68+	MariaDB 数据库服务器。
nvidia-cuda	11.0.194	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.1	Oozie 命令行客户端。
oozie-server	5.2.1	用于接受 Oozie  workflow 请求的服务。
opencv	4.5.0	开源计算机视觉库。
phoenix-library	5.1.2	服务器和客户端的 phoenix 库
phoenix-connectors	5.1.2	Apache Phoenix-Connectors for Spark-3
phoenix-query-server	5.1.2	轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限 API
presto-coordinator	0.272-amzn-0	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.272-amzn-0	用于执行查询的各个部分的服务。

组件	版本	描述
presto-client	0.272-amzn-0	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
trino-coordinator	378-amzn-0	用于在 trino-worker 之中接受查询并管理查询的服务。
trino-worker	378-amzn-0	用于执行查询的各个部分的服务。
trino-client	378-amzn-0	Trino 命令行客户端，安装在 HA 集群的备用主服务器（未启动 Trino 服务器）上。
pig-client	0.17.0	Pig 命令行客户端。
r	4.0.2	用于统计计算的 R 项目
ranger-kms-server	2.0.0	Apache Ranger 密钥管理系统
spark-client	3.2.1-amzn-0	Spark 命令行客户端。
spark-history-server	3.2.1-amzn-0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	3.2.1-amzn-0	的内存中执行引擎。YARN
spark-yarn-slave	3.2.1-amzn-0	YARN从属服务器需要的 Apache Spark 库。
spark-rapids	22.02.0-amzn-1	Nvidia Spark RAPIDS 插件可以加速 Apache Spark。GPUs
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。

组件	版本	描述
tensorflow	2.4.1	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.2	tez YARN 应用程序和库。
webserver	2.4.41+	Apache HTTP 服务器。
zeppelin-server	0.10.0	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.5.7	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.5.7	ZooKeeper 命令行客户端。

## 6.7.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

为处于运行状态的集群中的实例组指定配置时，将发生重新配置操作。Amazon EMR 仅对您修改的分类启动重新配置操作。有关更多信息，请参阅 [在正在运行的集群中重新配置实例组](#)。

### emr-6.7.0 分类

分类	描述	重新配置操作
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。	Restarts the ResourceManager service.
container-executor	更改 Hadoop YARN 的 container-executor.cfg 文件中的值。	Not available.

分类	描述	重新配置操作
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。	Not available.
core-site	更改 Hadoop 的 core-site.xml 文件中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	更改 docker 相关设置。	Not available.

分类	描述	重新配置操作
emrfs-site	更改EMRFS设置。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	更改 flink-conf.yaml 设置。	Restarts Flink history server.
flink-log4j	更改 Flink log4j.properties 设置。	Restarts Flink history server.
flink-log4j-session	为 Kubernetes/Yarn 会话更改 Flink log4j-session.properties 设置。	Restarts Flink history server.
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。	Restarts Flink history server.

分类	描述	重新配置操作
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	更改 hadoop ssl 服务器配置	Not available.
hadoop-ssl-client	更改 hadoop ssl 客户端配置	Not available.
hbase	亚马逊EMR精心策划的 Apache 设置。HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.

分类	描述	重新配置操作
hbase-env	更改环境HBase中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	更改 hadoop-metrics2-hbase base.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。	Not available.
hbase-site	更改HBase的 hbase-site.xml 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	配置HDFS加密区域。	This classification should not be reconfigured.
hdfs-env	更改HDFS环境中的值。	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	更改环境HCatalog中的值。	Restarts Hive HCatalog Server.

分类	描述	重新配置操作
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。	Restarts Hive WebHCat server.
hive	亚马逊EMR精心策划的 Apache Hive 设置。	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。	Not available.
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。	Not available.
hive-env	更改 Hive 环境中的值。	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	更改 Hive 的 hive-exec-log4j2.properties 文件中的值。	Not available.
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log4j2.properties 文件中的值。	Not available.



分类	描述	重新配置操作
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。	Not available.
hive-site	更改 Hive 的 hive-site.xml 文件中的值	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值	Not available.
hue-ini	更改 Hue 的 ini 文件中的值	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	更改HTTPFS环境中的值。	Restarts Hadoop Httpfs service.
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。	Restarts Hadoop Httpfs service.
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。	Not available.
hadoop-kms-env	在 Hadoop KMS 环境中更改值。	Restarts Hadoop-KMS service.
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。	Not available.
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。	Restarts Hadoop-KMS and Ranger-KMS service.

分类	描述	重新配置操作
hudi-env	更改 Hudi 环境中的值。	Not available.
hudi-defaults	更改 Hudi 的 hudi-defaults.conf 文件中的值。	Not available.
iceberg-defaults	更改 Iceberg 的 iceberg-defaults.conf 文件中的值。	Not available.
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。	Not available.
jupyter-hub-conf	更改 JupyterHubs 的 jupyterhub_config.py 文件中的值。	Not available.
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。	Not available.
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。	Not available.
livy-conf	更改 Livy 的 livy.conf 文件中的值。	Restarts Livy Server.
livy-env	更改 Livy 环境中的值。	Restarts Livy Server.
livy-log4j	更改 Livy log4j.properties 设置。	Restarts Livy Server.
mapred-env	更改 MapReduce 应用程序环境中的值。	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	更改 Oozie 的环境中的值。	Restarts Oozie.

分类	描述	重新配置操作
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。	Restarts Oozie.
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。	Restarts Oozie.
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。	Not available.
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。	Not available.
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。	Restarts Phoenix-QueryServer.
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。	Not available.
pig-env	更改 Pig 环境中的值。	Not available.
pig-properties	更改 Pig 的 pig.properties 文件中的值。	Restarts Oozie.
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。	Not available.
presto-log	更改 Presto 的 log.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-config	更改 Presto 的 config.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-env	更改 Presto 的 presto-env.sh 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-node	更改 Presto 的 node.properties 文件中的值。	Not available.
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。	Not available.
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。	Not available.
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。	Not available.
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。	Not available.
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。	Not available.
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。	Not available.
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。	Not available.
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。	Not available.
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。	Not available.
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。	Not available.
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。	Not available.
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。	Not available.
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。	Not available.
trino-log	更改 Trino 的 log.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-config	更改 Trino 的 config.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-password-authenticator	更改 Trino 的 password-authenticator.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-env	更改 Trino 的 trino-env.sh 文件中的值。	Restarts Trino-Server (for Trino)
trino-node	更改 Trino 的 node.properties 文件中的值。	Not available.
trino-connector-blackhole	更改 Trino 的 blackhole.properties 文件中的值。	Not available.
trino-connector-cassandra	更改 Trino 的 cassandra.properties 文件中的值。	Not available.
trino-connector-hive	更改 Trino 的 hive.properties 文件中的值。	Restarts Trino-Server (for Trino)

分类	描述	重新配置操作
trino-connector-iceberg	更改 Trino 的 iceberg.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-jmx	更改 Trino 的 jmx.properties 文件中的值。	Not available.
trino-connector-kafka	更改 Trino 的 kafka.properties 文件中的值。	Not available.
trino-connector-localfile	更改 Trino 的 localfile.properties 文件中的值。	Not available.
trino-connector-memory	更改 Trino 的 memory.properties 文件中的值。	Not available.
trino-connector-mongodb	更改 Trino 的 mongodb.properties 文件中的值。	Not available.
trino-connector-mysql	更改 Trino 的 mysql.properties 文件中的值。	Not available.
trino-connector-postgresql	更改 Trino 的 postgresql.properties 文件中的值。	Not available.
trino-connector-raptor	更改 Trino 的 raptor.properties 文件中的值。	Not available.
trino-connector-redis	更改 Trino 的 redis.properties 文件中的值。	Not available.
trino-connector-redshift	更改 Trino 的 redshift.properties 文件中的值。	Not available.
trino-connector-tpch	更改 Trino 的 tpch.properties 文件中的值。	Not available.
trino-connector-tpcds	更改 Trino 的 tpcds.properties 文件中的值。	Not available.

分类	描述	重新配置操作
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-site	更改 Ranger KMS 的 ranger-kms-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-env	在 Ranger KMS 环境中更改值。	Restarts Ranger KMS Server.
ranger-kms-log4j	更改 Ranger 的 kms-log4j.properties 文件中的值。KMS	Not available.
ranger-kms-db-ca	更改 S3 上我与 Ranger SQL SSL KMS 连接的 CA 文件值。	Not available.
spark	亚马逊EMR精心策划的 Apache Spark 设置。	This property modifies spark-defaults. See actions there.
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-env	更改 Spark 环境中的值。	Restarts Spark history server and Spark thrift server.
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值	Not available.
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
sqoop-env	更改 Sqoop 的环境中的值。	Not available.
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。	Not available.

分类	描述	重新配置操作
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。	Not available.
tez-site	更改 Tez 的 tez-site.xml 文件中的值。	Restart Oozie and HiveServer2.
yarn-env	更改YARN环境中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	更改YARN的 yarn-site.xml 文件中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	更改 Zeppelin 环境中的值。	Restarts Zeppelin.
zeppelin-site	更改 zeppelin-site.xml 中的配置设置。	Restarts Zeppelin.
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。	Restarts Zookeeper server.
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。	Restarts Zookeeper server.



## 亚马逊EMR版本 6.6.0

### 6.6.0 应用程序版本

此发行版支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[Iceberg](#)、[JupyterEnterpriseGateway](#)和[ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-6.6.0	emr-6.5.0	emr-6.4.0	emr-6.3.1
适用于 Java 的 AWS SDK	1.12.170	1.12.31	1.12.31	1.11.977
Python	2.7、3.7	2.7、3.7	2.7、3.7	2.7、3.7
Scala	2.12.10	2.12.10	2.12.10	2.12.10
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.14.2	1.14.0	1.13.1	1.12.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.4	2.4.4	2.4.4	2.2.6

	emr-6.6.0	emr-6.5.0	emr-6.4.0	emr-6.3.1
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.10.1-amzn-0	0.9.0-amzn-1	0.8.0-amzn-0	0.7.0-amzn-0
Hue	4.10.0	4.9.0	4.9.0	4.9.0
Iceberg	0.13.1	0.12.0	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.2.2
Livy	0.7.1	0.7.1	0.7.1	0.7.0
MXNet	1.8.0	1.8.0	1.8.0	1.7.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.0.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.267	0.261	0.254.1	0.245.1
Spark	3.2.0	3.1.2	3.1.2	3.1.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2

	emr-6.6.0	emr-6.5.0	emr-6.4.0	emr-6.3.1
Trino (PrestoSQL)	367	360	359	350
Zeppelin	0.10.0	0.10.0	0.9.0	0.9.0
ZooKeeper	3.5.7	3.5.7	3.5.7	3.4.14

## 6.6.0 发布说明

以下发行说明包含 Amazon 6.6. EMR 0 版本的信息。更改与 6.5.0 有关。

首次发布日期：2022 年 5 月 9 日

文档更新日期：2022 年 6 月 15 日

### 新功能

- 亚马逊 EMR 6.6 现在支持 Apache Spark 3.2、Apache Spark RAPIDS 22.02、CUDA 11、Apache Hudi 0.10.1、Apache Iceberg 0.13、Trino 0.367 和 PrestoDB 0.267。
- 当您启动带有最新补丁版本（亚马逊 EMR 5.36 或更高版本、6.6 或更高版本、7.0 或更高版本）的集群时，亚马逊将EMR使用最新的亚马逊 Linux 2023 或亚马逊 Linux 2 版本作为默认亚马逊 Linux 2 版本。EMR AMI有关更多信息，请参阅[使用亚马逊的默认 Amazon Linux AMI EMR](#)。

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 709.1	4.14.348	2024年7月23日	美国东部（弗吉尼亚北部）、美国东部（俄亥俄州）、美国西部（加利福尼亚北部）、美国西部（俄勒冈）、欧洲（斯德哥尔摩）、欧

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 223.0	4.14.336	2024 年 3 月 8 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 131.0	4.14.336	2024年2月14日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			中国 ( 北京 )、中国 ( 宁夏 )



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024124.0	4.14.336	2024 年 2 月 7 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 109.0	4.14.334	2024 年 1 月 24 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			中国(北京)、中国(宁夏)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 218.0	4.14.330	2024 年 1 月 2 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 206.0	4.14.330	2023 年 12 月 22 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			中国 ( 北京 )、中国 ( 宁夏 )



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023116.0	4.14.328	2023 年 12 月 11 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 101.0	4.14.327	2023 年 11 月 16 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			中国(北京)、中国(宁夏)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023020.1	4.14.326	2023 年 11 月 7 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023012.1	4.14.326	2023 年 10 月 26 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			中国 ( 北京 )、中国 ( 宁夏 )



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 926.0	4.14.322	2023 年 10 月 19 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 906.0	4.14.322	2023 年 10 月 4 日	中国 (北京)、中国 (宁夏) 美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、欧洲地区 (斯德哥尔摩)、欧洲地区 (米兰)、欧洲地区 (法兰克福)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 822.0	4.14.322	2023 年 8 月 30 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 米兰 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 808.0	4.14.320	2023 年 8 月 24 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、欧洲地区 (斯德哥尔摩)、欧洲地区 (米兰)、欧洲地区 (法兰克福)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 727.0	4.14.320	2023 年 8 月 14 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 米兰 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 719.0	4.14.320	2023 年 8 月 2 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 (巴林)、中东 (UAE)、加拿大 (中部)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 628.0	4.14.318	2023 年 7 月 12 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 612.0	4.14.314	2023 年 6 月 23 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )



OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 504.1	4.14.313	2023 年 5 月 16 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 418.0	4.14.311	2023 年 5 月 3 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 404.1	4.14.311	2023 年 4 月 18 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)
2.0.2023 404.0	4.14.311	2023 年 4 月 10 日	美国东部 (弗吉尼亚州北部)、欧洲地区 (巴黎)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 320.0	4.14.309	2023 年 3 月 30 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 307.0	4.14.305	2023 年 3 月 15 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 207.0	4.14.304	2023 年 2 月 22 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 119.1	4.14.301	2023 年 2 月 3 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022 210.1	4.14.301	2023 年 12 月 22 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )



OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022 103.3	4.14.296	2022 年 12 月 5 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022004.0	4.14.294	2022 年 11 月 2 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022 912.1	4.14.291	2022 年 10 月 7 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)
2.0.2022 805.0	4.14.287	2022 年 8 月 30 日	us-west-1

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022 719.0	4.14.287	2022 年 8 月 10 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022 426.0	4.14.281	2022 年 6 月 10 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022 406.1	4.14.275	2022 年 5 月 2 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

- [在 Amazon EMR 6.6 及更高版本中，使用 Log4j 1.x 和 Log4j 2.x 的应用程序将分别升级为使用 Log4j 1.2.17 \(或更高版本\) 和 Log4j 2.17.1 \(或更高版本\)，并且不需要使用提供的引导操作来缓解问题。](#) CVE
- [托管扩展] Spark shuffle 数据托管扩展优化——对于 Amazon EMR 5.34.0 及更高版本以及 6.4.0 及更高EMR版本，托管扩展现在支持 Spark 洗牌数据 (Spark 跨分区重新分配以执行特定操作的数据)。有关洗牌操作的更多信息，请参阅《[亚马逊EMREMR管理指南](#)》和《[Spark 编程指南](#)》EMR 中的“[在亚马逊使用托管扩展](#)”。

- 从亚马逊 EMR 5.32.0 和 6.5.0 开始，Apache Spark 的动态执行器大小设置默认处于启用状态。要打开或关闭此功能，您可以使用 `spark.yarn.heterogeneousExecutors.enabled` 配置参数。

## 更改、增强和解决的问题

- 对于使用EMR默认AMI选项且仅安装常见应用程序（例如 Apache Hadoop、Apache Spark 和 Apache Hive）的集群，亚马逊平均将集群启动时间EMR缩短多达 80 秒。

## 已知问题

- 当亚马逊EMR版本 6.5.0、6.6.0 或 6.7.0 通过 Apache Spark shell 读取 Apache Phoenix 表时，会发生这种情况NoSuchMethodError是因为亚马逊使用了错误的。EMR Hbase.compat.version亚马逊EMR版本 6.8.0 修复了这个问题。
- 当您在EMR亚马逊 6.0、6.7.0 和 6.8.0 版本上将 DynamoDB 连接器与 Spark 配合使用时，即使输入拆分引用非空数据，所有从表中读取的内容都会返回空结果。这是因为 Spark 3.2.0 将 `spark.hadoopRDD.ignoreEmptySplits` 默认设置为 `true`。解决方法是将 `spark.hadoopRDD.ignoreEmptySplits` 显式设置为 `false`。亚马逊 6.9.0 EMR 版本修复了这个问题。
- 在 Trino 长期运行的集群上，Amazon EMR 6.0 在 Trino `jvm.config` 中启用了垃圾收集日志参数，以便从垃圾收集日志中获得更好的见解。此更改会将许多垃圾回收日志附加到 `launcher.log (/var/log/trino/launcher.log)` 文件。如果您在 Amazon EMR 6.6.0 中运行 Trino 集群，则由于附加的日志，在集群运行了几天后，您可能会遇到节点磁盘空间耗尽的情况。

此问题的解决方法是在为 Amazon 6.0 创建或克隆集群时，将以下脚本作为引导操作运行，以禁用 `jvm.config` 中的垃圾收集日志参数。EMR

```
#!/bin/bash
set -ex
PRESTO_PUPPET_DIR='/var/aws/emr/bigtop-deploy/puppet/modules/trino'
sudo bash -c "sed -i '/-Xlog/d' ${PRESTO_PUPPET_DIR}/templates/jvm.config"
```

- 当你使用带有 Hive 分区位置格式的 Spark 来读取 Amazon S3 中的数据，并在亚马逊EMR版本 5.30.0 到 5.36.0 和 6.2.0 到 6.9.0 上运行 Spark 时，你可能会遇到一个问题，导致集群无法正确读取数据。如果您的分区具有以下所有特征，会发生这种情况：
  - 从同一个表扫描两个或多个分区。

- 至少有一个分区目录路径是至少一个其他分区目录路径的前缀，例如，`s3://bucket/table/p=a` 是 `s3://bucket/table/p=a b` 的前缀。
- 另一个分区目录中前缀之后的第一个字符的值为 UTF-8，小于该字符 (U+002F)。例如，在 `s3://bucket/table/p=a b` 中，`a` 和 `b` 之间出现的空格字符 (U+0020) 就属于此类。请注意，还有其他 14 个非控制字符：`!"#$%&'()*+,-`。有关更多信息，请参阅 [UTF-8 编码表和 Unicode 字符](#)。

解决方法是在 `spark-defaults` 分类中将

`spark.sql.sources.fastS3PartitionDiscovery.enabled` 配置设置为 `false`。

- 在 Amazon 5.36.0 和 6.6.0 到 6.9.0 EMR 版本中，由于 Log4 SecretAgent `j2` 属性中的文件名模式配置不正确，RecordServer 服务组件可能会丢失日志数据。错误的配置导致组件每天只生成一个日志文件。当应用轮换策略时，它会重写现有文件，而不是按预期生成新的日志文件。应变方法是使用引导操作每小时生成一次日志文件，并在文件名中附加一个自动增量的整数来处理轮换。

对于 Amazon EMR 6.0 到 6.9.0 版本，请在启动集群时使用以下引导操作。

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

对于 Amazon EMR 5.36.0，请在启动集群时使用以下引导操作。

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

## 6.6.0 组件版本

下面列出了 Amazon 随此版本 EMR 安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的 EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新 EMR 版本中的大数据应用程序包通常是社区中的最新版本。我们会 EMR 尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本 EMR 不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。



组件	版本	描述
aws-sagemaker-spark-sdk	1.4.1	亚马逊 SageMaker Spark SDK
emr-ddb	4.16.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	3.2.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.5.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-notebook-env	1.5.0	适用于 EMR Notebooks ( 可提供 jupyter 企业网关 ) 的 Conda env
emr-s3-dist-cp	2.20.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	2.1.0	EMRS3Select 连接器
emrfs	2.50.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.14.2	Apache Flink 命令行客户端脚本和应用程序。
flink-jobmanager-config	1.14.2	管理 Apache Flink EMR JobManager 节点上的资源。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。

组件	版本	描述
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	3.2.1-amzn-6	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	3.2.1-amzn-6	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	3.2.1-amzn-6	HDFS命令行客户端和库
hadoop-hdfs-namenode	3.2.1-amzn-6	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	3.2.1-amzn-6	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	3.2.1-amzn-6	HTTPHDFS操作端点。
hadoop-kms-server	3.2.1-amzn-6	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	3.2.1-amzn-6	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	3.2.1-amzn-6	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	3.2.1-amzn-6	YARN用于分配和管理群集资源和分布式应用程序的服务。

组件	版本	描述
hadoop-yarn-timeline-server	3.2.1-amzn-6	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	2.4.4-amzn-2	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	2.4.4-amzn-2	为一个或多个HBase地区提供服务的服务。
hbase-client	2.4.4-amzn-2	HBase命令行客户端。
hbase-rest-server	2.4.4-amzn-2	为其提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	2.4.4-amzn-2	为其提供 Thrift 端点的HBase服务。
hbase-operator-tools	2.4.4-amzn-2	适用于 Apache HBase 集群的修复工具。
hcatalog-client	3.1.2-amzn-7	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	3.1.2-amzn-7	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	3.1.2-amzn-7	HTTP端点为其提供REST接口HCatalog。
hive-client	3.1.2-amzn-7	Hive 命令行客户端。
hive-hbase	3.1.2-amzn-7	Hive-hbase 客户端。
hive-metastore-server	3.1.2-amzn-7	用于访问 Hive 元数据仓库的服务，这是一个语义存储库，用于存储 Hadoop 操作SQL的元数据。

组件	版本	描述
hive-server2	3.1.2-amzn-7	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.10.1-amzn-0	增量处理框架，以支持低延迟和高效率的数据管道。
hudi-presto	0.10.1-amzn-0	用于运行 Presto 以及 Hudi 的捆绑库。
hudi-trino	0.10.1-amzn-0	用于运行 Trino 以及 Hudi 的捆绑库。
hudi-spark	0.10.1-amzn-0	用于运行 Spark 以及 Hudi 的捆绑库。
hue-server	4.10.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
iceberg	0.13.1	Apache Iceberg 是一种适用于超大型分析数据集的开放表格式。
jupyterhub	1.4.1	Jupyter notebook 的多用户服务器
livy-server	0.7.1-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mxnet	1.8.0	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.68+	MariaDB 数据库服务器。

组件	版本	描述
nvidia-cuda	11.0.194	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.1	Oozie 命令行客户端。
oozie-server	5.2.1	用于接受 Oozie 工作流请求的服务。
opencv	4.5.0	开源计算机视觉库。
phoenix-library	5.1.2	服务器和客户端的 phoenix 库
phoenix-connectors	5.1.2	Apache Phoenix-Connectors for Spark-3
phoenix-query-server	5.1.2	轻量级服务器，提供JDBC对Avatica 的访问权限以及协议缓冲区和JSON格式访问权限 API
presto-coordinator	0.267-amzn-0	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.267-amzn-0	用于执行查询的各个部分的服务。
presto-client	0.267-amzn-0	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
trino-coordinator	367-amzn-0	用于在 trino-worker 之中接受查询并管理查询的服务。
trino-worker	367-amzn-0	用于执行查询的各个部分的服务。
trino-client	367-amzn-0	Trino 命令行客户端，安装在 HA 集群的备用主服务器（未启动 Trino 服务器）上。

组件	版本	描述
pig-client	0.17.0	Pig 命令行客户端。
r	4.0.2	用于统计计算的 R 项目
ranger-kms-server	2.0.0	Apache Ranger 密钥管理系统
spark-client	3.2.0-amzn-0	Spark 命令行客户端。
spark-history-server	3.2.0-amzn-0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	3.2.0-amzn-0	的内存中执行引擎。YARN
spark-yarn-slave	3.2.0-amzn-0	YARN从属服务器需要的 Apache Spark 库。
spark-rapids	22.02.0-amzn-0	Nvidia Spark RAPIDS 插件可以加速 Apache Spark。GPUs
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	2.4.1	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.2	tez YARN 应用程序和库。
webserver	2.4.41+	Apache HTTP 服务器。
zeppelin-server	0.10.0	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.5.7	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.5.7	ZooKeeper 命令行客户端。

## 6.6.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

为处于运行状态的集群中的实例组指定配置时，将发生重新配置操作。Amazon EMR 仅对您修改的分类启动重新配置操作。有关更多信息，请参阅 [在正在运行的集群中重新配置实例组](#)。

### emr-6.6.0 分类

分类	描述	重新配置操作
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。	Restarts the Resource Manager service.
container-executor	更改 Hadoop YARN 的 container-executor.cfg 文件中的值。	Not available.
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。	Not available.
core-site	更改 Hadoop 的 core-site.xml 文件中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	更改 docker 相关设置。	Not available.

分类	描述	重新配置操作
emrfs-site	更改EMRFS设置。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	更改 flink-conf.yaml 设置。	Restarts Flink history server.
flink-log4j	更改 Flink log4j.properties 设置。	Restarts Flink history server.
flink-log4j-session	为 Kubernetes/Yarn 会话更改 Flink log4j-session.properties 设置。	Restarts Flink history server.
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。	Restarts Flink history server.



分类	描述	重新配置操作
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	更改 hadoop ssl 服务器配置	Not available.
hadoop-ssl-client	更改 hadoop ssl 客户端配置	Not available.
hbase	亚马逊EMR精心策划的 Apache 设置。HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.

分类	描述	重新配置操作
hbase-env	更改环境HBase中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	更改 hadoop-metrics2-hbase base.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。	Not available.
hbase-site	更改HBase的 hbase-site.xml 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	配置HDFS加密区域。	This classification should not be reconfigured.
hdfs-env	更改HDFS环境中的值。	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	更改环境HCatalog中的值。	Restarts Hive HCatalog Server.

分类	描述	重新配置操作
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。	Restarts Hive WebHCat server.
hive	亚马逊EMR精心策划的 Apache Hive 设置。	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。	Not available.
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。	Not available.
hive-env	更改 Hive 环境中的值。	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	更改 Hive 的 hive-exec-log4j2.properties 文件中的值。	Not available.
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log4j2.properties 文件中的值。	Not available.

分类	描述	重新配置操作
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。	Not available.
hive-site	更改 Hive 的 hive-site.xml 文件中的值	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值	Not available.
hue-ini	更改 Hue 的 ini 文件中的值	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	更改HTTPFS环境中的值。	Restarts Hadoop Httpfs service.
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。	Restarts Hadoop Httpfs service.
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。	Not available.
hadoop-kms-env	在 Hadoop KMS 环境中更改值。	Restarts Hadoop-KMS service.
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。	Not available.
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。	Restarts Hadoop-KMS and Ranger-KMS service.

分类	描述	重新配置操作
hudi-env	更改 Hudi 环境中的值。	Not available.
hudi-defaults	更改 Hudi 的 hudi-defaults.conf 文件中的值。	Not available.
iceberg-defaults	更改 Iceberg 的 iceberg-defaults.conf 文件中的值。	Not available.
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。	Not available.
jupyter-hub-conf	更改 JupyterHubs 的 jupyterhub_config.py 文件中的值。	Not available.
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。	Not available.
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。	Not available.
livy-conf	更改 Livy 的 livy.conf 文件中的值。	Restarts Livy Server.
livy-env	更改 Livy 环境中的值。	Restarts Livy Server.
livy-log4j	更改 Livy log4j.properties 设置。	Restarts Livy Server.
mapred-env	更改 MapReduce 应用程序环境中的值。	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	更改 Oozie 的环境中的值。	Restarts Oozie.

分类	描述	重新配置操作
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。	Restarts Oozie.
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。	Restarts Oozie.
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。	Not available.
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。	Not available.
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。	Restarts Phoenix-QueryServer.
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。	Not available.
pig-env	更改 Pig 环境中的值。	Not available.
pig-properties	更改 Pig 的 pig.properties 文件中的值。	Restarts Oozie.
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。	Not available.
presto-log	更改 Presto 的 log.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-config	更改 Presto 的 config.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-env	更改 Presto 的 presto-env.sh 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-node	更改 Presto 的 node.properties 文件中的值。	Not available.
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。	Not available.
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。	Not available.
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。	Not available.
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。	Not available.
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。	Not available.
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。	Not available.
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。	Not available.
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。	Not available.
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。	Not available.
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。	Not available.
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。	Not available.
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。	Not available.
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。	Not available.
trino-log	更改 Trino 的 log.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-config	更改 Trino 的 config.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-password-authenticator	更改 Trino 的 password-authenticator.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-env	更改 Trino 的 trino-env.sh 文件中的值。	Restarts Trino-Server (for Trino)
trino-node	更改 Trino 的 node.properties 文件中的值。	Not available.
trino-connector-blackhole	更改 Trino 的 blackhole.properties 文件中的值。	Not available.
trino-connector-cassandra	更改 Trino 的 cassandra.properties 文件中的值。	Not available.
trino-connector-hive	更改 Trino 的 hive.properties 文件中的值。	Restarts Trino-Server (for Trino)



分类	描述	重新配置操作
trino-connector-iceberg	更改 Trino 的 iceberg.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-jmx	更改 Trino 的 jmx.properties 文件中的值。	Not available.
trino-connector-kafka	更改 Trino 的 kafka.properties 文件中的值。	Not available.
trino-connector-localfile	更改 Trino 的 localfile.properties 文件中的值。	Not available.
trino-connector-memory	更改 Trino 的 memory.properties 文件中的值。	Not available.
trino-connector-mongodb	更改 Trino 的 mongodb.properties 文件中的值。	Not available.
trino-connector-mysql	更改 Trino 的 mysql.properties 文件中的值。	Not available.
trino-connector-postgresql	更改 Trino 的 postgresql.properties 文件中的值。	Not available.
trino-connector-raptor	更改 Trino 的 raptor.properties 文件中的值。	Not available.
trino-connector-redis	更改 Trino 的 redis.properties 文件中的值。	Not available.
trino-connector-redshift	更改 Trino 的 redshift.properties 文件中的值。	Not available.
trino-connector-tpch	更改 Trino 的 tpch.properties 文件中的值。	Not available.
trino-connector-tpcds	更改 Trino 的 tpcds.properties 文件中的值。	Not available.

分类	描述	重新配置操作
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-site	更改 Ran ranger-kms-site KMS ger.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-env	在游侠KMS环境中更改值。	Restarts Ranger KMS Server.
ranger-kms-log4j	更改 Ranger 的 kms-log4j.properties 文件中的值。KMS	Not available.
ranger-kms-db-ca	更改 S3 上我与 Ranger SQL SSL KMS 连接的 CA 文件值。	Not available.
spark	亚马逊EMR精心策划的 Apache Spark 设置。	This property modifies spark-defaults. See actions there.
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-env	更改 Spark 环境中的值。	Restarts Spark history server and Spark thrift server.
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值	Not available.
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
sqoop-env	更改 Sqoop 的环境中的值。	Not available.
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。	Not available.

分类	描述	重新配置操作
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。	Not available.
tez-site	更改 Tez 的 tez-site.xml 文件中的值。	Restart Oozie and HiveServer2.
yarn-env	更改YARN环境中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	更改YARN的 yarn-site.xml 文件中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	更改 Zeppelin 环境中的值。	Restarts Zeppelin.
zeppelin-site	更改 zeppelin-site.xml 中的配置设置。	Restarts Zeppelin.
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。	Restarts Zookeeper server.
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。	Restarts Zookeeper server.

## 亚马逊EMR版本 6.5.0

### 6.5.0 应用程序版本

此发行版支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[Iceberg](#)、[JupyterEnterpriseGateway](#)和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-6.5.0	emr-6.4.0	emr-6.3.1	emr-6.3.0
适用于 Java 的 AWS SDK	1.12.31	1.12.31	1.11.977	1.11.977
Python	2.7、3.7	2.7、3.7	2.7、3.7	2.7、3.7
Scala	2.12.10	2.12.10	2.12.10	2.12.10
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.14.0	1.13.1	1.12.1	1.12.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.4	2.4.4	2.2.6	2.2.6

	emr-6.5.0	emr-6.4.0	emr-6.3.1	emr-6.3.0
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.9.0-amzn-1	0.8.0-amzn-0	0.7.0-amzn-0	0.7.0-amzn-0
Hue	4.9.0	4.9.0	4.9.0	4.9.0
Iceberg	0.12.0	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.2.2	1.2.2
Livy	0.7.1	0.7.1	0.7.0	0.7.0
MXNet	1.8.0	1.8.0	1.7.0	1.7.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.0.0	5.0.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.261	0.254.1	0.245.1	0.245.1
Spark	3.1.2	3.1.2	3.1.1	3.1.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2

	emr-6.5.0	emr-6.4.0	emr-6.3.1	emr-6.3.0
Trino (PrestoSQL)	360	359	350	350
Zeppelin	0.10.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.5.7	3.5.7	3.4.14	3.4.14

## 6.5.0 发布说明

以下发行说明包含 Amazon 6.5.0 EMR 版本的信息。更改与 6.4.0 有关。

首次发布日期：2022 年 1 月 20 日

发布更新日期：2022 年 3 月 21 日

### 新功能

- [托管扩展] Spark shuffle 数据托管扩展优化——对于 Amazon EMR 5.34.0 及更高版本以及 6.4.0 及更高EMR版本，托管扩展现在支持 Spark 洗牌数据（Spark 跨分区重新分配以执行特定操作的数据）。有关洗牌操作的更多信息，请参阅《[亚马逊EMREMR管理指南](#)》和《[Spark 编程指南](#)》EMR 中的“[在亚马逊使用托管扩展](#)”。
- 从亚马逊 EMR 5.32.0 和 6.5.0 开始，Apache Spark 的动态执行器大小设置默认处于启用状态。要打开或关闭此功能，您可以使用 `spark.yarn.heterogeneousExecutors.enabled` 配置参数。
- 支持 Apache Iceberg 开放表格式，用于大型分析数据集。
- 支持 ranger-trino-plugin 2.0.1-amzn-1
- 支持 toree 0.5.0

### 更改、增强和解决的问题

- 亚马逊 EMR 6.5 发行版现在支持 Apache Iceberg 0.12.0，并通过适用于 Apache Spark 的亚马逊 Runtime、适用于 Presto 的亚马逊 EMR Runtime 和 Apache Hive 的EMR亚马逊 Runtime 提供了运行时改进。EMR

- [Apache Iceberg](#) 是一种适用于 Amazon S3 中大型数据集的开放表格式，可针对大型表、原子提交、并发写入和SQL兼容表演提供快速查询性能。在 EMR 6.5 版本中，你可以使用带有 Iceberg 表格式的 Apache Spark 3.1.2。
- Apache Hudi 0.9 增加了 Spark SQL DDL 和支援。DML这允许您仅SQL使用语句创建、更新插入 Hudi 表。Apache Hudi 0.9 还包括查询端和写入器端的性能改进。
- Amazon EMR Runtime for Apache Hive 通过删除暂存操作期间的重命名操作提高了 Apache Hive 在 Amazon S3 上的性能，并提高了用于修复表的元数据仓检查 (MSCK) 命令的性能。

## 已知问题

- 当亚马逊EMR版本 6.5.0、6.6.0 或 6.7.0 通过 Apache Spark shell 读取 Apache Phoenix 表时，会发生这种情况NoSuchMethodError是因为亚马逊使用了错误的。EMR Hbase.compat.version亚马逊EMR版本 6.8.0 修复了这个问题。
- 高可用性 ( HA ) 的 Hbase 捆绑集群无法使用默认卷大小和实例类型进行预置。此问题的变通解决方法是增加根卷大小。
- 要将 Spark 操作与 Apache Oozie 一起使用，必须将以下配置添加到 Oozie workflow.xml 文件中。否则，Oozie 启动的 Spark 执行程序的路径中EMRFS将缺少几个关键库，例如 Hadoop 和。

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

- 当你使用带有 Hive 分区位置格式的 Spark 来读取 Amazon S3 中的数据，并在亚马逊EMR版本 5.30.0 到 5.36.0 和 6.2.0 到 6.9.0 上运行 Spark 时，你可能会遇到一个问题，导致集群无法正确读取数据。如果您的分区具有以下所有特征，会发生这种情况：
  - 从同一个表扫描两个或多个分区。
  - 至少有一个分区目录路径是至少一个其他分区目录路径的前缀，例如，s3://bucket/table/p=a 是 s3://bucket/table/p=a b 的前缀。
  - 另一个分区目录中前缀之后的第一个字符的值为 UTF -8，小于该/字符 (U+002F)。例如，在 s3://bucket/table/p=a b 中，a 和 b 之间出现的空格字符 (U+0020) 就属于此类。请注意，还有其他 14 个非控制字符：! "\$ % & ' ( ) \* + , - 。有关更多信息，请参阅 [UTF-8 编码表和 Unicode 字符](#)。

解决方法是在 spark-defaults 分类中将

spark.sql.sources.fastS3PartitionDiscovery.enabled 配置设置为 false。

## 6.5.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.4.1	亚马逊 SageMaker Spark SDK
emr-ddb	4.16.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	3.2.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.5.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-notebook-env	1.4.0	适用于 EMR Notebooks ( 可提供 jupyter 企业网关 ) 的 Conda env
emr-s3-dist-cp	2.19.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	2.1.0	EMRS3Select 连接器
emrfs	2.48.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。



组件	版本	描述
flink-client	1.14.0	Apache Flink 命令行客户端脚本和应用程序。
flink-jobmanager-config	1.14.0	管理 Apache Flink EMR JobManager 节点上的资源。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	3.2.1-amzn-5	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	3.2.1-amzn-5	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	3.2.1-amzn-5	HDFS命令行客户端和库
hadoop-hdfs-namenode	3.2.1-amzn-5	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	3.2.1-amzn-5	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	3.2.1-amzn-5	HTTPHDFS操作端点。
hadoop-kms-server	3.2.1-amzn-5	基于 Hadoop 的加密密钥管理服务。 KeyProvider API

组件	版本	描述
hadoop-mapred	3.2.1-amzn-5	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	3.2.1-amzn-5	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	3.2.1-amzn-5	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	3.2.1-amzn-5	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	2.4.4-amzn-1	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	2.4.4-amzn-1	为一个或多个HBase地区提供服务的服务。
hbase-client	2.4.4-amzn-1	HBase命令行客户端。
hbase-rest-server	2.4.4-amzn-1	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	2.4.4-amzn-1	为其提供 Thrift 端点的HBase服务。
hcatalog-client	3.1.2-amzn-6	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	3.1.2-amzn-6	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	3.1.2-amzn-6	HTTP端点为其提供REST接口HCatalog。
hive-client	3.1.2-amzn-6	Hive 命令行客户端。

组件	版本	描述
hive-hbase	3.1.2-amzn-6	Hive-hbase 客户端。
hive-metastore-server	3.1.2-amzn-6	用于访问 Hive 元数据仓库的服务，这是一个语义存储库，用于存储 Hadoop 操作SQL的元数据。
hive-server2	3.1.2-amzn-6	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.9.0-amzn-1	增量处理框架，以支持低延迟和高效率的数据管道。
hudi-presto	0.9.0-amzn-1	用于运行 Presto 以及 Hudi 的捆绑库。
hudi-trino	0.9.0-amzn-1	用于运行 Trino 以及 Hudi 的捆绑库。
hudi-spark	0.9.0-amzn-1	用于运行 Spark 以及 Hudi 的捆绑库。
hue-server	4.9.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
iceberg	0.12.0	Apache Iceberg 是一种适用于超大型分析数据集的开放表格式。
jupyterhub	1.4.1	Jupyter notebook 的多用户服务器
livy-server	0.7.1-incubating	REST用于与 Apache Spark 交互的接口

组件	版本	描述
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mxnet	1.8.0	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.68+	MariaDB 数据库服务器。
nvidia-cuda	10.1.243	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.1	Oozie 命令行客户端。
oozie-server	5.2.1	用于接受 Oozie 工作流请求的服务。
opencv	4.5.0	开源计算机视觉库。
phoenix-library	5.1.2	服务器和客户端的 phoenix 库
phoenix-query-server	5.1.2	轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限 API
presto-coordinator	0.261-amzn-0	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.261-amzn-0	用于执行查询的各个部分的服务。
presto-client	0.261-amzn-0	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
trino-coordinator	360	用于在 trino-worker 之中接受查询并管理查询的服务。

组件	版本	描述
trino-worker	360	用于执行查询的各个部分的服务。
trino-client	360	Trino 命令行客户端，安装在 HA 集群的备用主服务器（未启动 Trino 服务器）上。
pig-client	0.17.0	Pig 命令行客户端。
r	4.0.2	用于统计计算的 R 项目
ranger-kms-server	2.0.0	Apache Ranger 密钥管理系统
spark-client	3.1.2-amzn-1	Spark 命令行客户端。
spark-history-server	3.1.2-amzn-1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	3.1.2-amzn-1	的内存中执行引擎。YARN
spark-yarn-slave	3.1.2-amzn-1	YARN从属服务器需要的 Apache Spark 库。
spark-rapids	0.4.1	Nvidia Spark RAPIDS 插件可以加速 Apache Spark。GPUs
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	2.4.1	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.2	tez YARN 应用程序和库。
webserver	2.4.41+	Apache HTTP 服务器。

组件	版本	描述
zeppelin-server	0.10.0	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.5.7	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.5.7	ZooKeeper 命令行客户端。

## 6.5.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

为处于运行状态的集群中的实例组指定配置时，将发生重新配置操作。Amazon EMR 仅对您修改的分类启动重新配置操作。有关更多信息，请参阅 [在正在运行的集群中重新配置实例组](#)。

### emr-6.5.0 分类

分类	描述	重新配置操作
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。	Restarts the ResourceManager service.
container-executor	更改 Hadoop YARN 的 container-executor.cfg 文件中的值。	Not available.
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。	Not available.
core-site	更改 Hadoop 的 core-site.xml 文件中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

分类	描述	重新配置操作
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	更改 docker 相关设置。	Not available.
emrfs-site	更改EMRFS设置。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	更改 flink-conf.yaml 设置。	Restarts Flink history server.
flink-log4j	更改 Flink log4j.properties 设置。	Restarts Flink history server.
flink-log4j-session	为 Kubernetes/Yarn 会话更改 Flink log4j-session.properties 设置。	Restarts Flink history server.

分类	描述	重新配置操作
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。	Restarts Flink history server.
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	更改 hadoop ssl 服务器配置	Not available.
hadoop-ssl-client	更改 hadoop ssl 客户端配置	Not available.



分类	描述	重新配置操作
hbase	亚马逊EMR精心策划的Apache 设置。HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	更改环境HBase中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	更改 hadoop-metrics2-hbase base.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。	Not available.
hbase-site	更改HBase的 hbase-site.xml 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	配置HDFS加密区域。	This classification should not be reconfigured.
hdfs-env	更改HDFS环境中的值。	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

分类	描述	重新配置操作
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	更改环境HCatalog中的值。	Restarts Hive HCatalog Server.
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。	Restarts Hive WebHCat server.
hive	亚马逊EMR精心策划的 Apache Hive 设置。	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。	Not available.
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。	Not available.

分类	描述	重新配置操作
hive-env	更改 Hive 环境中的值。	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	更改 Hive 的 hive-exec-log4j2.properties 文件中的值。	Not available.
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log4j2.properties 文件中的值。	Not available.
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。	Not available.
hive-site	更改 Hive 的 hive-site.xml 文件中的值	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值	Not available.
hue-ini	更改 Hue 的 ini 文件中的值	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	更改HTTPFS环境中的值。	Restarts Hadoop Httpfs service.
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。	Restarts Hadoop Httpfs service.

分类	描述	重新配置操作
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。	Not available.
hadoop-kms-env	在 Hadoop KMS 环境中更改值。	Restarts Hadoop-KMS service.
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。	Not available.
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	更改 Hudi 环境中的值。	Not available.
hudi-defaults	更改 Hudi 的 hudi-defaults.conf 文件中的值。	Not available.
iceberg-defaults	更改 Iceberg 的 iceberg-defaults.conf 文件中的值。	Not available.
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。	Not available.
jupyter-hub-conf	更改 JupyterHubs 的 jupyterhub_config.py 文件中的值。	Not available.
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。	Not available.
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。	Not available.
livy-conf	更改 Livy 的 livy.conf 文件中的值。	Restarts Livy Server.
livy-env	更改 Livy 环境中的值。	Restarts Livy Server.

分类	描述	重新配置操作
livy-log4j	更改 Livy log4j.properties 设置。	Restarts Livy Server.
mapred-env	更改 MapReduce 应用程序环境中的值。	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	更改 Oozie 的环境中的值。	Restarts Oozie.
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。	Restarts Oozie.
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。	Restarts Oozie.
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。	Not available.
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。	Not available.
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。	Restarts Phoenix-QueryServer.
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。	Not available.
pig-env	更改 Pig 环境中的值。	Not available.
pig-properties	更改 Pig 的 pig.properties 文件中的值。	Restarts Oozie.
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-log	更改 Presto 的 log.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-config	更改 Presto 的 config.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。	Not available.
presto-env	更改 Presto 的 presto-env.sh 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-node	更改 Presto 的 node.properties 文件中的值。	Not available.
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。	Not available.
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。	Not available.
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。	Not available.
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。	Not available.
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。	Not available.
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。	Not available.
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。	Not available.
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。	Not available.
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。	Not available.
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。	Not available.
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。	Not available.
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。	Not available.
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。	Not available.
trino-log	更改 Trino 的 log.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-config	更改 Trino 的 config.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-password-authenticator	更改 Trino 的 password-authenticator.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-env	更改 Trino 的 trino-env.sh 文件中的值。	Restarts Trino-Server (for Trino)

分类	描述	重新配置操作
trino-node	更改 Trino 的 node.properties 文件中的值。	Not available.
trino-connector-blackhole	更改 Trino 的 blackhole.properties 文件中的值。	Not available.
trino-connector-cassandra	更改 Trino 的 cassandra.properties 文件中的值。	Not available.
trino-connector-hive	更改 Trino 的 hive.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-jmx	更改 Trino 的 jmx.properties 文件中的值。	Not available.
trino-connector-kafka	更改 Trino 的 kafka.properties 文件中的值。	Not available.
trino-connector-localfile	更改 Trino 的 localfile.properties 文件中的值。	Not available.
trino-connector-memory	更改 Trino 的 memory.properties 文件中的值。	Not available.
trino-connector-mongodb	更改 Trino 的 mongodb.properties 文件中的值。	Not available.
trino-connector-mysql	更改 Trino 的 mysql.properties 文件中的值。	Not available.
trino-connector-postgresql	更改 Trino 的 postgresql.properties 文件中的值。	Not available.
trino-connector-raptor	更改 Trino 的 raptor.properties 文件中的值。	Not available.
trino-connector-redis	更改 Trino 的 redis.properties 文件中的值。	Not available.



分类	描述	重新配置操作
trino-connector-redshift	更改 Trino 的 redshift.properties 文件中的值。	Not available.
trino-connector-tpch	更改 Trino 的 tpch.properties 文件中的值。	Not available.
trino-connector-tpcds	更改 Trino 的 tpcds.properties 文件中的值。	Not available.
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-site	更改 Ranger KMS 的 ranger-kms-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-env	在 Ranger KMS 环境中更改值。	Restarts Ranger KMS Server.
ranger-kms-log4j	更改 Ranger 的 kms-log4j.properties 文件中的值。KMS	Not available.
ranger-kms-db-ca	更改 S3 上我与 Ranger SQL SSL KMS 连接的 CA 文件值。	Not available.
spark	亚马逊EMR精心策划的 Apache Spark 设置。	This property modifies spark-defaults. See actions there.
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-env	更改 Spark 环境中的值。	Restarts Spark history server and Spark thrift server.
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值	Not available.
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。	Restarts Spark history server and Spark thrift server.

分类	描述	重新配置操作
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
sqoop-env	更改 Sqoop 的环境中的值。	Not available.
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。	Not available.
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。	Not available.
tez-site	更改 Tez 的 tez-site.xml 文件中的值。	Restart Oozie and HiveServer2.
yarn-env	更改YARN环境中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	更改YARN的 yarn-site.xml 文件中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	更改 Zeppelin 环境中的值。	Restarts Zeppelin.
zeppelin-site	更改 zeppelin-site.xml 中的配置设置。	Restarts Zeppelin.
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。	Restarts Zookeeper server.

分类	描述	重新配置操作
zookeeper-log4j	更改 ZooKeeper 的 log4j.properties 文件中的值。	Restarts Zookeeper server.

## 亚马逊EMR版本 6.4.0

### 6.4.0 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[JupyterEnterpriseGateway](#)、[Jupyter](#) 和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-6.4.0	emr-6.3.1	emr-6.3.0	emr-6.2.1
适用于 Java 的 AWS SDK	1.12.31	1.11.977	1.11.977	1.11.880
Python	2.7、3.7	2.7、3.7	2.7、3.7	2.7、3.7
Scala	2.12.10	2.12.10	2.12.10	2.12.10
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-

	emr-6.4.0	emr-6.3.1	emr-6.3.0	emr-6.2.1
Flink	1.13.1	1.12.1	1.12.1	1.11.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.4	2.2.6	2.2.6	2.2.6-amzn-0
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.8.0-amzn-0	0.7.0-amzn-0	0.7.0-amzn-0	0.6.0-amzn-1
Hue	4.9.0	4.9.0	4.9.0	4.8.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.2.2	1.2.2	1.1.0
Livy	0.7.1	0.7.0	0.7.0	0.7.0
MXNet	1.8.0	1.7.0	1.7.0	1.7.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.0
Phoenix	5.1.2	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.254.1	0.245.1	0.245.1	0.238.3
Spark	3.1.2	3.1.1	3.1.1	3.0.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7

	emr-6.4.0	emr-6.3.1	emr-6.3.0	emr-6.2.1
TensorFlow	2.4.1	2.4.1	2.4.1	2.3.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	359	350	350	343
Zeppelin	0.9.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.5.7	3.4.14	3.4.14	3.4.14

## 6.4.0 发布说明

以下发行说明包含有关亚马逊 6.4.0 EMR 版本的信息。更改与 6.3.0 有关。

首次发布日期：2021 年 9 月 20 日

发布更新日期：2022 年 3 月 21 日

### 受支持的应用程序

- AWS SDK for Java 版本 1.12.31
- CloudWatch 水槽版本 2.2.0
- DynamoDB 连接器 4.16.0
- EMRFS版本 2.47.0
- 亚马逊 EMR Goodies 版本 3.2.0
- 亚马逊 EMR Kinesis Connector 版本 3.5.0
- 亚马逊EMR录音服务器版本 2.1.0
- 亚马逊EMR脚本版本 2.5.0
- Flink 1.13.1
- Ganglia 3.7.2
- AWS Glue Hive Metastore 客户端版本 3.3.0
- Hadoop 3.2.1-amzn-4

- HBase版本 2.4.4-amzn-0
- HBase-操作员工具 1.1.0
- HCatalog版本 3.1.2-amzn-5
- Hive 3.1.2-amzn-5
- Hudi 0.8.0-amzn-0
- Hue 4.9.0
- Java JDK 版本 Corretto-8.302.08.1 ( 版本 1.8.0\_302-b08 )
- JupyterHub 版本 1.4.1
- Livy 0.7.1-incubating
- MXNet版本 1.8.0
- Oozie 5.2.1
- Phoenix 5.1.2
- Pig 0.17.0
- Presto 0.254.1-amzn-0
- Trino 359
- Apache RangerKMS ( 多主机透明加密 ) 版本 2.0.0
- ranger-plugins 2.0.1-amzn-0
- ranger-s3-plugin 1.2.0
- SageMaker Spark SDK 版本 1.4.1
- Scala 版本 2.12.10 ( 打开 JDK 64 位服务器虚拟机 , Java 1.8.0\_282 )
- Spark 3.1.2-amzn-0
- spark-rapids 0.4.1
- Sqoop 1.4.7
- TensorFlow 版本 2.4.1
- tez 0.9.2
- Zeppelin 0.9.0
- Zookeeper 3.5.7
- 连接器和驱动程序 : DynamoDB 连接器 4.16.0

## 新功能

- [托管扩展] Spark shuffle 数据托管扩展优化 ——对于 Amazon EMR 5.34.0 及更高版本以及 6.4.0 及更高EMR版本，托管扩展现在支持 Spark 洗牌数据（Spark 跨分区重新分配以执行特定操作的数据）。有关洗牌操作的更多信息，请参阅《[亚马逊EMREMR管理指南](#)》和《[Spark 编程指南](#)》EMR 中的“[在亚马逊使用托管扩展](#)”。
- 在支持 Apache Ranger 的EMR亚马逊集群上，你可以使用、和使用 Apache SQL Spark 在 Apache Hive 元数据仓表中插入数据或更新 Apache Hive 元数据仓表。INSERT INTO INSERT OVERWRITE ALTER TABLEALTERTABLE与 Spark 一起使用时SQL，分区位置必须是表位置的子目录。Amazon 目前EMR不支持在分区位置与表位置不同的分区中插入数据。
- Presto SQL 已[更名为 Trino](#)。
- Hive：SELECT获取到LIMIT子句中提到的记录数后，立即停止查询执行，从而加快带LIMIT子句的简单查询的执行速度。简单SELECT查询是指没有 BY/b GROUP y 子句ORDER的查询或没有 reducer 阶段的查询。例如，SELECT \* from <TABLE> WHERE <Condition> LIMIT <Number>。

## Hudi 并发控制

- Hudi 现在支持 Optimistic Concurrency Control (OCC)，它可以与写入操作（如UPSERT和）一起使用，INSERT以允许多个写入器对同一 Hudi 表进行更改。这是文件级的OCC，因此，如果任意两个提交（或写入者）的更改不冲突，则可以写入同一个表。有关更多信息，请参阅 [Hudi 并发性控制](#)。
- 亚马逊EMR集群安装了 Zookeeper，可以将其用作锁定提供程序。OCC为了便于使用此功能，Amazon EMR 集群已预先配置了以下属性：

```
hoodie.write.lock.provider=org.apache.hudi.client.transaction.lock.ZookeeperBasedLockProvider
hoodie.write.lock.zookeeper.url=<EMR Zookeeper URL>
hoodie.write.lock.zookeeper.port=<EMR Zookeeper Port>
hoodie.write.lock.zookeeper.base_path=/hudi
```

要启用OCC，您需要使用其 Hudi 任务选项或使用 Amazon 配置在集群级别配置以下属性：EMRAPI

```
hoodie.write.concurrency.mode=optimistic_concurrency_control
hoodie.cleaner.policy.failed.writes=LAZY (Performs cleaning of failed writes lazily
instead of inline with every write)
hoodie.write.lock.zookeeper.lock_key=<Key to uniquely identify the Hudi table> (Table
Name is a good option)
```

## Hudi 监控：亚马逊 CloudWatch 集成以报告 Hudi 指标

- 亚马逊EMR支持向亚马逊 CloudWatch发布 Hudi 指标。通过设置以下所需配置来启用：

```
hoodie.metrics.on=true
hoodie.metrics.reporter.type=CLOUDWATCH
```

- 以下是您可以更改的可选 Hudi 配置：

设置	描述	值
hoodie.metrics.cloudwatch.report.period.seconds	向 Amazon 报告指标的频率（以秒为单位）CloudWatch	默认值为 60s，这对于亚马逊提供的默认一分钟分辨率来说没问题 CloudWatch
hoodie.metrics.cloudwatch.metric.prefix	要添加到每个指标名称的前缀	默认值为空（无前缀）
hoodie.metrics.cloudwatch.namespace	发布指标时使用的 Amazon CloudWatch 命名空间	默认值为 Hudi
hoodie.metrics.cloudwatch.maxDatumsPer请求	向 Amazon 发出的单次请求中包含的最大数据数量 CloudWatch	默认值为 20，这与亚马逊的 CloudWatch 默认值相同

## Amazon EMR Hudi 配置支持和改进

- 客户现在可以利用EMR配置API和重新配置功能在集群级别配置 Hudi 配置。与 Spark 和 Hive 等其他应用程序一样，通过 /etc/hudi/CONF/hudi-defaults.conf 引入了基于文件的新配置支持。EMR配置一些默认值以改善用户体验：

— 已配置 `hoodie.datasource.hive_sync.jdbcurl` 为群集 Hive 服务器URL，不再需要指定。这在在 Spark 集群模式下运行作业时特别有用，您以前必须在此模式下指定 A EMR mazon 主 IP。

— HBase 特定配置，这对于在 Hudi 中使用HBase索引很有用。

— Zookeeper 锁定提供程序的特定配置，如并发控制下所述，这使得使用乐观并发控制 () 变得更加容易。OCC



- 还引入了其他更改，以减少需要通过的配置数量，并在可能的情况下自动推断：

- 该 `partitionBy` 关键字可用于指定分区列。

- 启用 Hive Sync 时，不再强制通过 `HIVE_TABLE_OPT_KEY`，`HIVE_PARTITION_FIELDS_OPT_KEY`，`HIVE_PARTITION_EXTRACTOR_CLASS_OPT_KEY`。这些值可以根据 Hudi 表名称和分区字段推断出来。

- `KEYGENERATOR_CLASS_OPT_KEY` 不强制通过，可以从更简单的 `SimpleKeyGenerator` 和 `ComplexKeyGenerator` 情况下推断。

## Hudi 注意事项

- Hudi 不支持在 Hive 中用于读取时合并 (MoR) 和 Bootstrap 表格中的矢量化执行。例如，当 `hive.vectorized.execution.enabled` 设置为 `true` 时，Hudi 实时表的 `count(*)` 失败。作为解决方法，您可以通过将 `hive.vectorized.execution.enabled` 设置为 `false` 禁用矢量化读入。
- 多写作器支持与 Hudi 引导启动功能不兼容。
- Flink Streamer 和 Flink SQL 是此版本中的实验性功能。建议不要在生产部署中使用这些功能。

## 更改、增强功能和解决的问题

此版本旨在修复 Amazon EMR Scaling 无法成功扩展/缩小集群或导致应用程序故障时出现的问题。

- 以前，在多主集群上手动重启资源管理器会导致 Amazon EMR on 集群上的守护程序（例如 Zookeeper）在 Zookeeper znode 文件中重新加载所有先前已停用或丢失的节点。在某些情况下，这会导致超出默认限制。Amazon EMR 现已从 Zookeeper 文件中删除已停用或丢失超过一小时的节点记录，并且内部限制也已提高。
- 修复了 Amazon EMR on 集群上守护程序运行状况检查活动（例如收集 YARN 节点状态和 HDFS 节点状态）时，对利用率很高的大型集群的扩展请求失败的问题。之所以发生这种情况，是因为集群上的守护程序无法将节点的运行状况数据传送到内部 Amazon EMR 组件。
- 改进了 EMR 集群上的守护程序，可在重复使用 IP 地址时正确跟踪节点状态，从而提高扩展操作期间的可靠性。
- [SPARK-29683](#)。修复了集群缩减期间出现任务失败的问题，因为 Spark 假定所有可用节点都被拒绝列出。
- [YARN-9011](#)。修复了集群尝试向上或向下扩展时，由于 YARN 停用过程中的争用条件而导致任务失败的问题。

- 通过确保 Amaz EMR on 集群上守护程序和 YARN /之间的节点状态始终保持一致，修复了集群扩展期间步骤或任务失败的问题。HDFS
- 修复了启用了 Kerberos 身份验证的 Amazon 集群的EMR集群操作（例如缩减和步骤提交）失败的问题。这是因为 Amaz EMR on 集群上守护程序没有续订 Kerberos 票证，而要与主节点上YARN运行的 HDFS /进行安全通信，则需要续订 Kerberos 票证。
- 配置群集以修复 Apache T YARN imeline Server 版本 1 和 1.5 的性能问题

Apache T YARN imeline Server 版本 1 和 1.5 可能会导致非常活跃的大型EMR集群出现性能问题`yarn.resourcemanager.system-metrics-publisher.enabled=true`，尤其是在使用 Amazon EMR 的默认设置时。开源YARN时间轴服务器 v2 解决了与时间YARN轴服务器可扩展性相关的性能问题。

此问题的其他解决方法包括：

- 配置 `yarn.resourceManager system-metrics-publisheryarn-site.xml` 中的 `.enabled=false`。
- 如下所述，在创建群集时启用此问题的修复程序。

以下 Amazon EMR 版本包含针对此YARN时间轴服务器性能问题的修复程序。

EMR5.30.2、5.31.1、5.32.1、5.33.1、5.34.x、6.0.1、6.1.1、6.2.1、6.3.1、6.3.1、6.3.1、6.4.x

要在上述任何指定的 Amazon EMR 版本上启用修复，请在使用[aws emr create-cluster](#)命令参数传true入的配置JSON文件中将这些属性设置为：`--configurations file:///./configurations.json`。或者使用[重新配置控制台 UI](#) 启用修复程序。

配置 .json 文件内容的示例：

```
[
{
  "Classification": "yarn-site",
  "Properties": {
    "yarn.resourcemanager.system-metrics-publisher.timeline-server-v1.enable-batch":
      "true",
    "yarn.resourcemanager.system-metrics-publisher.enabled": "true"
  },
  "Configurations": []
}
```

- 默认情况下HDFS，Web 和 HTTPFS 服务器处于禁用状态。您可以HDFS使用 Hadoop 配置重新启用 Web。dfs.webhdfs.enabledHTTPFS 服务器可以通过使用 `sudo systemctl start hadoop-httpfs` 启动。
- HTTPS现在，亚马逊 Linux 存储库已默认启用。如果您使用 Amazon S3 VPCE 策略来限制对特定存储桶的访问，则必须将新的 Amazon Linux 存储桶ARNarn:aws:s3:::amazonlinux-2-repos-\$region/\*添加到您的策略中（\$region替换为终端节点所在的区域）。有关更多信息，请在 AWS 讨论论坛中查看此主题。[公告：Amazon Linux 2 现在支持HTTPS在连接软件包存储库时使用。](#)
- Hive：通过为最后一个作业启用临时目录，可以提高写入查询性能。HDFS最终任务的临时数据是写入的，HDFS而不是 Amazon S3，而且由于数据是从HDFS最终表位置 (Amazon S3) 移动到最终表位置 (Amazon S3)，而不是在 Amazon S3 设备之间移动，因此性能得到了改善。
- Hive：使用 Glue 元存储分区修剪，查询编译时间最多可缩短 2.5 倍。
- 默认情况下，当 Hive 将内置UDFs内容传递给 Hive Metastore 服务器时，只有内置内容的子集会传递给 Glue Metastore，因UDFs为 Glue 仅支持有限的表达式运算符。如果您设置hive.glue.partition.pruning.client=true，则所有分区修剪发生在客户端。如果您设置hive.glue.partition.pruning.server=true，则所有分区修剪发生在服务器端。

## 已知问题

- Hue 查询在 Amazon EMR 6.4.0 中不起作用，因为 Apache Hadoop httpFS 服务器默认处于禁用状态。要在亚马逊 EMR 6.4.0 上使用 Hue，要么使用亚马逊步骤在亚马逊EMR主节点上手动启动 HttpFS 服务器`sudo systemctl start hadoop-httpfs`，要么使用[亚马逊步骤](#)。EMR
- 与 Livy 用户模拟一起使用的亚马逊EMR笔记本功能不起作用，因为 HTTPFS 在默认情况下处于禁用状态。在这种情况下，EMR笔记本无法连接到启用了 Livy 模拟的集群。解决方法是在使用将EMR笔记本连接到集群之前启动 HTTPFS 服务器。`sudo systemctl start hadoop-httpfs`
- 在亚马逊 6.4.0 EMR 版本中，Phoenix 不支持 Phoenix 连接器组件。
- 要将 Spark 操作与 Apache Oozie 一起使用，必须将以下配置添加到 Oozie workflow.xml 文件中。否则，Oozie 启动的 Spark 执行程序的路径中EMRFS将缺少几个关键库，例如 Hadoop 和。

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

- 当你使用带有 Hive 分区位置格式的 Spark 来读取 Amazon S3 中的数据，并在亚马逊EMR版本 5.30.0 到 5.36.0 和 6.2.0 到 6.9.0 上运行 Spark 时，你可能会遇到一个问题，导致集群无法正确读取数据。如果您的分区具有以下所有特征，会发生这种情况：
  - 从同一个表扫描两个或多个分区。

- 至少有一个分区目录路径是至少一个其他分区目录路径的前缀，例如，`s3://bucket/table/p=a` 是 `s3://bucket/table/p=a b` 的前缀。
- 另一个分区目录中前缀之后的第一个字符的值为 UTF -8，小于该/字符 (U+002F)。例如，在 `s3://bucket/table/p=a b` 中，`a` 和 `b` 之间出现的空格字符 (U+0020) 就属于此类。请注意，还有其他 14 个非控制字符：`!"#$%&'()*+,-`。有关更多信息，请参阅 [UTF-8 编码表和 Unicode 字符](#)。

解决方法是在 `spark-defaults` 分类中将

`spark.sql.sources.fastS3PartitionDiscovery.enabled` 配置设置为 `false`。

## 6.4.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>aws-sagemaker-spark-sdk</code>	1.4.1	亚马逊 SageMaker Spark SDK
<code>emr-ddb</code>	4.16.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	3.2.0	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.5.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。

组件	版本	描述
emr-notebook-env	1.3.0	适用于 EMR Notebooks ( 可提供 jupyter 企业网关 ) 的 Conda env
emr-s3-dist-cp	2.18.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	2.1.0	EMRS3Select 连接器
emrfs	2.47.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.13.1	Apache Flink 命令行客户端脚本和应用程序。
flink-jobmanager-config	1.13.1	管理 Apache Flink EMR JobManager 节点上的资源。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	3.2.1-amzn-4	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	3.2.1-amzn-4	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	3.2.1-amzn-4	HDFS命令行客户端和库

组件	版本	描述
hadoop-hdfs-namenode	3.2.1-amzn-4	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	3.2.1-amzn-4	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	3.2.1-amzn-4	HTTPHDFS操作端点。
hadoop-kms-server	3.2.1-amzn-4	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	3.2.1-amzn-4	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	3.2.1-amzn-4	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	3.2.1-amzn-4	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	3.2.1-amzn-4	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	2.4.4-amzn-0	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	2.4.4-amzn-0	为一个或多个HBase地区提供服务的服务。
hbase-client	2.4.4-amzn-0	HBase命令行客户端。
hbase-rest-server	2.4.4-amzn-0	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	2.4.4-amzn-0	为其提供 Thrift 端点的HBase服务。

组件	版本	描述
hcatalog-client	3.1.2-amzn-5	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	3.1.2-amzn-5	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	3.1.2-amzn-5	HTTP端点为其提供REST接口HCatalog。
hive-client	3.1.2-amzn-5	Hive 命令行客户端。
hive-hbase	3.1.2-amzn-5	Hive-hbase 客户端。
hive-metastore-server	3.1.2-amzn-5	用于访问 Hive 元数据仓库的服务，这是一个语义存储库，用于存储 Hadoop 操作SQL的元数据。
hive-server2	3.1.2-amzn-5	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.8.0-amzn-0	增量处理框架，以支持低延迟和高效率的数据管道。
hudi-presto	0.8.0-amzn-0	用于运行 Presto 以及 Hudi 的捆绑库。
hudi-trino	0.8.0-amzn-0	用于运行 Trino 以及 Hudi 的捆绑库。
hudi-spark	0.8.0-amzn-0	用于运行 Spark 以及 Hudi 的捆绑库。
hue-server	4.9.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序

组件	版本	描述
jupyterhub	1.4.1	Jupyter notebook 的多用户服务器
livy-server	0.7.1-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mxnet	1.8.0	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.68+	MariaDB 数据库服务器。
nvidia-cuda	10.1.243	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.1	Oozie 命令行客户端。
oozie-server	5.2.1	用于接受 Oozie  workflow 请求的服务。
opencv	4.5.0	开源计算机视觉库。
phoenix-library	5.1.2	服务器和客户端的 phoenix 库
phoenix-query-server	5.1.2	轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限 API
presto-coordinator	0.254.1-amzn-0	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.254.1-amzn-0	用于执行查询的各个部分的服务。



组件	版本	描述
presto-client	0.254.1-amzn-0	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
trino-coordinator	359	用于在 trino-worker 之中接受查询并管理查询的服务。
trino-worker	359	用于执行查询的各个部分的服务。
trino-client	359	Trino 命令行客户端，安装在 HA 集群的备用主服务器（未启动 Trino 服务器）上。
pig-client	0.17.0	Pig 命令行客户端。
r	4.0.2	用于统计计算的 R 项目
ranger-kms-server	2.0.0	Apache Ranger 密钥管理系统
spark-client	3.1.2-amzn-0	Spark 命令行客户端。
spark-history-server	3.1.2-amzn-0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	3.1.2-amzn-0	的内存中执行引擎。YARN
spark-yarn-slave	3.1.2-amzn-0	YARN从属服务器需要的 Apache Spark 库。
spark-rapids	0.4.1	Nvidia Spark RAPIDS 插件可以加速 Apache Spark。GPUs
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。

组件	版本	描述
tensorflow	2.4.1	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.2	tez YARN 应用程序和库。
webserver	2.4.41+	Apache HTTP 服务器。
zeppelin-server	0.9.0	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.5.7	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.5.7	ZooKeeper 命令行客户端。

## 6.4.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

为处于运行状态的集群中的实例组指定配置时，将发生重新配置操作。Amazon EMR 仅对您修改的分类启动重新配置操作。有关更多信息，请参阅 [在正在运行的集群中重新配置实例组](#)。

### emr-6.4.0 分类

分类	描述	重新配置操作
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。	Restarts the ResourceManager service.
container-executor	更改 Hadoop YARN 的 container-executor.cfg 文件中的值。	Not available.

分类	描述	重新配置操作
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。	Not available.
core-site	更改 Hadoop 的 core-site.xml 文件中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	更改 docker 相关设置。	Not available.

分类	描述	重新配置操作
emrfs-site	更改EMRFS设置。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	更改 flink-conf.yaml 设置。	Restarts Flink history server.
flink-log4j	更改 Flink log4j.properties 设置。	Restarts Flink history server.
flink-log4j-session	为 Kubernetes/Yarn 会话更改 Flink log4j-session.properties 设置。	Restarts Flink history server.
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。	Restarts Flink history server.

分类	描述	重新配置操作
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	更改 hadoop ssl 服务器配置	Not available.
hadoop-ssl-client	更改 hadoop ssl 客户端配置	Not available.
hbase	亚马逊EMR精心策划的 Apache 设置。HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.

分类	描述	重新配置操作
hbase-env	更改环境HBase中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	更改 hadoop-metrics2-hbase base.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。	Not available.
hbase-site	更改HBase的 hbase-site.xml 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	配置HDFS加密区域。	This classification should not be reconfigured.
hdfs-env	更改HDFS环境中的值。	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	更改环境HCatalog中的值。	Restarts Hive HCatalog Server.

分类	描述	重新配置操作
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。	Restarts Hive WebHCat server.
hive	亚马逊EMR精心策划的 Apache Hive 设置。	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。	Not available.
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。	Not available.
hive-env	更改 Hive 环境中的值。	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	更改 Hive 的 hive-exec-log4j2.properties 文件中的值。	Not available.
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log4j2.properties 文件中的值。	Not available.

分类	描述	重新配置操作
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。	Not available.
hive-site	更改 Hive 的 hive-site.xml 文件中的值	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值	Not available.
hue-ini	更改 Hue 的 ini 文件中的值	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	更改HTTPFS环境中的值。	Restarts Hadoop Httpfs service.
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。	Restarts Hadoop Httpfs service.
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。	Not available.
hadoop-kms-env	在 Hadoop KMS 环境中更改值。	Restarts Hadoop-KMS service.
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。	Not available.
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。	Restarts Hadoop-KMS and Ranger-KMS service.



分类	描述	重新配置操作
hudi-env	更改 Hudi 环境中的值。	Not available.
hudi-defaults	更改 Hudi 的 hudi-defaults.conf 文件中的值。	Not available.
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。	Not available.
jupyter-hub-conf	更改 JupyterHubs 的 jupyterhub_config.py 文件中的值。	Not available.
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。	Not available.
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。	Not available.
livy-conf	更改 Livy 的 livy.conf 文件中的值。	Restarts Livy Server.
livy-env	更改 Livy 环境中的值。	Restarts Livy Server.
livy-log4j	更改 Livy log4j.properties 设置。	Restarts Livy Server.
mapred-env	更改 MapReduce 应用程序环境中的值。	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	更改 Oozie 的环境中的值。	Restarts Oozie.
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。	Restarts Oozie.

分类	描述	重新配置操作
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。	Restarts Oozie.
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。	Not available.
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。	Not available.
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。	Restarts Phoenix-QueryServer.
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。	Not available.
pig-env	更改 Pig 环境中的值。	Not available.
pig-properties	更改 Pig 的 pig.properties 文件中的值。	Restarts Oozie.
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。	Not available.
presto-log	更改 Presto 的 log.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-config	更改 Presto 的 config.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。	Not available.
presto-env	更改 Presto 的 presto-env.sh 文件中的值。	Restarts Presto-Server (for PrestoDB)

分类	描述	重新配置操作
presto-node	更改 Presto 的 node.properties 文件中的值。	Not available.
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。	Not available.
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。	Not available.
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。	Not available.
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。	Not available.
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。	Not available.
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。	Not available.
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。	Not available.
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。	Not available.
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。	Not available.
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。	Not available.
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。	Not available.
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。	Not available.
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。	Not available.
trino-log	更改 Trino 的 log.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-config	更改 Trino 的 config.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-password-authenticator	更改 Trino 的 password-authenticator.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-env	更改 Trino 的 trino-env.sh 文件中的值。	Restarts Trino-Server (for Trino)
trino-node	更改 Trino 的 node.properties 文件中的值。	Not available.
trino-connector-blackhole	更改 Trino 的 blackhole.properties 文件中的值。	Not available.
trino-connector-cassandra	更改 Trino 的 cassandra.properties 文件中的值。	Not available.
trino-connector-hive	更改 Trino 的 hive.properties 文件中的值。	Restarts Trino-Server (for Trino)
trino-connector-jmx	更改 Trino 的 jmx.properties 文件中的值。	Not available.

分类	描述	重新配置操作
trino-connector-kafka	更改 Trino 的 kafka.properties 文件中的值。	Not available.
trino-connector-localfile	更改 Trino 的 localfile.properties 文件中的值。	Not available.
trino-connector-memory	更改 Trino 的 memory.properties 文件中的值。	Not available.
trino-connector-mongodb	更改 Trino 的 mongodb.properties 文件中的值。	Not available.
trino-connector-mysql	更改 Trino 的 mysql.properties 文件中的值。	Not available.
trino-connector-postgresql	更改 Trino 的 postgresql.properties 文件中的值。	Not available.
trino-connector-raptor	更改 Trino 的 raptor.properties 文件中的值。	Not available.
trino-connector-redis	更改 Trino 的 redis.properties 文件中的值。	Not available.
trino-connector-redshift	更改 Trino 的 redshift.properties 文件中的值。	Not available.
trino-connector-tpch	更改 Trino 的 tpch.properties 文件中的值。	Not available.
trino-connector-tpcds	更改 Trino 的 tpcds.properties 文件中的值。	Not available.
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-site	更改 Ranger KMS 的 ranger-kms-site.xml 文件中的值。	Restarts Ranger KMS Server.

分类	描述	重新配置操作
ranger-kms-env	在游侠KMS环境中更改值。	Restarts Ranger KMS Server.
ranger-kms-log4j	更改 Ranger 的 kms-log4j .properties 文件中的值。KMS	Not available.
ranger-kms-db-ca	更改 S3 上我与 Ranger SQL SSL KMS 连接的 CA 文件值。	Not available.
spark	亚马逊EMR精心策划的 Apache Spark 设置。	This property modifies spark-defaults. See actions there.
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-env	更改 Spark 环境中的值。	Restarts Spark history server and Spark thrift server.
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值	Not available.
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
sqoop-env	更改 Sqoop 的环境中的值。	Not available.
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。	Not available.
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。	Not available.
tez-site	更改 Tez 的 tez-site.xml 文件中的值。	Restart Oozie and HiveServer2.

分类	描述	重新配置操作
yarn-env	更改YARN环境中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	更改YARN的 yarn-site.xml 文件中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	更改 Zeppelin 环境中的值。	Restarts Zeppelin.
zeppelin-site	更改 zeppelin-site.xml 中的配置设置。	Restarts Zeppelin.
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。	Restarts Zookeeper server.
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。	Restarts Zookeeper server.

## 亚马逊EMR版本 6.3.1

### 6.3.1 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[JupyterEnterpriseGateway](#)、[Jupyter](#) 和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本 ( 如果适用 )。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-6.3.1	emr-6.3.0	emr-6.2.1	emr-6.2.0
适用于 Java 的 AWS SDK	1.11.977	1.11.977	1.11.880	1.11.880
Python	2.7、3.7	2.7、3.7	2.7、3.7	2.7、3.7
Scala	2.12.10	2.12.10	2.12.10	2.12.10
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.12.1	1.12.1	1.11.2	1.11.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.2.6	2.2.6	2.2.6-amzn-0	2.2.6-amzn-0
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.7.0-amzn-0	0.7.0-amzn-0	0.6.0-amzn-1	0.6.0-amzn-1
Hue	4.9.0	4.9.0	4.8.0	4.8.0
Iceberg	-	-	-	-



	emr-6.3.1	emr-6.3.0	emr-6.2.1	emr-6.2.0
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.2.2	1.2.2	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.7.0	1.7.0	1.7.0	1.7.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.0	5.2.0
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.245.1	0.245.1	0.238.3	0.238.3
Spark	3.1.1	3.1.1	3.0.1	3.0.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.3.1	2.3.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	350	350	343	343
Zeppelin	0.9.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

### 6.3.1 发布说明

此版本旨在修复 Amazon EMR Scaling 无法成功扩展/缩小集群或导致应用程序故障时出现的问题。

## 更改、增强和解决的问题

- 修复了 Amaz EMR on 集群上守护程序运行状况检查活动 ( 例如收集YARN节点状态和HDFS节点状态 ) 时, 对利用率很高的大型集群的扩展请求失败的问题。之所以发生这种情况, 是因为集群上的守护程序无法将节点的运行状况数据传送到内部 Ama EMR zon 组件。
- 改进了EMR集群上的守护程序, 可在重复使用 IP 地址时正确跟踪节点状态, 从而提高扩展操作期间的可靠性。
- [SPARK-29683](#)。修复了集群缩减期间出现任务失败的问题, 因为 Spark 假定所有可用节点都被拒绝列出。
- [YARN-9011](#)。修复了集群尝试向上或向下扩展时, 由于YARN停用过程中的争用条件而导致任务失败的问题。
- 通过确保 Amaz EMR on 集群上守护程序和 YARN /之间的节点状态始终保持一致, 修复了集群扩展期间步骤或任务失败的问题。HDFS
- 修复了启用了 Kerberos 身份验证的 Amazon 集群的EMR集群操作 ( 例如缩减和步骤提交 ) 失败的问题。这是因为 Amaz EMR on 集群上守护程序没有续订 Kerberos 票证, 而要与主节点上YARN运行的 HDFS /进行安全通信, 则需要续订 Kerberos 票证。
- HTTPS现在, 亚马逊 Linux 存储库已默认启用。如果您使用 Amazon S3 VPCE 策略来限制对特定存储桶的访问, 则必须将新的 Amazon Linux 存储桶ARNarn:aws:s3:::amazonlinux-2-repos-\$region/\*添加到您的策略中 ( \$region替换为终端节点所在的区域 )。有关更多信息, 请在 AWS 讨论论坛中查看此主题。[公告 : Amazon Linux 2 现在支持HTTPS在连接软件包存储库时使用。](#)

## 已知问题

- 当你使用带有 Hive 分区位置格式的 Spark 来读取 Amazon S3 中的数据, 并在亚马逊EMR版本 5.30.0 到 5.36.0 和 6.2.0 到 6.9.0 上运行 Spark 时, 你可能会遇到一个问题, 导致集群无法正确读取数据。如果您的分区具有以下所有特征, 会发生这种情况 :
  - 从同一个表扫描两个或多个分区。
  - 至少有一个分区目录路径是至少一个其他分区目录路径的前缀, 例如, s3://bucket/table/p=a 是 s3://bucket/table/p=a b 的前缀。
  - 另一个分区目录中前缀之后的第一个字符的值为 UTF -8, 小于该/字符 (U+002F)。例如, 在 s3://bucket/table/p=a b 中, a 和 b 之间出现的空格字符 (U+0020) 就属于此类。请注意, 还有其他 14 个非控制字符 : !"#%&'()\*+,-。有关更多信息, 请参阅 [UTF-8 编码表和 Unicode 字符](#)。

解决方法是在 `spark-defaults` 分类中将 `spark.sql.sources.fastS3PartitionDiscovery.enabled` 配置设置为 `false`。

### 6.3.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>aws-sagemaker-spark-sdk</code>	1.4.1	亚马逊 SageMaker Spark SDK
<code>emr-ddb</code>	4.16.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	3.2.0	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.5.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
<code>emr-notebook-env</code>	1.2.0	适用于 EMR Notebooks ( 可提供 jupyter 企业网关 ) 的 Conda env
<code>emr-s3-dist-cp</code>	2.18.0	针对 Amazon S3 优化的分布式复制应用程序。
<code>emr-s3-select</code>	2.1.0	EMRS3Select 连接器

组件	版本	描述
emrfs	2.46.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.12.1	Apache Flink 命令行客户端脚本和应用程序。
flink-jobmanager-config	1.12.1	管理 Apache Flink EMR JobManager 节点上的资源。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	3.2.1-amzn-3.1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	3.2.1-amzn-3.1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	3.2.1-amzn-3.1	HDFS命令行客户端和库
hadoop-hdfs-namenode	3.2.1-amzn-3.1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	3.2.1-amzn-3.1	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	3.2.1-amzn-3.1	HTTPHDFS操作端点。

组件	版本	描述
hadoop-kms-server	3.2.1-amzn-3.1	基于 Hadoop 的加密密钥管理服务。KeyProvider API
hadoop-mapred	3.2.1-amzn-3.1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	3.2.1-amzn-3.1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	3.2.1-amzn-3.1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	3.2.1-amzn-3.1	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	2.2.6-amzn-1	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	2.2.6-amzn-1	为一个或多个HBase地区提供服务的服务。
hbase-client	2.2.6-amzn-1	HBase命令行客户端。
hbase-rest-server	2.2.6-amzn-1	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	2.2.6-amzn-1	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	3.1.2-amzn-4	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	3.1.2-amzn-4	提供服务HCatalog，为分布式应用程序提供表和存储管理层。

组件	版本	描述
hcatalog-webhcat-server	3.1.2-amzn-4	HTTP端点为其提供REST接口HCatalog。
hive-client	3.1.2-amzn-4	Hive 命令行客户端。
hive-hbase	3.1.2-amzn-4	Hive-hbase 客户端。
hive-metastore-server	3.1.2-amzn-4	用于访问 Hive 元数据仓库的服务，这是一个语义存储库，用于存储 Hadoop 操作SQL的元数据。
hive-server2	3.1.2-amzn-4	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.7.0-amzn-0	增量处理框架，以支持低延迟和高效率的数据管道。
hudi-presto	0.7.0-amzn-0	用于运行 Presto 以及 Hudi 的捆绑库。
hudi-prestosql	0.7.0-amzn-0	用于使用 Hudi 运行 Presto SQL 的捆绑库。
hudi-spark	0.7.0-amzn-0	用于运行 Spark 以及 Hudi 的捆绑库。
hue-server	4.9.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	1.2.2	Jupyter notebook 的多用户服务器
livy-server	0.7.0-incubating	REST用于与 Apache Spark 交互的接口

组件	版本	描述
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mxnet	1.7.0	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.68+	MariaDB 数据库服务器。
nvidia-cuda	10.1.243	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.1	Oozie 命令行客户端。
oozie-server	5.2.1	用于接受 Oozie 工作流请求的服务。
opencv	4.5.0	开源计算机视觉库。
phoenix-library	5.0.0--2.0 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	5.0.0--2.0 HBase	轻量级服务器，提供JDBC对Avatica 的访问权限以及协议缓冲区和JSON格式访问权限 API
presto-coordinator	0.245.1-amzn-0	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.245.1-amzn-0	用于执行查询的各个部分的服务。
presto-client	0.245.1-amzn-0	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
prestosql-coordinator	350	用于在 prestosql-worker 之中接受查询并管理查询执行的服务。

组件	版本	描述
prestoql-worker	350	用于执行查询的各个部分的服务。
prestoql-client	350	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
pig-client	0.17.0	Pig 命令行客户端。
r	4.0.2	用于统计计算的 R 项目
ranger-kms-server	2.0.0	Apache Ranger 密钥管理系统
spark-client	3.1.1-amzn-0.1	Spark 命令行客户端。
spark-history-server	3.1.1-amzn-0.1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	3.1.1-amzn-0.1	的内存中执行引擎。YARN
spark-yarn-slave	3.1.1-amzn-0.1	YARN从属服务器需要的 Apache Spark 库。
spark-rapids	0.4.1	Nvidia Spark RAPIDS 插件可以加速 Apache Spark。GPUs
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	2.4.1	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.2	tez YARN 应用程序和库。
webserver	2.4.41+	Apache HTTP 服务器。



组件	版本	描述
zeppelin-server	0.9.0	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.14	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.14	ZooKeeper 命令行客户端。

### 6.3.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

为处于运行状态的集群中的实例组指定配置时，将发生重新配置操作。Amazon EMR 仅对您修改的分类启动重新配置操作。有关更多信息，请参阅 [在正在运行的集群中重新配置实例组](#)。

#### emr-6.3.1 分类

分类	描述	重新配置操作
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。	Restarts the ResourceManager service.
container-executor	更改 Hadoop YARN 的 container-executor.cfg 文件中的值。	Not available.
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。	Not available.
core-site	更改 Hadoop 的 core-site.xml 文件中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

分类	描述	重新配置操作
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	更改 docker 相关设置。	Not available.
emrfs-site	更改EMRFS设置。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	更改 flink-conf.yaml 设置。	Restarts Flink history server.
flink-log4j	更改 Flink log4j.properties 设置。	Restarts Flink history server.
flink-log4j-session	为 Kubernetes/Yarn 会话更改 Flink log4j-session.properties 设置。	Restarts Flink history server.

分类	描述	重新配置操作
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。	Restarts Flink history server.
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	更改 hadoop ssl 服务器配置	Not available.
hadoop-ssl-client	更改 hadoop ssl 客户端配置	Not available.

分类	描述	重新配置操作
hbase	亚马逊EMR精心策划的Apache 设置。HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	更改环境HBase中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	更改 hadoop-metrics2-hbase base.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。	Not available.
hbase-site	更改HBase的 hbase-site.xml 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	配置HDFS加密区域。	This classification should not be reconfigured.
hdfs-env	更改HDFS环境中的值。	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

分类	描述	重新配置操作
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	更改环境HCatalog中的值。	Restarts Hive HCatalog Server.
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。	Restarts Hive WebHCat server.
hive	亚马逊EMR精心策划的 Apache Hive 设置。	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。	Not available.
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。	Not available.

分类	描述	重新配置操作
hive-env	更改 Hive 环境中的值。	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	更改 Hive 的 hive-exec-log4j2.properties 文件中的值。	Not available.
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log4j2.properties 文件中的值。	Not available.
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。	Not available.
hive-site	更改 Hive 的 hive-site.xml 文件中的值	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值	Not available.
hue-ini	更改 Hue 的 ini 文件中的值	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	更改HTTPFS环境中的值。	Restarts Hadoop Httpfs service.
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。	Restarts Hadoop Httpfs service.

分类	描述	重新配置操作
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。	Not available.
hadoop-kms-env	在 Hadoop KMS 环境中更改值。	Restarts Hadoop-KMS service.
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。	Not available.
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	更改 Hudi 环境中的值。	Not available.
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。	Not available.
jupyter-hub-conf	更改 JupyterHubs 的 jupyterhub_config.py 文件中的值。	Not available.
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。	Not available.
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。	Not available.
livy-conf	更改 Livy 的 livy.conf 文件中的值。	Restarts Livy Server.
livy-env	更改 Livy 环境中的值。	Restarts Livy Server.
livy-log4j	更改 Livy log4j.properties 设置。	Restarts Livy Server.
mapred-env	更改 MapReduce 应用程序环境中的值。	Restarts Hadoop MapReduce-HistoryServer.

分类	描述	重新配置操作
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	更改 Oozie 的环境中的值。	Restarts Oozie.
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。	Restarts Oozie.
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。	Restarts Oozie.
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。	Not available.
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。	Not available.
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。	Restarts Phoenix-QueryServer.
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。	Not available.
pig-env	更改 Pig 环境中的值。	Not available.
pig-properties	更改 Pig 的 pig.properties 文件中的值。	Restarts Oozie.
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。	Not available.
presto-log	更改 Presto 的 log.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-config	更改 Presto 的 config.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)



分类	描述	重新配置操作
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。	Not available.
presto-env	更改 Presto 的 presto-env.sh 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-node	更改 Presto 的 node.properties 文件中的值。	Not available.
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。	Not available.
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。	Not available.
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。	Not available.
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。	Not available.
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。	Not available.
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。	Not available.
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。	Not available.
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。	Not available.
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。	Not available.
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。	Not available.
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。	Not available.
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。	Not available.
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。	Not available.
prestosql-log	更改 Presto 的 log.properties 文件中的值。	Restarts Presto-Server (for PrestoSQL)
prestosql-config	更改 Presto 的 config.properties 文件中的值。	Restarts Presto-Server (for PrestoSQL)
prestosql-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。	Restarts Presto-Server (for PrestoSQL)
prestosql-env	更改 Presto 的 presto-env.sh 文件中的值。	Restarts Presto-Server (for PrestoSQL)
prestosql-node	更改 Presto 的 node.properties 文件中的值。	Not available.
prestosql-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。	Not available.

分类	描述	重新配置操作
prestosql-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。	Not available.
prestosql-connector-hive	更改 Presto 的 hive.properties 文件中的值。	Restarts Presto-Server (for PrestoSQL)
prestosql-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。	Not available.
prestosql-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。	Not available.
prestosql-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。	Not available.
prestosql-connector-memory	更改 Presto 的内存.properties 文件中的值。	Not available.
prestosql-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。	Not available.
prestosql-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。	Not available.
prestosql-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。	Not available.
prestosql-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。	Not available.
prestosql-connector-redis	更改 Presto 的 redis.properties 文件中的值。	Not available.
prestosql-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。	Not available.
prestosql-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。	Not available.

分类	描述	重新配置操作
prestosql-connector-tpcds	更改 Presto 的 tpcds.prop SQL erties 文件中的值。	Not available.
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site .xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-site	更改 Ran ranger-kms-site KMS ger .xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-env	在游侠KMS环境中更改值。	Restarts Ranger KMS Server.
ranger-kms-log4j	更改 Ranger 的 kms-log4j .properties 文件中的值。KMS	Not available.
ranger-kms-db-ca	更改 S3 上我与 Ranger SQL SSL KMS 连接的 CA 文件值。	Not available.
spark	亚马逊EMR精心策划的 Apache Spark 设置。	This property modifies spark-defaults. See actions there.
spark-defaults	更改 Spark 的 spark-def aults.conf 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-env	更改 Spark 环境中的值。	Restarts Spark history server and Spark thrift server.
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值	Not available.
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-metrics	更改 Spark 的 metrics.p roperties 文件中的值。	Restarts Spark history server and Spark thrift server.
sqoop-env	更改 Sqoop 的环境中的值。	Not available.

分类	描述	重新配置操作
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。	Not available.
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。	Not available.
tez-site	更改 Tez 的 tez-site.xml 文件中的值。	Restart Oozie and HiveServer2.
yarn-env	更改YARN环境中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	更改YARN的 yarn-site.xml 文件中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	更改 Zeppelin 环境中的值。	Restarts Zeppelin.
zeppelin-site	更改 zeppelin-site.xml 中的配置设置。	Restarts Zeppelin.
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。	Restarts Zookeeper server.
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。	Restarts Zookeeper server.

## 亚马逊EMR版本 6.3.0

### 6.3.0 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[JupyterEnterpriseGateway](#)、[Jupyter](#)和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-6.3.0	emr-6.2.1	emr-6.2.0	emr-6.1.1
适用于 Java 的 AWS SDK	1.11.977	1.11.880	1.11.880	1.11.828
Python	2.7、3.7	2.7、3.7	2.7、3.7	2.7、3.7
Scala	2.12.10	2.12.10	2.12.10	2.12.10
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.12.1	1.11.2	1.11.2	1.11.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.2.6	2.2.6-amzn-0	2.2.6-amzn-0	2.2.5

	emr-6.3.0	emr-6.2.1	emr-6.2.0	emr-6.1.1
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.7.0-amzn-0	0.6.0-amzn-1	0.6.0-amzn-1	0.5.2-incubating-amzn-2
Hue	4.9.0	4.8.0	4.8.0	4.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	-
JupyterHub	1.2.2	1.1.0	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.7.0	1.7.0	1.7.0	1.6.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.0	5.2.0	5.2.0
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.245.1	0.238.3	0.238.3	0.232
Spark	3.1.1	3.0.1	3.0.1	3.0.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.3.1	2.3.1	2.1.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2

	emr-6.3.0	emr-6.2.1	emr-6.2.0	emr-6.1.1
Trino (PrestoSQL)	350	343	343	338
Zeppelin	0.9.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 6.3.0 发布说明

以下发行说明包含 Amazon 6.3.0 EMR 版本的信息。更改与 6.2.0 有关。

首次发布日期：2021 年 5 月 12 日

上次更新日期：2021 年 8 月 9 日

### 受支持的应用程序

- AWS SDK for Java 版本 1.11.977
- CloudWatch 水槽版本 2.1.0
- DynamoDB 连接器 4.16.0
- EMRFS版本 2.46.0
- 亚马逊 EMR Goodies 版本 3.2.0
- 亚马逊 EMR Kinesis Connector 版本 3.5.0
- 亚马逊EMR录音服务器版本 2.0.0
- 亚马逊EMR脚本版本 2.5.0
- Flink 1.12.1
- Ganglia 3.7.2
- AWS Glue Hive Metastore 客户端版本 3.2.0
- Hadoop 3.2.1-amzn-3
- HBase版本 2.2.6-amzn-1
- HBase-operator-tools 1.0.0
- HCatalog版本 3.1.2-amzn-0



- Hive 3.1.2-amzn-4
- Hudi 0.7.0-amzn-0
- Hue 4.9.0
- Java JDK 版本 Corretto-8.282.08.1 ( 版本 1.8.0\_282-b08 )
- JupyterHub 版本 1.2.0
- Livy 0.7.0-incubating
- MXNet版本 1.7.0
- Oozie 5.2.1
- Phoenix 5.0.0
- Pig 0.17.0
- Presto 0.245.1-amzn-0
- Presto SQL 版本 350
- Apache RangerKMS ( 多主机透明加密 ) 版本 2.0.0
- ranger-plugins 2.0.1-amzn-0
- ranger-s3-plugin 1.1.0
- SageMaker Spark SDK 版本 1.4.1
- Scala 版本 2.12.10 ( 打开 JDK 64 位服务器虚拟机 , Java 1.8.0\_282 )
- Spark 3.1.1-amzn-0
- spark-rapids 0.4.1
- Sqoop 1.4.7
- TensorFlow 版本 2.4.1
- tez 0.9.2
- Zeppelin 0.9.0
- Zookeeper 3.4.14
- 连接器和驱动程序 : DynamoDB 连接器 4.16.0

## 新功能

- 亚马逊EMR支持 Amazon S3 接入点，这是 Amazon S3 的一项功能，可让您轻松管理共享数据湖的访问权限。使用您的 Amazon S3 接入点别名，您可以在 Amazon 上大规模简化数据访问

EMR。EMR在提供亚马逊服务的所有 AWS 地区，您可以将 Amazon S3 接入点用于所有版本EMR的 Amazon，无需支付额外费用。要了解有关 Amazon S3 访问点和访问点别名的详细信息，请参阅《Amazon S3 用户指南》中的[为接入点使用存储桶式别名](#)。

- 新增DescribeReleaseLabel和ListReleaseLabelAPI参数提供了 Amazon EMR 发行标签详情。您可以通过编程方式列出运行API请求的区域中可用的版本，并列出特定的 Amazon EMR 发行标签的可用应用程序。发布标签参数还列出了支持指定应用程序（例如 Spark）的 Amazon EMR 版本。此信息可用于以编程方式启动 Amazon EMR 集群。例如，您可以使用 ListReleaseLabel 结果中的最新发行版启动集群。有关更多信息，请参阅 Amazon EMR API 参考[ListReleaseLabels](#)中的[DescribeReleaseLabel](#)和。
- 在 Amazon EMR 6.3.0 中，您可以启动与 Apache Ranger 原生集成的集群。Apache Ranger 是一个开源框架，可跨 Hadoop 平台启用、监控和管理全面的数据安全。有关更多信息，请参阅 [Apache Ranger](#)。通过原生集成，您可以自带Apache Ranger来在亚马逊上实施精细的数据访问控制。EMR 参见《[亚马逊EMR管理指南](#)》中的[将亚马逊EMR与 Apache Ranger 集成](#)。
- 限定范围的托管策略：为了与 AWS 最佳实践保持一致，Amazon 引入EMR了 v2 EMR 范围的默认托管策略，以取代即将弃用的策略。参见 [Amazon EMR 托管政策](#)。
- 实例元数据服务 (IMDS) V2 支持状态：对于 Amazon EMR 6.2 或更高版本，Amazon EMR 组件IMDSv2用于所有IMDS调用。对于应用程序代码中的IMDS调用，您可以同时使用IMDSv1和IMDSv2，或者将配置IMDS为仅IMDSv2用于增强安全性。如果您IMDSv1在早期的 Amazon EMR 6.x 版本中禁用，则会导致集群启动失败。

## 更改、增强功能和解决的问题

- 此版本旨在修复 Amazon EMR Scaling 无法成功扩展/缩小集群或导致应用程序故障时出现的问题。
- 修复了 Amaz EMR on 集群上守护程序运行状况检查活动（例如收集YARN节点状态和HDFS节点状态）时，对利用率很高的大型集群的扩展请求失败的问题。之所以发生这种情况，是因为集群上的守护程序无法将节点的运行状况数据传送到内部 Ama EMR zon 组件。
- 改进了EMR集群上的守护程序，可在重复使用 IP 地址时正确跟踪节点状态，从而提高扩展操作期间的可靠性。
- [SPARK-29683](#)。修复了集群缩减期间出现任务失败的问题，因为 Spark 假定所有可用节点都被拒绝列出。
- [YARN-9011](#)。修复了集群尝试向上或向下扩展时，由于YARN停用过程中的争用条件而导致任务失败的问题。
- 通过确保 Amaz EMR on 集群上守护程序和 YARN /之间的节点状态始终保持一致，修复了集群扩展期间步骤或任务失败的问题。HDFS

- 修复了启用了 Kerberos 身份验证的 Amazon 集群的EMR集群操作（例如缩减和步骤提交）失败的问题。这是因为 Amaz EMR on 集群上守护程序没有续订 Kerberos 票证，而要与主节点上YARN运行的 HDFS /进行安全通信，则需要续订 Kerberos 票证。
- 较新的Amazon EMR 版本通过降低亚马逊较旧AL2版本的“最大打开文件数”限制来解决这个问题EMR。亚马逊EMR发布的 5.30.1、5.30.2、5.31.1、5.32.1、5.32.1、6.0.1、6.1.1、6.2.1、5.33.0、6.3.0及更高版本现在包含一个永久修复程序，其中包含一个具有更高的“最大打开文件数”设置的永久修复。
- 在 Spark [3.1](#) 中，[Spark SQL](#) 用户界面解释模式formatted的默认值从extended更改为。亚马逊将其EMR恢复为extended在 Spark SQL 用户界面中包含逻辑计划信息。可以通过将 spark.sql.ui.explainMode 设置为 formatted 进行还原。
- 以下提交是从 Spark 主分支逆向移植的。

-[\[SPARK-34752\]](#) [\[BUILD\]](#) 将 Jetty 推至 9.4.37，地址为 -2020-27223。CVE

-[\[SPARK-34534\]](#) 修复了用于获取方 blockIds 块时的顺序 FetchShuffleBlocks。

-[\[SPARK-34681\]](#) [\[SQL\]](#) 修复了在不相等条件下构建左侧时完全外部洗牌哈希链接的错误。

-[\[SPARK-34497\]](#) [\[SQL\]](#) 修复内置JDBC连接提供程序以恢复JVM安全上下文更改。

- 为了提高与 Nvidia Spark RAPIDS 插件的互操作性，添加了解决方法，以解决在禁 RAPIDS用自适应查询执行的情况下使用 Nvidia Spark 时无法触发动态分区修剪的问题，请参阅[RAPIDS问题 #1378](#) 和[RAPIDS问题 # #1386](#)。有关新配置的详细信息spark.sql.optimizer.dynamicPartitionPruning.enforceBroadcastReuse，请参阅[RAPIDS问题 # #1386](#)。
- 文件输出提交程序默认算法已在开源 Spark 3.1 中将 v2 算法更改为 v1 算法。有关更多信息，请参阅 [Amazon EMR 优化 Spark 性能-动态分区修剪](#)。
- 为了防止EMR性能下降，亚马逊恢复了 v2 算法，这是之前的 Amazon EMR 6.x 版本中使用的默认算法。要恢复开源 Spark 3.1 行为，请将 spark.hadoop.mapreduce.fileoutputcommitter.algorithm.version 设置为 1。开源 Spark 进行了此更改，因为文件输出提交程序算法 v2 中的任务提交不是原子操作，在某些情况下可能会导致输出数据正确性问题。不过，算法 v1 中的任务提交也不是原子操作。在某些情况下，任务提交会包括在重命名之前执行的删除。这可能会导致出现无提示的数据正确性问题。
- 修复了早期 Amazon EMR 版本中的托管扩展问题，并进行了改进，从而显著降低了应用程序失败率。
- 已在每个新集群上安装了 AWS Java SDK 捆绑包。这是一个包含所有服务SDKs及其依赖关系的单个 jar，而不是单个组件 jar。有关更多信息，请参阅 [Java SDK 捆绑依赖关系](#)。

## 已知问题

- 对于 Amazon EMR 6.3.0 和 6.2.0 私有子网集群，您无法访问 Ganglia 网页用户界面。您将收到“access denied (403)”错误。其他网络UIs，例如 Spark、Hue JupyterHub、Zeppelin、Livy 和 Tez，都正常运行。公有子网集群上的 Ganglia Web UI 访问也正常工作。要解决该问题，请在具有 `sudo systemctl restart httpd` 的主节点上重新启动 httpd 服务。此问题已在 Amazon EMR 6.4.0 中修复。
- 启用 AWS Glue 数据目录后，使用 Spark 访问字符串位置为空的 AWS Glue 数据库URI可能会失败。亚马逊的早期EMR版本会发生这种情况，但是 SPARK -31709 (<https://issues.apache.org/jira/browse/SPARK-31709>) 使其适用于更多案例。例如，在默认 AWS Glue DB 中创建位置URI 为空字符串的表时，`spark.sql("CREATE TABLE mytest (key string) location '/table_path';")` 会失败，并显示消息“无法从空字符串创建路径”。要解决此问题，请手动设置 AWS Glue 数据库的位置URI，然后使用 Spark 在这些数据库中创建表。
- 在亚马逊 EMR 6.3.0 中，Presto SQL 已从 343 版升级到 350 版。开源中有两个与安全相关的更改与此版本更改相关。未定义表、架构或会话属性规则时，基于文件的目录访问控制已从 deny 更改为 allow。此外，基于文件的系统访问控制已更改为支持目录规则未定义的文件。在这种情况下，允许完全访问目录。

有关更多信息，请参阅[发行版 344 \(2020 年 10 月 9 日\)](#)。

- 请注意，所有人都可读取 Hadoop 用户目录 (`/home/hadoop`)。它具有 Unix 755 (`drwxr-xr-x`) 目录权限，允许像 Hive 这样的框架进行读取访问。您可以将文件放入 `/home/hadoop` 及其子目录中，但请注意这些目录的权限，做好对敏感信息的保护。
- 降低旧版本的“最大打开文件数”限制 AL2 [在新版本中已修复]。亚马逊EMR版本：`emr-5.30.x`、`emr-5.31.0`、`emr-5.32.0`、`emr-6.0.0`、`emr-6.1.0` 和 `emr-6.2.0` 基于较旧版本的 Linux ofAmazon 2 ()AL2，当使用默认版本创建亚马逊集群时，“最大打开文件数”的用户限制较低。EMR AMI亚马逊EMR发布的版本为5.30.1、5.30.2、5.31.1、5.32.1、6.0.1、6.1.1、6.2.1、5.33.0、6.3.0 及更高版本，包括具有更高的“最大打开文件数”设置的永久修复。如果使用打开文件数限制较低的发行版，会在提交 Spark 任务时导致“Too many open files”（打开的文件过多）错误。在受影响的版本中，亚马逊EMR默认AMI的“最大打开文件数”的ulimit设置为4096，低于Linux 2中的65536个文件限制。latestAmazon AMISpark 驱动程序和执行程序尝试打开超过 4096 个文件时，“打开的最大文件数”的较低 ulimit 设置会导致 Spark 任务失败。为了解决这个问题，Amazon EMR 提供了一个引导操作 (BA) 脚本，可以在创建集群时调整ulimit设置。

如果您使用的是无法永久修复此问题的旧EMR版 Amazon，则可以使用以下解决方法将实例控制器 ulimit 明确设置为最多 65536 个文件。

## 从命令行显式设置 ulimit

1. 编辑 `/etc/systemd/system/instance-controller.service` , 将以下参数添加到 Service (服务) 部分。

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. 重启 InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

## 使用引导操作 (BA) 设置 ulimit

您还可以在创建集群时使用引导操作 ( BA ) 脚本将实例控制器 ulimit 配置为 65536 个文件。

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

### Important

EMR运行 Amazon Linux 或 Amazon Linux 2 Amazon 机器映像 (AMIs) 的集群使用默认 Amazon Linux 行为 , 不会自动下载和安装需要重启的重要关键内核更新。这与运行默认

Amazon Linux 的其他亚马逊EC2实例的行为相同AMI。如果在 Amazon EMR 版本发布后需要重启的新 Amazon Linux 软件CUDA更新（例如内核和更新）可用，则默认运行的EMR集群实例AMI不会自动下载和安装这些更新。NVIDIA要获取内核更新，您可以[自定义您的亚马逊EMRAMI](#)以[使用最新的亚马逊 Linux AMI](#)。

- 要将 Spark 操作与 Apache Oozie 一起使用，必须将以下配置添加到 Oozie workflow.xml 文件中。否则，Oozie 启动的 Spark 执行程序的路径中EMRFS将缺少几个关键库，例如 Hadoop 和。

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

- 当你使用带有 Hive 分区位置格式的 Spark 来读取 Amazon S3 中的数据，并在亚马逊EMR版本 5.30.0 到 5.36.0 和 6.2.0 到 6.9.0 上运行 Spark 时，你可能会遇到一个问题，导致集群无法正确读取数据。如果您的分区具有以下所有特征，会发生这种情况：
  - 从同一个表扫描两个或多个分区。
  - 至少有一个分区目录路径是至少一个其他分区目录路径的前缀，例如，s3://bucket/table/p=a 是 s3://bucket/table/p=a b 的前缀。
  - 另一个分区目录中前缀之后的第一个字符的值为 UTF -8，小于该/字符 (U+002F)。例如，在 s3://bucket/table/p=a b 中，a 和 b 之间出现的空格字符 (U+0020) 就属于此类。请注意，还有其他 14 个非控制字符：!"#\$%&'()\*+,-。有关更多信息，请参阅[UTF-8 编码表和 Unicode 字符](#)。

解决方法是在 spark-defaults 分类中将

spark.sql.sources.fastS3PartitionDiscovery.enabled 配置设置为 false。

### 6.3.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 emr 或 aws 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.4.1	亚马逊 SageMaker Spark SDK
emr-ddb	4.16.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	3.2.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.5.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-notebook-env	1.2.0	适用于 EMR Notebooks ( 可提供 jupyter 企业网关 ) 的 Conda env
emr-s3-dist-cp	2.18.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	2.1.0	EMRS3Select 连接器
emrfs	2.46.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.12.1	Apache Flink 命令行客户端脚本和应用程序。
flink-jobmanager-config	1.12.1	管理 Apache Flink EMR JobManager 节点上的资源。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。

组件	版本	描述
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	3.2.1-amzn-3	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	3.2.1-amzn-3	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	3.2.1-amzn-3	HDFS命令行客户端和库
hadoop-hdfs-namenode	3.2.1-amzn-3	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	3.2.1-amzn-3	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	3.2.1-amzn-3	HTTPHDFS操作端点。
hadoop-kms-server	3.2.1-amzn-3	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	3.2.1-amzn-3	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	3.2.1-amzn-3	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	3.2.1-amzn-3	YARN用于分配和管理群集资源和分布式应用程序的服务。



组件	版本	描述
hadoop-yarn-timeline-server	3.2.1-amzn-3	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	2.2.6-amzn-1	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	2.2.6-amzn-1	为一个或多个HBase地区提供服务的服务。
hbase-client	2.2.6-amzn-1	HBase命令行客户端。
hbase-rest-server	2.2.6-amzn-1	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	2.2.6-amzn-1	为其提供 Thrift 端点的HBase服务。
hcatalog-client	3.1.2-amzn-4	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	3.1.2-amzn-4	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	3.1.2-amzn-4	HTTP端点为其提供REST接口HCatalog。
hive-client	3.1.2-amzn-4	Hive 命令行客户端。
hive-hbase	3.1.2-amzn-4	Hive-hbase 客户端。
hive-metastore-server	3.1.2-amzn-4	用于访问 Hive 元数据仓库的服务，这是一个语义存储库，用于存储 Hadoop 操作SQL的元数据。
hive-server2	3.1.2-amzn-4	用于将 Hive 查询作为 Web 请求接受的服务。

组件	版本	描述
hudi	0.7.0-amzn-0	增量处理框架，以支持低延迟和高效率的数据管道。
hudi-presto	0.7.0-amzn-0	用于运行 Presto 以及 Hudi 的捆绑库。
hudi-prestosql	0.7.0-amzn-0	用于使用 Hudi 运行 Presto SQL 的捆绑库。
hudi-spark	0.7.0-amzn-0	用于运行 Spark 以及 Hudi 的捆绑库。
hue-server	4.9.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	1.2.2	Jupyter notebook 的多用户服务器
livy-server	0.7.0-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mxnet	1.7.0	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.68+	MariaDB 数据库服务器。
nvidia-cuda	10.1.243	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.1	Oozie 命令行客户端。
oozie-server	5.2.1	用于接受 Oozie 工作流请求的服务。

组件	版本	描述
opencv	4.5.0	开源计算机视觉库。
phoenix-library	5.0.0--2.0 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	5.0.0--2.0 HBase	轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限 API
presto-coordinator	0.245.1-amzn-0	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.245.1-amzn-0	用于执行查询的各个部分的服务。
presto-client	0.245.1-amzn-0	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
prestosql-coordinator	350	用于在 prestosql-worker 之中接受查询并管理查询执行的服务。
prestosql-worker	350	用于执行查询的各个部分的服务。
prestosql-client	350	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
pig-client	0.17.0	Pig 命令行客户端。
r	4.0.2	用于统计计算的 R 项目
ranger-kms-server	2.0.0	Apache Ranger 密钥管理系统
spark-client	3.1.1-amzn-0	Spark 命令行客户端。

组件	版本	描述
spark-history-server	3.1.1-amzn-0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	3.1.1-amzn-0	的内存中执行引擎。YARN
spark-yarn-slave	3.1.1-amzn-0	YARN从属服务器需要的 Apache Spark 库。
spark-rapids	0.4.1	Nvidia Spark RAPIDS 插件可以加速 Apache Spark。GPUs
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	2.4.1	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.2	tez YARN 应用程序和库。
webserver	2.4.41+	Apache HTTP 服务器。
zeppelin-server	0.9.0	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.14	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.14	ZooKeeper 命令行客户端。

### 6.3.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

为处于运行状态的集群中的实例组指定配置时，将发生重新配置操作。Amazon EMR 仅对您修改的分类启动重新配置操作。有关更多信息，请参阅 [在正在运行的集群中重新配置实例组](#)。

## emr-6.3.0 分类

分类	描述	重新配置操作
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。	Restarts the Resource Manager service.
container-executor	更改 Hadoop YARN 的 container-executor.cfg 文件中的值。	Not available.
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。	Not available.
core-site	更改 Hadoop 的 core-site.xml 文件中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	更改 docker 相关设置。	Not available.
emrfs-site	更改EMRFS设置。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager,

分类	描述	重新配置操作
		NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	更改 flink-conf.yaml 设置。	Restarts Flink history server.
flink-log4j	更改 Flink log4j.properties 设置。	Restarts Flink history server.
flink-log4j-session	为 Kubernetes/Yarn 会话更改 Flink log4j-session.properties 设置。	Restarts Flink history server.
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。	Restarts Flink history server.
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.

分类	描述	重新配置操作
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	更改 hadoop ssl 服务器配置	Not available.
hadoop-ssl-client	更改 hadoop ssl 客户端配置	Not available.
hbase	亚马逊EMR精心策划的 Apache 设置。HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	更改环境HBase中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。	Not available.

分类	描述	重新配置操作
hbase-site	更改HBase的 hbase-site.xml 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	配置HDFS加密区域。	This classification should not be reconfigured.
hdfs-env	更改HDFS环境中的值。	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	更改环境HCatalog中的值。	Restarts Hive HCatalog Server.
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。	Restarts Hive WebHCat server.



分类	描述	重新配置操作
hcatalog-webhcat-site	更改 HCatalog WebHCat 的 webhcat-site.xml 文件中的值。	Restarts Hive WebHCat server.
hive	亚马逊EMR精心策划的 Apache Hive 设置。	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。	Not available.
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。	Not available.
hive-env	更改 Hive 环境中的值。	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	更改 Hive 的 hive-exec-log4j2.properties 文件中的值。	Not available.
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log4j2.properties 文件中的值。	Not available.
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。	Not available.
hive-site	更改 Hive 的 hive-site.xml 文件中的值	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.

分类	描述	重新配置操作
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值	Not available.
hue-ini	更改 Hue 的 ini 文件中的值	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	更改HTTPFS环境中的值。	Restarts Hadoop Httpfs service.
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。	Restarts Hadoop Httpfs service.
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。	Not available.
hadoop-kms-env	在 Hadoop KMS 环境中更改值。	Restarts Hadoop-KMS service.
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。	Not available.
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	更改 Hudi 环境中的值。	Not available.
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。	Not available.
jupyter-hub-conf	更改 JupyterHubs的 jupyterhub_config.py 文件中的值。	Not available.

分类	描述	重新配置操作
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。	Not available.
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.js on 文件中的值。	Not available.
livy-conf	更改 Livy 的 livy.conf 文件中的值。	Restarts Livy Server.
livy-env	更改 Livy 环境中的值。	Restarts Livy Server.
livy-log4j	更改 Livy log4j.properties 设置。	Restarts Livy Server.
mapred-env	更改 MapReduce 应用程序环境中的值。	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	更改 Oozie 的环境中的值。	Restarts Oozie.
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。	Restarts Oozie.
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。	Restarts Oozie.
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。	Not available.
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。	Not available.
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。	Restarts Phoenix-QueryServer.

分类	描述	重新配置操作
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。	Not available.
pig-env	更改 Pig 环境中的值。	Not available.
pig-properties	更改 Pig 的 pig.properties 文件中的值。	Restarts Oozie.
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。	Not available.
presto-log	更改 Presto 的 log.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-config	更改 Presto 的 config.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。	Not available.
presto-env	更改 Presto 的 presto-env.sh 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-node	更改 Presto 的 node.properties 文件中的值。	Not available.
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。	Not available.
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。	Not available.
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)

分类	描述	重新配置操作
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。	Not available.
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。	Not available.
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。	Not available.
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。	Not available.
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。	Not available.
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。	Not available.
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。	Not available.
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。	Not available.
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。	Not available.
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。	Not available.
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。	Not available.
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。	Not available.
prestosql-log	更改 Presto 的 log.properties 文件中的值。	Restarts Presto-Server (for PrestoSQL)

分类	描述	重新配置操作
prestosql-config	更改 Presto 的 config.properties 文件中的值。	Restarts Presto-Server (for PrestoSQL)
prestosql-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。	Restarts Presto-Server (for PrestoSQL)
prestosql-env	更改 Presto 的 presto-env.sh 文件中的值。	Restarts Presto-Server (for PrestoSQL)
prestosql-node	更改 Presto 的 node.properties SQL 文件中的值。	Not available.
prestosql-connector-blackhole	更改 Presto 的 blackhole.properties SQL 文件中的值。	Not available.
prestosql-connector-cassandra	更改 Presto 的 cassandra.properties SQL 文件中的值。	Not available.
prestosql-connector-hive	更改 Presto 的 hive.properties SQL 文件中的值。	Restarts Presto-Server (for PrestoSQL)
prestosql-connector-jmx	更改 Presto 的 jmx.properties SQL 文件中的值。	Not available.
prestosql-connector-kafka	更改 Presto 的 kafka.properties SQL 文件中的值。	Not available.
prestosql-connector-localfile	更改 Presto 的 localfile.properties SQL 文件中的值。	Not available.
prestosql-connector-memory	更改 Presto 的内存.properties SQL 文件中的值。	Not available.
prestosql-connector-mongodb	更改 Presto 的 mongodb.properties SQL 文件中的值。	Not available.

分类	描述	重新配置操作
prestosql-connector-mysql	更改 Presto 的 mysql.prop SQL erties 文件中的值。	Not available.
prestosql-connector-postgresql	更改 Presto 的 postgresql.prop SQL erties 文件中的值。	Not available.
prestosql-connector-raptor	更改 Presto 的 raptor.prop SQL erties 文件中的值。	Not available.
prestosql-connector-redis	更改 Presto 的 redis.prop SQL erties 文件中的值。	Not available.
prestosql-connector-redshift	更改 Presto 的 reds SQL hift.properties 文件中的值。	Not available.
prestosql-connector-tpch	更改 Presto 的 tpch.prop SQL erties 文件中的值。	Not available.
prestosql-connector-tpcds	更改 Presto 的 tpcds.prop SQL erties 文件中的值。	Not available.
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site .xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-site	更改 Ran ranger-kms-site KMS ger .xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-env	在游侠KMS环境中更改值。	Restarts Ranger KMS Server.
ranger-kms-log4j	更改 Ranger 的 kms-log4j .properties 文件中的值。KMS	Not available.
ranger-kms-db-ca	更改 S3 上我与 Ranger SQL SSL KMS 连接的 CA 文件值。	Not available.
spark	亚马逊EMR精心策划的 Apache Spark 设置。	This property modifies spark-defaults. See actions there.

分类	描述	重新配置操作
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-env	更改 Spark 环境中的值。	Restarts Spark history server and Spark thrift server.
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值	Not available.
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
sqoop-env	更改 Sqoop 的环境中的值。	Not available.
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。	Not available.
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。	Not available.
tez-site	更改 Tez 的 tez-site.xml 文件中的值。	Restart Oozie and HiveServer2.
yarn-env	更改YARN环境中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.



分类	描述	重新配置操作
yarn-site	更改YARN的 yarn-site.xml 文件中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	更改 Zeppelin 环境中的值。	Restarts Zeppelin.
zeppelin-site	更改 zeppelin-site.xml 中的配置设置。	Restarts Zeppelin.
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。	Restarts Zookeeper server.
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。	Restarts Zookeeper server.

## 亚马逊EMR版本 6.2.1

### 6.2.1 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[JupyterEnterpriseGateway](#)、[Jupyter](#)和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

## 应用程序版本信息

	emr-6.2.1	emr-6.2.0	emr-6.1.1	emr-6.1.0
适用于 Java 的 AWS SDK	1.11.880	1.11.880	1.11.828	1.11.828
Python	2.7、3.7	2.7、3.7	2.7、3.7	2.7、3.7
Scala	2.12.10	2.12.10	2.12.10	2.12.10
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.11.2	1.11.2	1.11.0	1.11.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.2.6-amzn-0	2.2.6-amzn-0	2.2.5	2.2.5
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.6.0-amzn-1	0.6.0-amzn-1	0.5.2-incubating-amzn-2	0.5.2-incubating-amzn-2
Hue	4.8.0	4.8.0	4.7.1	4.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	-	-
JupyterHub	1.1.0	1.1.0	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0

	emr-6.2.1	emr-6.2.0	emr-6.1.1	emr-6.1.0
MXNet	1.7.0	1.7.0	1.6.0	1.6.0
Mahout	-	-	-	-
Oozie	5.2.0	5.2.0	5.2.0	5.2.0
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.238.3	0.238.3	0.232	0.232
Spark	3.0.1	3.0.1	3.0.0	3.0.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.3.1	2.3.1	2.1.0	2.1.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	343	343	338	338
Zeppelin	0.9.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 6.2.1 发布说明

此版本旨在修复 Amazon EMR Scaling 无法成功扩展/缩小集群或导致应用程序故障时出现的问题。

### 更改、增强和解决的问题

- 修复了 Amaz EMR on 集群上守护程序运行状况检查活动（例如收集YARN节点状态和HDFS节点状态）时，对利用率很高的大型集群的扩展请求失败的问题。之所以发生这种情况，是因为集群上的守护程序无法将节点的运行状况数据传送到内部 Ama EMR zon 组件。
- 改进了EMR集群上的守护程序，可在重复使用 IP 地址时正确跟踪节点状态，从而提高扩展操作期间的可靠性。

- [SPARK-29683](#)。修复了集群缩减期间出现任务失败的问题，因为 Spark 假定所有可用节点都被拒绝列出。
- [YARN-9011](#)。修复了集群尝试向上或向下扩展时，由于YARN停用过程中的争用条件而导致任务失败的问题。
- 通过确保 Amaz EMR on 集群上守护程序和 YARN /之间的节点状态始终保持一致，修复了集群扩展期间步骤或任务失败的问题。HDFS
- 修复了启用了 Kerberos 身份验证的 Amazon 集群的EMR集群操作（例如缩减和步骤提交）失败的问题。这是因为 Amaz EMR on 集群上守护程序没有续订 Kerberos 票证，而要与主节点上YARN运行的 HDFS /进行安全通信，则需要续订 Kerberos 票证。
- 较新的Amazon EMR 版本通过降低亚马逊较旧AL2版本的“最大打开文件数”限制来解决这个问题 EMR。亚马逊EMR发布的 5.30.1、5.30.2、5.31.1、5.32.1、5.32.1、6.0.1、6.1.1、6.2.1、5.33.0、6.3.0及更高版本现在包含一个永久修复程序，其中包含一个具有更高的“最大打开文件数”设置的永久修复。
- HTTPS现在，亚马逊 Linux 存储库已默认启用。如果您使用 Amazon S3 VPCE 策略来限制对特定存储桶的访问，则必须将新的 Amazon Linux 存储桶ARNarn:aws:s3:::amazonlinux-2-repos-\$region/\*添加到您的策略中（\$region替换为终端节点所在的区域）。有关更多信息，请在 AWS 讨论论坛中查看此主题。[公告：Amazon Linux 2 现在支持HTTPS在连接软件包存储库时使用](#)。

## 已知问题

- 当你使用带有 Hive 分区位置格式的 Spark 来读取 Amazon S3 中的数据，并在亚马逊EMR版本 5.30.0 到 5.36.0 和 6.2.0 到 6.9.0 上运行 Spark 时，你可能会遇到一个问题，导致集群无法正确读取数据。如果您的分区具有以下所有特征，会发生这种情况：
  - 从同一个表扫描两个或多个分区。
  - 至少有一个分区目录路径是至少一个其他分区目录路径的前缀，例如，s3://bucket/table/p=a 是 s3://bucket/table/p=a b 的前缀。
  - 另一个分区目录中前缀之后的第一个字符的值为 UTF -8，小于该/字符 (U+002F)。例如，在 s3://bucket/table/p=a b 中，a 和 b 之间出现的空格字符 (U+0020) 就属于此类。请注意，还有其他 14 个非控制字符：! "\$ % & ' ( ) \* + , -。有关更多信息，请参阅 [UTF-8 编码表和 Unicode 字符](#)。

解决方法是在 spark-defaults 分类中将

spark.sql.sources.fastS3PartitionDiscovery.enabled 配置设置为 false。

## 6.2.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>aws-sagemaker-spark-sdk</code>	1.4.1	亚马逊 SageMaker Spark SDK
<code>emr-ddb</code>	4.16.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	3.1.0	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.5.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
<code>emr-notebook-env</code>	1.0.0	适用于 EMR Notebooks ( 可提供 jupyter 企业网关 ) 的 Conda env
<code>emr-s3-dist-cp</code>	2.16.0	针对 Amazon S3 优化的分布式复制应用程序。
<code>emr-s3-select</code>	2.0.0	EMRS3Select 连接器
<code>emrfs</code>	2.44.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。

组件	版本	描述
flink-client	1.11.2	Apache Flink 命令行客户端脚本和应用程序。
flink-jobmanager-config	1.11.2	管理 Apache Flink EMR JobManager 节点上的资源。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	3.2.1-amzn-2.1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	3.2.1-amzn-2.1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	3.2.1-amzn-2.1	HDFS命令行客户端和库
hadoop-hdfs-namenode	3.2.1-amzn-2.1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	3.2.1-amzn-2.1	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	3.2.1-amzn-2.1	HTTPHDFS操作端点。
hadoop-kms-server	3.2.1-amzn-2.1	基于 Hadoop 的加密密钥管理服务。 KeyProvider API

组件	版本	描述
hadoop-mapred	3.2.1-amzn-2.1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	3.2.1-amzn-2.1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	3.2.1-amzn-2.1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	3.2.1-amzn-2.1	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	2.2.6-amzn-0	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	2.2.6-amzn-0	为一个或多个HBase地区提供服务的服务。
hbase-client	2.2.6-amzn-0	HBase命令行客户端。
hbase-rest-server	2.2.6-amzn-0	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	2.2.6-amzn-0	为其提供 Thrift 端点的HBase服务。
hcatalog-client	3.1.2-amzn-3	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	3.1.2-amzn-3	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	3.1.2-amzn-3	HTTP端点为其提供REST接口HCatalog。
hive-client	3.1.2-amzn-3	Hive 命令行客户端。

组件	版本	描述
hive-hbase	3.1.2-amzn-3	Hive-hbase 客户端。
hive-metastore-server	3.1.2-amzn-3	用于访问 Hive 元数据仓库的服务，这是一个语义存储库，用于存储 Hadoop 操作 SQL 的元数据。
hive-server2	3.1.2-amzn-3	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.6.0-amzn-1	增量处理框架，以支持低延迟和高效率的数据管道。
hudi-presto	0.6.0-amzn-1	用于运行 Presto 以及 Hudi 的捆绑库。
hudi-prestosql	0.6.0-amzn-1	用于使用 Hudi 运行 Presto SQL 的捆绑库。
hudi-spark	0.6.0-amzn-1	用于运行 Spark 以及 Hudi 的捆绑库。
hue-server	4.8.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	1.1.0	Jupyter notebook 的多用户服务器
livy-server	0.7.0-incubating	REST 用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mxnet	1.7.0	用于深度学习的灵活的、可扩展且高效的库。



组件	版本	描述
mariadb-server	5.5.64+	MariaDB 数据库服务器。
nvidia-cuda	10.1.243	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.0	Oozie 命令行客户端。
oozie-server	5.2.0	用于接受 Oozie 工作流请求的服务。
opencv	4.4.0	开源计算机视觉库。
phoenix-library	5.0.0--2.0 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	5.0.0--2.0 HBase	轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限 API
presto-coordinator	0.238.3-amzn-1	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.238.3-amzn-1	用于执行查询的各个部分的服务。
presto-client	0.238.3-amzn-1	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
prestosql-coordinator	343	用于在 prestosql-worker 之中接受查询并管理查询执行的服务。
prestosql-worker	343	用于执行查询的各个部分的服务。
prestosql-client	343	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。

组件	版本	描述
pig-client	0.17.0	Pig 命令行客户端。
r	3.4.3	用于统计计算的 R 项目
ranger-kms-server	2.0.0	Apache Ranger 密钥管理系统
spark-client	3.0.1-amzn-0.1	Spark 命令行客户端。
spark-history-server	3.0.1-amzn-0.1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	3.0.1-amzn-0.1	的内存中执行引擎。YARN
spark-yarn-slave	3.0.1-amzn-0.1	YARN从属服务器需要的 Apache Spark 库。
spark-rapids	0.2.0	Nvidia Spark RAPIDS 插件可以加速 Apache Spark。GPUs
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	2.3.1	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.2	tez YARN 应用程序和库。
webserver	2.4.41+	Apache HTTP 服务器。
zeppelin-server	0.9.0-preview1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.14	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.14	ZooKeeper 命令行客户端。

## 6.2.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

为处于运行状态的集群中的实例组指定配置时，将发生重新配置操作。Amazon EMR 仅对您修改的分类启动重新配置操作。有关更多信息，请参阅 [在正在运行的集群中重新配置实例组](#)。

### emr-6.2.1 分类

分类	描述	重新配置操作
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。	Restarts the Resource Manager service.
container-executor	更改 Hadoop YARN 的 container-executor.cfg 文件中的值。	Not available.
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。	Not available.
core-site	更改 Hadoop 的 core-site.xml 文件中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	更改 docker 相关设置。	Not available.

分类	描述	重新配置操作
emrfs-site	更改EMRFS设置。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	更改 flink-conf.yaml 设置。	Not available.
flink-log4j	更改 Flink log4j.properties 设置。	Not available.
flink-log4j-yarn-session	更改 Flink log4j-yarn-session.properties 设置。	Not available.
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。	Not available.

分类	描述	重新配置操作
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	更改 hadoop ssl 服务器配置	Not available.
hadoop-ssl-client	更改 hadoop ssl 客户端配置	Not available.
hbase	亚马逊EMR精心策划的 Apache 设置。HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.

分类	描述	重新配置操作
hbase-env	更改环境HBase中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	更改 hadoop-metrics2-hbase base.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。	Not available.
hbase-site	更改HBase的 hbase-site.xml 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	配置HDFS加密区域。	This classification should not be reconfigured.
hdfs-env	更改HDFS环境中的值。	Restarts Hadoop HDFS ZKFC.
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	更改环境HCatalog中的值。	Restarts Hive HCatalog Server.

分类	描述	重新配置操作
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。	Restarts Hive WebHCat server.
hive	亚马逊EMR精心策划的 Apache Hive 设置。	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。	Not available.
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。	Not available.
hive-env	更改 Hive 环境中的值。	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	更改 Hive 的 hive-exec-log4j2.properties 文件中的值。	Not available.
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log4j2.properties 文件中的值。	Not available.

分类	描述	重新配置操作
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。	Not available.
hive-site	更改 Hive 的 hive-site.xml 文件中的值	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值	Not available.
hue-ini	更改 Hue 的 ini 文件中的值	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	更改HTTPFS环境中的值。	Restarts Hadoop Httpfs service.
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。	Restarts Hadoop Httpfs service.
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。	Not available.
hadoop-kms-env	在 Hadoop KMS 环境中更改值。	Restarts Hadoop-KMS service.
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。	Not available.
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。	Restarts Hadoop-KMS and Ranger-KMS service.



分类	描述	重新配置操作
hudi-env	更改 Hudi 环境中的值。	Not available.
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。	Not available.
jupyter-hub-conf	更改 JupyterHubs的 jupyterhub_config.py 文件中的值。	Not available.
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。	Not available.
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。	Not available.
livy-conf	更改 Livy 的 livy.conf 文件中的值。	Restarts Livy Server.
livy-env	更改 Livy 环境中的值。	Restarts Livy Server.
livy-log4j	更改 Livy log4j.properties 设置。	Restarts Livy Server.
mapred-env	更改 MapReduce 应用程序环境中的值。	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	更改 Oozie 的环境中的值。	Restarts Oozie.
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。	Restarts Oozie.
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。	Restarts Oozie.

分类	描述	重新配置操作
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。	Not available.
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。	Not available.
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。	Restarts Phoenix-QueryServer.
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。	Not available.
pig-env	更改 Pig 环境中的值。	Not available.
pig-properties	更改 Pig 的 pig.properties 文件中的值。	Restarts Oozie.
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。	Not available.
presto-log	更改 Presto 的 log.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-config	更改 Presto 的 config.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。	Not available.
presto-env	更改 Presto 的 presto-env.sh 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-node	更改 Presto 的 node.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。	Not available.
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。	Not available.
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。	Not available.
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。	Not available.
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。	Not available.
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。	Not available.
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。	Not available.
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。	Not available.
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。	Not available.
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。	Not available.
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。	Not available.
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。	Not available.
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。	Not available.
prestoql-log	更改 Presto 的 log.properties 文件中的值。	Restarts Presto-Server (for PrestoSQL)
prestoql-config	更改 Presto 的 config.properties 文件中的值。	Restarts Presto-Server (for PrestoSQL)
prestoql-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。	Restarts Presto-Server (for PrestoSQL)
prestoql-env	更改 Presto 的 presto-env.sh 文件中的值。	Restarts Presto-Server (for PrestoSQL)
prestoql-node	更改 Presto 的 node.properties SQL 文件中的值。	Not available.
prestoql-connector-blackhole	更改 Presto 的 blackhole.properties SQL 文件中的值。	Not available.
prestoql-connector-cassandra	更改 Presto 的 cassandra.properties SQL 文件中的值。	Not available.
prestoql-connector-hive	更改 Presto 的 hive.properties SQL 文件中的值。	Restarts Presto-Server (for PrestoSQL)
prestoql-connector-jmx	更改 Presto 的 jmx.properties SQL 文件中的值。	Not available.
prestoql-connector-kafka	更改 Presto 的 kafka.properties SQL 文件中的值。	Not available.

分类	描述	重新配置操作
prestosql-connector-localfile	更改 Presto 的 localfile.prop SQL erties 文件中的值。	Not available.
prestosql-connector-memory	更改 Presto 的内存.properties SQL erties 文件中的值。	Not available.
prestosql-connector-mongodb	更改 Presto 的 mongodb.properties SQL erties 文件中的值。	Not available.
prestosql-connector-mysql	更改 Presto 的 mysql.properties SQL erties 文件中的值。	Not available.
prestosql-connector-postgresql	更改 Presto 的 postgresql.properties SQL erties 文件中的值。	Not available.
prestosql-connector-raptor	更改 Presto 的 raptor.properties SQL erties 文件中的值。	Not available.
prestosql-connector-redis	更改 Presto 的 redis.properties SQL erties 文件中的值。	Not available.
prestosql-connector-redshift	更改 Presto 的 redshift.properties SQL erties 文件中的值。	Not available.
prestosql-connector-tpch	更改 Presto 的 tpch.properties SQL erties 文件中的值。	Not available.
prestosql-connector-tpcds	更改 Presto 的 tpcds.properties SQL erties 文件中的值。	Not available.
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-site	更改 Ranger KMS 的 ranger-kms-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-env	在 Ranger KMS 环境中更改值。	Restarts Ranger KMS Server.

分类	描述	重新配置操作
ranger-kms-log4j	更改 Ranger 的 kms-log4j .properties 文件中的值。KMS	Not available.
ranger-kms-db-ca	更改 S3 上我与 Ranger SQL SSL KMS 连接的 CA 文件值。	Not available.
spark	亚马逊EMR精心策划的 Apache Spark 设置。	This property modifies spark-defaults. See actions there.
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-env	更改 Spark 环境中的值。	Restarts Spark history server and Spark thrift server.
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值	Not available.
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
sqoop-env	更改 Sqoop 的环境中的值。	Not available.
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。	Not available.
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。	Not available.
tez-site	更改 Tez 的 tez-site.xml 文件中的值。	Restart Oozie.

分类	描述	重新配置操作
yarn-env	更改YARN环境中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	更改YARN的 yarn-site.xml 文件中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	更改 Zeppelin 环境中的值。	Restarts Zeppelin.
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。	Restarts Zookeeper server.
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。	Restarts Zookeeper server.

## 亚马逊EMR版本 6.2.0

### 6.2.0 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[JupyterEnterpriseGateway](#)、[Jupyter](#)和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)

- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

## 应用程序版本信息

	emr-6.2.0	emr-6.1.1	emr-6.1.0	emr-6.0.1
适用于 Java 的 AWS SDK	1.11.880	1.11.828	1.11.828	1.11.711
Python	2.7、3.7	2.7、3.7	2.7、3.7	2.7、3.7
Scala	2.12.10	2.12.10	2.12.10	2.12.10
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.11.2	1.11.0	1.11.0	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.2.6-amzn-0	2.2.5	2.2.5	2.2.3
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.6.0-amzn-1	0.5.2-incubating-amzn-2	0.5.2-incubating-amzn-2	0.5.0-incubating-amzn-1
Hue	4.8.0	4.7.1	4.7.1	4.4.0
Iceberg	-	-	-	-



	emr-6.2.0	emr-6.1.1	emr-6.1.0	emr-6.0.1
JupyterEnterpriseGateway	2.1.0	-	-	-
JupyterHub	1.1.0	1.1.0	1.1.0	1.0.0
Livy	0.7.0	0.7.0	0.7.0	0.6.0
MXNet	1.7.0	1.6.0	1.6.0	1.5.1
Mahout	-	-	-	-
Oozie	5.2.0	5.2.0	5.2.0	5.1.0
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	0.17.0	-
Presto	0.238.3	0.232	0.232	0.230
Spark	3.0.1	3.0.0	3.0.0	2.4.4
Sqoop	1.4.7	1.4.7	1.4.7	-
TensorFlow	2.3.1	2.1.0	2.1.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	343	338	338	-
Zeppelin	0.9.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 6.2.0 发布说明

以下发行说明包含有关亚马逊 6.2.0 EMR 版本的信息。更改与 6.1.0 有关。

首次发布日期：2020 年 12 月 9 日

上次更新日期：2021 年 10 月 4 日

## 受支持的应用程序

- AWS SDK for Java 版本 1.11.828
- emr-record-server 版本 1.7.0
- Flink 1.11.2
- Ganglia 3.7.2
- Hadoop 3.2.1-amzn-1
- HBase版本 2.2.6-amzn-0
- HBase-operator-tools 1.0.0
- HCatalog版本 3.1.2-amzn-0
- Hive 3.1.2-amzn-3
- Hudi 0.6.0-amzn-1
- Hue 4.8.0
- JupyterHub 版本 1.1.0
- Livy 0.7.0
- MXNet版本 1.7.0
- Oozie 5.2.0
- Phoenix 5.0.0
- Pig 0.17.0
- Presto 0.238.3-amzn-1
- Presto SQL 版本 343
- Spark 3.0.1-amzn-0
- spark-rapids 0.2.0
- TensorFlow 版本 2.3.1
- Zeppelin 0.9.0-preview1
- Zookeeper 3.4.14
- 连接器和驱动程序：DynamoDB 连接器 4.16.0

## 新功能

- HBase: 删除了提交阶段的重命名并添加了持续HFile跟踪。请参阅《Amazon EMR 发布指南》中的[持续HFile跟踪](#)。
- HBase: 向后移植[创建一个在压缩时强制缓存块的配置](#)。
- PrestoDB : 改进了动态分区修剪。基于规则的连接重新排序对未分区数据运行。
- 限定范围的托管策略 : 为了与 AWS 最佳实践保持一致 , Amazon 引入EMR了 v2 EMR 范围的默认托管策略 , 以取代即将弃用的策略。参见 [Amazon EMR 托管政策](#)。
- 实例元数据服务 (IMDS) V2 支持状态 : 对于 Amazon EMR 6.2 或更高版本 , Amazon EMR 组件IMDSv2用于所有IMDS调用。对于应用程序代码中的IMDS调用 , 您可以同时使用IMDSv1和IMDSv2 , 或者将配置IMDS为仅IMDSv2用于增强安全性。如果您IMDSv1在早期的 Amazon EMR 6.x 版本中禁用 , 则会导致集群启动失败。

## 更改、增强功能和解决的问题

- 此版本旨在修复 Amazon EMR Scaling 无法成功扩展/缩小集群或导致应用程序故障时出现的问题。
- 修复了 Amaz EMR on 集群上守护程序运行状况检查活动 ( 例如收集YARN节点状态和HDFS节点状态 ) 时 , 对利用率很高的大型集群的扩展请求失败的问题。之所以发生这种情况 , 是因为集群上的守护程序无法将节点的运行状况数据传送到内部 Ama EMR zon 组件。
- 改进了EMR集群上的守护程序 , 可在重复使用 IP 地址时正确跟踪节点状态 , 从而提高扩展操作期间的可靠性。
- [SPARK-29683](#)。修复了集群缩减期间出现任务失败的问题 , 因为 Spark 假定所有可用节点都被拒绝列出。
- [YARN-9011](#)。修复了集群尝试向上或向下扩展时 , 由于YARN停用过程中的争用条件而导致任务失败的问题。
- 通过确保 Amaz EMR on 集群上守护程序和 YARN /之间的节点状态始终保持一致 , 修复了集群扩展期间步骤或任务失败的问题。HDFS
- 修复了启用了 Kerberos 身份验证的 Amazon 集群的EMR集群操作 ( 例如缩减和步骤提交 ) 失败的问题。这是因为 Amaz EMR on 集群上守护程序没有续订 Kerberos 票证 , 而要与主节点上YARN运行的 HDFS /进行安全通信 , 则需要续订 Kerberos 票证。
- 较新的Amazon EMR 版本通过降低亚马逊较旧AL2版本的“最大打开文件数”限制来解决这个问题 EMR。亚马逊EMR发布的 5.30.1、5.30.2、5.31.1、5.32.1、5.32.1、6.0.1、6.1.1、6.2.1、5.33.0、6.3.0及更高版本现在包含一个永久修复程序 , 其中包含一个具有更高的“最大打开文件数”设置的永久修复。

- Spark : 改进了 Spark 运行时的性能。

## 已知问题

- 亚马逊 EMR 6.2 在 6.2.0 版本中对 `/etc/cron/libinstance-controller-java .d/` 文件设置的权限不正确。EMR 文件的权限应为 644 (`-rw-r--r--`) 时，它们为 645 (`-rw-r--r-x`)。因此，Amazon EMR 版本 6.2 不记录实例状态日志，并且 `/emr/instance-logs` 目录为空。此问题已在 Amazon EMR 6.3.0 及更高版本中修复。

要解决此问题，请在集群启动时将以下脚本作为引导操作运行。

```
#!/bin/bash
sudo chmod 644 /etc/cron.d/libinstance-controller-java
```

- 对于 Amazon EMR 6.2.0 和 6.3.0 私有子网集群，您无法访问 Ganglia 网页用户界面。您将收到“access denied (403)”错误。其他网络UIs，例如 Spark、Hue JupyterHub、Zeppelin、Livy 和 Tez，都正常运行。公有子网集群上的 Ganglia Web UI 访问也正常工作。要解决该问题，请在具有 `sudo systemctl restart httpd` 的主节点上重新启动 httpd 服务。此问题已在 Amazon EMR 6.4.0 中修复。
- 亚马逊 EMR 6.2.0 中存在一个问题，即 httpd 持续失败，导致 Ganglia 不可用。您会收到“cannot connect to the server (无法连接到服务器)”错误。要修复已在运行但存在此问题的群集，SSH 请访问群集主节点，然后将该行 `Listen 80` 添加到 `httpd.conf` 位于的文件中 `/etc/httpd/conf/httpd.conf`。此问题已在 Amazon EMR 6.3.0 中修复。
- HTTPD 使用安全配置时，在 EMR 6.2.0 集群上失败。因此，Ganglia Web 应用程序用户界面不可用。要访问 Ganglia Web 应用程序用户界面，请将 `Listen 80` 添加到集群主节点上的 `/etc/httpd/conf/httpd.conf` 文件中。有关连接到集群的信息，请参阅[使用连接到主节点SSH](#)。

EMR 当您使用安全配置时，笔记本也无法与 EMR 6.2.0 集群建立连接。笔记本将无法列出内核和提交 Spark 任务。我们建议您 EMR 改用其他版本的 Amazon 的 EMR 笔记本电脑。

- 降低旧版本的“最大打开文件数”限制 AL2 [在新版本中已修复]。亚马逊 EMR 版本：`emr-5.30.x`、`emr-5.31.0`、`emr-5.32.0`、`emr-6.0.0`、`emr-6.1.0` 和 `emr-6.2.0` 基于较旧版本的 Linux of Amazon 2 () AL2，当使用默认版本创建亚马逊集群时，“最大打开文件数”的用户限制较低。EMR AMI 亚马逊 EMR 发布的版本为 `5.30.1`、`5.30.2`、`5.31.1`、`5.32.1`、`6.0.1`、`6.1.1`、`6.2.1`、`5.33.0`、`6.3.0` 及更高版本，包括具有更高的“最大打开文件数”设置的永久修复。如果使用打开文件数限制较低的发行版，会在提交 Spark 任务时导致“Too many open files”（打开的文件过多）错误。在受影响的版本中，亚马逊 EMR 默认 AMI 的“最大打开文件数”的 `ulimit` 设置为 4096，低于 Linux 2 中的 65536 个文件限

制。latestAmazon AMISpark 驱动程序和执行程序尝试打开超过 4096 个文件时，“打开的最大文件数”的较低 ulimit 设置会导致 Spark 任务失败。为了解决这个问题，Amazon EMR 提供了一个引导操作 (BA) 脚本，可以在创建集群时调整ulimit设置。

如果您使用的是无法永久修复此问题的旧EMR版 Amazon，则可以使用以下解决方法将实例控制器 ulimit 明确设置为最多 65536 个文件。

#### 从命令行显式设置 ulimit

1. 编辑 `/etc/systemd/system/instance-controller.service`，将以下参数添加到 Service (服务) 部分。

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. 重启 InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

#### 使用引导操作 (BA) 设置 ulimit

您还可以在创建集群时使用引导操作 ( BA ) 脚本将实例控制器 ulimit 配置为 65536 个文件。

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
```

```
done
sudo systemctl daemon-reload
```

### ⚠ Important

Amazon EMR 6.1.0 和 6.2.0 包含一个性能问题，可能会严重影响 Hudi 的所有插入、更新和删除操作。如果您计划在亚马逊 EMR 6.1.0 或 6.2.0 中使用 Hudi，则应联系 AWS 支持人员获取已修补的 Hudi。RPM

### ⚠ Important

EMR运行 Amazon Linux 或 Amazon Linux 2 Amazon 机器映像 (AMIs) 的集群使用默认 Amazon Linux 行为，不会自动下载和安装需要重启的重要关键内核更新。这与运行默认 Amazon Linux 的其他亚马逊EC2实例的行为相同AMI。如果在 Amazon EMR 版本发布后需要重启的新 Amazon Linux 软件CUDA更新（例如内核和更新）可用，则默认运行的EMR集群实例AMI不会自动下载和安装这些更新。NVIDIA要获取内核更新，您可以[自定义您的亚马逊EMRAMI](#)以[使用最新的亚马逊 Linux AMI](#)。

- 亚马逊 EMR 6.2.0 Maven 工件尚未发布。它们将与 Amazon 的未来版本一起发布EMR。
- 使用 HBase storefile 系统表进行持续HFile跟踪不支持HBase区域复制功能。有关HBase区域复制的更多信息，请参阅[时间轴一致的高可用读取](#)。
- 亚马逊 EMR 6.x 和 EMR 5.x Hive 存储桶版本的区别

EMR5.x 使用 OOS Apache Hive 2，而在 EMR 6.x 中使用 OOS Apache Hive 3。开源 Hive2 使用分桶版本 1，而开源 Hive3 使用分桶版本 2。Hive 2 (EMR5.x) 和 Hive 3 (EMR6.x) 之间的这种存储分区版本差异意味着 Hive 存储分区哈希功能不同。请参见以下示例。

下表分别是在 EMR 6.x 和 EMR 5.x 中创建的示例。

```
-- Using following LOCATION in EMR 6.x
CREATE TABLE test_bucketing (id INT, desc STRING)
PARTITIONED BY (day STRING)
CLUSTERED BY(id) INTO 128 BUCKETS
LOCATION 's3://your-own-s3-bucket/emr-6-bucketing/';

-- Using following LOCATION in EMR 5.x
LOCATION 's3://your-own-s3-bucket/emr-5-bucketing/';
```

在 EMR 6.x 和 EMR 5.x 中插入相同的数据。

```
INSERT INTO test_bucketing PARTITION (day='01') VALUES(66, 'some_data');
INSERT INTO test_bucketing PARTITION (day='01') VALUES(200, 'some_data');
```

检查 S3 位置会显示存储桶文件名不同，因为哈希函数在 EMR 6.x ( Hive 3 ) 和 EMR 5.x ( Hive 2 ) 之间有所不同。

```
[hadoop@ip-10-0-0-122 ~]$ aws s3 ls s3://your-own-s3-bucket/emr-6-bucketing/day=01/
2020-10-21 20:35:16          13 000025_0
2020-10-21 20:35:22          14 000121_0
[hadoop@ip-10-0-0-122 ~]$ aws s3 ls s3://your-own-s3-bucket/emr-5-bucketing/day=01/
2020-10-21 20:32:07          13 000066_0
2020-10-21 20:32:51          14 000072_0
```

您还可以通过在 EMR 6.x 的 Hive CLI 中运行以下命令来查看版本差异。请注意，它将返回分桶版本 2。

```
hive> DESCRIBE FORMATTED test_bucketing;
...
Table Parameters:
  bucketing_version      2
...
```

- 具有多个主节点的集群和 Kerberos 身份验证中的已知问题

如果您在 Amazon 5.20.0 及更高EMR版本中运行具有多个主节点和 Kerberos 身份验证的集群，则在集群运行一段时间后，集群操作可能会遇到问题，例如缩小规模或步骤提交。具体时间段取决于您定义的 Kerberos 票证有效期。缩减问题会影响您提交的自动缩减和显式缩减请求。其它集群操作也可能受到影响。

解决办法：

- SSH以hadoop用户身份访问具有多个主节点的EMR群集的主主节点。
- 运行以下命令，为 hadoop 用户续订 Kerberos 票证。

```
kinit -kt <keytab_file> <principal>
```

通常情况下，keytab 文件位于 /etc/hadoop.keytab，而 principal 为 hadoop/<hostname>@<REALM> 格式。

**Note**

此解决方法将在 Kerberos 票证有效期内生效。默认情况下，此持续时间为 10 个小时，但可以通过 Kerberos 设置进行配置。Kerberos 票证过期后，您必须重新运行上述命令。

- 当你使用带有 Hive 分区位置格式的 Spark 来读取 Amazon S3 中的数据，并在亚马逊EMR版本 5.30.0 到 5.36.0 和 6.2.0 到 6.9.0 上运行 Spark 时，你可能会遇到一个问题，导致集群无法正确读取数据。如果您的分区具有以下所有特征，会发生这种情况：
  - 从同一个表扫描两个或多个分区。
  - 至少有一个分区目录路径是至少一个其他分区目录路径的前缀，例如，s3://bucket/table/p=a 是 s3://bucket/table/p=a b 的前缀。
  - 另一个分区目录中前缀之后的第一个字符的值为 UTF -8，小于该/字符 (U+002F)。例如，在 s3://bucket/table/p=a b 中，a 和 b 之间出现的空格字符 (U+0020) 就属于此类。请注意，还有其他 14 个非控制字符：!"#\$%&'()\*+,-。有关更多信息，请参阅 [UTF-8 编码表和 Unicode 字符](#)。

解决方法是在 spark-defaults 分类中将 spark.sql.sources.fastS3PartitionDiscovery.enabled 配置设置为 false。

## 6.2.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 emr 或 aws 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.4.1	亚马逊 SageMaker Spark SDK



组件	版本	描述
emr-ddb	4.16.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	3.1.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.5.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-notebook-env	1.0.0	适用于 EMR Notebooks ( 可提供 jupyter 企业网关 ) 的 Conda env
emr-s3-dist-cp	2.16.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	2.0.0	EMRS3Select 连接器
emrfs	2.44.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.11.2	Apache Flink 命令行客户端脚本和应用程序。
flink-jobmanager-config	1.11.2	管理 Apache Flink EMR JobManager 节点上的资源。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。

组件	版本	描述
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	3.2.1-amzn-2	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	3.2.1-amzn-2	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	3.2.1-amzn-2	HDFS命令行客户端和库
hadoop-hdfs-namenode	3.2.1-amzn-2	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	3.2.1-amzn-2	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	3.2.1-amzn-2	HTTPHDFS操作端点。
hadoop-kms-server	3.2.1-amzn-2	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	3.2.1-amzn-2	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	3.2.1-amzn-2	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	3.2.1-amzn-2	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	3.2.1-amzn-2	用于检索YARN应用程序的当前和历史信息的服务。

组件	版本	描述
hbase-hmaster	2.2.6-amzn-0	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	2.2.6-amzn-0	为一个或多个HBase地区提供服务的服务器。
hbase-client	2.2.6-amzn-0	HBase命令行客户端。
hbase-rest-server	2.2.6-amzn-0	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	2.2.6-amzn-0	为其提供 Thrift 端点的HBase服务。
hcatalog-client	3.1.2-amzn-3	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	3.1.2-amzn-3	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	3.1.2-amzn-3	HTTP端点为其提供REST接口HCatalog。
hive-client	3.1.2-amzn-3	Hive 命令行客户端。
hive-hbase	3.1.2-amzn-3	Hive-hbase 客户端。
hive-metastore-server	3.1.2-amzn-3	用于访问 Hive 元数据仓库的服务，这是一个语义存储库，用于存储 Hadoop 操作SQL的元数据。
hive-server2	3.1.2-amzn-3	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.6.0-amzn-1	增量处理框架，以支持低延迟和高效率的数据管道。

组件	版本	描述
hudi-presto	0.6.0-amzn-1	用于运行 Presto 以及 Hudi 的捆绑库。
hudi-prestosql	0.6.0-amzn-1	用于使用 Hudi 运行 Presto SQL 的捆绑库。
hudi-spark	0.6.0-amzn-1	用于运行 Spark 以及 Hudi 的捆绑库。
hue-server	4.8.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	1.1.0	Jupyter notebook 的多用户服务器
livy-server	0.7.0-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mxnet	1.7.0	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.64+	MariaDB 数据库服务器。
nvidia-cuda	10.1.243	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.0	Oozie 命令行客户端。
oozie-server	5.2.0	用于接受 Oozie  workflow 请求的服务。
opencv	4.4.0	开源计算机视觉库。
phoenix-library	5.0.0--2.0 HBase	服务器和客户端的 phoenix 库

组件	版本	描述
phoenix-query-server	5.0.0--2.0 HBase	轻量级服务器，提供JDBC对Avatica 的访问权限以及协议缓冲区和JSON格式访问权限 API
presto-coordinator	0.238.3-amzn-1	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.238.3-amzn-1	用于执行查询的各个部分的服务。
presto-client	0.238.3-amzn-1	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
prestosql-coordinator	343	用于在 prestosql-worker 之中接受查询并管理查询执行的服务。
prestosql-worker	343	用于执行查询的各个部分的服务。
prestosql-client	343	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
pig-client	0.17.0	Pig 命令行客户端。
r	3.4.3	用于统计计算的 R 项目
ranger-kms-server	2.0.0	Apache Ranger 密钥管理系统
spark-client	3.0.1-amzn-0	Spark 命令行客户端。
spark-history-server	3.0.1-amzn-0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	3.0.1-amzn-0	的内存中执行引擎。YARN

组件	版本	描述
spark-yarn-slave	3.0.1-amzn-0	YARN从属服务器需要的 Apache Spark 库。
spark-rapids	0.2.0	Nvidia Spark RAPIDS 插件可以加速 Apache Spark。GPUs
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	2.3.1	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.2	tez YARN 应用程序和库。
webserver	2.4.41+	Apache HTTP 服务器。
zeppelin-server	0.9.0-preview1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.14	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.14	ZooKeeper 命令行客户端。

## 6.2.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

为处于运行状态的集群中的实例组指定配置时，将发生重新配置操作。Amazon EMR 仅对您修改的分类启动重新配置操作。有关更多信息，请参阅 [在正在运行的集群中重新配置实例组](#)。

## emr-6.2.0 分类

分类	描述	重新配置操作
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。	Restarts the Resource Manager service.
container-executor	更改 Hadoop YARN 的 container-executor.cfg 文件中的值。	Not available.
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。	Not available.
core-site	更改 Hadoop 的 core-site.xml 文件中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	更改 docker 相关设置。	Not available.
emrfs-site	更改EMRFS设置。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager,

分类	描述	重新配置操作
		NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	更改 flink-conf.yaml 设置。	Not available.
flink-log4j	更改 Flink log4j.properties 设置。	Not available.
flink-log4j-yarn-session	更改 Flink log4j-yarn-session.properties 设置。	Not available.
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。	Not available.
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.



分类	描述	重新配置操作
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	更改 hadoop ssl 服务器配置	Not available.
hadoop-ssl-client	更改 hadoop ssl 客户端配置	Not available.
hbase	亚马逊EMR精心策划的 Apache 设置。HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	更改环境HBase中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	更改 hadoop-metrics2-hbase base.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。	Not available.

分类	描述	重新配置操作
hbase-site	更改HBase的 hbase-site.xml 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	配置HDFS加密区域。	This classification should not be reconfigured.
hdfs-env	更改HDFS环境中的值。	Restarts Hadoop HDFS ZKFC.
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	更改环境HCatalog中的值。	Restarts Hive HCatalog Server.
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。	Restarts Hive WebHCat server.
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。	Restarts Hive WebHCat server.

分类	描述	重新配置操作
hive	亚马逊EMR精心策划的 Apache Hive 设置。	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。	Not available.
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。	Not available.
hive-env	更改 Hive 环境中的值。	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	更改 Hive 的 hive-exec-log4j2.properties 文件中的值。	Not available.
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log4j2.properties 文件中的值。	Not available.
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。	Not available.
hive-site	更改 Hive 的 hive-site.xml 文件中的值	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值	Not available.

分类	描述	重新配置操作
hue-ini	更改 Hue 的 ini 文件中的值	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	更改HTTPFS环境中的值。	Restarts Hadoop Httpfs service.
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。	Restarts Hadoop Httpfs service.
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。	Not available.
hadoop-kms-env	在 Hadoop KMS 环境中更改值。	Restarts Hadoop-KMS service.
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。	Not available.
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	更改 Hudi 环境中的值。	Not available.
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。	Not available.
jupyter-hub-conf	更改 JupyterHubs的 jupyterhub_config.py 文件中的值。	Not available.
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。	Not available.
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。	Not available.

分类	描述	重新配置操作
livy-conf	更改 Livy 的 livy.conf 文件中的值。	Restarts Livy Server.
livy-env	更改 Livy 环境中的值。	Restarts Livy Server.
livy-log4j	更改 Livy log4j.properties 设置。	Restarts Livy Server.
mapred-env	更改 MapReduce 应用程序环境中的值。	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	更改 Oozie 的环境中的值。	Restarts Oozie.
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。	Restarts Oozie.
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。	Restarts Oozie.
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。	Not available.
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。	Not available.
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。	Restarts Phoenix-QueryServer.
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。	Not available.
pig-env	更改 Pig 环境中的值。	Not available.

分类	描述	重新配置操作
pig-properties	更改 Pig 的 pig.properties 文件中的值。	Restarts Oozie.
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。	Not available.
presto-log	更改 Presto 的 log.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-config	更改 Presto 的 config.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。	Not available.
presto-env	更改 Presto 的 presto-env.sh 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-node	更改 Presto 的 node.properties 文件中的值。	Not available.
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。	Not available.
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。	Not available.
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。	Not available.
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。	Not available.
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。	Not available.
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。	Not available.
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。	Not available.
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。	Not available.
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。	Not available.
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。	Not available.
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。	Not available.
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。	Not available.
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。	Not available.
prestoql-log	更改 Presto 的 log.properties 文件中的值。	Restarts Presto-Server (for PrestoSQL)
prestoql-config	更改 Presto 的 config.properties 文件中的值。	Restarts Presto-Server (for PrestoSQL)

分类	描述	重新配置操作
prestosql-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。	Restarts Presto-Server (for PrestoSQL)
prestosql-env	更改 Presto 的 presto-env.sh 文件中的值。	Restarts Presto-Server (for PrestoSQL)
prestosql-node	更改 Presto 的 node.properties 文件中的值。	Not available.
prestosql-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。	Not available.
prestosql-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。	Not available.
prestosql-connector-hive	更改 Presto 的 hive.properties 文件中的值。	Restarts Presto-Server (for PrestoSQL)
prestosql-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。	Not available.
prestosql-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。	Not available.
prestosql-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。	Not available.
prestosql-connector-memory	更改 Presto 的内存.properties 文件中的值。	Not available.
prestosql-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。	Not available.
prestosql-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。	Not available.



分类	描述	重新配置操作
prestosql-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。	Not available.
prestosql-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。	Not available.
prestosql-connector-redis	更改 Presto 的 redis.properties 文件中的值。	Not available.
prestosql-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。	Not available.
prestosql-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。	Not available.
prestosql-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。	Not available.
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-site	更改 Ranger KMS 的 ranger-kms-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-env	在 Ranger KMS 环境中更改值。	Restarts Ranger KMS Server.
ranger-kms-log4j	更改 Ranger 的 kms-log4j.properties 文件中的值。KMS	Not available.
ranger-kms-db-ca	更改 S3 上我与 Ranger SQL SSL KMS 连接的 CA 文件值。	Not available.
spark	亚马逊EMR精心策划的 Apache Spark 设置。	This property modifies spark-defaults. See actions there.
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。	Restarts Spark history server and Spark thrift server.

分类	描述	重新配置操作
spark-env	更改 Spark 环境中的值。	Restarts Spark history server and Spark thrift server.
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值	Not available.
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
sqoop-env	更改 Sqoop 的环境中的值。	Not available.
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。	Not available.
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。	Not available.
tez-site	更改 Tez 的 tez-site.xml 文件中的值。	Restart Oozie.
yarn-env	更改YARN环境中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	更改YARN的 yarn-site.xml 文件中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.

分类	描述	重新配置操作
zeppelin-env	更改 Zeppelin 环境中的值。	Restarts Zeppelin.
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。	Restarts Zookeeper server.
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。	Restarts Zookeeper server.

## 亚马逊EMR版本 6.1.1

### 6.1.1 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Oozie](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
适用于 Java 的 AWS SDK	1.11.828	1.11.828	1.11.711	1.11.711
Python	2.7、3.7	2.7、3.7	2.7、3.7	2.7、3.7
Scala	2.12.10	2.12.10	2.12.10	2.11.12

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.11.0	1.11.0	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.2.5	2.2.5	2.2.3	2.2.3
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.5.2-incubating-amzn-2	0.5.2-incubating-amzn-2	0.5.0-incubating-amzn-1	0.5.0-incubating-amzn-1
Hue	4.7.1	4.7.1	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.0.0	1.0.0
Livy	0.7.0	0.7.0	0.6.0	0.6.0
MXNet	1.6.0	1.6.0	1.5.1	1.5.1
Mahout	-	-	-	-
Oozie	5.2.0	5.2.0	5.1.0	5.1.0
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	-	-

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
Presto	0.232	0.232	0.230	0.230
Spark	3.0.0	3.0.0	2.4.4	2.4.4
Sqoop	1.4.7	1.4.7	-	-
TensorFlow	2.1.0	2.1.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	338	338	-	-
Zeppelin	0.9.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

### 6.1.1 发布说明

此版本旨在修复 Amazon EMR Scaling 无法成功扩展/缩小集群或导致应用程序故障时出现的问题。

#### 更改、增强和解决的问题

- 修复了 Amaz EMR on 集群上守护程序运行状况检查活动（例如收集YARN节点状态和HDFS节点状态）时，对利用率很高的大型集群的扩展请求失败的问题。之所以发生这种情况，是因为集群上的守护程序无法将节点的运行状况数据传送到内部 Ama EMR zon 组件。
- 改进了EMR集群上的守护程序，可在重复使用 IP 地址时正确跟踪节点状态，从而提高扩展操作期间的可靠性。
- [SPARK-29683](#)。修复了集群缩减期间出现任务失败的问题，因为 Spark 假定所有可用节点都被拒绝列出。
- [YARN-9011](#)。修复了集群尝试向上或向下扩展时，由于YARN停用过程中的争用条件而导致任务失败的问题。
- 通过确保 Amaz EMR on 集群上守护程序和 YARN /之间的节点状态始终保持一致，修复了集群扩展期间步骤或任务失败的问题。HDFS

- 修复了启用了 Kerberos 身份验证的 Amazon 集群的EMR集群操作（例如缩减和步骤提交）失败的问题。这是因为 Amaz EMR on 集群上守护程序没有续订 Kerberos 票证，而要与主节点上YARN运行的 HDFS /进行安全通信，则需要续订 Kerberos 票证。
- 较新的Amazon EMR 版本通过降低亚马逊较旧AL2版本的“最大打开文件数”限制来解决这个问题。EMR。亚马逊EMR发布的 5.30.1、5.30.2、5.31.1、5.32.1、5.32.1、6.0.1、6.1.1、6.2.1、5.33.0、6.3.0及更高版本现在包含一个永久修复程序，其中包含一个具有更高的“最大打开文件数”设置的永久修复。
- HTTPS现在，亚马逊 Linux 存储库已默认启用。如果您使用 Amazon S3 VPCE 策略来限制对特定存储桶的访问，则必须将新的 Amazon Linux 存储桶ARNarn:aws:s3:::amazonlinux-2-repos-\$region/\*添加到您的策略中（\$region替换为终端节点所在的区域）。有关更多信息，请在 AWS 讨论论坛中查看此主题。[公告：Amazon Linux 2 现在支持HTTPS在连接软件包存储库时使用。](#)

### 6.1.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 emr 或 aws 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.3.0	亚马逊 SageMaker Spark SDK
emr-ddb	4.14.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	3.1.0	适用于 Hadoop 生态系统的方便易用的库。

组件	版本	描述
emr-kinesis	3.5.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	emrfs	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	2.0.0	EMRS3Select 连接器
emrfs	2.42.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.11.0	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	3.2.1-amzn-1.1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	3.2.1-amzn-1.1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	3.2.1-amzn-1.1	HDFS命令行客户端和库
hadoop-hdfs-namenode	3.2.1-amzn-1.1	HDFS用于跟踪文件名和区块位置的服务。

组件	版本	描述
hadoop-hdfs-journalnode	3.2.1-amzn-1.1	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httfs-server	3.2.1-amzn-1.1	HTTPHDFS操作端点。
hadoop-kms-server	3.2.1-amzn-1.1	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	3.2.1-amzn-1.1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	3.2.1-amzn-1.1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	3.2.1-amzn-1.1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	3.2.1-amzn-1.1	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	2.2.5	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	2.2.5	为一个或多个HBase地区提供服务的服务。
hbase-client	2.2.5	HBase命令行客户端。
hbase-rest-server	2.2.5	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	2.2.5	为其提供 Thrift 端点的HBase服务。
hcatalog-client	3.1.2-amzn-2	用于操作 hcatalog-server 的“hcat”命令行客户端。



组件	版本	描述
hcatalog-server	3.1.2-amzn-2	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	3.1.2-amzn-2	HTTP端点为其提供REST接口HCatalog。
hive-client	3.1.2-amzn-2	Hive 命令行客户端。
hive-hbase	3.1.2-amzn-2	Hive-hbase 客户端。
hive-metastore-server	3.1.2-amzn-2	用于访问 Hive 元数据仓库的服务，这是一个语义存储库，用于存储 Hadoop 操作SQL的元数据。
hive-server2	3.1.2-amzn-2	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.5.2-incubating-amzn-2	增量处理框架，以支持低延迟和高效率的数据管道。
hudi-presto	0.5.2-incubating-amzn-2	用于运行 Presto 以及 Hudi 的捆绑库。
hudi-prestosql	0.5.2-incubating-amzn-2	用于使用 Hudi 运行 Presto SQL 的捆绑库。
hudi-spark	0.5.2-incubating-amzn-2	用于运行 Spark 以及 Hudi 的捆绑库。
hue-server	4.7.1	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	1.1.0	Jupyter notebook 的多用户服务器

组件	版本	描述
livy-server	0.7.0-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mxnet	1.6.0	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.64+	MariaDB 数据库服务器。
nvidia-cuda	9.2.88	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.0	Oozie 命令行客户端。
oozie-server	5.2.0	用于接受 Oozie  workflow 请求的服务。
opencv	4.3.0	开源计算机视觉库。
phoenix-library	5.0.0--2.0 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	5.0.0--2.0 HBase	轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限 API
presto-coordinator	0.232	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.232	用于执行查询的各个部分的服务。
presto-client	0.232	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。

组件	版本	描述
prestoql-coordinator	338	用于在 prestoql-worker 之中接受查询并管理查询执行的服务。
prestoql-worker	338	用于执行查询的各个部分的服务。
prestoql-client	338	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
pig-client	0.17.0	Pig 命令行客户端。
r	3.4.3	用于统计计算的 R 项目
ranger-kms-server	2.0.0	Apache Ranger 密钥管理系统
spark-client	3.0.0-amzn-0.1	Spark 命令行客户端。
spark-history-server	3.0.0-amzn-0.1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	3.0.0-amzn-0.1	的内存中执行引擎。YARN
spark-yarn-slave	3.0.0-amzn-0.1	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	2.1.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.2	tez YARN 应用程序和库。
webserver	2.4.41+	Apache HTTP 服务器。

组件	版本	描述
zeppelin-server	0.9.0-preview1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.14	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.14	ZooKeeper 命令行客户端。

### 6.1.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-6.1.1 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-executor	更改 Hadoop YARN 的 container-executor.cfg 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。

分类	描述
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-env	更改HDFS环境中的值。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。

分类	描述
hcatalog-webhcat-log4j2	更改 HCatalog WebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog WebHCat 的 webhcat-site.xml 文件中的值。
hive	亚马逊EMR精心策划的 Apache Hive 设置。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。

分类	描述
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
hudi-env	更改 Hudi 环境中的值。
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 JupyterHubs 的 jupyterhub_config.py 文件中的值。
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。

分类	描述
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。



分类	描述
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。
prestoql-log	更改 Presto 的 log.properties 文件中的值。
prestoql-config	更改 Presto 的 config.properties 文件中的值。
prestoql-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
prestoql-env	更改 Presto 的 presto-env.sh 文件中的值。
prestoql-node	更改 Presto 的 node.properties 文件中的值。
prestoql-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
prestoql-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
prestoql-connector-hive	更改 Presto 的 hive.properties 文件中的值。

分类	描述
prestosql-connector-jmx	更改 Presto 的 jmx.prop SQL erties 文件中的值。
prestosql-connector-kafka	更改 Presto 的 kafka.prop SQL erties 文件中的值。
prestosql-connector-localfile	更改 Presto 的 localfile.prop SQL erties 文件中的值。
prestosql-connector-memory	更改 Presto 的内存.properties SQL erties 文件中的值。
prestosql-connector-mongodb	更改 Presto 的 mon SQL godb.properties 文件中的值。
prestosql-connector-mysql	更改 Presto 的 mysql.prop SQL erties 文件中的值。
prestosql-connector-postgresql	更改 Presto 的 postgresql.prop SQL erties 文件中的值。
prestosql-connector-raptor	更改 Presto 的 raptor.prop SQL erties 文件中的值。
prestosql-connector-redis	更改 Presto 的 redis.prop SQL erties 文件中的值。
prestosql-connector-redshift	更改 Presto 的 reds SQL hift.properties 文件中的值。
prestosql-connector-tpch	更改 Presto 的 tpch.prop SQL erties 文件中的值。
prestosql-connector-tpcds	更改 Presto 的 tpcds.prop SQL erties 文件中的值。
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。

分类	描述
ranger-kms-site	更改 Ran ranger-kms-site KMS ger .xml 文件中的值。
ranger-kms-env	在游侠KMS环境中更改值。
ranger-kms-log4j	更改 Ranger 的 kms-log4j.properties 文件中的值。KMS
ranger-kms-db-ca	更改 S3 上我与 Ranger SQL SSL KMS 连接的 CA 文件值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。

分类	描述
zookeeper-log4j	更改 ZooKeeper 的 log4j.properties 文件中的值。

## 亚马逊EMR版本 6.1.0

### 6.1.0 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Oozie](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
适用于 Java 的 AWS SDK	1.11.828	1.11.828	1.11.711	1.11.711
Python	2.7、3.7	2.7、3.7	2.7、3.7	2.7、3.7
Scala	2.12.10	2.12.10	2.12.10	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
Flink	1.11.0	1.11.0	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.2.5	2.2.5	2.2.3	2.2.3
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.5.2-incubating-amzn-2	0.5.2-incubating-amzn-2	0.5.0-incubating-amzn-1	0.5.0-incubating-amzn-1
Hue	4.7.1	4.7.1	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.0.0	1.0.0
Livy	0.7.0	0.7.0	0.6.0	0.6.0
MXNet	1.6.0	1.6.0	1.5.1	1.5.1
Mahout	-	-	-	-
Oozie	5.2.0	5.2.0	5.1.0	5.1.0
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	-	-
Presto	0.232	0.232	0.230	0.230
Spark	3.0.0	3.0.0	2.4.4	2.4.4

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
Sqoop	1.4.7	1.4.7	-	-
TensorFlow	2.1.0	2.1.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	338	338	-	-
Zeppelin	0.9.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 6.1.0 发布说明

以下发行说明包含有关亚马逊 6.1.0 EMR 版本的信息。更改与 6.0.0 有关。

首次发布日期：2020 年 9 月 4 日

上次更新日期：2020 年 10 月 15 日

受支持的应用程序

- AWS SDK for Java 版本 1.11.828
- Flink 1.11.0
- Ganglia 3.7.2
- Hadoop 3.2.1-amzn-1
- HBase版本 2.2.5
- HBase-operator-tools 1.0.0
- HCatalog版本 3.1.2-amzn-0
- Hive 3.1.2-amzn-1
- Hudi 0.5.2-incubating
- Hue 4.7.1
- JupyterHub 版本 1.1.0

- Livy 0.7.0
- MXNet版本 1.6.0
- Oozie 5.2.0
- Phoenix 5.0.0
- Presto 0.232
- Presto SQL 版本 338
- Spark 3.0.0-amzn-0
- TensorFlow 版本 2.1.0
- Zeppelin 0.9.0-preview1
- Zookeeper 3.4.14
- 连接器和驱动程序：DynamoDB 连接器 4.14.0

## 新功能

- ARM从亚马逊EMR版本 5.30.0 和亚马逊EMR 6.1.0 版开始支持实例类型。
- 从 Amazon 6.1.0 和 5.30.0 EMR 版本开始，支持 M6g 通用实例类型。有关更多信息，请参阅 Amazon EMR 管理指南中的[支持的实例类型](#)。
- 从 Amazon EMR 版本 5.23.0 开始，支持EC2置放群组功能作为多个主节点集群的选项。目前，置放群组功能仅支持主节点类型，并将 SPREAD 策略应用于这些主节点。SPREAD 策略将一小组实例放置在单独的基础硬件上，以防止发生硬件故障时出现多个主节点丢失的问题。有关更多信息，请参阅 Amazon EMR 管理指南中的[与EC2置放群组EMR集成](#)。
- 托管扩展 — 在 Amazon 6.1.0 EMR 版本中，您可以启用 Amazon EMR 托管扩展，根据工作负载自动增加或减少集群中的实例或单元数量。Amazon 会持续评估集群指标，以做出扩展决策，从而优化集群的成本和速度。托管扩展也适用于亚马逊 5.30.0 及更高EMR版本，6.0.0 除外。有关更多信息，请参阅 Amazon EMR 管理指南中的[扩展集群资源](#)。
- 6.1.0 支持 Presto SQL 版本 338。EMR有关更多信息，请参阅 [Presto](#)。
  - 只有 EMR 6.1.0 及更高版本支持 Presto SQL，EMR 6.0.0 或 5.x 不支持。
  - 您可以继续使用应用程序名称 Presto 在集群上安装 PrestoDB。要在集群上安装 Presto，请使用应用程序名称 PrestoSQL。
  - 您可以安装 PrestoDB 或 Presto，但不能在单个集群上同时安装两者。如果在尝试创建集群时同时指定了 PrestoDB 和 SQL Presto，则会发生验证错误，集群创建请求将失败。
  - 单主集群和多主集群都支持 Presto。在多主集群上，需要外部 Hive 元数据仓库才能运行 Presto 或 PrestoDB。SQL 请参阅[具有多个主节点的EMR集群中支持的应用程序](#)。

- ECR Apache Hadoop 和带有 Docker 的 Apache Spark 支持自动身份验证：Spark 用户可以使用 Docker Hub 和亚马逊弹性容器注册表（亚马逊 ECR）中的 Docker 镜像来定义环境和库依赖关系。

使用 [Ama EMR zon 6.x 配置 Docker 并使用 Docker 运行 Spark 应用程序](#)。

- EMR支持 Apache Hive ACID 事务：Amazon EMR 6.1.0 增加了对 Hive ACID 事务的支持，因此它符合数据库的ACID属性。借助此功能，您可以使用 Amazon Simple Storage Service（Amazon S3）中的数据在 Hive 托管表中运行INSERT，UPDATE，DELETE，和MERGE操作。对于流媒体摄取、数据重述、使用批量更新和缓慢更改维度等用例来说MERGE，这是一项关键功能。有关更多信息，包括配置示例和用例，请参阅[亚马逊EMR支持 Apache Hiv ACID e 事务](#)。

## 更改、增强功能和解决的问题

- 此版本旨在修复 Amazon EMR Scaling 无法成功扩展/缩小集群或导致应用程序故障时出现的问题。
- 修复了 Amaz EMR on 集群上守护程序运行状况检查活动（例如收集YARN节点状态和HDFS节点状态）时，对利用率很高的大型集群的扩展请求失败的问题。之所以发生这种情况，是因为集群上的守护程序无法将节点的运行状况数据传送到内部 Ama EMR zon 组件。
- 改进了EMR集群上的守护程序，可在重复使用 IP 地址时正确跟踪节点状态，从而提高扩展操作期间的可靠性。
- [SPARK-29683](#)。修复了集群缩减期间出现任务失败的问题，因为 Spark 假定所有可用节点都被拒绝列出。
- [YARN-9011](#)。修复了集群尝试向上或向下扩展时，由于YARN停用过程中的争用条件而导致任务失败的问题。
- 通过确保 Amaz EMR on 集群上守护程序和 YARN /之间的节点状态始终保持一致，修复了集群扩展期间步骤或任务失败的问题。HDFS
- 修复了启用了 Kerberos 身份验证的 Amazon 集群的EMR集群操作（例如缩减和步骤提交）失败的问题。这是因为 Amaz EMR on 集群上守护程序没有续订 Kerberos 票证，而要与主节点上YARN运行的 HDFS /进行安全通信，则需要续订 Kerberos 票证。
- 较新的Amazon EMR 版本通过降低亚马逊较旧AL2版本的“最大打开文件数”限制来解决这个问题  
EMR。亚马逊EMR发布的  
5.30.1、5.30.2、5.31.1、5.32.1、5.32.1、6.0.1、6.1.1、6.2.1、5.33.0、6.3.0及更高版本现在包含一个永久修复程序，其中包含一个具有更高的“最大打开文件数”设置的永久修复。
- EMR6.0.0 不支持 Apache Flink，但在 6.1.0 和 Flink 1.11.0 上EMR支持 Apache Flink。这是首个正式支持 Hadoop 3 的 Fink 版本。请参阅 [Apache Flink 1.11.0 发布公告](#)。
- Ganglia 已从默认 EMR 6.1.0 软件包捆绑包中移除。



## 已知问题

- 降低旧版本的“最大打开文件数”限制 AL2 [在新版本中已修复]。亚马逊EMR版本：  
emr-5.30.x、emr-5.31.0、emr-5.32.0、emr-6.0.0、emr-6.1.0 和 emr-6.2.0 基于较旧版本的 Linux of Amazon 2 ( )AL2，当使用默认版本创建亚马逊集群时，“最大打开文件数”的用户限制较低。EMR AMI 亚马逊EMR发布的版本为 5.30.1、5.30.2、5.31.1、5.32.1、6.0.1、6.1.1、6.2.1、5.33.0、6.3.0 及更高版本，包括具有更高的“最大打开文件数”设置的永久修复。如果使用打开文件数限制较低的发行版，会在提交 Spark 任务时导致“Too many open files”（打开的文件过多）错误。在受影响的版本中，亚马逊EMR默认AMI的“最大打开文件数”的ulimit设置为4096，低于Linux 2中的65536个文件限制。latest Amazon AMI Spark 驱动程序和执行程序尝试打开超过 4096 个文件时，“打开的最大文件数”的较低 ulimit 设置会导致 Spark 任务失败。为了解决这个问题，Amazon EMR 提供了一个引导操作 (BA) 脚本，可以在创建集群时调整ulimit设置。

如果您使用的是无法永久修复此问题的旧EMR版 Amazon，则可以使用以下解决方法将实例控制器 ulimit 明确设置为最多 65536 个文件。

### 从命令行显式设置 ulimit

1. 编辑 `/etc/systemd/system/instance-controller.service`，将以下参数添加到 Service (服务) 部分。

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. 重启 InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

### 使用引导操作 (BA) 设置 ulimit

您还可以在创建集群时使用引导操作 ( BA ) 脚本将实例控制器 ulimit 配置为 65536 个文件。

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
```

```
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

### ⚠ Important

Amazon EMR 6.1.0 和 6.2.0 包含一个性能问题，可能会严重影响 Hudi 的所有插入、更新和删除操作。如果您计划在亚马逊 EMR 6.1.0 或 6.2.0 中使用 Hudi，则应联系 AWS 支持人员获取已修补的 Hudi。RPM

- 如果您使用 `spark.driver.extraJavaOptions` 和设置自定义垃圾收集配置 `spark.executor.extraJavaOptions`，则由于垃圾收集配置冲突，将导致 EMR 6.1 版的驱动程序/执行程序启动失败。在 6.1.0 EMR 版本中，您应该使用属性 `spark.driver.defaultJavaOptions` 为驱动程序和执行程序指定自定义 Spark 垃圾收集配置。 `spark.executor.defaultJavaOptions` 在 [Apache Spark 运行时环境](#) 和在 [亚马逊 EMR 6.1.0 上配置 Spark 垃圾回收](#) 中阅读更多内容。
- 在 Oozie 中使用 Pig ( 以及在 Hue 中，因为 Hue 使用 Oozie 操作来运行 Pig 脚本 ) 会生成一个错误，即无法加载 native-lzo 库。此错误消息是信息性的，不会阻止 Pig 运行。
- Hudi 并发支持：目前 Hudi 不支持并发写入单个 Hudi 表。此外，Hudi 会回滚处于运行状态的写入器所做的所有更改后再允许新写入器启动。并发写入可能会干扰此机制并引入竞争条件，这会导致数据损坏。您应确保作为数据处理工作流程的一部分，任何时候都只有一个 Hudi 写入器对 Hudi 表进行操作。Hudi 支持多个并发读取器对同一 Hudi 表进行操作。
- 具有多个主节点的集群和 Kerberos 身份验证中的已知问题

如果您在 Amazon 5.20.0 及更高 EMR 版本中运行具有多个主节点和 Kerberos 身份验证的集群，则在集群运行一段时间后，集群操作可能会遇到问题，例如缩小规模或步骤提交。具体时间段取决于您定义的 Kerberos 票证有效期。缩减问题会影响您提交的自动缩减和显式缩减请求。其它集群操作也可能受到影响。

解决办法：

- SSH以hadoop用户身份访问具有多个主节点的EMR群集的主节点。
- 运行以下命令，为 hadoop 用户续订 Kerberos 票证。

```
kinit -kt <keytab_file> <principal>
```

通常情况下，keytab 文件位于 `/etc/hadoop.keytab`，而 principal 为 `hadoop/<hostname>@<REALM>` 格式。

### Note

此解决方法将在 Kerberos 票证有效期内生效。默认情况下，此持续时间为 10 个小时，但可以通过 Kerberos 设置进行配置。Kerberos 票证过期后，您必须重新运行上述命令。

- Amazon EMR 6.1.0 中存在一个影响运行 Presto 的集群的问题。在较长时间（天）后，集群可能会引发错误，例如“su: failed to execute /bin/bash: Resource temporarily unavailable”或“shell request failed on channel 0”。此问题是由亚马逊内部EMR进程 (InstanceController) 引起的，该进程生成了太多的轻量级进程 (LWP)，最终导致 Hadoop 用户超出其 nproc 限制。这可以阻止用户打开其它进程。此问题的解决方案是升级到 EMR 6.2.0。

## 6.1.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>aws-sagemaker-spark-sdk</code>	1.3.0	亚马逊 SageMaker Spark SDK

组件	版本	描述
emr-ddb	4.14.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	3.1.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.5.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	emrfs	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	2.0.0	EMRS3Select 连接器
emrfs	2.42.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.11.0	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	3.2.1-amzn-1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。

组件	版本	描述
hadoop-hdfs-datanode	3.2.1-amzn-1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	3.2.1-amzn-1	HDFS命令行客户端和库
hadoop-hdfs-namenode	3.2.1-amzn-1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	3.2.1-amzn-1	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	3.2.1-amzn-1	HTTPHDFS操作端点。
hadoop-kms-server	3.2.1-amzn-1	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	3.2.1-amzn-1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	3.2.1-amzn-1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	3.2.1-amzn-1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	3.2.1-amzn-1	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	2.2.5	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	2.2.5	为一个或多个HBase地区提供服务的服务。
hbase-client	2.2.5	HBase命令行客户端。

组件	版本	描述
hbase-rest-server	2.2.5	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	2.2.5	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	3.1.2-amzn-2	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	3.1.2-amzn-2	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	3.1.2-amzn-2	HTTP端点为其提供REST接口HCatalog。
hive-client	3.1.2-amzn-2	Hive 命令行客户端。
hive-hbase	3.1.2-amzn-2	Hive-hbase 客户端。
hive-metastore-server	3.1.2-amzn-2	用于访问 Hive 元数据仓库的服务，这是一个语义存储库，用于存储 Hadoop 操作SQL的元数据。
hive-server2	3.1.2-amzn-2	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.5.2-incubating-amzn-2	增量处理框架，以支持低延迟和高效率的数据管道。
hudi-presto	0.5.2-incubating-amzn-2	用于运行 Presto 以及 Hudi 的捆绑库。
hudi-prestosql	0.5.2-incubating-amzn-2	用于使用 Hudi 运行 Presto SQL 的捆绑库。

组件	版本	描述
hudi-spark	0.5.2-incubating-amzn-2	用于运行 Spark 以及 Hudi 的捆绑库。
hue-server	4.7.1	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	1.1.0	Jupyter notebook 的多用户服务器
livy-server	0.7.0-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mxnet	1.6.0	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.64+	MariaDB 数据库服务器。
nvidia-cuda	9.2.88	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.0	Oozie 命令行客户端。
oozie-server	5.2.0	用于接受 Oozie  workflow 请求的服务。
opencv	4.3.0	开源计算机视觉库。
phoenix-library	5.0.0--2.0 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	5.0.0--2.0 HBase	轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限 API

组件	版本	描述
presto-coordinator	0.232	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.232	用于执行查询的各个部分的服务。
presto-client	0.232	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
prestosql-coordinator	338	用于在 prestosql-worker 之中接受查询并管理查询执行的服务。
prestosql-worker	338	用于执行查询的各个部分的服务。
prestosql-client	338	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
pig-client	0.17.0	Pig 命令行客户端。
r	3.4.3	用于统计计算的 R 项目
ranger-kms-server	2.0.0	Apache Ranger 密钥管理系统
spark-client	3.0.0-amzn-0	Spark 命令行客户端。
spark-history-server	3.0.0-amzn-0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	3.0.0-amzn-0	的内存中执行引擎。YARN
spark-yarn-slave	3.0.0-amzn-0	YARN从属服务器需要的 Apache Spark 库。



组件	版本	描述
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	2.1.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.2	tez YARN 应用程序和库。
webserver	2.4.41+	Apache HTTP 服务器。
zeppelin-server	0.9.0-preview1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.14	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.14	ZooKeeper 命令行客户端。

## 6.1.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

### emr-6.1.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-executor	更改 Hadoop YARN 的 container-executor.cfg 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。

分类	描述
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-env	更改HDFS环境中的值。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。

分类	描述
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive	亚马逊EMR精心策划的 Apache Hive 设置。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值

分类	描述
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
hudi-env	更改 Hudi 环境中的值。
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 JupyterHubs的 jupyterhub_config.py 文件中的值。
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。

分类	描述
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。

分类	描述
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。
prestosql-log	更改 Presto 的 log.properties 文件中的值。
prestosql-config	更改 Presto 的 config.properties 文件中的值。
prestosql-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
prestosql-env	更改 Presto 的 presto-env.sh 文件中的值。
prestosql-node	更改 Presto 的 node.properties 文件中的值。

分类	描述
prestosql-connector-blackhole	更改 Presto 的 blackhole.prop SQL erties 文件中的值。
prestosql-connector-cassandra	更改 Presto 的 cassandra.prop SQL erties 文件中的值。
prestosql-connector-hive	更改 Presto 的 hive.prop SQL erties 文件中的值。
prestosql-connector-jmx	更改 Presto 的 jmx.prop SQL erties 文件中的值。
prestosql-connector-kafka	更改 Presto 的 kafka.prop SQL erties 文件中的值。
prestosql-connector-localfile	更改 Presto 的 localfile.prop SQL erties 文件中的值。
prestosql-connector-memory	更改 Presto 的内存.properties SQL erties 文件中的值。
prestosql-connector-mongodb	更改 Presto 的 mon SQL godb.properties 文件中的值。
prestosql-connector-mysql	更改 Presto 的 mysql.prop SQL erties 文件中的值。
prestosql-connector-postgresql	更改 Presto 的 postgresql.prop SQL erties 文件中的值。
prestosql-connector-raptor	更改 Presto 的 raptor.prop SQL erties 文件中的值。
prestosql-connector-redis	更改 Presto 的 redis.prop SQL erties 文件中的值。
prestosql-connector-redshift	更改 Presto 的 reds SQL hift.properties 文件中的值。

分类	描述
prestosql-connector-tpch	更改 Presto 的 tpch.prop SQL erties 文件中的值。
prestosql-connector-tpcds	更改 Presto 的 tpcds.prop SQL erties 文件中的值。
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。
ranger-kms-site	更改 Ran ranger-kms-site KMS ger .xml 文件中的值。
ranger-kms-env	在游侠KMS环境中更改值。
ranger-kms-log4j	更改 Ranger 的 kms-log4j.properties 文件中的值。KMS
ranger-kms-db-ca	更改 S3 上我与 Ranger SQL SSL KMS 连接的 CA 文件值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。



分类	描述
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 6.0.1

### 6.0.1 应用程序版本

此版本支持以下应用程序

序：[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Oozie](#)、[Phoenix](#)、[Presto](#)、[Spark](#)、[TensorFlow](#)、[Tez](#)、[Zeppelin](#) 和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

## 应用程序版本信息

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
适用于 Java 的 AWS SDK	1.11.828	1.11.828	1.11.711	1.11.711
Python	2.7、3.7	2.7、3.7	2.7、3.7	2.7、3.7
Scala	2.12.10	2.12.10	2.12.10	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.11.0	1.11.0	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.2.5	2.2.5	2.2.3	2.2.3
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.5.2-incubating-amzn-2	0.5.2-incubating-amzn-2	0.5.0-incubating-amzn-1	0.5.0-incubating-amzn-1
Hue	4.7.1	4.7.1	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.0.0	1.0.0
Livy	0.7.0	0.7.0	0.6.0	0.6.0

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
MXNet	1.6.0	1.6.0	1.5.1	1.5.1
Mahout	-	-	-	-
Oozie	5.2.0	5.2.0	5.1.0	5.1.0
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	-	-
Presto	0.232	0.232	0.230	0.230
Spark	3.0.0	3.0.0	2.4.4	2.4.4
Sqoop	1.4.7	1.4.7	-	-
TensorFlow	2.1.0	2.1.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	338	338	-	-
Zeppelin	0.9.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 6.0.1 发布说明

此版本旨在修复 Amazon EMR Scaling 无法成功扩展/缩小集群或导致应用程序故障时出现的问题。

### 更改、增强和解决的问题

- 修复了 Amaz EMR on 集群上守护程序运行状况检查活动（例如收集YARN节点状态和HDFS节点状态）时，对利用率很高的大型集群的扩展请求失败的问题。之所以发生这种情况，是因为集群上的守护程序无法将节点的运行状况数据传送到内部 Ama EMR zon 组件。
- 改进了EMR集群上的守护程序，可在重复使用 IP 地址时正确跟踪节点状态，从而提高扩展操作期间的可靠性。

- [SPARK-29683](#)。修复了集群缩减期间出现任务失败的问题，因为 Spark 假定所有可用节点都被拒绝列出。
- [YARN-9011](#)。修复了集群尝试向上或向下扩展时，由于YARN停用过程中的争用条件而导致任务失败的问题。
- 通过确保 Amaz EMR on 集群上守护程序和 YARN /之间的节点状态始终保持一致，修复了集群扩展期间步骤或任务失败的问题。HDFS
- 修复了启用了 Kerberos 身份验证的 Amazon 集群的EMR集群操作（例如缩减和步骤提交）失败的问题。这是因为 Amaz EMR on 集群上守护程序没有续订 Kerberos 票证，而要与主节点上YARN运行的 HDFS /进行安全通信，则需要续订 Kerberos 票证。
- 较新的Amazon EMR 版本通过降低亚马逊较旧AL2版本的“最大打开文件数”限制来解决这个问题EMR。亚马逊EMR发布的 5.30.1、5.30.2、5.31.1、5.32.1、5.32.1、6.0.1、6.1.1、6.2.1、5.33.0、6.3.0及更高版本现在包含一个永久修复程序，其中包含一个具有更高的“最大打开文件数”设置的永久修复。
- HTTPS现在，亚马逊 Linux 存储库已默认启用。如果您使用 Amazon S3 VPCE 策略来限制对特定存储桶的访问，则必须将新的 Amazon Linux 存储桶ARNarn:aws:s3:::amazonlinux-2-repos-\$region/\*添加到您的策略中（\$region替换为终端节点所在的区域）。有关更多信息，请在 AWS 讨论论坛中查看此主题。[公告：Amazon Linux 2 现在支持HTTPS在连接软件包存储库时使用](#)。

## 6.0.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 emr 或 aws 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.2.6	亚马逊 SageMaker Spark SDK

组件	版本	描述
emr-ddb	4.14.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	3.0.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.5.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	emrfs	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	1.5.0	EMRS3Select 连接器
emrfs	2.39.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	3.2.1-amzn-0.1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	3.2.1-amzn-0.1	HDFS用于存储区块的节点级服务。

组件	版本	描述
hadoop-hdfs-library	3.2.1-amzn-0.1	HDFS命令行客户端和库
hadoop-hdfs-namenode	3.2.1-amzn-0.1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	3.2.1-amzn-0.1	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	3.2.1-amzn-0.1	HTTPHDFS操作端点。
hadoop-kms-server	3.2.1-amzn-0.1	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	3.2.1-amzn-0.1	MapReduce 用于运行 MapReduce 应用程序的执行引擎。
hadoop-yarn-nodemanager	3.2.1-amzn-0.1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	3.2.1-amzn-0.1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	3.2.1-amzn-0.1	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	2.2.3	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	2.2.3	为一个或多个HBase地区提供服务的服务。
hbase-client	2.2.3	HBase命令行客户端。
hbase-rest-server	2.2.3	为提供RESTfulHTTP终端节点的服务HBase。

组件	版本	描述
hbase-thrift-server	2.2.3	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	3.1.2-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	3.1.2-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	3.1.2-amzn-0	HTTP端点为其提供REST接口 HCatalog。
hive-client	3.1.2-amzn-0	Hive 命令行客户端。
hive-hbase	3.1.2-amzn-0	Hive-hbase 客户端。
hive-metastore-server	3.1.2-amzn-0	用于访问 Hive 元数据仓库的服务，这是一个语义存储库，用于存储 Hadoop 操作SQL的元数据。
hive-server2	3.1.2-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.5.0-incubating-amzn-1	增量处理框架，以支持低延迟和高效率的数据管道。
hudi-presto	0.5.0-incubating-amzn-1	用于运行 Presto 以及 Hudi 的捆绑库。
hue-server	4.4.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	1.0.0	Jupyter notebook 的多用户服务器

组件	版本	描述
livy-server	0.6.0-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mxnet	1.5.1	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.64+	MariaDB 数据库服务器。
nvidia-cuda	9.2.88	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.1.0	Oozie 命令行客户端。
oozie-server	5.1.0	用于接受 Oozie  workflow 请求的服务。
opencv	3.4.0	开源计算机视觉库。
phoenix-library	5.0.0--2.0 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	5.0.0--2.0 HBase	轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限 API
presto-coordinator	0.230	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.230	用于执行查询的各个部分的服务。
presto-client	0.230	Presto 命令行客户端，安装在 HA 集群的备用主服务器（未启动 Presto 服务器）上。
r	3.4.3	用于统计计算的 R 项目



组件	版本	描述
spark-client	2.4.4	Spark 命令行客户端。
spark-history-server	2.4.4	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.4.4	的内存中执行引擎。YARN
spark-yarn-slave	2.4.4	YARN从属服务器需要的 Apache Spark 库。
tensorflow	1.14.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.2	tez YARN 应用程序和库。
webserver	2.4.41+	Apache HTTP 服务器。
zeppelin-server	0.9.0-SNAPSHOT	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.14	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.14	ZooKeeper 命令行客户端。

## 6.0.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

### emr-6.0.1 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。

分类	描述
container-executor	更改 Hadoop YARN 的 container-executor.cfg 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-env	更改HDFS环境中的值。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。

分类	描述
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive	亚马逊EMR精心策划的 Apache Hive 设置。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值

分类	描述
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 JupyterHubs的 jupyterhub_config.py 文件中的值。
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。

分类	描述
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。

分类	描述
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。
ranger-kms-site	更改 Ranger-kms-site KMS ger .xml 文件中的值。
ranger-kms-env	在游侠KMS环境中更改值。
ranger-kms-log4j	更改 Ranger 的 kms-log4j.properties 文件中的值。KMS
ranger-kms-db-ca	更改 S3 上我与 Ranger SQL SSL KMS 连接的 CA 文件值。
recordserver-env	更改EMR RecordServer 环境中的值。
recordserver-conf	更改 server.proper EMR RecordServer ties 文件中的值。
recordserver-log4j	更改EMR RecordServer的 log4j.properties 文件中的值。

分类	描述
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 6.0.0

### 6.0.0 应用程序版本

此版本支持以下应用程序

序：[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Oozie](#)、[Phoenix](#)、[Presto](#)、[Spark](#)、[TensorFlow](#)、[Tez](#)、[Zeppelin](#) 和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)

- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

## 应用程序版本信息

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
适用于 Java 的 AWS SDK	1.11.828	1.11.828	1.11.711	1.11.711
Python	2.7、3.7	2.7、3.7	2.7、3.7	2.7、3.7
Scala	2.12.10	2.12.10	2.12.10	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.11.0	1.11.0	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.2.5	2.2.5	2.2.3	2.2.3
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.5.2-incubating-amzn-2	0.5.2-incubating-amzn-2	0.5.0-incubating-amzn-1	0.5.0-incubating-amzn-1
Hue	4.7.1	4.7.1	4.4.0	4.4.0
Iceberg	-	-	-	-



	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.0.0	1.0.0
Livy	0.7.0	0.7.0	0.6.0	0.6.0
MXNet	1.6.0	1.6.0	1.5.1	1.5.1
Mahout	-	-	-	-
Oozie	5.2.0	5.2.0	5.1.0	5.1.0
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	-	-
Presto	0.232	0.232	0.230	0.230
Spark	3.0.0	3.0.0	2.4.4	2.4.4
Sqoop	1.4.7	1.4.7	-	-
TensorFlow	2.1.0	2.1.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	338	338	-	-
Zeppelin	0.9.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 6.0.0 发布说明

以下发行说明包含 Amazon EMR 版本 6.0.0 的信息。

首次发布日期：2020 年 3 月 10 日

## 受支持的应用程序

- AWS SDK for Java 版本 1.11.711
- Ganglia 3.7.2
- Hadoop 3.2.1
- HBase版本 2.2.3
- HCatalog版本 3.1.2
- Hive 3.1.2
- Hudi 0.5.0-incubating
- Hue 4.4.0
- JupyterHub 版本 1.0.0
- Livy 0.6.0
- MXNet版本 1.5.1
- Oozie 5.1.0
- Phoenix 5.0.0
- Presto 0.230
- Spark 2.4.4
- TensorFlow 版本 1.14.0
- 齐柏林飞艇版本 0.9.0-SNAPSHOT
- Zookeeper 3.4.14
- 连接器和驱动程序 : DynamoDB 连接器 4.14.0

### Note

Flink、Sqoop、Pig 和 Mahout 在亚马逊版本 6.0.0 中不可用。EMR

## 新功能

- YARN Docker Runtime Support——诸如 Spark 作业之类的YARN应用程序现在可以在 Docker 容器的环境中运行。这使您可以轻松地在 Docker 映像中定义依赖关系，而无需在 Amazon EMR 集群上安装自定义库。有关更多信息，请参阅[使用 Amazon EMR 版本 6.0.0 配置 Docker 集成和使用 Docker 运行 Spark 应用程序](#)。

- Hive LLAP Support-Hive 现在支持LLAP执行模式，以提高查询性能。有关更多信息，请参阅[使用 Hive。LLAP](#)

## 更改、增强功能和解决的问题

- 此版本旨在修复 Amazon EMR Scaling 无法成功扩展/缩小集群或导致应用程序故障时出现的问题。
- 修复了 Amaz EMR on 集群上守护程序运行状况检查活动（例如收集YARN节点状态和HDFS节点状态）时，对利用率很高的大型集群的扩展请求失败的问题。之所以发生这种情况，是因为集群上的守护程序无法将节点的运行状况数据传送到内部 Ama EMR zon 组件。
- 改进了EMR集群上的守护程序，可在重复使用 IP 地址时正确跟踪节点状态，从而提高扩展操作期间的可靠性。
- [SPARK-29683](#)。修复了集群缩减期间出现任务失败的问题，因为 Spark 假定所有可用节点都被拒绝列出。
- [YARN-9011](#)。修复了集群尝试向上或向下扩展时，由于YARN停用过程中的争用条件而导致任务失败的问题。
- 通过确保 Amaz EMR on 集群上守护程序和 YARN /之间的节点状态始终保持一致，修复了集群扩展期间步骤或任务失败的问题。HDFS
- 修复了启用了 Kerberos 身份验证的 Amazon 集群的EMR集群操作（例如缩减和步骤提交）失败的问题。这是因为 Amaz EMR on 集群上守护程序没有续订 Kerberos 票证，而要与主节点上YARN运行的 HDFS /进行安全通信，则需要续订 Kerberos 票证。
- 较新的Amazon EMR 版本通过降低亚马逊较旧AL2版本的“最大打开文件数”限制来解决这个问题  
EMR。亚马逊EMR发布的  
5.30.1、5.30.2、5.31.1、5.32.1、5.32.1、6.0.1、6.1.1、6.2.1、5.33.0、6.3.0及更高版本现在包含一个永久修复程序，其中包含一个具有更高的“最大打开文件数”设置的永久修复。
- Amazon Linux
  - 亚马逊 Linux 2 是 EMR 6.x 版本系列的操作系统。
  - systemd用于服务管理，而不是upstart使用的 inAmazon Linux 1。
- Java 开发套件 (JDK)
  - C JDK orretto 8 是 6.x 版本系列EMR的JDK默认版本。
- Scala
  - Scala 2.12 与 Apache Spark 和 Apache Livy 一起使用。
- Python 3
  - Python 3 现在是中默认版本的 Python EMR。

- YARN节点标签

- 从 Amazon EMR 6.x 版本系列开始，默认情况下YARN节点标签功能处于禁用状态。默认情况下，应用程序主进程可以在核心节点和任务节点上运行。您可以通过配置以下属性来启用YARN节点标签功能：`yarn.node-labels.enabled`和`yarn.node-labels.am.default-node-label-expression`。有关更多信息，请参阅[了解主节点、核心节点和任务节点](#)。

## 已知问题

- 降低旧版本的“最大打开文件数”限制 AL2 [在新版本中已修复]。亚马逊EMR版本：`emr-5.30.x`、`emr-5.31.0`、`emr-5.32.0`、`emr-6.0.0`、`emr-6.1.0` 和 `emr-6.2.0` 基于较旧版本的 Linux of Amazon 2 ( )AL2，当使用默认版本创建亚马逊集群时，“最大打开文件数”的用户限制较低。EMR AMI亚马逊EMR发布的版本为5.30.1、5.30.2、5.31.1、5.32.1、6.0.1、6.1.1、6.2.1、5.33.0、6.3.0 及更高版本，包括具有更高的“最大打开文件数”设置的永久修复。如果使用打开文件数限制较低的发行版，会在提交 Spark 任务时导致“Too many open files”（打开的文件过多）错误。在受影响的版本中，亚马逊EMR默认AMI的“最大打开文件数”的`ulimit`设置为4096，低于Linux 2中的65536个文件限制。latest Amazon AMI Spark 驱动程序和执行程序尝试打开超过 4096 个文件时，“打开的最大文件数”的较低 `ulimit` 设置会导致 Spark 任务失败。为了解决这个问题，Amazon EMR 提供了一个引导操作 (BA) 脚本，可以在创建集群时调整`ulimit`设置。

如果您使用的是无法永久修复此问题的旧EMR版 Amazon，则可以使用以下解决方法将实例控制器 `ulimit` 明确设置为最多 65536 个文件。

### 从命令行显式设置 `ulimit`

1. 编辑 `/etc/systemd/system/instance-controller.service`，将以下参数添加到 Service (服务) 部分。

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. 重启 InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

### 使用引导操作 (BA) 设置 `ulimit`

您还可以在创建集群时使用引导操作 ( BA ) 脚本将实例控制器 ulimit 配置为 65536 个文件。

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- Spark 交互式外壳 PySpark，包括 SparkR 和 spark-shell，不支持将 Docker 与其他库一起使用。
- 要在亚马逊 6.0.0 EMR 版本中使用 Python 3，必须添加PATH到。yarn.nodemanager.env-whitelist
- 当您使用 AWS Glue 数据目录作为 Hive 的元存储时，不支持 Live Long 和 Process (LLAP) 功能。
- 使用集成 Spark 和 Docker 的 Amazon EMR 6.0.0 时，您需要在集群中配置具有相同实例类型和相同数量的EBS卷的实例，以避免在提交 Docker 运行时的 Spark 任务时失败。
- 在亚马逊 EMR 6.0.0 中，在亚马逊 S3 HBase 上，存储模式受到 [HBASE-24](#) 286. 问题的影响。HBase使用现有 S3 数据创建集群时，master 无法初始化。
- 具有多个主节点的集群和 Kerberos 身份验证中的已知问题

如果您在 Amazon 5.20.0 及更高EMR版本中运行具有多个主节点和 Kerberos 身份验证的集群，则在集群运行一段时间后，集群操作可能会遇到问题，例如缩小规模或步骤提交。具体时间段取决于您定义的 Kerberos 票证有效期。缩减问题会影响您提交的自动缩减和显式缩减请求。其它集群操作也可能受到影响。

解决办法：

- SSH以hadoop用户身份访问具有多个主节点的EMR群集的主主节点。

- 运行以下命令，为 hadoop 用户续订 Kerberos 票证。

```
kinit -kt <keytab_file> <principal>
```

通常情况下，keytab 文件位于 `/etc/hadoop.keytab`，而 principal 为 `hadoop/<hostname>@<REALM>` 格式。

#### Note

此解决方法将在 Kerberos 票证有效期内生效。默认情况下，此持续时间为 10 个小时，但可以通过 Kerberos 设置进行配置。Kerberos 票证过期后，您必须重新运行上述命令。

## 6.0.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>aws-sagemaker-spark-sdk</code>	1.2.6	亚马逊 SageMaker Spark SDK
<code>emr-ddb</code>	4.14.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	3.0.0	适用于 Hadoop 生态系统的方便易用的库。

组件	版本	描述
emr-kinesis	3.5.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	emrfs	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	1.5.0	EMRS3Select 连接器
emrfs	2.39.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	3.2.1-amzn-0	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	3.2.1-amzn-0	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	3.2.1-amzn-0	HDFS命令行客户端和库
hadoop-hdfs-namenode	3.2.1-amzn-0	HDFS用于跟踪文件名和区块位置的服务。

组件	版本	描述
hadoop-hdfs-journalnode	3.2.1-amzn-0	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httfs-server	3.2.1-amzn-0	HTTPHDFS操作端点。
hadoop-kms-server	3.2.1-amzn-0	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	3.2.1-amzn-0	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	3.2.1-amzn-0	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	3.2.1-amzn-0	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	3.2.1-amzn-0	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	2.2.3	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	2.2.3	为一个或多个HBase地区提供服务的服务。
hbase-client	2.2.3	HBase命令行客户端。
hbase-rest-server	2.2.3	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	2.2.3	为其提供 Thrift 端点的HBase服务。
hcatalog-client	3.1.2-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。



组件	版本	描述
hcatalog-server	3.1.2-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	3.1.2-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	3.1.2-amzn-0	Hive 命令行客户端。
hive-hbase	3.1.2-amzn-0	Hive-hbase 客户端。
hive-metastore-server	3.1.2-amzn-0	用于访问 Hive 元数据仓库的服务，这是一个语义存储库，用于存储 Hadoop 操作SQL的元数据。
hive-server2	3.1.2-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.5.0-incubating-amzn-1	增量处理框架，以支持低延迟和高效率的数据管道。
hudi-presto	0.5.0-incubating-amzn-1	用于运行 Presto 以及 Hudi 的捆绑库。
hue-server	4.4.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	1.0.0	Jupyter notebook 的多用户服务器
livy-server	0.6.0-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器

组件	版本	描述
mxnet	1.5.1	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.64+	MariaDB 数据库服务器。
nvidia-cuda	9.2.88	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.1.0	Oozie 命令行客户端。
oozie-server	5.1.0	用于接受 Oozie 工作流请求的服务。
opencv	3.4.0	开源计算机视觉库。
phoenix-library	5.0.0--2.0 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	5.0.0--2.0 HBase	轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限 API
presto-coordinator	0.230	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.230	用于执行查询的各个部分的服务。
presto-client	0.230	Presto 命令行客户端，安装在 HA 集群的备用主服务器（未启动 Presto 服务器）上。
r	3.4.3	用于统计计算的 R 项目
spark-client	2.4.4	Spark 命令行客户端。
spark-history-server	2.4.4	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。

组件	版本	描述
spark-on-yarn	2.4.4	的内存中执行引擎。YARN
spark-yarn-slave	2.4.4	YARN从属服务器需要的 Apache Spark 库。
tensorflow	1.14.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.2	tez YARN 应用程序和库。
webserver	2.4.41+	Apache HTTP 服务器。
zeppelin-server	0.9.0-SNAPSHOT	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.14	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.14	ZooKeeper 命令行客户端。

## 6.0.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

### emr-6.0.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-executor	更改 Hadoop YARN 的 container-executor.cfg 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。

分类	描述
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-env	更改HDFS环境中的值。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。

分类	描述
hcatalog-webhcat-log4j2	更改 HCatalog WebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog WebHCat 的 webhcat-site.xml 文件中的值。
hive	亚马逊EMR精心策划的 Apache Hive 设置。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。

分类	描述
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 JupyterHubs的 jupyterhub_config.py 文件中的值。
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。

分类	描述
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。

分类	描述
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。
ranger-kms-site	更改 Ranger KMS 的 ranger-kms-site.xml 文件中的值。
ranger-kms-env	在 Ranger KMS 环境中更改值。
ranger-kms-log4j	更改 Ranger 的 kms-log4j.properties 文件中的值。KMS
ranger-kms-db-ca	更改 S3 上我与 Ranger SQL SSL KMS 连接的 CA 文件值。
recordserver-env	更改 EMR RecordServer 环境中的值。
recordserver-conf	更改 server.properties EMR RecordServer 文件中的值。
recordserver-log4j	更改 EMR RecordServer 的 log4j.properties 文件中的值。
spark	亚马逊 EMR 精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值。



分类	描述
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊 EMR 5.x 发布版本

本部分包含每个 Amazon EMR 5.x 发行版本中可用的应用程序版本、发行说明、组件版本和配置分类。

启动集群时，您可以从 Amazon 的多个版本中进行选择EMR。这允许您测试和使用满足您解决方案兼容性需求的应用程序版本。您可以使用发行版标注指定版本号。版本标签的格式是 `emr-x.x.x`。例如，`emr-7.2.0`。

Amazon 的新EMR版本将在几天内在不同地区上市，从最初发布日期的第一个地区开始。在此期间，您所在区域可能无法提供最新发行版。

有关每个 Amazon EMR 5.x 版本中应用程序版本的完整表，请参阅[Amazon EMR 5.x 发行版中的应用程序版本](#)。

### 主题

- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [亚马逊EMR版本 5.36.2](#)
- [亚马逊EMR版本 5.36.1](#)
- [亚马逊EMR版本 5.36.0](#)

- [亚马逊EMR版本 5.35.0](#)
- [亚马逊EMR版本 5.34.0](#)
- [亚马逊EMR版本 5.33.1](#)
- [亚马逊EMR发布 5.33.0](#)
- [亚马逊EMR版本 5.32.1](#)
- [亚马逊EMR版本 5.32.0](#)
- [亚马逊EMR版本 5.31.1](#)
- [亚马逊EMR版本 5.31.0](#)
- [亚马逊EMR版本 5.30.2](#)
- [亚马逊EMR版本 5.30.1](#)
- [亚马逊EMR版本 5.30.0](#)
- [亚马逊EMR版本 5.29.0](#)
- [亚马逊EMR版本 5.28.1](#)
- [亚马逊EMR版本 5.28.0](#)
- [亚马逊EMR版本 5.27.1](#)
- [亚马逊EMR版本 5.27.0](#)
- [亚马逊EMR版本 5.26.0](#)
- [亚马逊EMR版本 5.25.0](#)
- [亚马逊EMR版本 5.24.1](#)
- [亚马逊EMR版本 5.24.0](#)
- [亚马逊EMR版本 5.23.1](#)
- [亚马逊EMR发布 5.23.0](#)
- [亚马逊EMR版本 5.22.0](#)
- [亚马逊EMR版本 5.21.2](#)
- [亚马逊EMR版本 5.21.1](#)
- [亚马逊EMR版本 5.21.0](#)
- [亚马逊EMR版本 5.20.1](#)
- [亚马逊EMR版本 5.20.0](#)
- [亚马逊EMR版本 5.19.1](#)
- [亚马逊EMR版本 5.19.0](#)

- [亚马逊EMR版本 5.18.1](#)
- [亚马逊EMR版本 5.18.0](#)
- [亚马逊EMR版本 5.17.2](#)
- [亚马逊EMR版本 5.17.1](#)
- [亚马逊EMR版本 5.17.0](#)
- [亚马逊EMR版本 5.16.1](#)
- [亚马逊EMR版本 5.16.0](#)
- [亚马逊EMR版本 5.15.1](#)
- [亚马逊EMR版本 5.15.0](#)
- [亚马逊EMR版本 5.14.2](#)
- [亚马逊EMR版本 5.14.1](#)
- [亚马逊EMR版本 5.14.0](#)
- [亚马逊EMR版本 5.13.1](#)
- [亚马逊EMR版本 5.13.0](#)
- [亚马逊EMR版本 5.12.3](#)
- [亚马逊EMR版本 5.12.2](#)
- [亚马逊EMR版本 5.12.1](#)
- [亚马逊EMR版本 5.12.0](#)
- [亚马逊EMR版本 5.11.4](#)
- [亚马逊EMR版本 5.11.3](#)
- [亚马逊EMR版本 5.11.2](#)
- [亚马逊EMR版本 5.11.1](#)
- [亚马逊EMR版本 5.11.0](#)
- [亚马逊EMR版本 5.10.1](#)
- [亚马逊EMR版本 5.10.0](#)
- [亚马逊EMR版本 5.9.1](#)
- [亚马逊EMR版本 5.9.0](#)
- [亚马逊EMR版本 5.8.3](#)
- [亚马逊EMR版本 5.8.2](#)
- [亚马逊EMR版本 5.8.1](#)

- [亚马逊EMR版本 5.8.0](#)
- [亚马逊EMR版本 5.7.1](#)
- [亚马逊EMR版本 5.7.0](#)
- [亚马逊EMR版本 5.6.1](#)
- [亚马逊EMR版本 5.6.0](#)
- [亚马逊EMR版本 5.5.4](#)
- [亚马逊EMR版本 5.5.3](#)
- [亚马逊EMR版本 5.5.2](#)
- [亚马逊EMR版本 5.5.1](#)
- [亚马逊EMR版本 5.5.0](#)
- [亚马逊EMR版本 5.4.1](#)
- [亚马逊EMR版本 5.4.0](#)
- [亚马逊EMR版本 5.3.2](#)
- [亚马逊EMR版本 5.3.1](#)
- [亚马逊EMR版本 5.3.0](#)
- [亚马逊EMR版本 5.2.3](#)
- [亚马逊EMR版本 5.2.2](#)
- [亚马逊EMR版本 5.2.1](#)
- [亚马逊EMR版本 5.2.0](#)
- [亚马逊EMR版本 5.1.1](#)
- [亚马逊EMR版本 5.1.0](#)
- [亚马逊EMR版本 5.0.3](#)
- [亚马逊EMR版本 5.0.2](#)
- [亚马逊EMR版本 5.0.1](#)
- [亚马逊EMR版本 5.0.0](#)

## Amazon EMR 5.x 发行版中的应用程序版本

有关列出每个 Amazon EMR 5.x 发行版中可用的应用程序版本的综合表格，请在浏览器打开 [Amazon EMR 5.x 发行版中的应用程序版本](#)。

## 亚马逊EMR版本 5.36.2

### 5.36.2 应用程序版本

此发行版支持以下应用程序

序：[Delta](#)、[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[Iceberg](#)、[JupyterEnterpriseC](#)和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-5.36.2	emr-5.36.1	emr-5.36.0	emr-5.35.0
适用于 Java 的 AWS SDK	1.12.206	1.12.206	1.12.206	1.12.159
Python	2.7、3.7	2.7、3.7	2.7、3.7	2.7、3.7
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.14.2	1.14.2	1.14.2	1.14.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13

	emr-5.36.2	emr-5.36.1	emr-5.36.0	emr-5.35.0
HCatalog	2.3.9	2.3.9	2.3.9	2.3.9
Hadoop	2.10.1	2.10.1	2.10.1	2.10.1
Hive	2.3.9	2.3.9	2.3.9	2.3.9
Hudi	0.10.1-amzn-1	0.10.1-amzn-1	0.10.1-amzn-1	0.9.0-amzn-2
Hue	4.10.0	4.10.0	4.10.0	4.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.6.0	2.6.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.8.0	1.8.0	1.8.0	1.8.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.267	0.267	0.267	0.266
Spark	2.4.8	2.4.8	2.4.8	2.4.8
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2

	emr-5.36.2	emr-5.36.1	emr-5.36.0	emr-5.35.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.10.0	0.10.0	0.10.0	0.10.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 5.36.2 发行说明

以下发行说明包含 Amazon EMR 版本 5.36.2 的信息。变化是相对于 5.36.1 的。有关发布时间表的信息，请参阅[更改日志](#)。

### 更改、增强功能和解决的问题

- 此版本改进了集群扩容逻辑，因此 Amazon EMR 不会将核心节点缩小到低于集群的HDFS重复因子设置。这一改进满足了数据冗余要求，并降低了扩展操作可能停滞的机会。
- 此版本为运行 Presto 或 Trino 的集群扩展工作流程添加了新的重试机制。这一改进降低了由于单个调整大小操作失败而导致集群调整无限期运行的风险。它还可以提高集群利用率，因为您的集群可以更快地向上和向下扩展。
- 修复了以下问题：当 Amazon EMR 优雅地停用核心节点时，集群扩容操作可能会停滞不前，核心节点在完全停用之前就会变为运行不正常。
- 当 Amazon EMR 重启单个节点时，可提高具有多个主节点的高可用性集群中节点的稳定性。
- 通过 Amazon 在亚马逊EC2上EMR运行，优化日志管理。因此，您可能会看到集群日志的存储成本略有降低。
- 改进了对主节点上维护的 ZooKeeper 事务日志文件的管理，以最大限度地减少日志文件越界增长并中断集群操作的情况。
- 修复了一个罕见的错误，该错误可能导致具有多个主节点的高可用性集群因无法与 Yarn ResourceManager 通信而发生故障。
- 当您启动带有最新补丁版本（亚马逊 EMR 5.36 或更高版本、6.6 或更高版本、7.0 或更高版本）的集群时，亚马逊将EMR使用最新的亚马逊 Linux 2023 或亚马逊 Linux 2 版本作为默认亚马逊 Linux 2 版本。EMR AMI有关更多信息，请参阅[使用亚马逊的默认 Amazon Linux AMI EMR](#)。

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 709.1	4.14.348	2024年7月23日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、AWS GovCloud (美国西部)、AWS GovCloud (美国



OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			东部)、(美国东部)、中国(北京)、中国(宁夏)、亚洲太平洋(海得拉巴)、中东(UAE)、欧洲(西班牙)、欧洲(苏黎世)、亚太地区(墨尔本)、以色列(特拉维夫)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024503.0	4.14.343	xxxxxx, 2024	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			中国 (北京)、中国 (宁夏)

### 5.36.2 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>aws-sagemaker-spark-sdk</code>	1.4.1	亚马逊 SageMaker Spark SDK
<code>emr-ddb</code>	4.16.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.16.0	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.5.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。

组件	版本	描述
emr-notebook-env	1.5.0	适用于 EMR Notebooks ( 可提供 jupyter 企业网关 ) 的 Conda env
emr-s3-dist-cp	2.21.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	1.7.0	EMRS3Select 连接器
emrfs	2.51.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.14.2	Apache Flink 命令行客户端脚本和应用程序。
flink-jobmanager-config	1.14.2	管理 Apache Flink EMR JobManager 节点上的资源。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.10.1-amzn-4	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.10.1-amzn-4	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.10.1-amzn-4	HDFS命令行客户端和库

组件	版本	描述
hadoop-hdfs-namenode	2.10.1-amzn-4	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	2.10.1-amzn-4	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	2.10.1-amzn-4	HTTPHDFS操作端点。
hadoop-kms-server	2.10.1-amzn-4	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.10.1-amzn-4	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.10.1-amzn-4	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.10.1-amzn-4	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.10.1-amzn-4	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.13	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.13	为一个或多个HBase地区提供服务的服务。
hbase-client	1.4.13	HBase命令行客户端。
hbase-rest-server	1.4.13	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.13	为其提供 Thrift 端点的HBase服务。

组件	版本	描述
hcatalog-client	2.3.9-amzn-2	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.9-amzn-2	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.9-amzn-2	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.9-amzn-2	Hive 命令行客户端。
hive-hbase	2.3.9-amzn-2	Hive-hbase 客户端。
hive-metastore-server	2.3.9-amzn-2	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.9-amzn-2	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.10.1-amzn-1	增量处理框架，以支持低延迟和高效率的数据管道。
hudi-spark	0.10.1-amzn-1	用于运行 Spark 以及 Hudi 的捆绑库。
hudi-presto	0.10.1-amzn-1	用于运行 Presto 以及 Hudi 的捆绑库。
hue-server	4.10.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	1.4.1	Jupyter notebook 的多用户服务器

组件	版本	描述
livy-server	0.7.1-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.13.1	nginx [engine x] HTTP 是一个反向代理服务器
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.8.0	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.68	我的SQL数据库服务器。
nvidia-cuda	11.0.194	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.1	Oozie 命令行客户端。
oozie-server	5.2.1	用于接受 Oozie 工作流请求的服务。
opencv	4.5.0	开源计算机视觉库。
phoenix-library	4.14.3--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.14.3--1.4 HBase	轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限 API
presto-coordinator	0.267-amzn-1	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.267-amzn-1	用于执行查询的各个部分的服务。
presto-client	0.267-amzn-1	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。

组件	版本	描述
pig-client	0.17.0	Pig 命令行客户端。
r	4.0.2	用于统计计算的 R 项目
ranger-kms-server	1.2.0	Apache Ranger 密钥管理系统
spark-client	2.4.8-amzn-2	Spark 命令行客户端。
spark-history-server	2.4.8-amzn-2	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.4.8-amzn-2	的内存中执行引擎。YARN
spark-yarn-slave	2.4.8-amzn-2	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	2.4.1	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.2	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.10.0	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.14	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.14	ZooKeeper 命令行客户端。



## 5.36.2 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

为处于运行状态的集群中的实例组指定配置时，将发生重新配置操作。Amazon EMR 仅对您修改的分类启动重新配置操作。有关更多信息，请参阅 [在正在运行的集群中重新配置实例组](#)。

### emr-5.36.2 分类

分类	描述	重新配置操作
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。	Restarts the Resource Manager service.
container-executor	更改 Hadoop YARN 的 container-executor.cfg 文件中的值。	Not available.
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。	Not available.
core-site	更改 Hadoop 的 core-site.xml 文件中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	更改 docker 相关设置。	Not available.

分类	描述	重新配置操作
emrfs-site	更改EMRFS设置。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	更改 flink-conf.yaml 设置。	Restarts Flink history server.
flink-log4j	更改 Flink log4j.properties 设置。	Restarts Flink history server.
flink-log4j-session	为 Kubernetes/Yarn 会话更改 Flink log4j-session.properties 设置。	Not available.
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。	Restarts Flink history server.

分类	描述	重新配置操作
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	更改 hadoop ssl 服务器配置	Not available.
hadoop-ssl-client	更改 hadoop ssl 客户端配置	Not available.
hbase	亚马逊EMR精心策划的 Apache 设置。HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.

分类	描述	重新配置操作
hbase-env	更改环境HBase中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	更改 hadoop-metrics2-hbase base.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。	Not available.
hbase-site	更改HBase的 hbase-site.xml 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	配置HDFS加密区域。	Should not be reconfigured.
hdfs-site	更改 hdfs-site.xml HDFS 中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	更改环境HCatalog中的值。	Restarts Hive HCatalog Server.
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。	Restarts Hive HCatalog Server.

分类	描述	重新配置操作
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat 中的值。	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。	Restarts Hive WebHCat Server.
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。	Not available.
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。	Not available.
hive-env	更改 Hive 环境中的值。	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	更改 Hive 的 hive-exec-log4j2.properties 文件中的值。	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log4j2.properties 文件中的值。	Not available.
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。	Not available.

分类	描述	重新配置操作
hive-site	更改 Hive 的 hive-site.xml 文件中的值	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值	Not available.
hue-ini	更改 Hue 的 ini 文件中的值	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	更改HTTPFS环境中的值。	Restarts Hadoop Httpfs service.
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。	Restarts Hadoop Httpfs service.
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。	Not available.
hadoop-kms-env	在 Hadoop KMS 环境中更改值。	Restarts Hadoop-KMS service.
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。	Not available.
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	更改 Hudi 环境中的值。	Not available.
hudi-defaults	更改 Hudi 的 hudi-defaults.conf 文件中的值。	Not available.

分类	描述	重新配置操作
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。	Not available.
jupyter-hub-conf	更改 jupyterhub_config.py 文件 JupyterHubs 中的值。	Not available.
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。	Not available.
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.js on 文件中的值。	Not available.
livy-conf	更改 Livy 的 livy.conf 文件中的值。	Restarts Livy Server.
livy-env	更改 Livy 环境中的值。	Restarts Livy Server.
livy-log4j	更改 Livy log4j.properties 设置。	Restarts Livy Server.
mapred-env	更改 MapReduce 应用程序环境中的值。	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	更改 Oozie 的环境中的值。	Restarts Oozie.
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。	Restarts Oozie.
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。	Restarts Oozie.
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。	Not available.

分类	描述	重新配置操作
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。	Not available.
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。	Restarts Phoenix-QueryServer.
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。	Not available.
pig-env	更改 Pig 环境中的值。	Not available.
pig-properties	更改 Pig 的 pig.properties 文件中的值。	Restarts Oozie.
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。	Not available.
presto-log	更改 Presto 的 log.properties 文件中的值。	Restarts Presto-Server.
presto-config	更改 Presto 的 config.properties 文件中的值。	Restarts Presto-Server.
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。	Not available.
presto-env	更改 Presto 的 presto-env.sh 文件中的值。	Restarts Presto-Server.
presto-node	更改 Presto 的 node.properties 文件中的值。	Not available.
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。	Not available.



分类	描述	重新配置操作
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。	Not available.
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。	Restarts Presto-Server.
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。	Not available.
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。	Not available.
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。	Not available.
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。	Not available.
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。	Not available.
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。	Not available.
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。	Not available.
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。	Not available.
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。	Not available.
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。	Not available.
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。	Not available.
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-site	更改 Ranger KMS 的 ranger-kms-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-env	在 Ranger KMS 环境中更改值。	Restarts Ranger KMS Server.
ranger-kms-log4j	更改 Ranger 的 kms-log4j.properties 文件中的值。KMS	Not available.
ranger-kms-db-ca	更改 S3 上 Ranger SQL SSL KMS 连接的 CA 文件值。	Not available.
recordserver-env	更改 EMR RecordServer 环境中的值。	Restarts EMR record server.
recordserver-conf	更改 EMR RecordServer 的 proper EMR RecordServer ties 文件中的值。	Restarts EMR record server.
recordserver-log4j	更改 EMR RecordServer 的 log4j.properties 文件中的值。	Restarts EMR record server.
spark	亚马逊 EMR 精心策划的 Apache Spark 设置。	This property modifies spark-defaults. See actions there.
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-env	更改 Spark 环境中的值。	Restarts Spark history server and Spark thrift server.
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值	Not available.

分类	描述	重新配置操作
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
sqoop-env	更改 Sqoop 的环境中的值。	Not available.
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。	Not available.
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。	Not available.
tez-site	更改 Tez 的 tez-site.xml 文件中的值。	Restarts Oozie.
yarn-env	更改YARN环境中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	更改YARN的 yarn-site.xml 文件中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	更改 Zeppelin 环境中的值。	Restarts Zeppelin.
zeppelin-site	更改 zeppelin-site.xml 中的配置设置。	Restarts Zeppelin.

分类	描述	重新配置操作
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。	Restarts Zookeeper server.
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。	Restarts Zookeeper server.

## 5.36.2 更改日志

### 5.36.2 版本和发行说明的更改日志

Date	事件	描述
2024-05-28	部署完成	<a href="#">Amazon EMR 5.36.2 已完全部署到所有支持的区域</a>
2024-05-28	文档发布	亚马逊 EMR 5.36.2 发行说明首次发布
2024-05-20	初始版本	亚马逊 EMR 5.36.2 首次部署到有限的商业区域

## 亚马逊EMR版本 5.36.1

### 5.36.1 应用程序版本

此发行版支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[Iceberg](#)、[JupyterEnterpriseGateway](#) 和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)

- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-5.36.1	emr-5.36.0	emr-5.35.0	emr-5.34.0
适用于 Java 的 AWS SDK	1.12.206	1.12.206	1.12.159	1.11.970
Python	2.7、3.7	2.7、3.7	2.7、3.7	2.7、3.7
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.14.2	1.14.2	1.14.2	1.13.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13
HCatalog	2.3.9	2.3.9	2.3.9	2.3.8
Hadoop	2.10.1	2.10.1	2.10.1	2.10.1
Hive	2.3.9	2.3.9	2.3.9	2.3.8
Hudi	0.10.1-amzn-1	0.10.1-amzn-1	0.9.0-amzn-2	0.9.0-amzn-0
Hue	4.10.0	4.10.0	4.10.0	4.9.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.6.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.4.1

	emr-5.36.1	emr-5.36.0	emr-5.35.0	emr-5.34.0
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.8.0	1.8.0	1.8.0	1.8.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.267	0.267	0.266	0.261
Spark	2.4.8	2.4.8	2.4.8	2.4.8
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.10.0	0.10.0	0.10.0	0.10.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

### 5.36.1 发布说明

以下发行说明包含 Amazon EMR 版本 5.36.1 的信息。更改与 5.36.0 有关。有关发布时间表的信息，请参阅[更改日志](#)。

#### 更改、增强功能和解决的问题

- Amazon EMR 版本 5.36.1 增加了对集群缩小期间将日志存档到 Amazon S3 的支持。在之前的 5.x 版本中，您只能在集群终止期间将日志文件存档到 Amazon S3。这项改进可确保即使在节点终止

后，集群上生成的日志文件仍保留在 Amazon S3 上。有关更多信息，请参阅[配置集群日志记录和调试](#)。

- 5.36.1 版本改进了集群上的日志管理守护程序，可以监视集群中的其他日志文件夹。EMR 这一改进最大限度地减少了磁盘过度使用情况。
- 5.36.1 版本在集群上日志管理进程守护程序停止后会自动重启该守护程序。这一改进降低了由于磁盘过度使用而导致节点出现运行状况不佳的风险。
- 5.36.1 版本修复了主节点上的 Amazon EMR 守护程序会维护集群中已终止实例的过时元数据的问题。维护陈旧的数据可能会导致集群 CPU 和内存使用量无限增长，并最终导致集群故障。
- 对于使用多个主节点启动的集群，5.36.1 版本修复了其中一个主节点上的 Amazon EC2 硬件故障可能导致第二个主节点出现故障并导致集群不稳定的问题。
- 对于配置了传输中加密的集群，托管扩展现在支持 Spark shuffle 数据感知。Spark shuffle 数据是 Spark 跨分区重新分配以执行特定操作的数据。在缩减期间，托管扩展会忽略带有随机数据的实例。这样可以防止任务的重新尝试和重新计算，这些都会给价格和性能带来高昂的代价。有关随机排序操作的更多信息，请参阅[Spark 编程指南](#)。
- 当您启动带有最新补丁版本（亚马逊 EMR 5.36 或更高版本、6.6 或更高版本、7.0 或更高版本）的集群时，亚马逊将 EMR 使用最新的亚马逊 Linux 2023 或亚马逊 Linux 2 版本作为默认亚马逊 Linux 2 版本。EMR AMI 有关更多信息，请参阅[使用亚马逊的默认 Amazon Linux AMI EMR](#)。

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 709.1	4.14.348	2024年7月23日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、AWS GovCloud ( 美国东部 )、( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )、亚洲太平洋 ( 海得拉巴 )、中东 ( UAE )、欧洲 ( 西班牙 )、欧洲 ( 苏黎世 )、亚太地区 ( 墨尔本 )、以色列 ( 特拉维夫 )、加拿大西部 ( 卡尔加里 )



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 223.0	4.14.336	2024 年 3 月 8 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 131.0	4.14.336	2024年2月14日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024124.0	4.14.336	2024 年 2 月 7 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 109.0	4.14.334	2024 年 1 月 24 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			中国 ( 北京 )、中国 ( 宁夏 )



OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 218.0	4.14.330	2024 年 1 月 2 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 206.0	4.14.330	2023 年 12 月 22 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023116.0	4.14.328	2023 年 12 月 11 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 101.0	4.14.327	2023 年 11 月 16 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			中国 ( 北京 )、中国 ( 宁夏 )



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023020.1	4.14.326	2023 年 11 月 7 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023012.1	4.14.326	2023 年 10 月 26 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			中国(北京)、中国(宁夏)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 926.0	4.14.322	2023 年 10 月 19 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 906.0	4.14.322	2023 年 10 月 4 日	中国 ( 北京 )、中国 ( 宁夏 )  美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 米兰 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.20230822.0	4.14.322	2023 年 8 月 30 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、欧洲地区 (斯德哥尔摩)、欧洲地区 (米兰)、欧洲地区 (法兰克福)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 808.0	4.14.320	2023 年 8 月 24 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、欧洲地区 (斯德哥尔摩)、欧洲地区 (米兰)、欧洲地区 (法兰克福)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)



OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 727.0	4.14.320	2023 年 8 月 14 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 719.0	4.14.320	2023 年 8 月 2 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 628.0	4.14.318	2023 年 7 月 12 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 612.0	4.14.314	2023 年 6 月 23 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 404.1	4.14.311	2023 年 4 月 18 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)

### 5.36.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.4.1	亚马逊 SageMaker Spark SDK
emr-ddb	4.16.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.16.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.5.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-notebook-env	1.5.0	适用于 EMR Notebooks ( 可提供 jupyter 企业网关 ) 的 Conda env
emr-s3-dist-cp	2.21.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	1.7.0	EMRS3Select 连接器
emrfs	2.51.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.14.2	Apache Flink 命令行客户端脚本和应用程序。
flink-jobmanager-config	1.14.2	管理 Apache Flink EMR JobManager 节点上的资源。

组件	版本	描述
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.10.1-amzn-4	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.10.1-amzn-4	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.10.1-amzn-4	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.10.1-amzn-4	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	2.10.1-amzn-4	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	2.10.1-amzn-4	HTTPHDFS操作端点。
hadoop-kms-server	2.10.1-amzn-4	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.10.1-amzn-4	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。

组件	版本	描述
hadoop-yarn-nodemanager	2.10.1-amzn-4	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.10.1-amzn-4	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.10.1-amzn-4	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.13	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.13	为一个或多个HBase地区提供服务的服务。
hbase-client	1.4.13	HBase命令行客户端。
hbase-rest-server	1.4.13	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.13	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.3.9-amzn-2	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.9-amzn-2	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.9-amzn-2	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.9-amzn-2	Hive 命令行客户端。
hive-hbase	2.3.9-amzn-2	Hive-hbase 客户端。



组件	版本	描述
hive-metastore-server	2.3.9-amzn-2	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的 SQL 元数据。
hive-server2	2.3.9-amzn-2	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.10.1-amzn-1	增量处理框架，以支持低延迟和高效率的数据管道。
hudi-spark	0.10.1-amzn-1	用于运行 Spark 以及 Hudi 的捆绑库。
hudi-presto	0.10.1-amzn-1	用于运行 Presto 以及 Hudi 的捆绑库。
hue-server	4.10.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	1.4.1	Jupyter notebook 的多用户服务器
livy-server	0.7.1-incubating	REST 用于与 Apache Spark 交互的接口
nginx	1.13.1	nginx [engine x] HTTP 是一个反向代理服务器
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.8.0	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.68	我的 SQL 数据库服务器。

组件	版本	描述
nvidia-cuda	11.0.194	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.1	Oozie 命令行客户端。
oozie-server	5.2.1	用于接受 Oozie 工作流请求的服务。
opencv	4.5.0	开源计算机视觉库。
phoenix-library	4.14.3--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.14.3--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.267-amzn-1	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.267-amzn-1	用于执行查询的各个部分的服务。
presto-client	0.267-amzn-1	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
pig-client	0.17.0	Pig 命令行客户端。
r	4.0.2	用于统计计算的 R 项目
ranger-kms-server	1.2.0	Apache Ranger 密钥管理系统
spark-client	2.4.8-amzn-2	Spark 命令行客户端。
spark-history-server	2.4.8-amzn-2	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。

组件	版本	描述
spark-on-yarn	2.4.8-amzn-2	的内存中执行引擎。YARN
spark-yarn-slave	2.4.8-amzn-2	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	2.4.1	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.2	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.10.0	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.14	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.14	ZooKeeper 命令行客户端。

### 5.36.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

为处于运行状态的集群中的实例组指定配置时，将发生重新配置操作。Amazon EMR 仅对您修改的分类启动重新配置操作。有关更多信息，请参阅 [在正在运行的集群中重新配置实例组](#)。

#### emr-5.36.1 分类

分类	描述	重新配置操作
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。	Restarts the ResourceM anager service.

分类	描述	重新配置操作
container-executor	更改 Hadoop YARN 的 container-executor.cfg 文件中的值。	Not available.
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。	Not available.
core-site	更改 Hadoop 的 core-site.xml 文件中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	更改 docker 相关设置。	Not available.

分类	描述	重新配置操作
emrfs-site	更改EMRFS设置。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	更改 flink-conf.yaml 设置。	Restarts Flink history server.
flink-log4j	更改 Flink log4j.properties 设置。	Restarts Flink history server.
flink-log4j-session	为 Kubernetes/Yarn 会话更改 Flink log4j-session.properties 设置。	Not available.
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。	Restarts Flink history server.

分类	描述	重新配置操作
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	更改 hadoop ssl 服务器配置	Not available.
hadoop-ssl-client	更改 hadoop ssl 客户端配置	Not available.
hbase	亚马逊EMR精心策划的 Apache 设置。HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.

分类	描述	重新配置操作
hbase-env	更改环境HBase中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	更改 hadoop-metrics2-hbase base.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。	Not available.
hbase-site	更改HBase的 hbase-site.xml 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	配置HDFS加密区域。	Should not be reconfigured.
hdfs-site	更改 hdfs-site.xml HDFS 中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	更改环境HCatalog中的值。	Restarts Hive HCatalog Server.
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。	Restarts Hive HCatalog Server.

分类	描述	重新配置操作
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat 中的值。	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。	Restarts Hive WebHCat Server.
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。	Not available.
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。	Not available.
hive-env	更改 Hive 环境中的值。	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	更改 Hive 的 hive-exec-log4j2.properties 文件中的值。	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log4j2.properties 文件中的值。	Not available.
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。	Not available.



分类	描述	重新配置操作
hive-site	更改 Hive 的 hive-site.xml 文件中的值	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值	Not available.
hue-ini	更改 Hue 的 ini 文件中的值	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	更改HTTPFS环境中的值。	Restarts Hadoop Httpfs service.
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。	Restarts Hadoop Httpfs service.
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。	Not available.
hadoop-kms-env	在 Hadoop KMS 环境中更改值。	Restarts Hadoop-KMS service.
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。	Not available.
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	更改 Hudi 环境中的值。	Not available.
hudi-defaults	更改 Hudi 的 hudi-defaults.conf 文件中的值。	Not available.

分类	描述	重新配置操作
jupyter-notebook-conf	更改 Jupyter notebook 的 <code>jupyter_notebook_config.py</code> 文件中的值。	Not available.
jupyter-hub-conf	更改 <code>jupyterhub_config.py</code> 文件 JupyterHubs 中的值。	Not available.
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。	Not available.
jupyter-sparkmagic-conf	更改 Sparkmagic 的 <code>config.js on</code> 文件中的值。	Not available.
livy-conf	更改 Livy 的 <code>livy.conf</code> 文件中的值。	Restarts Livy Server.
livy-env	更改 Livy 环境中的值。	Restarts Livy Server.
livy-log4j	更改 Livy <code>log4j.properties</code> 设置。	Restarts Livy Server.
mapred-env	更改 MapReduce 应用程序环境中的值。	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	更改 MapReduce 应用程序的 <code>mapred-site.xml</code> 文件中的值。	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	更改 Oozie 的环境中的值。	Restarts Oozie.
oozie-log4j	更改 Oozie 的 <code>oozie-log 4j.properties</code> 文件中的值。	Restarts Oozie.
oozie-site	更改 Oozie 的 <code>oozie-site.xml</code> 文件中的值。	Restarts Oozie.
phoenix-hbase-metrics	更改 Phoenix 的 <code>hadoop-me trics2-hbase.properties</code> 文件中的值。	Not available.

分类	描述	重新配置操作
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。	Not available.
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。	Restarts Phoenix-QueryServer.
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。	Not available.
pig-env	更改 Pig 环境中的值。	Not available.
pig-properties	更改 Pig 的 pig.properties 文件中的值。	Restarts Oozie.
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。	Not available.
presto-log	更改 Presto 的 log.properties 文件中的值。	Restarts Presto-Server.
presto-config	更改 Presto 的 config.properties 文件中的值。	Restarts Presto-Server.
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。	Not available.
presto-env	更改 Presto 的 presto-env.sh 文件中的值。	Restarts Presto-Server.
presto-node	更改 Presto 的 node.properties 文件中的值。	Not available.
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。	Not available.
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。	Restarts Presto-Server.
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。	Not available.
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。	Not available.
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。	Not available.
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。	Not available.
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。	Not available.
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。	Not available.
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。	Not available.
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。	Not available.
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。	Not available.
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。	Not available.
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。	Not available.
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-site	更改 Ranger KMS 的 ranger-kms-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-env	在 Ranger KMS 环境中更改值。	Restarts Ranger KMS Server.
ranger-kms-log4j	更改 Ranger 的 kms-log4j.properties 文件中的值。KMS	Not available.
ranger-kms-db-ca	更改 S3 上 Ranger SQL SSL KMS 连接的 CA 文件值。	Not available.
recordserver-env	更改 EMR RecordServer 环境中的值。	Restarts EMR record server.
recordserver-conf	更改 EMR RecordServer 的 proper EMR RecordServer ties 文件中的值。	Restarts EMR record server.
recordserver-log4j	更改 EMR RecordServer 的 log4j.properties 文件中的值。	Restarts EMR record server.
spark	亚马逊 EMR 精心策划的 Apache Spark 设置。	This property modifies spark-defaults. See actions there.
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-env	更改 Spark 环境中的值。	Restarts Spark history server and Spark thrift server.
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值	Not available.

分类	描述	重新配置操作
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
sqoop-env	更改 Sqoop 的环境中的值。	Not available.
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。	Not available.
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。	Not available.
tez-site	更改 Tez 的 tez-site.xml 文件中的值。	Restarts Oozie.
yarn-env	更改YARN环境中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	更改YARN的 yarn-site.xml 文件中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	更改 Zeppelin 环境中的值。	Restarts Zeppelin.
zeppelin-site	更改 zeppelin-site.xml 中的配置设置。	Restarts Zeppelin.

分类	描述	重新配置操作
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。	Restarts Zookeeper server.
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。	Restarts Zookeeper server.

### 5.36.1 更新日志

发行版 5.36.1 的更改日志和发布说明

Date	事件	描述
2023-07-26	更新	新的操作系统版本标签 2.0.20230612.0 和 2.0.20230628.0 。
2023-05-25	部署完成	<a href="#">亚马逊 EMR 5.36.1 已全面部署到所有支持的区域</a>
2023-05-09	文档发布	亚马逊 EMR 5.36.1 发行说明首次发布
2023-05-04	初始版本	亚马逊 EMR 5.36.1 首次部署到有限的商业区域

## 亚马逊EMR版本 5.36.0

### 5.36.0 应用程序版本

此发行版支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[Iceberg](#)、[JupyterEnterpriseGateway](#) 和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

## 应用程序版本信息

	emr-5.36.0	emr-5.35.0	emr-5.34.0	emr-5.33.1
适用于 Java 的 AWS SDK	1.12.206	1.12.159	1.11.970	1.11.970
Python	2.7、3.7	2.7、3.7	2.7、3.7	2.7、3.7
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.14.2	1.14.2	1.13.1	1.12.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13
HCatalog	2.3.9	2.3.9	2.3.8	2.3.7
Hadoop	2.10.1	2.10.1	2.10.1	2.10.1
Hive	2.3.9	2.3.9	2.3.8	2.3.7
Hudi	0.10.1-amzn-1	0.9.0-amzn-2	0.9.0-amzn-0	0.7.0-amzn-1
Hue	4.10.0	4.10.0	4.9.0	4.9.0
Iceberg	-	-	-	-



	emr-5.36.0	emr-5.35.0	emr-5.34.0	emr-5.33.1
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.2.2
Livy	0.7.1	0.7.1	0.7.1	0.7.0
MXNet	1.8.0	1.8.0	1.8.0	1.7.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.1	5.2.1	5.2.1	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.267	0.266	0.261	0.245.1
Spark	2.4.8	2.4.8	2.4.8	2.4.7
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.10.0	0.10.0	0.10.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 5.36.0 发布说明

以下发行说明包含 Amazon EMR 版本 5.36.0 的信息。更改与 5.35.0 有关。

首次发布日期：2022 年 6 月 15 日

## 新功能

- 亚马逊EMR版本 5.36.0 增加了对数据定义语言 (DDL) 的支持，在支持 Apache Ranger 的集群上使用 Apache Spark。这使您可以使用 Apache Ranger 来管理诸如创建、更改和删除亚马逊集群中的数据库和表等操作的访问权限。EMR
- 当您启动带有最新补丁版本 ( 亚马逊 EMR 5.36 或更高版本、6.6 或更高版本、7.0 或更高版本 ) 的集群时，亚马逊将EMR使用最新的亚马逊 Linux 2023 或亚马逊 Linux 2 版本作为默认亚马逊 Linux 2 版本。EMR AMI有关更多信息，请参阅[使用亚马逊的默认 Amazon Linux AMI EMR](#)。

### Note

此版本不再获得自动AMI更新，因为又有 1 个补丁版本接替了该版本。补丁版本以第二位小数点后的数字 (6.8.1) 表示。要查看您是否使用的是最新补丁版本，请查看[发布指南](#)中的可用版本，或者在控制台中创建集群时查看 Amazon EMR 版本下拉列表，或者使用[ListReleaseLabelsAPI](#)或[list-release-labels](#)CLI操作。要获取有关新版本的更新，请订阅“[最新消息?](#)”上的RSS提要 页面。

OsReleaseLabel ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 709.1	4.14.348	2024年7月23日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪) (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、AWS GovCloud (美国西部)、AWS GovCloud (美国东部)、(美国东部)、中国 (北京)、中国 (宁夏)、亚洲太平洋 (海得拉巴)、中东 (UAE)、欧洲 (西班牙)、欧洲 (苏黎世)、亚太地区 (墨尔本)、以色列 (特拉维夫)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 504.1	4.14.313	2023 年 5 月 16 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 418.0	4.14.311	2023 年 5 月 3 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 404.1	4.14.311	2023 年 4 月 18 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)
2.0.2023 404.0	4.14.311	2023 年 4 月 10 日	美国东部 (弗吉尼亚州北部)、欧洲地区 (巴黎)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023320.0	4.14.309	2023 年 3 月 30 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 307.0	4.14.305	2023 年 3 月 15 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 207.0	4.14.304	2023 年 2 月 22 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 119.1	4.14.301	2023 年 2 月 3 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022 210.1	4.14.301	2023 年 12 月 22 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022 103.3	4.14.296	2022 年 12 月 5 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022004.0	4.14.294	2022 年 11 月 2 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022 912.1	4.14.291	2022 年 10 月 7 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)
2.0.2022 719.0	4.14.287	2022 年 8 月 10 日	美国西部 (北加利福尼亚)、欧洲地区 (巴黎)、欧洲地区 (斯德哥尔摩)、欧洲地区 (法兰克福)、亚太地区 (孟买)、中东 (巴林)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022 426.0	4.14.281	2022 年 6 月 14 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

### 更改、增强和解决的问题

- 亚马逊 EMR 5.36.0 升级现在支持 : aws-sdk 1.12.206、Hadoop 2.10.1-amzn-4、Hive 2.3.9-amzn-2、Hudi 0.10.1-amzn-1、Spark 2.4.8-amzn-2、Presto 0.267-amzn-1、Amazon Glue 连接器 1.18.0、2.51.0。EMRFS

## 已知问题

- 当你使用带有 Hive 分区位置格式的 Spark 来读取 Amazon S3 中的数据，并在亚马逊EMR版本 5.30.0 到 5.36.0 和 6.2.0 到 6.9.0 上运行 Spark 时，你可能会遇到一个问题，导致集群无法正确读取数据。如果您的分区具有以下所有特征，会发生这种情况：
  - 从同一个表扫描两个或多个分区。
  - 至少有一个分区目录路径是至少一个其他分区目录路径的前缀，例如，s3://bucket/table/p=a 是 s3://bucket/table/p=a b 的前缀。
  - 另一个分区目录中前缀之后的第一个字符的值为 UTF -8，小于该/字符 (U+002F)。例如，在 s3://bucket/table/p=a b 中，a 和 b 之间出现的空格字符 (U+0020) 就属于此类。请注意，还有其他 14 个非控制字符：!"#\$%&'()\*+,-。有关更多信息，请参阅 [UTF-8 编码表和 Unicode 字符](#)。

解决方法是在 spark-defaults 分类中将

spark.sql.sources.fastS3PartitionDiscovery.enabled 配置设置为 false。

- 在 Amazon 5.36.0 和 6.6.0 到 6.9.0 EMR 版本中，由于 Log4 SecretAgent j2 属性中的文件名模式配置不正确，RecordServer服务组件可能会丢失日志数据。错误的配置导致组件每天只生成一个日志文件。当应用轮换策略时，它会重写现有文件，而不是按预期生成新的日志文件。应变方法是使用引导操作每小时生成一次日志文件，并在文件名中附加一个自动增量的整数来处理轮换。

对于 Amazon EMR 6.0 到 6.9.0 版本，请在启动集群时使用以下引导操作。

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

对于 Amazon EMR 5.36.0，请在启动集群时使用以下引导操作。

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

## 5.36.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 emr 或 aws 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。



Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.4.1	亚马逊 SageMaker Spark SDK
emr-ddb	4.16.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.16.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.5.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-notebook-env	1.5.0	适用于 EMR Notebooks ( 可提供 jupyter 企业网关 ) 的 Conda env
emr-s3-dist-cp	2.21.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	1.7.0	EMRS3Select 连接器
emrfs	2.51.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.14.2	Apache Flink 命令行客户端脚本和应用程序。
flink-jobmanager-config	1.14.2	管理 Apache Flink EMR JobManager 节点上的资源。

组件	版本	描述
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.10.1-amzn-4	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.10.1-amzn-4	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.10.1-amzn-4	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.10.1-amzn-4	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	2.10.1-amzn-4	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	2.10.1-amzn-4	HTTPHDFS操作端点。
hadoop-kms-server	2.10.1-amzn-4	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.10.1-amzn-4	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。

组件	版本	描述
hadoop-yarn-nodemanager	2.10.1-amzn-4	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.10.1-amzn-4	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.10.1-amzn-4	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.13	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.13	为一个或多个HBase地区提供服务的服务。
hbase-client	1.4.13	HBase命令行客户端。
hbase-rest-server	1.4.13	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.13	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.3.9-amzn-2	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.9-amzn-2	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.9-amzn-2	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.9-amzn-2	Hive 命令行客户端。
hive-hbase	2.3.9-amzn-2	Hive-hbase 客户端。

组件	版本	描述
hive-metastore-server	2.3.9-amzn-2	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的 SQL 元数据。
hive-server2	2.3.9-amzn-2	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.10.1-amzn-1	增量处理框架，以支持低延迟和高效率的数据管道。
hudi-spark	0.10.1-amzn-1	用于运行 Spark 以及 Hudi 的捆绑库。
hudi-presto	0.10.1-amzn-1	用于运行 Presto 以及 Hudi 的捆绑库。
hue-server	4.10.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	1.4.1	Jupyter notebook 的多用户服务器
livy-server	0.7.1-incubating	REST 用于与 Apache Spark 交互的接口
nginx	1.13.1	nginx [engine x] HTTP 是一个反向代理服务器
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.8.0	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.68	我的 SQL 数据库服务器。

组件	版本	描述
nvidia-cuda	11.0.194	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.1	Oozie 命令行客户端。
oozie-server	5.2.1	用于接受 Oozie 工作流请求的服务。
opencv	4.5.0	开源计算机视觉库。
phoenix-library	4.14.3--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.14.3--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.267-amzn-1	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.267-amzn-1	用于执行查询的各个部分的服务。
presto-client	0.267-amzn-1	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
pig-client	0.17.0	Pig 命令行客户端。
r	4.0.2	用于统计计算的 R 项目
ranger-kms-server	1.2.0	Apache Ranger 密钥管理系统
spark-client	2.4.8-amzn-2	Spark 命令行客户端。
spark-history-server	2.4.8-amzn-2	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。

组件	版本	描述
spark-on-yarn	2.4.8-amzn-2	的内存中执行引擎。YARN
spark-yarn-slave	2.4.8-amzn-2	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	2.4.1	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.2	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.10.0	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.14	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.14	ZooKeeper 命令行客户端。

### 5.36.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

为处于运行状态的集群中的实例组指定配置时，将发生重新配置操作。Amazon EMR 仅对您修改的分类启动重新配置操作。有关更多信息，请参阅 [在正在运行的集群中重新配置实例组](#)。

#### emr-5.36.0 分类

分类	描述	重新配置操作
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。	Restarts the ResourceM anager service.

分类	描述	重新配置操作
container-executor	更改 Hadoop YARN 的 container-executor.cfg 文件中的值。	Not available.
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。	Not available.
core-site	更改 Hadoop 的 core-site.xml 文件中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	更改 docker 相关设置。	Not available.

分类	描述	重新配置操作
emrfs-site	更改EMRFS设置。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	更改 flink-conf.yaml 设置。	Restarts Flink history server.
flink-log4j	更改 Flink log4j.properties 设置。	Restarts Flink history server.
flink-log4j-session	为 Kubernetes/Yarn 会话更改 Flink log4j-session.properties 设置。	Not available.
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。	Restarts Flink history server.



分类	描述	重新配置操作
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	更改 hadoop ssl 服务器配置	Not available.
hadoop-ssl-client	更改 hadoop ssl 客户端配置	Not available.
hbase	亚马逊EMR精心策划的 Apache 设置。HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.

分类	描述	重新配置操作
hbase-env	更改环境HBase中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	更改 hadoop-metrics2-hbase base.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。	Not available.
hbase-site	更改HBase的 hbase-site.xml 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	配置HDFS加密区域。	Should not be reconfigured.
hdfs-site	更改 hdfs-site.xml HDFS 中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	更改环境HCatalog中的值。	Restarts Hive HCatalog Server.
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。	Restarts Hive HCatalog Server.

分类	描述	重新配置操作
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat 中的值。	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。	Restarts Hive WebHCat Server.
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。	Not available.
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。	Not available.
hive-env	更改 Hive 环境中的值。	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	更改 Hive 的 hive-exec-log4j2.properties 文件中的值。	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log4j2.properties 文件中的值。	Not available.
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。	Not available.

分类	描述	重新配置操作
hive-site	更改 Hive 的 hive-site.xml 文件中的值	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值	Not available.
hue-ini	更改 Hue 的 ini 文件中的值	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	更改HTTPFS环境中的值。	Restarts Hadoop Httpfs service.
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。	Restarts Hadoop Httpfs service.
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。	Not available.
hadoop-kms-env	在 Hadoop KMS 环境中更改值。	Restarts Hadoop-KMS service.
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。	Not available.
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	更改 Hudi 环境中的值。	Not available.

分类	描述	重新配置操作
jupyter-notebook-conf	更改 Jupyter notebook 的 <code>jupyter_notebook_config.py</code> 文件中的值。	Not available.
jupyter-hub-conf	更改 <code>jupyterhub_config.py</code> 文件 JupyterHubs 中的值。	Not available.
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。	Not available.
jupyter-sparkmagic-conf	更改 Sparkmagic 的 <code>config.json</code> 文件中的值。	Not available.
livy-conf	更改 Livy 的 <code>livy.conf</code> 文件中的值。	Restarts Livy Server.
livy-env	更改 Livy 环境中的值。	Restarts Livy Server.
livy-log4j	更改 Livy <code>log4j.properties</code> 设置。	Restarts Livy Server.
mapred-env	更改 MapReduce 应用程序环境中的值。	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	更改 MapReduce 应用程序的 <code>mapred-site.xml</code> 文件中的值。	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	更改 Oozie 的环境中的值。	Restarts Oozie.
oozie-log4j	更改 Oozie 的 <code>oozie-log4j.properties</code> 文件中的值。	Restarts Oozie.
oozie-site	更改 Oozie 的 <code>oozie-site.xml</code> 文件中的值。	Restarts Oozie.
phoenix-hbase-metrics	更改 Phoenix 的 <code>hadoop-metrics2-hbase.properties</code> 文件中的值。	Not available.

分类	描述	重新配置操作
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。	Not available.
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。	Restarts Phoenix-QueryServer.
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。	Not available.
pig-env	更改 Pig 环境中的值。	Not available.
pig-properties	更改 Pig 的 pig.properties 文件中的值。	Restarts Oozie.
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。	Not available.
presto-log	更改 Presto 的 log.properties 文件中的值。	Restarts Presto-Server.
presto-config	更改 Presto 的 config.properties 文件中的值。	Restarts Presto-Server.
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。	Not available.
presto-env	更改 Presto 的 presto-env.sh 文件中的值。	Restarts Presto-Server.
presto-node	更改 Presto 的 node.properties 文件中的值。	Not available.
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。	Not available.
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。	Restarts Presto-Server.
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。	Not available.
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。	Not available.
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。	Not available.
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。	Not available.
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。	Not available.
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。	Not available.
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。	Not available.
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。	Not available.
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。	Not available.
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。	Not available.
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。	Not available.
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-site	更改 Ranger KMS 的 ranger-kms-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-env	在 Ranger KMS 环境中更改值。	Restarts Ranger KMS Server.
ranger-kms-log4j	更改 Ranger 的 kms-log4j.properties 文件中的值。KMS	Not available.
ranger-kms-db-ca	更改 S3 上 Ranger SQL SSL KMS 连接的 CA 文件值。	Not available.
recordserver-env	更改 EMR RecordServer 环境中的值。	Restarts EMR record server.
recordserver-conf	更改 EMR RecordServer 的 proper EMR RecordServer ties 文件中的值。	Restarts EMR record server.
recordserver-log4j	更改 EMR RecordServer 的 log4j.properties 文件中的值。	Restarts EMR record server.
spark	亚马逊 EMR 精心策划的 Apache Spark 设置。	This property modifies spark-defaults. See actions there.
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-env	更改 Spark 环境中的值。	Restarts Spark history server and Spark thrift server.
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值	Not available.



分类	描述	重新配置操作
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
sqoop-env	更改 Sqoop 的环境中的值。	Not available.
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。	Not available.
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。	Not available.
tez-site	更改 Tez 的 tez-site.xml 文件中的值。	Restarts Oozie.
yarn-env	更改YARN环境中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	更改YARN的 yarn-site.xml 文件中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	更改 Zeppelin 环境中的值。	Restarts Zeppelin.
zeppelin-site	更改 zeppelin-site.xml 中的配置设置。	Restarts Zeppelin.

分类	描述	重新配置操作
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。	Restarts Zookeeper server.
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。	Restarts Zookeeper server.

## 亚马逊EMR版本 5.35.0

### 5.35.0 应用程序版本

此发行版支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[Iceberg](#)、[JupyterEnterpriseGateway](#) 和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-5.35.0	emr-5.34.0	emr-5.33.1	emr-5.33.0
适用于 Java 的 AWS SDK	1.12.159	1.11.970	1.11.970	1.11.970
Python	2.7、3.7	2.7、3.7	2.7、3.7	2.7、3.7
Scala	2.11.12	2.11.12	2.11.12	2.11.12

	emr-5.35.0	emr-5.34.0	emr-5.33.1	emr-5.33.0
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.14.2	1.13.1	1.12.1	1.12.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13
HCatalog	2.3.9	2.3.8	2.3.7	2.3.7
Hadoop	2.10.1	2.10.1	2.10.1	2.10.1
Hive	2.3.9	2.3.8	2.3.7	2.3.7
Hudi	0.9.0-amzn-2	0.9.0-amzn-0	0.7.0-amzn-1	0.7.0-amzn-1
Hue	4.10.0	4.9.0	4.9.0	4.9.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.2.2	1.2.2
Livy	0.7.1	0.7.1	0.7.0	0.7.0
MXNet	1.8.0	1.8.0	1.7.0	1.7.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.1	5.2.1	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0

	emr-5.35.0	emr-5.34.0	emr-5.33.1	emr-5.33.0
Presto	0.266	0.261	0.245.1	0.245.1
Spark	2.4.8	2.4.8	2.4.7	2.4.7
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.10.0	0.10.0	0.9.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 5.35.0 发布说明

这是亚马逊 5.35.0 EMR 版本的发行说明。

以下发行说明包含有关亚马逊 5.35.0 EMR 版本的信息。更改与 5.34.0 有关。

首次发布日期: 2022 年 3 月 30 日

### 新功能

- 使用 Log4j 1.x 和 Log4j 2.x 的亚马逊EMR版本 5.35 应用程序已分别升级为使用 Log4j 1.2.17 ( 或更高版本 ) 和 Log4j 2.17.1 ( 或更高版本 ) ，并且不需要使用引导操作来缓解先前版本中的问题。CVE 请参阅 [缓解方法 CVE -2021-44228](#)。

### 更改、增强和解决的问题

#### Flink 更改

更改类型	描述
升级	<ul style="list-style-type: none"> <li>将 Flink 版本更新到 1.14.2。</li> </ul>

更改类型	描述
	<ul style="list-style-type: none"> <li>log4j 升级到 2.17.1。</li> </ul>

## Hadoop 更改

更改类型	描述
自 5.34.0 起的 Hadoop 开源向后移植 EMR	<ul style="list-style-type: none"> <li><a href="#">YARN-10438</a> : containerId 在 ClientRMService #中处理空值 () getContainerReport</li> <li><a href="#">YARN-7266</a> : 时间轴服务器事件处理程序线程已锁定</li> <li><a href="#">YARN-10438</a> : 如果 RollingLevelDb 文件损坏或丢失, 则无法启动 ATS 1.5</li> <li><a href="#">HADOOP-13500</a> : 同步配置属性对象的小版本</li> <li><a href="#">YARN-10651</a> : 在 in 中 CapacityScheduler 坠毁。NPE AbstractYarnScheduler updateNodeResource()</li> <li><a href="#">HDFS-12221</a> : 替换中的 xerces XmlEditsVisitor</li> <li><a href="#">HDFS-16410</a> : <a href="#">解析中的 X ml 不安全</a> OfflineEditsXmlLoader</li> </ul>
Hadoop 更改和修复	<ul style="list-style-type: none"> <li>中KMS使用的 Tomcat , HttpFS 已升级到 8.5.75</li> <li>在 FileSystemOptimizedCommitter V2 中, 成功标记写在创建提交者时定义的 commitJob 输出路径中。由于任务级别 commitJob 的输出路径可能不同, 因此已将路径更正为使用清单文件中定义的路径。对于 Hive 作业, 这会导致在执行诸如动态分区或UNIONALL之类的操作时正确写入成功标记。</li> </ul>

## Hive 更改

更改类型	描述
Hive 已升级到开源 <a href="#">版本 2.3.9</a> ，包括这些修复 JIRA	<ul style="list-style-type: none"> <li>• <a href="#">HIVE-17155</a>: HiveConf .java 中的 findConfFile () 配置路径存在一些问题</li> <li>• <a href="#">HIVE-24797</a> : 在解析 Avro 架构时禁用验证默认值</li> <li>• <a href="#">HIVE-21563</a> : 通过禁用 Once 来提高 Table#性能 getEmptyTable registerAllFunctions</li> <li>• <a href="#">HIVE-18147</a> : 使用 java.net 时，测试可能会失败。BindException: 地址已在使用中</li> <li>• <a href="#">HIVE-24608</a> : 在 Hive 2.3.x 的客户端中切换回 get_table HMS</li> <li>• <a href="#">HIVE-211200</a> : 矢量化——日期列抛出 java.lang。UnsupportedOperationException 用于实木复合地板</li> <li>• <a href="#">HIVE-19228</a> : 移除 commons-httpclient 3.x 用法</li> </ul>
自 5.34.0 起的 Hive 开源向后移植 EMR	<ul style="list-style-type: none"> <li>• <a href="#">HIVE-19990</a> : 在连接条件下使用间隔文字进行查询失败</li> <li>• <a href="#">HIVE-25824</a> : 将 branch-2.3 升级到 log4j 2.17.0</li> <li>• <a href="#">TEZ-4062</a> : 任务完成后应中止推测尝试调度</li> <li>• <a href="#">TEZ-4108</a> : NullPointerException 在投机执行期间竞赛条件</li> <li>• <a href="#">TEZ-3918</a> : 设置 tez.task.log.level 不起作用</li> </ul>
Hive 升级和修复	<ul style="list-style-type: none"> <li>• 将 Log4j 版本升级到 2.17.1</li> <li>• 将ORC版本升级到 1.4.3</li> <li>• 修复了由于处罚线程而导致的死锁 ShuffleScheduler</li> </ul>
新功能	<ul style="list-style-type: none"> <li>• 添加了在 AM 日志中打印 Hive 查询的功能 默认情况下，将禁用该功能。标记/配置</li> </ul>

更改类型	描述
	: tez.am.emr.print.hive.query.in.log 。状态 ( 默认 ) : FALSE。

## Oozie 更改

更改类型	描述
Oozie 自 5.34.0 起的开源向后移植 EMR	<ul style="list-style-type: none"> <li><a href="#">OOZIE-3652</a> : Oozie 启动器应在出现目录列表时重试目录列表 NoSuchFileException</li> </ul>

## Pig 更改

更改类型	描述
升级	<ul style="list-style-type: none"> <li>log4j 升级到 1.2.17。</li> </ul>

## 已知问题

- 当你使用带有 Hive 分区位置格式的 Spark 来读取 Amazon S3 中的数据，并在亚马逊EMR版本 5.30.0 到 5.36.0 和 6.2.0 到 6.9.0 上运行 Spark 时，你可能会遇到一个问题，导致集群无法正确读取数据。如果您的分区具有以下所有特征，会发生这种情况：
  - 从同一个表扫描两个或多个分区。
  - 至少有一个分区目录路径是至少一个其他分区目录路径的前缀，例如，s3://bucket/table/p=a 是 s3://bucket/table/p=a b 的前缀。
  - 另一个分区目录中前缀之后的第一个字符的值为 UTF -8，小于该/字符 (U+002F)。例如，在 s3://bucket/table/p=a b 中，a 和 b 之间出现的空格字符 (U+0020) 就属于此类。请注意，还有其他 14 个非控制字符：! "\$ % & ' ( ) \* + , - 。有关更多信息，请参阅 [UTF-8 编码表和 Unicode 字符](#)。

解决方法是在 spark-defaults 分类中将 spark.sql.sources.fastS3PartitionDiscovery.enabled 配置设置为 false。

## 5.35.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>aws-sagemaker-spark-sdk</code>	1.4.1	亚马逊 SageMaker Spark SDK
<code>emr-ddb</code>	4.16.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.15.0	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.5.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
<code>emr-notebook-env</code>	1.5.0	适用于 EMR Notebooks ( 可提供 jupyter 企业网关 ) 的 Conda env
<code>emr-s3-dist-cp</code>	2.20.0	针对 Amazon S3 优化的分布式复制应用程序。
<code>emr-s3-select</code>	1.7.0	EMRS3Select 连接器
<code>emrfs</code>	2.49.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。



组件	版本	描述
flink-client	1.14.2	Apache Flink 命令行客户端脚本和应用程序。
flink-jobmanager-config	1.14.2	管理 Apache Flink EMR JobManager 节点上的资源。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.10.1-amzn-3	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.10.1-amzn-3	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.10.1-amzn-3	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.10.1-amzn-3	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	2.10.1-amzn-3	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	2.10.1-amzn-3	HTTPHDFS操作端点。
hadoop-kms-server	2.10.1-amzn-3	基于 Hadoop 的加密密钥管理服务。 KeyProvider API

组件	版本	描述
hadoop-mapred	2.10.1-amzn-3	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.10.1-amzn-3	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.10.1-amzn-3	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.10.1-amzn-3	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.13	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.13	为一个或多个HBase地区提供服务的服务。
hbase-client	1.4.13	HBase命令行客户端。
hbase-rest-server	1.4.13	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.13	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	2.3.9-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.9-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.9-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.9-amzn-0	Hive 命令行客户端。

组件	版本	描述
hive-hbase	2.3.9-amzn-0	Hive-hbase 客户端。
hive-metastore-server	2.3.9-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的 SQL 元数据。
hive-server2	2.3.9-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.9.0-amzn-2	增量处理框架，以支持低延迟和高效率的数据管道。
hudi-spark	0.9.0-amzn-2	用于运行 Spark 以及 Hudi 的捆绑库。
hudi-presto	0.9.0-amzn-2	用于运行 Presto 以及 Hudi 的捆绑库。
hue-server	4.10.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	1.4.1	Jupyter notebook 的多用户服务器
livy-server	0.7.1-incubating	REST 用于与 Apache Spark 交互的接口
nginx	1.13.1	nginx [engine x] HTTP 是一个反向代理服务器
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.8.0	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.68	我的 SQL 数据库服务器。

组件	版本	描述
nvidia-cuda	10.1.243	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.1	Oozie 命令行客户端。
oozie-server	5.2.1	用于接受 Oozie 工作流请求的服务。
opencv	4.5.0	开源计算机视觉库。
phoenix-library	4.14.3--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.14.3--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.266-amzn-0	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.266-amzn-0	用于执行查询的各个部分的服务。
presto-client	0.266-amzn-0	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
pig-client	0.17.0	Pig 命令行客户端。
r	4.0.2	用于统计计算的 R 项目
ranger-kms-server	1.2.0	Apache Ranger 密钥管理系统
spark-client	2.4.8-amzn-1	Spark 命令行客户端。
spark-history-server	2.4.8-amzn-1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。

组件	版本	描述
spark-on-yarn	2.4.8-amzn-1	的内存中执行引擎。YARN
spark-yarn-slave	2.4.8-amzn-1	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	2.4.1	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.2	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.10.0	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.14	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.14	ZooKeeper 命令行客户端。

### 5.35.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

为处于运行状态的集群中的实例组指定配置时，将发生重新配置操作。Amazon EMR 仅对您修改的分类启动重新配置操作。有关更多信息，请参阅 [在正在运行的集群中重新配置实例组](#)。

#### emr-5.35.0 分类

分类	描述	重新配置操作
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。	Restarts the ResourceM anager service.

分类	描述	重新配置操作
container-executor	更改 Hadoop YARN 的 container-executor.cfg 文件中的值。	Not available.
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。	Not available.
core-site	更改 Hadoop 的 core-site.xml 文件中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	更改 docker 相关设置。	Not available.

分类	描述	重新配置操作
emrfs-site	更改EMRFS设置。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	更改 flink-conf.yaml 设置。	Restarts Flink history server.
flink-log4j	更改 Flink log4j.properties 设置。	Restarts Flink history server.
flink-log4j-session	为 Kubernetes/Yarn 会话更改 Flink log4j-session.properties 设置。	Not available.
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。	Restarts Flink history server.

分类	描述	重新配置操作
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	更改 hadoop ssl 服务器配置	Not available.
hadoop-ssl-client	更改 hadoop ssl 客户端配置	Not available.
hbase	亚马逊EMR精心策划的 Apache 设置。HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.



分类	描述	重新配置操作
hbase-env	更改环境HBase中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	更改 hadoop-metrics2-hbase base.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。	Not available.
hbase-site	更改HBase的 hbase-site.xml 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	配置HDFS加密区域。	Should not be reconfigured.
hdfs-site	更改 hdfs-site.xml HDFS 中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	更改环境HCatalog中的值。	Restarts Hive HCatalog Server.
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。	Restarts Hive HCatalog Server.

分类	描述	重新配置操作
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat 中的值。	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。	Restarts Hive WebHCat Server.
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。	Not available.
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。	Not available.
hive-env	更改 Hive 环境中的值。	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	更改 Hive 的 hive-exec-log4j2.properties 文件中的值。	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log4j2.properties 文件中的值。	Not available.
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。	Not available.

分类	描述	重新配置操作
hive-site	更改 Hive 的 hive-site.xml 文件中的值	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值	Not available.
hue-ini	更改 Hue 的 ini 文件中的值	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	更改HTTPFS环境中的值。	Restarts Hadoop Httpfs service.
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。	Restarts Hadoop Httpfs service.
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。	Not available.
hadoop-kms-env	在 Hadoop KMS 环境中更改值。	Restarts Hadoop-KMS service.
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。	Not available.
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	更改 Hudi 环境中的值。	Not available.

分类	描述	重新配置操作
jupyter-notebook-conf	更改 Jupyter notebook 的 <code>jupyter_notebook_config.py</code> 文件中的值。	Not available.
jupyter-hub-conf	更改 <code>jupyterhub_config.py</code> 文件 JupyterHubs 中的值。	Not available.
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。	Not available.
jupyter-sparkmagic-conf	更改 Sparkmagic 的 <code>config.json</code> 文件中的值。	Not available.
livy-conf	更改 Livy 的 <code>livy.conf</code> 文件中的值。	Restarts Livy Server.
livy-env	更改 Livy 环境中的值。	Restarts Livy Server.
livy-log4j	更改 Livy <code>log4j.properties</code> 设置。	Restarts Livy Server.
mapred-env	更改 MapReduce 应用程序环境中的值。	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	更改 MapReduce 应用程序的 <code>mapred-site.xml</code> 文件中的值。	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	更改 Oozie 的环境中的值。	Restarts Oozie.
oozie-log4j	更改 Oozie 的 <code>oozie-log4j.properties</code> 文件中的值。	Restarts Oozie.
oozie-site	更改 Oozie 的 <code>oozie-site.xml</code> 文件中的值。	Restarts Oozie.
phoenix-hbase-metrics	更改 Phoenix 的 <code>hadoop-metrics2-hbase.properties</code> 文件中的值。	Not available.

分类	描述	重新配置操作
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。	Not available.
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。	Restarts Phoenix-QueryServer.
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。	Not available.
pig-env	更改 Pig 环境中的值。	Not available.
pig-properties	更改 Pig 的 pig.properties 文件中的值。	Restarts Oozie.
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。	Not available.
presto-log	更改 Presto 的 log.properties 文件中的值。	Restarts Presto-Server.
presto-config	更改 Presto 的 config.properties 文件中的值。	Restarts Presto-Server.
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。	Not available.
presto-env	更改 Presto 的 presto-env.sh 文件中的值。	Restarts Presto-Server.
presto-node	更改 Presto 的 node.properties 文件中的值。	Not available.
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。	Not available.
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。	Restarts Presto-Server.
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。	Not available.
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。	Not available.
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。	Not available.
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。	Not available.
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。	Not available.
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。	Not available.
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。	Not available.
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。	Not available.
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。	Not available.
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。	Not available.
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。	Not available.
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-site	更改 Ranger KMS 的 ranger-kms-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-env	在 Ranger KMS 环境中更改值。	Restarts Ranger KMS Server.
ranger-kms-log4j	更改 Ranger 的 kms-log4j.properties 文件中的值。KMS	Not available.
ranger-kms-db-ca	更改 S3 上 Ranger SQL SSL KMS 连接的 CA 文件值。	Not available.
recordserver-env	更改 EMR RecordServer 环境中的值。	Restarts EMR record server.
recordserver-conf	更改 EMR RecordServer 的 proper EMR RecordServer ties 文件中的值。	Restarts EMR record server.
recordserver-log4j	更改 EMR RecordServer 的 log4j.properties 文件中的值。	Restarts EMR record server.
spark	亚马逊 EMR 精心策划的 Apache Spark 设置。	This property modifies spark-defaults. See actions there.
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-env	更改 Spark 环境中的值。	Restarts Spark history server and Spark thrift server.
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值	Not available.

分类	描述	重新配置操作
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
sqoop-env	更改 Sqoop 的环境中的值。	Not available.
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。	Not available.
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。	Not available.
tez-site	更改 Tez 的 tez-site.xml 文件中的值。	Restarts Oozie.
yarn-env	更改YARN环境中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	更改YARN的 yarn-site.xml 文件中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	更改 Zeppelin 环境中的值。	Restarts Zeppelin.
zeppelin-site	更改 zeppelin-site.xml 中的配置设置。	Restarts Zeppelin.



分类	描述	重新配置操作
zookeeper-config	更改 ZooKeeper 的 zoo.cfg 文件中的值。	Restarts Zookeeper server.
zookeeper-log4j	更改 ZooKeeper 的 log4j.properties 文件中的值。	Restarts Zookeeper server.

## 亚马逊EMR版本 5.34.0

### 5.34.0 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[JupyterEnterpriseGateway](#)、[Jupyter](#) 和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-5.34.0	emr-5.33.1	emr-5.33.0	emr-5.32.1
适用于 Java 的 AWS SDK	1.11.970	1.11.970	1.11.970	1.11.890
Python	2.7、3.7	2.7、3.7	2.7、3.7	2.7、3.7
Scala	2.11.12	2.11.12	2.11.12	2.11.12

	emr-5.34.0	emr-5.33.1	emr-5.33.0	emr-5.32.1
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.13.1	1.12.1	1.12.1	1.11.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13
HCatalog	2.3.8	2.3.7	2.3.7	2.3.7
Hadoop	2.10.1	2.10.1	2.10.1	2.10.1
Hive	2.3.8	2.3.7	2.3.7	2.3.7
Hudi	0.9.0-amzn-0	0.7.0-amzn-1	0.7.0-amzn-1	0.6.0-amzn-0
Hue	4.9.0	4.9.0	4.9.0	4.8.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.2.2	1.2.2	1.1.0
Livy	0.7.1	0.7.0	0.7.0	0.7.0
MXNet	1.8.0	1.7.0	1.7.0	1.7.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.1	5.2.0	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0

	emr-5.34.0	emr-5.33.1	emr-5.33.0	emr-5.32.1
Presto	0.261	0.245.1	0.245.1	0.240.1
Spark	2.4.8	2.4.7	2.4.7	2.4.7
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.3.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.10.0	0.9.0	0.9.0	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 5.34.0 发布说明

以下发行说明包含有关亚马逊 5.34.0 EMR 版本的信息。更改与 5.33.1 有关。

首次发布日期：2022 年 1 月 20 日

发布更新日期：2022 年 3 月 21 日

### 新功能

- [托管扩展] Spark shuffle 数据托管扩展优化 ——对于 Amazon EMR 5.34.0 及更高版本以及 6.4.0 及更高EMR版本，托管扩展现在支持 Spark 洗牌数据（Spark 跨分区重新分配以执行特定操作的数据）。有关洗牌操作的更多信息，请参阅《[亚马逊EMREMR管理指南](#)》和《[Spark 编程指南](#)》EMR 中的“[在亚马逊使用托管扩展](#)”。
- [Hudi] 简化了 Hudi 配置的改进。预设情况下禁用乐观并发控制。

### 更改、增强和解决的问题

- 此版本旨在修复 Amazon EMR Scaling 无法成功扩展/缩小集群或导致应用程序故障时出现的问题。
- 以前，在多主集群上手动重启资源管理器会导致 Amaz EMR on 集群上的守护程序（例如 Zookeeper）在 Zookeeper znode 文件中重新加载所有先前已停用或丢失的节点。在某些情况下，

这会导致超出默认限制。Amazon EMR 现已从 Zookeeper 文件中删除已停用或丢失超过一小时的节点记录，并且内部限制也已提高。

- 修复了 Amaz EMR on 集群上守护程序运行状况检查活动（例如收集YARN节点状态和HDFS节点状态）时，对利用率很高的大型集群的扩展请求失败的问题。之所以发生这种情况，是因为集群上的守护程序无法将节点的运行状况数据传送到内部 Ama EMR zon 组件。
- 改进了EMR集群上的守护程序，可在重复使用 IP 地址时正确跟踪节点状态，从而提高扩展操作期间的可靠性。
- [SPARK-29683](#)。修复了集群缩减期间出现任务失败的问题，因为 Spark 假定所有可用节点都被拒绝列出。
- [YARN-9011](#)。修复了集群尝试向上或向下扩展时，由于YARN停用过程中的争用条件而导致任务失败的问题。
- 通过确保 Amaz EMR on 集群上守护程序和 YARN /之间的节点状态始终保持一致，修复了集群扩展期间步骤或任务失败的问题。HDFS
- 修复了启用了 Kerberos 身份验证的 Amazon 集群的EMR集群操作（例如缩减和步骤提交）失败的问题。这是因为 Amaz EMR on 集群上守护程序没有续订 Kerberos 票证，而要与主节点上YARN运行的 HDFS /进行安全通信，则需要续订 Kerberos 票证。
- Zeppelin 已升级到版本 0.10.0。
- Livy 修复 - 已升级到 0.7.1
- Spark 性能改进-在 5.34.0 中重写某些 Spark 配置值时，异构执行器将被禁用。EMR
- 默认情况下HDFS，Web 和 HTTPFS 服务器处于禁用状态。您可以HDFS使用 Hadoop 配置重新启用 Web。dfs.webhdfs.enabledHTTPFS 服务器可以通过使用 `sudo systemctl start hadoop-httpfs` 启动。

## 已知问题

- 与 Livy 用户模拟一起使用的亚马逊EMR笔记本功能不起作用，因为 HTTPFS 在默认情况下处于禁用状态。在这种情况下，EMR笔记本无法连接到启用了 Livy 模拟的集群。解决方法是在使用将EMR笔记本连接到集群之前启动 HTTPFS 服务器。`sudo systemctl start hadoop-httpfs`
- Hue 查询在 Amazon EMR 6.4.0 中不起作用，因为 Apache Hadoop httpFS 服务器默认处于禁用状态。要在亚马逊 EMR 6.4.0 上使用 Hue，要么使用亚马逊步骤在亚马逊EMR主节点上手动启动 HttpFS 服务器`sudo systemctl start hadoop-httpfs`，要么使用[亚马逊步骤](#)。EMR
- 与 Livy 用户模拟一起使用的亚马逊EMR笔记本功能不起作用，因为 HTTPFS 在默认情况下处于禁用状态。在这种情况下，EMR笔记本无法连接到启用了 Livy 模拟的集群。解决方法是在使用将EMR笔记本连接到集群之前启动 HTTPFS 服务器。`sudo systemctl start hadoop-httpfs`

- 当你使用带有 Hive 分区位置格式的 Spark 来读取 Amazon S3 中的数据，并在亚马逊EMR版本 5.30.0 到 5.36.0 和 6.2.0 到 6.9.0 上运行 Spark 时，你可能会遇到一个问题，导致集群无法正确读取数据。如果您的分区具有以下所有特征，会发生这种情况：
  - 从同一个表扫描两个或多个分区。
  - 至少有一个分区目录路径是至少一个其他分区目录路径的前缀，例如，s3://bucket/table/p=a 是 s3://bucket/table/p=a b 的前缀。
  - 另一个分区目录中前缀之后的第一个字符的值为 UTF -8，小于该/字符 (U+002F)。例如，在 s3://bucket/table/p=a b 中，a 和 b 之间出现的空格字符 (U+0020) 就属于此类。请注意，还有其他 14 个非控制字符：!"#\$%&'()\*+,-。有关更多信息，请参阅 [UTF-8 编码表和 Unicode 字符](#)。

解决方法是在 spark-defaults 分类中将 `spark.sql.sources.fastS3PartitionDiscovery.enabled` 配置设置为 `false`。

## 5.34.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>aws-sagemaker-spark-sdk</code>	1.4.1	亚马逊 SageMaker Spark SDK
<code>emr-ddb</code>	4.16.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	<code>emrfs</code>	适用于 Hadoop 生态系统的方便易用的库。

组件	版本	描述
emr-kinesis	3.5.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-notebook-env	1.4.0	适用于 EMR Notebooks ( 可提供 jupyter 企业网关 ) 的 Conda env
emr-s3-dist-cp	2.18.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	1.7.0	EMRS3Select 连接器
emrfs	2.48.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.13.1	Apache Flink 命令行客户端脚本和应用程序。
flink-jobmanager-config	1.13.1	管理 Apache Flink EMR JobManager 节点上的资源。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.10.1-amzn-2	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。

组件	版本	描述
hadoop-hdfs-datanode	2.10.1-amzn-2	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.10.1-amzn-2	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.10.1-amzn-2	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	2.10.1-amzn-2	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	2.10.1-amzn-2	HTTPHDFS操作端点。
hadoop-kms-server	2.10.1-amzn-2	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.10.1-amzn-2	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.10.1-amzn-2	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.10.1-amzn-2	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.10.1-amzn-2	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.13	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.13	为一个或多个HBase地区提供服务的服务。
hbase-client	1.4.13	HBase命令行客户端。

组件	版本	描述
hbase-rest-server	1.4.13	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.13	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	2.3.8-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.8-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.8-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.8-amzn-0	Hive 命令行客户端。
hive-hbase	2.3.8-amzn-0	Hive-hbase 客户端。
hive-metastore-server	2.3.8-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.8-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.9.0-amzn-0	增量处理框架，以支持低延迟和高效率的数据管道。
hudi-spark	0.9.0-amzn-0	用于运行 Spark 以及 Hudi 的捆绑库。
hudi-presto	0.9.0-amzn-0	用于运行 Presto 以及 Hudi 的捆绑库。



组件	版本	描述
hue-server	4.9.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	1.4.1	Jupyter notebook 的多用户服务器
livy-server	0.7.1-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.13.1	nginx [engine x] HTTP 是一个反向代理服务器
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.8.0	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.68	我的SQL数据库服务器。
nvidia-cuda	10.1.243	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.1	Oozie 命令行客户端。
oozie-server	5.2.1	用于接受 Oozie 工作流请求的服务。
opencv	4.5.0	开源计算机视觉库。
phoenix-library	4.14.3--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.14.3--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API

组件	版本	描述
presto-coordinator	0.261-amzn-0	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.261-amzn-0	用于执行查询的各个部分的服务。
presto-client	0.261-amzn-0	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
pig-client	0.17.0	Pig 命令行客户端。
r	4.0.2	用于统计计算的 R 项目
ranger-kms-server	1.2.0	Apache Ranger 密钥管理系统
spark-client	2.4.8-amzn-0	Spark 命令行客户端。
spark-history-server	2.4.8-amzn-0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.4.8-amzn-0	的内存中执行引擎。YARN
spark-yarn-slave	2.4.8-amzn-0	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	2.4.1	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.2	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。

组件	版本	描述
zeppelin-server	0.10.0	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.14	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.14	ZooKeeper 命令行客户端。

### 5.34.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

为处于运行状态的集群中的实例组指定配置时，将发生重新配置操作。Amazon EMR 仅对您修改的分类启动重新配置操作。有关更多信息，请参阅 [在正在运行的集群中重新配置实例组](#)。

#### emr-5.34.0 分类

分类	描述	重新配置操作
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。	Restarts the ResourceManager service.
container-executor	更改 Hadoop YARN 的 container-executor.cfg 文件中的值。	Not available.
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。	Not available.
core-site	更改 Hadoop 的 core-site.xml 文件中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

分类	描述	重新配置操作
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	更改 docker 相关设置。	Not available.
emrfs-site	更改EMRFS设置。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	更改 flink-conf.yaml 设置。	Restarts Flink history server.
flink-log4j	更改 Flink log4j.properties 设置。	Restarts Flink history server.
flink-log4j-session	为 Kubernetes/Yarn 会话更改 Flink log4j-session.properties 设置。	Not available.

分类	描述	重新配置操作
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。	Restarts Flink history server.
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	更改 hadoop ssl 服务器配置	Not available.
hadoop-ssl-client	更改 hadoop ssl 客户端配置	Not available.

分类	描述	重新配置操作
hbase	亚马逊EMR精心策划的Apache 设置。HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	更改环境HBase中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	更改 hadoop-metrics2-hbase base.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。	Not available.
hbase-site	更改HBase的 hbase-site.xml 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	配置HDFS加密区域。	Should not be reconfigured.
hdfs-site	更改 hdfs-site.xml HDFS 中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	更改环境HCatalog中的值。	Restarts Hive HCatalog Server.

分类	描述	重新配置操作
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat 中的值。	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。	Restarts Hive WebHCat Server.
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。	Not available.
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。	Not available.
hive-env	更改 Hive 环境中的值。	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	更改 Hive 的 hive-exec-log4j2.properties 文件中的值。	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log4j2.properties 文件中的值。	Not available.
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。	Not available.

分类	描述	重新配置操作
hive-site	更改 Hive 的 hive-site.xml 文件中的值	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值	Not available.
hue-ini	更改 Hue 的 ini 文件中的值	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	更改HTTPFS环境中的值。	Restarts Hadoop Httpfs service.
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。	Restarts Hadoop Httpfs service.
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。	Not available.
hadoop-kms-env	在 Hadoop KMS 环境中更改值。	Restarts Hadoop-KMS service.
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。	Not available.
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	更改 Hudi 环境中的值。	Not available.



分类	描述	重新配置操作
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。	Not available.
jupyter-hub-conf	更改 jupyterhub_config.py 文件 JupyterHubs 中的值。	Not available.
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。	Not available.
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。	Not available.
livy-conf	更改 Livy 的 livy.conf 文件中的值。	Restarts Livy Server.
livy-env	更改 Livy 环境中的值。	Restarts Livy Server.
livy-log4j	更改 Livy log4j.properties 设置。	Restarts Livy Server.
mapred-env	更改 MapReduce 应用程序环境中的值。	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	更改 Oozie 的环境中的值。	Restarts Oozie.
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。	Restarts Oozie.
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。	Restarts Oozie.
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。	Not available.

分类	描述	重新配置操作
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。	Not available.
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。	Restarts Phoenix-QueryServer.
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。	Not available.
pig-env	更改 Pig 环境中的值。	Not available.
pig-properties	更改 Pig 的 pig.properties 文件中的值。	Restarts Oozie.
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。	Not available.
presto-log	更改 Presto 的 log.properties 文件中的值。	Restarts Presto-Server.
presto-config	更改 Presto 的 config.properties 文件中的值。	Restarts Presto-Server.
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。	Not available.
presto-env	更改 Presto 的 presto-env.sh 文件中的值。	Restarts Presto-Server.
presto-node	更改 Presto 的 node.properties 文件中的值。	Not available.
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。	Not available.
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。	Restarts Presto-Server.
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。	Not available.
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。	Not available.
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。	Not available.
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。	Not available.
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。	Not available.
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。	Not available.
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。	Not available.
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。	Not available.
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。	Not available.
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。	Not available.
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。	Not available.
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-site	更改 Ranger KMS 的 ranger-kms-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-env	在 Ranger KMS 环境中更改值。	Restarts Ranger KMS Server.
ranger-kms-log4j	更改 Ranger 的 kms-log4j.properties 文件中的值。KMS	Not available.
ranger-kms-db-ca	更改 S3 上 Ranger SQL SSL KMS 连接的 CA 文件值。	Not available.
recordserver-env	更改 EMR RecordServer 环境中的值。	Restarts EMR record server.
recordserver-conf	更改 EMR RecordServer 的 proper EMR RecordServer ties 文件中的值。	Restarts EMR record server.
recordserver-log4j	更改 EMR RecordServer 的 log4j.properties 文件中的值。	Restarts EMR record server.
spark	亚马逊 EMR 精心策划的 Apache Spark 设置。	This property modifies spark-defaults. See actions there.
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-env	更改 Spark 环境中的值。	Restarts Spark history server and Spark thrift server.
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值	Not available.

分类	描述	重新配置操作
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
sqoop-env	更改 Sqoop 的环境中的值。	Not available.
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。	Not available.
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。	Not available.
tez-site	更改 Tez 的 tez-site.xml 文件中的值。	Restarts Oozie.
yarn-env	更改YARN环境中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	更改YARN的 yarn-site.xml 文件中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	更改 Zeppelin 环境中的值。	Restarts Zeppelin.
zeppelin-site	更改 zeppelin-site.xml 中的配置设置。	Restarts Zeppelin.

分类	描述	重新配置操作
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。	Restarts Zookeeper server.
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。	Restarts Zookeeper server.

## 亚马逊EMR版本 5.33.1

### 5.33.1 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[JupyterEnterpriseGateway](#)、[Jupyter](#)和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-5.33.1	emr-5.33.0	emr-5.32.1	emr-5.32.0
适用于 Java 的 AWS SDK	1.11.970	1.11.970	1.11.890	1.11.890
Python	2.7、3.7	2.7、3.7	2.7、3.7	2.7、3.7
Scala	2.11.12	2.11.12	2.11.12	2.11.12

	emr-5.33.1	emr-5.33.0	emr-5.32.1	emr-5.32.0
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.12.1	1.12.1	1.11.2	1.11.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13
HCatalog	2.3.7	2.3.7	2.3.7	2.3.7
Hadoop	2.10.1	2.10.1	2.10.1	2.10.1
Hive	2.3.7	2.3.7	2.3.7	2.3.7
Hudi	0.7.0-amzn-1	0.7.0-amzn-1	0.6.0-amzn-0	0.6.0-amzn-0
Hue	4.9.0	4.9.0	4.8.0	4.8.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.2.2	1.2.2	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.7.0	1.7.0	1.7.0	1.7.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.2.0	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0

	emr-5.33.1	emr-5.33.0	emr-5.32.1	emr-5.32.0
Presto	0.245.1	0.245.1	0.240.1	0.240.1
Spark	2.4.7	2.4.7	2.4.7	2.4.7
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.3.1	2.3.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.9.0	0.9.0	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

### 5.33.1 发布说明

以下发行说明包含亚马逊EMR版本 5.33.0/5.33.1 的信息。更改与 5.32.0 有关。

首次发布日期：2021 年 4 月 19 日

上次更新日期：2021 年 8 月 9 日

#### 升级

- 已将 Amazon Glue 连接器升级到版本 1.15.0
- 已升级 AWS SDK for Java 到 1.11.970 版本
- 已升级EMRFS到 2.46.0 版
- 已将 EMR Goodies 升级到 2.14.0 版
- 已将 R EMR ecord Server 升级到 1.9.0 版
- 已EMR将 S3 Dist CP 升级到版本 2.18.0
- 已将 S EMR ecret Agent 升级到 1.8.0 版
- 已将 Flink 升级到版本 1.12.1
- 已将 Hadoop 升级到版本 2.10.1-amzn-1



- 已将 Hive 升级到版本 2.3.7-amzn-4
- 已将 Hudi 升级到版本 0.7.0
- 已将 Hue 升级到版本 4.9.0
- 已将 OpenCV 升级到版本 4.5.0
- 已将 Presto 升级到版本 0.245.1-amzn-0
- 已将 R 升级到版本 4.0.2
- 已将 Spark 升级到版本 2.4.7-amzn-1
- 已升级 TensorFlow 到 2.4.1 版
- 已将 Zeppelin 升级到版本 0.9.0

### 更改、增强功能和解决的问题

- 此版本旨在修复 Amazon EMR Scaling 无法成功扩展/缩小集群或导致应用程序故障时出现的问题。
- 修复了 Amaz EMR on 集群上守护程序运行状况检查活动 ( 例如收集YARN节点状态和HDFS节点状态 ) 时, 对利用率很高的大型集群的扩展请求失败的问题。之所以发生这种情况, 是因为集群上的守护程序无法将节点的运行状况数据传送到内部 Ama EMR zon 组件。
- 改进了EMR集群上的守护程序, 可在重复使用 IP 地址时正确跟踪节点状态, 从而提高扩展操作期间的可靠性。
- [SPARK-29683](#)。修复了集群缩减期间出现任务失败的问题, 因为 Spark 假定所有可用节点都被拒绝列出。
- [YARN-9011](#)。修复了集群尝试向上或向下扩展时, 由于YARN停用过程中的争用条件而导致任务失败的问题。
- 通过确保 Amaz EMR on 集群上守护程序和 YARN /之间的节点状态始终保持一致, 修复了集群扩展期间步骤或任务失败的问题。HDFS
- 修复了启用了 Kerberos 身份验证的 Amazon 集群的EMR集群操作 ( 例如缩减和步骤提交 ) 失败的问题。这是因为 Amaz EMR on 集群上守护程序没有续订 Kerberos 票证, 而要与主节点上YARN运行的 HDFS /进行安全通信, 则需要续订 Kerberos 票证。
- 较新的Amazon EMR 版本通过降低亚马逊较旧AL2版本的“最大打开文件数”限制来解决这个问题 EMR。亚马逊EMR发布的 5.30.1、5.30.2、5.31.1、5.32.1、5.32.1、6.0.1、6.1.1、6.2.1、5.33.0、6.3.0及更高版本现在包含一个永久修复程序, 其中包含一个具有更高的“最大打开文件数”设置的永久修复。
- 配置群集以修复 Apache T YARN imeline Server 版本 1 和 1.5 的性能问题

Apache T YARN imeline Server 版本 1 和 1.5 可能会导致非常活跃的大型EMR集群出现性能问题`yarn.resourcemanager.system-metrics-publisher.enabled=true`，尤其是在使用 Amazon EMR 的默认设置时。开源YARN时间轴服务器 v2 解决了与时间YARN轴服务器可扩展性相关的性能问题。

此问题的其他解决方法包括：

- 配置 `yarn.resourceManager system-metrics-publisheryarn-site.xml` 中的 `.enabled=false`。
- 如下所述，在创建群集时启用此问题的修复程序。

以下 Amazon EMR 版本包含针对此YARN时间轴服务器性能问题的修复程序。

EMR5.30.2、5.31.1、5.32.1、5.33.1、5.34.x、6.0.1、6.1.1、6.2.1、6.3.1、6.3.1、6.3.1、6.4.x

要在上述任何指定的 Amazon EMR 版本上启用修复，请在使用[aws emr create-cluster](#)命令参数传`true`入的[配置JSON文件](#)中将[这些属性](#)设置为：`--configurations file://./configurations.json`。或者使用[重新配置控制台 UI](#) 启用修复程序。

配置 `.json` 文件内容的示例：

```
[
  {
    "Classification": "yarn-site",
    "Properties": {
      "yarn.resourcemanager.system-metrics-publisher.timeline-server-v1.enable-batch":
        "true",
      "yarn.resourcemanager.system-metrics-publisher.enabled": "true"
    },
    "Configurations": []
  }
]
```

- 现在，从 Hive 元存储中获取分区位置进行 Spark 插入查询时，Spark 运行时的速度更快。
- 升级了组件版本。有关组件版本的列表，请参阅本指南中的[关于 Amazon EMR 版本](#)。
- 在每个新集群上安装了 AWS Java SDK 捆绑包。这是一个包含所有服务SDKs及其依赖关系的单个 jar，而不是单个组件 jar。有关更多信息，请参阅 [Java SDK 捆绑依赖关系](#)。
- 修复了早期 Amazon EMR 版本中的托管扩展问题，并进行了改进，从而显著降低了应用程序失败率。

- HTTPS现在，亚马逊 Linux 存储库已默认启用。如果您使用 Amazon S3 VPC 策略来限制对特定存储桶的访问，则必须将新的 Amazon Linux 存储桶ARNarn:aws:s3:::amazonlinux-2-repos-\$region/\*添加到您的策略中（\$region替换为终端节点所在的区域）。有关更多信息，请在 AWS 讨论论坛中查看此主题。[公告：Amazon Linux 2 现在支持HTTPS在连接软件包存储库时使用。](#)

## 新功能

- 亚马逊EMR支持 Amazon S3 接入点，这是 Amazon S3 的一项功能，可让您轻松管理共享数据湖的访问权限。使用您的 Amazon S3 接入点别名，您可以在 Amazon 上大规模简化数据访问 EMR。EMR在提供亚马逊服务的所有 AWS 地区，您可以将 Amazon S3 接入点用于所有版本EMR的 Amazon，无需支付额外费用。要了解有关 Amazon S3 访问点和访问点别名的详细信息，请参阅《Amazon S3 用户指南》中的[为接入点使用存储桶式别名](#)。
- 亚马逊 EMR -5.33 支持新的亚马逊EC2实例类型：  
c5a、c5ad、c6gn、c6gd、m6gd、d3、d3en、m5zn、r5b、r6gd。请参阅[支持的实例类型](#)。

## 已知问题

- 降低旧版本的“最大打开文件数”限制 AL2 [在新版本中已修复]。亚马逊EMR版本：  
emr-5.30.x、emr-5.31.0、emr-5.32.0、emr-6.0.0、emr-6.1.0 和 emr-6.2.0 基于较旧版本的 Linux ofAmazon 2 ()AL2，当使用默认版本创建亚马逊集群时，“最大打开文件数”的用户限制较低。EMR AMI亚马逊EMR发布的版本为5.30.1、5.30.2、5.31.1、5.32.1、6.0.1、6.1.1、6.2.1、5.33.0、6.3.0 及更高版本，包括具有更高的“最大打开文件数”设置的永久修复。如果使用打开文件数限制较低的发行版，会在提交 Spark 任务时导致“Too many open files”（打开的文件过多）错误。在受影响的版本中，亚马逊EMR默认AMI认“最大打开文件数”的ulimit设置为4096，低于Linux 2中的65536个文件限制。latestAmazon AMISpark 驱动程序和执行程序尝试打开超过 4096 个文件时，“打开的最大文件数”的较低 ulimit 设置会导致 Spark 任务失败。为了解决这个问题，Amazon EMR 提供了一个引导操作 (BA) 脚本，可以在创建集群时调整ulimit设置。

如果您使用的是无法永久修复此问题的旧EMR版 Amazon，则可以使用以下解决方法将实例控制器 ulimit 明确设置为最多 65536 个文件。

### 从命令行显式设置 ulimit

1. 编辑 /etc/systemd/system/instance-controller.service，将以下参数添加到 Service (服务) 部分。

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

## 2. 重启 InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

### 使用引导操作 (BA) 设置 ulimit

您还可以在创建集群时使用引导操作 ( BA ) 脚本将实例控制器 ulimit 配置为 65536 个文件。

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- 对于 Amazon EMR 6.3.0 和 6.2.0 私有子网集群，您无法访问 Ganglia 网页用户界面。您将收到“access denied (403)”错误。其他网络UIs，例如 Spark、Hue JupyterHub、Zeppelin、Livy 和 Tez，都正常运行。公有子网集群上的 Ganglia Web UI 访问也正常工作。要解决该问题，请在具有 `sudo systemctl restart httpd` 的主节点上重新启动 httpd 服务。此问题已在 Amazon EMR 6.4.0 中修复。

#### Important

EMR运行 Amazon Linux 或 Amazon Linux 2 Amazon 机器映像 (AMIs) 的集群使用默认 Amazon Linux 行为，不会自动下载和安装需要重启的重要关键内核更新。这与运行默认

Amazon Linux 的其他亚马逊EC2实例的行为相同AMI。如果在 Amazon EMR 版本发布后需要重启的新 Amazon Linux 软件CUDA更新（例如内核和更新）可用，则默认运行的EMR集群实例AMI不会自动下载和安装这些更新。NVIDIA要获取内核更新，您可以[自定义您的亚马逊EMRAMI](#)以[使用最新的亚马逊 Linux AMI](#)。

- 该 GovCloud 区域目前不支持控制台支持创建指定 AWS Ranger 集成选项的安全配置。可以使用来完成安全配置CLI。请参阅 [《Amazon EMR 管理指南》中的创建EMR安全配置](#)。
- 限定范围的托管策略：为了与 AWS 最佳实践保持一致，Amazon 引入EMR了 v2 EMR 范围的默认托管策略，以取代即将弃用的策略。参见 [Amazon EMR 托管政策](#)。
- 当你使用带有 Hive 分区位置格式的 Spark 来读取 Amazon S3 中的数据，并在亚马逊EMR版本 5.30.0 到 5.36.0 和 6.2.0 到 6.9.0 上运行 Spark 时，你可能会遇到一个问题，导致集群无法正确读取数据。如果您的分区具有以下所有特征，会发生这种情况：
  - 从同一个表扫描两个或多个分区。
  - 至少有一个分区目录路径是至少一个其他分区目录路径的前缀，例如，s3://bucket/table/p=a 是 s3://bucket/table/p=a b 的前缀。
  - 另一个分区目录中前缀之后的第一个字符的值为 UTF -8，小于该/字符 (U+002F)。例如，在 s3://bucket/table/p=a b 中，a 和 b 之间出现的空格字符 (U+0020) 就属于此类。请注意，还有其他 14 个非控制字符：!"#\$%&'()\*+,-。有关更多信息，请参阅 [UTF-8 编码表和 Unicode 字符](#)。

解决方法是在 spark-defaults 分类中将

`spark.sql.sources.fastS3PartitionDiscovery.enabled` 配置设置为 `false`。

### 5.33.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
aws-sagemaker-spark-sdk	1.4.1	亚马逊 SageMaker Spark SDK
emr-ddb	4.16.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	emrfs	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.5.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-notebook-env	1.2.0	适用于 EMR Notebooks ( 可提供 jupyter 企业网关 ) 的 Conda env
emr-s3-dist-cp	2.18.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	1.6.0	EMRS3Select 连接器
emrfs	2.46.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.12.1	Apache Flink 命令行客户端脚本和应用程序。
flink-jobmanager-config	1.12.1	管理 Apache Flink EMR JobManager 节点上的资源。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。

组件	版本	描述
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.10.1-amzn-1.1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.10.1-amzn-1.1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.10.1-amzn-1.1	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.10.1-amzn-1.1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	2.10.1-amzn-1.1	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	2.10.1-amzn-1.1	HTTPHDFS操作端点。
hadoop-kms-server	2.10.1-amzn-1.1	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.10.1-amzn-1.1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.10.1-amzn-1.1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.10.1-amzn-1.1	YARN用于分配和管理群集资源和分布式应用程序的服务。

组件	版本	描述
hadoop-yarn-timeline-server	2.10.1-amzn-1.1	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.13	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.13	为一个或多个HBase地区提供服务的服务。
hbase-client	1.4.13	HBase命令行客户端。
hbase-rest-server	1.4.13	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.13	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.3.7-amzn-4	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.7-amzn-4	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.7-amzn-4	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.7-amzn-4	Hive 命令行客户端。
hive-hbase	2.3.7-amzn-4	Hive-hbase 客户端。
hive-metastore-server	2.3.7-amzn-4	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.7-amzn-4	用于将 Hive 查询作为 Web 请求接受的服务。



组件	版本	描述
hudi	0.7.0-amzn-1	增量处理框架，以支持低延迟和高效率的数据管道。
hudi-spark	0.7.0-amzn-1	用于运行 Spark 以及 Hudi 的捆绑库。
hudi-presto	0.7.0-amzn-1	用于运行 Presto 以及 Hudi 的捆绑库。
hue-server	4.9.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	1.2.2	Jupyter notebook 的多用户服务器
livy-server	0.7.0-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.7.0	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.68+	我的SQL数据库服务器。
nvidia-cuda	10.1.243	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.0	Oozie 命令行客户端。
oozie-server	5.2.0	用于接受 Oozie 工作流请求的服务。
opencv	4.5.0	开源计算机视觉库。

组件	版本	描述
phoenix-library	4.14.3--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.14.3--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.245.1-amzn-0	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.245.1-amzn-0	用于执行查询的各个部分的服务。
presto-client	0.245.1-amzn-0	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
pig-client	0.17.0	Pig 命令行客户端。
r	4.0.2	用于统计计算的 R 项目
ranger-kms-server	1.2.0	Apache Ranger 密钥管理系统
spark-client	2.4.7-amzn-1.1	Spark 命令行客户端。
spark-history-server	2.4.7-amzn-1.1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.4.7-amzn-1.1	的内存中执行引擎。YARN
spark-yarn-slave	2.4.7-amzn-1.1	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。

组件	版本	描述
tensorflow	2.4.1	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.2	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.9.0	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.14	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.14	ZooKeeper 命令行客户端。

### 5.33.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

为处于运行状态的集群中的实例组指定配置时，将发生重新配置操作。Amazon EMR 仅对您修改的分类启动重新配置操作。有关更多信息，请参阅 [在正在运行的集群中重新配置实例组](#)。

#### emr-5.33.1 分类

分类	描述	重新配置操作
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。	Restarts the ResourceManager service.
container-executor	更改 Hadoop YARN 的 container-executor.cfg 文件中的值。	Not available.

分类	描述	重新配置操作
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。	Not available.
core-site	更改 Hadoop 的 core-site.xml 文件中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	更改 docker 相关设置。	Not available.

分类	描述	重新配置操作
emrfs-site	更改EMRFS设置。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	更改 flink-conf.yaml 设置。	Restarts Flink history server.
flink-log4j	更改 Flink log4j.properties 设置。	Restarts Flink history server.
flink-log4j-session	为 Kubernetes/Yarn 会话更改 Flink log4j-session.properties 设置。	Not available.
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。	Restarts Flink history server.

分类	描述	重新配置操作
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	更改 hadoop ssl 服务器配置	Not available.
hadoop-ssl-client	更改 hadoop ssl 客户端配置	Not available.
hbase	亚马逊EMR精心策划的 Apache 设置。HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.

分类	描述	重新配置操作
hbase-env	更改环境HBase中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	更改 hadoop-metrics2-hbase base.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。	Not available.
hbase-site	更改HBase的 hbase-site.xml 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	配置HDFS加密区域。	Should not be reconfigured.
hdfs-site	更改 hdfs-site.xml HDFS 中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	更改环境HCatalog中的值。	Restarts Hive HCatalog Server.
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。	Restarts Hive HCatalog Server.

分类	描述	重新配置操作
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat 中的值。	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。	Restarts Hive WebHCat Server.
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。	Not available.
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。	Not available.
hive-env	更改 Hive 环境中的值。	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	更改 Hive 的 hive-exec-log4j2.properties 文件中的值。	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log4j2.properties 文件中的值。	Not available.
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。	Not available.



分类	描述	重新配置操作
hive-site	更改 Hive 的 hive-site.xml 文件中的值	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值	Not available.
hue-ini	更改 Hue 的 ini 文件中的值	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	更改HTTPFS环境中的值。	Restarts Hadoop Httpfs service.
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。	Restarts Hadoop Httpfs service.
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。	Not available.
hadoop-kms-env	在 Hadoop KMS 环境中更改值。	Restarts Hadoop-KMS service.
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。	Not available.
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	更改 Hudi 环境中的值。	Not available.

分类	描述	重新配置操作
jupyter-notebook-conf	更改 Jupyter notebook 的 <code>jupyter_notebook_config.py</code> 文件中的值。	Not available.
jupyter-hub-conf	更改 <code>jupyterhub_config.py</code> 文件 JupyterHubs 中的值。	Not available.
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。	Not available.
jupyter-sparkmagic-conf	更改 Sparkmagic 的 <code>config.json</code> 文件中的值。	Not available.
livy-conf	更改 Livy 的 <code>livy.conf</code> 文件中的值。	Restarts Livy Server.
livy-env	更改 Livy 环境中的值。	Restarts Livy Server.
livy-log4j	更改 Livy <code>log4j.properties</code> 设置。	Restarts Livy Server.
mapred-env	更改 MapReduce 应用程序环境中的值。	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	更改 MapReduce 应用程序的 <code>mapred-site.xml</code> 文件中的值。	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	更改 Oozie 的环境中的值。	Restarts Oozie.
oozie-log4j	更改 Oozie 的 <code>oozie-log4j.properties</code> 文件中的值。	Restarts Oozie.
oozie-site	更改 Oozie 的 <code>oozie-site.xml</code> 文件中的值。	Restarts Oozie.
phoenix-hbase-metrics	更改 Phoenix 的 <code>hadoop-metrics2-hbase.properties</code> 文件中的值。	Not available.

分类	描述	重新配置操作
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。	Not available.
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。	Restarts Phoenix-QueryServer.
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。	Not available.
pig-env	更改 Pig 环境中的值。	Not available.
pig-properties	更改 Pig 的 pig.properties 文件中的值。	Restarts Oozie.
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。	Not available.
presto-log	更改 Presto 的 log.properties 文件中的值。	Restarts Presto-Server.
presto-config	更改 Presto 的 config.properties 文件中的值。	Restarts Presto-Server.
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。	Not available.
presto-env	更改 Presto 的 presto-env.sh 文件中的值。	Restarts Presto-Server.
presto-node	更改 Presto 的 node.properties 文件中的值。	Not available.
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。	Not available.
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。	Restarts Presto-Server.
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。	Not available.
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。	Not available.
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。	Not available.
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。	Not available.
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。	Not available.
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。	Not available.
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。	Not available.
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。	Not available.
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。	Not available.
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。	Not available.
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。	Not available.
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-site	更改 Ranger KMS 的 ranger-kms-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-env	在 Ranger KMS 环境中更改值。	Restarts Ranger KMS Server.
ranger-kms-log4j	更改 Ranger 的 kms-log4j.properties 文件中的值。KMS	Not available.
ranger-kms-db-ca	更改 S3 上 Ranger SQL SSL KMS 连接的 CA 文件值。	Not available.
recordserver-env	更改 EMR RecordServer 环境中的值。	Restarts EMR record server.
recordserver-conf	更改 EMR RecordServer 的 proper EMR RecordServer ties 文件中的值。	Restarts EMR record server.
recordserver-log4j	更改 EMR RecordServer 的 log4j.properties 文件中的值。	Restarts EMR record server.
spark	亚马逊 EMR 精心策划的 Apache Spark 设置。	This property modifies spark-defaults. See actions there.
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-env	更改 Spark 环境中的值。	Restarts Spark history server and Spark thrift server.
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值	Not available.

分类	描述	重新配置操作
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
sqoop-env	更改 Sqoop 的环境中的值。	Not available.
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。	Not available.
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。	Not available.
tez-site	更改 Tez 的 tez-site.xml 文件中的值。	Restarts Oozie.
yarn-env	更改YARN环境中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	更改YARN的 yarn-site.xml 文件中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	更改 Zeppelin 环境中的值。	Restarts Zeppelin.
zeppelin-site	更改 zeppelin-site.xml 中的配置设置。	Restarts Zeppelin.

分类	描述	重新配置操作
zookeeper-config	更改 ZooKeeper 的 zoo.cfg 文件中的值。	Restarts Zookeeper server.
zookeeper-log4j	更改 ZooKeeper 的 log4j.properties 文件中的值。	Restarts Zookeeper server.

## 亚马逊EMR发布 5.33.0

### 5.33.0 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[JupyterEnterpriseGateway](#)、[Jupyter](#) 和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-5.33.0	emr-5.32.1	emr-5.32.0	emr-5.31.1
适用于 Java 的 AWS SDK	1.11.970	1.11.890	1.11.890	1.11.852
Python	2.7、3.7	2.7、3.7	2.7、3.7	2.7、3.7
Scala	2.11.12	2.11.12	2.11.12	2.11.12

	emr-5.33.0	emr-5.32.1	emr-5.32.0	emr-5.31.1
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.12.1	1.11.2	1.11.2	1.11.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13
HCatalog	2.3.7	2.3.7	2.3.7	2.3.7
Hadoop	2.10.1	2.10.1	2.10.1	2.10.0
Hive	2.3.7	2.3.7	2.3.7	2.3.7
Hudi	0.7.0-amzn-1	0.6.0-amzn-0	0.6.0-amzn-0	0.6.0-amzn-0
Hue	4.9.0	4.8.0	4.8.0	4.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	-
JupyterHub	1.2.2	1.1.0	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.7.0	1.7.0	1.7.0	1.6.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.2.0	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0



	emr-5.33.0	emr-5.32.1	emr-5.32.0	emr-5.31.1
Presto	0.245.1	0.240.1	0.240.1	0.238.3
Spark	2.4.7	2.4.7	2.4.7	2.4.6
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.3.1	2.3.1	2.1.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.9.0	0.8.2	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 5.33.0 发布说明

### 5.33.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>aws-sagemaker-spark-sdk</code>	1.4.1	亚马逊 SageMaker Spark SDK

组件	版本	描述
emr-ddb	4.16.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	emrfs	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.5.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-notebook-env	1.2.0	适用于 EMR Notebooks ( 可提供 jupyter 企业网关 ) 的 Conda env
emr-s3-dist-cp	2.18.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	1.6.0	EMRS3Select 连接器
emrfs	2.46.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.12.1	Apache Flink 命令行客户端脚本和应用程序。
flink-jobmanager-config	1.12.1	管理 Apache Flink EMR JobManager 节点上的资源。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。

组件	版本	描述
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.10.1-amzn-1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.10.1-amzn-1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.10.1-amzn-1	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.10.1-amzn-1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	2.10.1-amzn-1	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	2.10.1-amzn-1	HTTPHDFS操作端点。
hadoop-kms-server	2.10.1-amzn-1	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.10.1-amzn-1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.10.1-amzn-1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.10.1-amzn-1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.10.1-amzn-1	用于检索YARN应用程序的当前和历史信息的服务。

组件	版本	描述
hbase-hmaster	1.4.13	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.13	为一个或多个HBase地区提供服务的服务器。
hbase-client	1.4.13	HBase命令行客户端。
hbase-rest-server	1.4.13	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.13	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.3.7-amzn-4	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.7-amzn-4	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.7-amzn-4	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.7-amzn-4	Hive 命令行客户端。
hive-hbase	2.3.7-amzn-4	Hive-hbase 客户端。
hive-metastore-server	2.3.7-amzn-4	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.7-amzn-4	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.7.0-amzn-1	增量处理框架，以支持低延迟和高效率的数据管道。

组件	版本	描述
hudi-spark	0.7.0-amzn-1	用于运行 Spark 以及 Hudi 的捆绑库。
hudi-presto	0.7.0-amzn-1	用于运行 Presto 以及 Hudi 的捆绑库。
hue-server	4.9.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	1.2.2	Jupyter notebook 的多用户服务器
livy-server	0.7.0-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.7.0	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.68	我的SQL数据库服务器。
nvidia-cuda	10.1.243	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.0	Oozie 命令行客户端。
oozie-server	5.2.0	用于接受 Oozie 工作流请求的服务。
opencv	4.5.0	开源计算机视觉库。
phoenix-library	4.14.3--1.4 HBase	服务器和客户端的 phoenix 库

组件	版本	描述
phoenix-query-server	4.14.3--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.245.1-amzn-0	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.245.1-amzn-0	用于执行查询的各个部分的服务。
presto-client	0.245.1-amzn-0	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
pig-client	0.17.0	Pig 命令行客户端。
r	4.0.2	用于统计计算的 R 项目
ranger-kms-server	1.2.0	Apache Ranger 密钥管理系统
spark-client	2.4.7-amzn-1	Spark 命令行客户端。
spark-history-server	2.4.7-amzn-1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.4.7-amzn-1	的内存中执行引擎。YARN
spark-yarn-slave	2.4.7-amzn-1	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	2.4.1	TensorFlow 用于高性能数值计算的开源软件库。

组件	版本	描述
tez-on-yarn	0.9.2	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.9.0	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.14	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.14	ZooKeeper 命令行客户端。

### 5.33.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

为处于运行状态的集群中的实例组指定配置时，将发生重新配置操作。Amazon EMR 仅对您修改的分类启动重新配置操作。有关更多信息，请参阅 [在正在运行的集群中重新配置实例组](#)。

#### emr-5.33.0 分类

分类	描述	重新配置操作
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。	Restarts the ResourceManager service.
container-executor	更改 Hadoop YARN 的 container-executor.cfg 文件中的值。	Not available.
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。	Not available.
core-site	更改 Hadoop 的 core-site.xml 文件中的值。	Restarts the Hadoop HDFS services Namenode,

分类	描述	重新配置操作
		SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	更改 docker 相关设置。	Not available.
emrfs-site	更改EMRFS设置。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	更改 flink-conf.yaml 设置。	Restarts Flink history server.
flink-log4j	更改 Flink log4j.properties 设置。	Restarts Flink history server.



分类	描述	重新配置操作
flink-log4j-session	为 Kubernetes/Yarn 会话更改 Flink log4j-session.properties 设置。	Not available.
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。	Restarts Flink history server.
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	更改 hadoop ssl 服务器配置	Not available.

分类	描述	重新配置操作
hadoop-ssl-client	更改 hadoop ssl 客户端配置	Not available.
hbase	亚马逊EMR精心策划的 Apache 设置。HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	更改环境HBase中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	更改 hadoop-metrics2-hbase base.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。	Not available.
hbase-site	更改HBase的 hbase-site.xml 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	配置HDFS加密区域。	Should not be reconfigured.
hdfs-site	更改 hdfs-site.xml HDFS 中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

分类	描述	重新配置操作
hcatalog-env	更改环境HCatalog中的值。	Restarts Hive HCatalog Server.
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat 中的值。	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。	Restarts Hive WebHCat Server.
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。	Not available.
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。	Not available.
hive-env	更改 Hive 环境中的值。	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	更改 Hive 的 hive-exec-log4j2.properties 文件中的值。	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log4j2.properties 文件中的值。	Not available.

分类	描述	重新配置操作
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。	Not available.
hive-site	更改 Hive 的 hive-site.xml 文件中的值	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值	Not available.
hue-ini	更改 Hue 的 ini 文件中的值	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	更改HTTPFS环境中的值。	Restarts Hadoop Httpfs service.
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。	Restarts Hadoop Httpfs service.
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。	Not available.
hadoop-kms-env	在 Hadoop KMS 环境中更改值。	Restarts Hadoop-KMS service.
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。	Not available.
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	更改 Hudi 环境中的值。	Not available.

分类	描述	重新配置操作
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。	Not available.
jupyter-hub-conf	更改 jupyterhub_config.py 文件 JupyterHubs 中的值。	Not available.
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。	Not available.
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。	Not available.
livy-conf	更改 Livy 的 livy.conf 文件中的值。	Restarts Livy Server.
livy-env	更改 Livy 环境中的值。	Restarts Livy Server.
livy-log4j	更改 Livy log4j.properties 设置。	Restarts Livy Server.
mapred-env	更改 MapReduce 应用程序环境中的值。	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	更改 Oozie 的环境中的值。	Restarts Oozie.
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。	Restarts Oozie.
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。	Restarts Oozie.
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。	Not available.

分类	描述	重新配置操作
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。	Not available.
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。	Restarts Phoenix-QueryServer.
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。	Not available.
pig-env	更改 Pig 环境中的值。	Not available.
pig-properties	更改 Pig 的 pig.properties 文件中的值。	Restarts Oozie.
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。	Not available.
presto-log	更改 Presto 的 log.properties 文件中的值。	Restarts Presto-Server.
presto-config	更改 Presto 的 config.properties 文件中的值。	Restarts Presto-Server.
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。	Not available.
presto-env	更改 Presto 的 presto-env.sh 文件中的值。	Restarts Presto-Server.
presto-node	更改 Presto 的 node.properties 文件中的值。	Not available.
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。	Not available.
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。	Restarts Presto-Server.
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。	Not available.
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。	Not available.
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。	Not available.
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。	Not available.
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。	Not available.
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。	Not available.
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。	Not available.
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。	Not available.
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。	Not available.
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。	Not available.
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。	Not available.
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-site	更改 Ranger KMS 的 ranger-kms-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-env	在 Ranger KMS 环境中更改值。	Restarts Ranger KMS Server.
ranger-kms-log4j	更改 Ranger 的 kms-log4j.properties 文件中的值。KMS	Not available.
ranger-kms-db-ca	更改 S3 上 Ranger SQL SSL KMS 连接的 CA 文件值。	Not available.
recordserver-env	更改 EMR RecordServer 环境中的值。	Restarts EMR record server.
recordserver-conf	更改 EMR RecordServer 的 proper EMR RecordServer ties 文件中的值。	Restarts EMR record server.
recordserver-log4j	更改 EMR RecordServer 的 log4j.properties 文件中的值。	Restarts EMR record server.
spark	亚马逊 EMR 精心策划的 Apache Spark 设置。	This property modifies spark-defaults. See actions there.
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-env	更改 Spark 环境中的值。	Restarts Spark history server and Spark thrift server.
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值	Not available.



分类	描述	重新配置操作
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
sqoop-env	更改 Sqoop 的环境中的值。	Not available.
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。	Not available.
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。	Not available.
tez-site	更改 Tez 的 tez-site.xml 文件中的值。	Restarts Oozie.
yarn-env	更改YARN环境中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	更改YARN的 yarn-site.xml 文件中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	更改 Zeppelin 环境中的值。	Restarts Zeppelin.
zeppelin-site	更改 zeppelin-site.xml 中的配置设置。	Restarts Zeppelin.

分类	描述	重新配置操作
zookeeper-config	更改 ZooKeeper 的 zoo.cfg 文件中的值。	Restarts Zookeeper server.
zookeeper-log4j	更改 ZooKeeper 的 log4j.properties 文件中的值。	Restarts Zookeeper server.

## 亚马逊EMR版本 5.32.1

### 5.32.1 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[JupyterEnterpriseGateway](#)、[Jupyter](#) 和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-5.32.1	emr-5.32.0	emr-5.31.1	emr-5.31.0
适用于 Java 的 AWS SDK	1.11.890	1.11.890	1.11.852	1.11.852
Python	2.7、3.7	2.7、3.7	2.7、3.7	2.7、3.7
Scala	2.11.12	2.11.12	2.11.12	2.11.12

	emr-5.32.1	emr-5.32.0	emr-5.31.1	emr-5.31.0
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.11.2	1.11.2	1.11.0	1.11.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13
HCatalog	2.3.7	2.3.7	2.3.7	2.3.7
Hadoop	2.10.1	2.10.1	2.10.0	2.10.0
Hive	2.3.7	2.3.7	2.3.7	2.3.7
Hudi	0.6.0-amzn-0	0.6.0-amzn-0	0.6.0-amzn-0	0.6.0-amzn-0
Hue	4.8.0	4.8.0	4.7.1	4.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	-	-
JupyterHub	1.1.0	1.1.0	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.7.0	1.7.0	1.6.0	1.6.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.2.0	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0

	emr-5.32.1	emr-5.32.0	emr-5.31.1	emr-5.31.0
Presto	0.240.1	0.240.1	0.238.3	0.238.3
Spark	2.4.7	2.4.7	2.4.6	2.4.6
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.3.1	2.3.1	2.1.0	2.1.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

### 5.32.1 发布说明

此版本旨在修复 Amazon EMR Scaling 无法成功扩展/缩小集群或导致应用程序故障时出现的问题。

#### 更改、增强和解决的问题

- 修复了 Amaz EMR on 集群上守护程序运行状况检查活动 ( 例如收集YARN节点状态和HDFS节点状态 ) 时，对利用率很高的大型集群的扩展请求失败的问题。之所以发生这种情况，是因为集群上的守护程序无法将节点的运行状况数据传送到内部 Ama EMR zon 组件。
- 改进了EMR集群上的守护程序，可在重复使用 IP 地址时正确跟踪节点状态，从而提高扩展操作期间的可靠性。
- [SPARK-29683](#)。修复了集群缩减期间出现任务失败的问题，因为 Spark 假定所有可用节点都被拒绝列出。
- [YARN-9011](#)。修复了集群尝试向上或向下扩展时，由于YARN停用过程中的争用条件而导致任务失败的问题。
- 通过确保 Amaz EMR on 集群上守护程序和 YARN /之间的节点状态始终保持一致，修复了集群扩展期间步骤或任务失败的问题。HDFS

- 修复了启用了 Kerberos 身份验证的 Amazon 集群的EMR集群操作（例如缩减和步骤提交）失败的问题。这是因为 Amaz EMR on 集群上守护程序没有续订 Kerberos 票证，而要与主节点上YARN运行的 HDFS /进行安全通信，则需要续订 Kerberos 票证。
- 较新的Amazon EMR 版本通过降低亚马逊较旧AL2版本的“最大打开文件数”限制来解决这个问题EMR。亚马逊EMR发布的 5.30.1、5.30.2、5.31.1、5.32.1、5.32.1、6.0.1、6.1.1、6.2.1、5.33.0、6.3.0及更高版本现在包含一个永久修复程序，其中包含一个具有更高的“最大打开文件数”设置的永久修复。
- HTTPS现在，亚马逊 Linux 存储库已默认启用。如果您使用 Amazon S3 VPCE 策略来限制对特定存储桶的访问，则必须将新的 Amazon Linux 存储桶ARNarn:aws:s3:::amazonlinux-2-repos-\$region/\*添加到您的策略中（\$region替换为终端节点所在的区域）。有关更多信息，请在 AWS 讨论论坛中查看此主题。[公告：Amazon Linux 2 现在支持HTTPS在连接软件包存储库时使用。](#)

## 已知问题

- 当你使用带有 Hive 分区位置格式的 Spark 来读取 Amazon S3 中的数据，并在亚马逊EMR版本 5.30.0 到 5.36.0 和 6.2.0 到 6.9.0 上运行 Spark 时，你可能会遇到一个问题，导致集群无法正确读取数据。如果您的分区具有以下所有特征，会发生这种情况：
  - 从同一个表扫描两个或多个分区。
  - 至少有一个分区目录路径是至少一个其他分区目录路径的前缀，例如，s3://bucket/table/p=a 是 s3://bucket/table/p=a b 的前缀。
  - 另一个分区目录中前缀之后的第一个字符的值为 UTF -8，小于该/字符 (U+002F)。例如，在 s3://bucket/table/p=a b 中，a 和 b 之间出现的空格字符 (U+0020) 就属于此类。请注意，还有其他 14 个非控制字符：! "\$ % & ' ( ) \* + , - 。有关更多信息，请参阅 [UTF-8 编码表和 Unicode 字符](#)。

解决方法是在 spark-defaults 分类中将

```
spark.sql.sources.fastS3PartitionDiscovery.enabled 配置设置为 false。
```

### 5.32.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 emr 或 aws 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.4.1	亚马逊 SageMaker Spark SDK
emr-ddb	4.16.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.13.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.5.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-notebook-env	1.1.0	适用于 EMR Notebooks ( 可提供 jupyter 企业网关 ) 的 Conda env
emr-s3-dist-cp	2.17.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	1.6.0	EMRS3Select 连接器
emrfs	2.45.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.11.2	Apache Flink 命令行客户端脚本和应用程序。
flink-jobmanager-config	1.11.2	管理 Apache Flink EMR JobManager 节点上的资源。

组件	版本	描述
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.10.1-amzn-0.1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.10.1-amzn-0.1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.10.1-amzn-0.1	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.10.1-amzn-0.1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	2.10.1-amzn-0.1	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	2.10.1-amzn-0.1	HTTPHDFS操作端点。
hadoop-kms-server	2.10.1-amzn-0.1	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.10.1-amzn-0.1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。

组件	版本	描述
hadoop-yarn-nodemanager	2.10.1-amzn-0.1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.10.1-amzn-0.1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.10.1-amzn-0.1	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.13	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.13	为一个或多个HBase地区提供服务的服务。
hbase-client	1.4.13	HBase命令行客户端。
hbase-rest-server	1.4.13	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.13	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.3.7-amzn-3	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.7-amzn-3	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.7-amzn-3	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.7-amzn-3	Hive 命令行客户端。
hive-hbase	2.3.7-amzn-3	Hive-hbase 客户端。



组件	版本	描述
hive-metastore-server	2.3.7-amzn-3	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的 SQL 元数据。
hive-server2	2.3.7-amzn-3	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.6.0-amzn-0	增量处理框架，以支持低延迟和高效率的数据管道。
hudi-spark	0.6.0-amzn-0	用于运行 Spark 以及 Hudi 的捆绑库。
hudi-presto	0.6.0-amzn-0	用于运行 Presto 以及 Hudi 的捆绑库。
hue-server	4.8.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	1.1.0	Jupyter notebook 的多用户服务器
livy-server	0.7.0-incubating	REST 用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.7.0	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.68+	我的 SQL 数据库服务器。

组件	版本	描述
nvidia-cuda	10.1.243	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.0	Oozie 命令行客户端。
oozie-server	5.2.0	用于接受 Oozie 工作流请求的服务。
opencv	4.4.0	开源计算机视觉库。
phoenix-library	4.14.3--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.14.3--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.240.1-amzn-0	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.240.1-amzn-0	用于执行查询的各个部分的服务。
presto-client	0.240.1-amzn-0	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
pig-client	0.17.0	Pig 命令行客户端。
r	3.4.3	用于统计计算的 R 项目
ranger-kms-server	1.2.0	Apache Ranger 密钥管理系统
spark-client	2.4.7-amzn-0.1	Spark 命令行客户端。
spark-history-server	2.4.7-amzn-0.1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。

组件	版本	描述
spark-on-yarn	2.4.7-amzn-0.1	的内存中执行引擎。YARN
spark-yarn-slave	2.4.7-amzn-0.1	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	2.3.1	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.2	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.8.2	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.14	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.14	ZooKeeper 命令行客户端。

### 5.32.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

为处于运行状态的集群中的实例组指定配置时，将发生重新配置操作。Amazon EMR 仅对您修改的分类启动重新配置操作。有关更多信息，请参阅 [在正在运行的集群中重新配置实例组](#)。

#### emr-5.32.1 分类

分类	描述	重新配置操作
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。	Restarts the ResourceM anager service.

分类	描述	重新配置操作
container-executor	更改 Hadoop YARN 的 container-executor.cfg 文件中的值。	Not available.
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。	Not available.
core-site	更改 Hadoop 的 core-site.xml 文件中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	更改 docker 相关设置。	Not available.

分类	描述	重新配置操作
emrfs-site	更改EMRFS设置。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	更改 flink-conf.yaml 设置。	Restarts Flink history server.
flink-log4j	更改 Flink log4j.properties 设置。	Restarts Flink history server.
flink-log4j-yarn-session	更改 Flink log4j-yarn-session.properties 设置。	Not available.
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。	Restarts Flink history server.

分类	描述	重新配置操作
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	更改 hadoop ssl 服务器配置	Not available.
hadoop-ssl-client	更改 hadoop ssl 客户端配置	Not available.
hbase	亚马逊EMR精心策划的 Apache 设置。HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.

分类	描述	重新配置操作
hbase-env	更改环境HBase中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	更改 hadoop-metrics2-hbase base.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。	Not available.
hbase-site	更改HBase的 hbase-site.xml 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	配置HDFS加密区域。	Should not be reconfigured.
hdfs-site	更改 hdfs-site.xml HDFS 中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	更改环境HCatalog中的值。	Restarts Hive HCatalog Server.
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。	Restarts Hive HCatalog Server.

分类	描述	重新配置操作
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat 中的值。	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。	Restarts Hive WebHCat Server.
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。	Not available.
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。	Not available.
hive-env	更改 Hive 环境中的值。	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	更改 Hive 的 hive-exec-log4j2.properties 文件中的值。	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log4j2.properties 文件中的值。	Not available.
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。	Not available.



分类	描述	重新配置操作
hive-site	更改 Hive 的 hive-site.xml 文件中的值	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值	Not available.
hue-ini	更改 Hue 的 ini 文件中的值	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	更改HTTPFS环境中的值。	Restarts Hadoop Httpfs service.
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。	Restarts Hadoop Httpfs service.
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。	Not available.
hadoop-kms-env	在 Hadoop KMS 环境中更改值。	Restarts Hadoop-KMS service.
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。	Not available.
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	更改 Hudi 环境中的值。	Not available.

分类	描述	重新配置操作
jupyter-notebook-conf	更改 Jupyter notebook 的 <code>jupyter_notebook_config.py</code> 文件中的值。	Not available.
jupyter-hub-conf	更改 <code>jupyterhub_config.py</code> 文件 JupyterHubs 中的值。	Not available.
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。	Not available.
jupyter-sparkmagic-conf	更改 Sparkmagic 的 <code>config.json</code> 文件中的值。	Not available.
livy-conf	更改 Livy 的 <code>livy.conf</code> 文件中的值。	Restarts Livy Server.
livy-env	更改 Livy 环境中的值。	Restarts Livy Server.
livy-log4j	更改 Livy <code>log4j.properties</code> 设置。	Restarts Livy Server.
mapred-env	更改 MapReduce 应用程序环境中的值。	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	更改 MapReduce 应用程序的 <code>mapred-site.xml</code> 文件中的值。	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	更改 Oozie 的环境中的值。	Restarts Oozie.
oozie-log4j	更改 Oozie 的 <code>oozie-log4j.properties</code> 文件中的值。	Restarts Oozie.
oozie-site	更改 Oozie 的 <code>oozie-site.xml</code> 文件中的值。	Restarts Oozie.
phoenix-hbase-metrics	更改 Phoenix 的 <code>hadoop-metrics2-hbase.properties</code> 文件中的值。	Not available.

分类	描述	重新配置操作
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。	Not available.
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。	Restarts Phoenix-QueryServer.
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。	Not available.
pig-env	更改 Pig 环境中的值。	Not available.
pig-properties	更改 Pig 的 pig.properties 文件中的值。	Restarts Oozie.
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。	Not available.
presto-log	更改 Presto 的 log.properties 文件中的值。	Restarts Presto-Server.
presto-config	更改 Presto 的 config.properties 文件中的值。	Restarts Presto-Server.
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。	Not available.
presto-env	更改 Presto 的 presto-env.sh 文件中的值。	Restarts Presto-Server.
presto-node	更改 Presto 的 node.properties 文件中的值。	Not available.
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。	Not available.
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。	Restarts Presto-Server.
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。	Not available.
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。	Not available.
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。	Not available.
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。	Not available.
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。	Not available.
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。	Not available.
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。	Not available.
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。	Not available.
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。	Not available.
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。	Not available.
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。	Not available.
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-site	更改 Ranger KMS 的 ranger-kms-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-env	在 Ranger KMS 环境中更改值。	Restarts Ranger KMS Server.
ranger-kms-log4j	更改 Ranger 的 kms-log4j.properties 文件中的值。KMS	Not available.
ranger-kms-db-ca	更改 S3 上 Ranger SQL SSL KMS 连接的 CA 文件值。	Not available.
recordserver-env	更改 EMR RecordServer 环境中的值。	Restarts EMR record server.
recordserver-conf	更改 EMR RecordServer 的 proper EMR RecordServer ties 文件中的值。	Restarts EMR record server.
recordserver-log4j	更改 EMR RecordServer 的 log4j.properties 文件中的值。	Restarts EMR record server.
spark	亚马逊 EMR 精心策划的 Apache Spark 设置。	This property modifies spark-defaults. See actions there.
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-env	更改 Spark 环境中的值。	Restarts Spark history server and Spark thrift server.
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值	Not available.

分类	描述	重新配置操作
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
sqoop-env	更改 Sqoop 的环境中的值。	Not available.
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。	Not available.
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。	Not available.
tez-site	更改 Tez 的 tez-site.xml 文件中的值。	Restarts Oozie.
yarn-env	更改YARN环境中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	更改YARN的 yarn-site.xml 文件中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	更改 Zeppelin 环境中的值。	Restarts Zeppelin.
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。	Restarts Zookeeper server.

分类	描述	重新配置操作
zookeeper-log4j	更改 ZooKeeper 的 log4j.properties 文件中的值。	Restarts Zookeeper server.

## 亚马逊EMR版本 5.32.0

### 5.32.0 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[JupyterEnterpriseGateway](#)、[Jupyter](#) 和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-5.32.0	emr-5.31.1	emr-5.31.0	emr-5.30.2
适用于 Java 的 AWS SDK	1.11.890	1.11.852	1.11.852	1.11.759
Python	2.7、3.7	2.7、3.7	2.7、3.7	2.7、3.7
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-

	emr-5.32.0	emr-5.31.1	emr-5.31.0	emr-5.30.2
Flink	1.11.2	1.11.0	1.11.0	1.10.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13
HCatalog	2.3.7	2.3.7	2.3.7	2.3.6
Hadoop	2.10.1	2.10.0	2.10.0	2.8.5
Hive	2.3.7	2.3.7	2.3.7	2.3.6
Hudi	0.6.0-amzn-0	0.6.0-amzn-0	0.6.0-amzn-0	0.5.2-incubating
Hue	4.8.0	4.7.1	4.7.1	4.6.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	-	-	-
JupyterHub	1.1.0	1.1.0	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.7.0	1.6.0	1.6.0	1.5.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.2.0	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.240.1	0.238.3	0.238.3	0.232
Spark	2.4.7	2.4.6	2.4.6	2.4.5
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7



	emr-5.32.0	emr-5.31.1	emr-5.31.0	emr-5.30.2
TensorFlow	2.3.1	2.1.0	2.1.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 5.32.0 发布说明

以下发行说明包含有关亚马逊 5.32.0 EMR 版本的信息。更改与 5.31.0 有关。

首次发布日期：2021 年 1 月 8 日

### 升级

- 已将 Amazon Glue 连接器升级到 1.14.0
- 已将 Amazon SageMaker Spark 升级 SDK 到 1.4.1 版
- 已升级 AWS SDK for Java 到 1.11.890 版本
- 升级后的 EMR DynamoDB 连接器版本 4.16.0
- 已升级 EMRFS 到 2.45.0 版
- 已将 EMR 日志分析指标升级到 1.18.0 版
- 已将 EMR MetricsAndEventsApiGateway 客户端升级到 1.5.0 版
- 已将 R EMR record Server 升级到 1.8.0 版
- 已 EMR 将 S3 Dist CP 升级到版本 2.17.0
- 已将 S EMR secret Agent 升级到 1.7.0 版
- 已将 Flink 升级到版本 1.11.2
- 已将 Hadoop 升级到版本 2.10.1-amzn-0
- 已将 Hive 升级到版本 2.3.7-amzn-3
- 已将 Hue 升级到版本 4.8.0
- 已将 Mxnet 升级到版本 1.7.0

- 已将 OpenCV 升级到版本 4.4.0
- 已将 Presto 升级到版本 0.240.1-amzn-0
- 已将 Spark 升级到版本 2.4.7-amzn-0
- 已升级 TensorFlow 到 2.3.1 版

## 更改、增强功能和解决的问题

- 此版本旨在修复 Amazon EMR Scaling 无法成功扩展/缩小集群或导致应用程序故障时出现的问题。
- 修复了 Amaz EMR on 集群上守护程序运行状况检查活动（例如收集YARN节点状态和HDFS节点状态）时，对利用率很高的大型集群的扩展请求失败的问题。之所以发生这种情况，是因为集群上的守护程序无法将节点的运行状况数据传送到内部 Ama EMR zon 组件。
- 改进了EMR集群上的守护程序，可在重复使用 IP 地址时正确跟踪节点状态，从而提高扩展操作期间的可靠性。
- [SPARK-29683](#)。修复了集群缩减期间出现任务失败的问题，因为 Spark 假定所有可用节点都被拒绝列出。
- [YARN-9011](#)。修复了集群尝试向上或向下扩展时，由于YARN停用过程中的争用条件而导致任务失败的问题。
- 通过确保 Amaz EMR on 集群上守护程序和 YARN /之间的节点状态始终保持一致，修复了集群扩展期间步骤或任务失败的问题。HDFS
- 修复了启用了 Kerberos 身份验证的 Amazon 集群的EMR集群操作（例如缩减和步骤提交）失败的问题。这是因为 Amaz EMR on 集群上守护程序没有续订 Kerberos 票证，而要与主节点上YARN运行的 HDFS /进行安全通信，则需要续订 Kerberos 票证。
- 较新的Amazon EMR 版本通过降低亚马逊较旧AL2版本的“最大打开文件数”限制来解决这个问题 EMR。亚马逊EMR发布的 5.30.1、5.30.2、5.31.1、5.32.1、5.32.1、6.0.1、6.1.1、6.2.1、5.33.0、6.3.0及更高版本现在包含一个永久修复程序，其中包含一个具有更高的“最大打开文件数”设置的永久修复。
- 升级了组件版本。
- 有关组件版本的列表，请参阅本指南中的[关于 Amazon EMR 版本](#)。

## 新功能

- 从亚马逊 EMR 5.32.0 和 6.5.0 开始，Apache Spark 的动态执行器大小设置默认处于启用状态。要打开或关闭此功能，您可以使用 `spark.yarn.heterogeneousExecutors.enabled` 配置参数。

- 实例元数据服务 (IMDS) V2 支持状态：Amazon EMR 5.23.1、5.27.1 和 5.32 或更高版本的组件用于所有调用。IMDSv2 IMDS对于应用程序代码中的IMDS调用，您可以同时使用IMDSv1和IMDSv2，或者将配置IMDS为仅IMDSv2用于增强安全性。对于其他 5.x EMR 版本，禁用IMDSv1会导致集群启动失败。
- 从 Amazon EMR 5.32.0 开始，您可以启动与 Apache Ranger 原生集成的集群。Apache Ranger 是一个开源框架，可跨 Hadoop 平台启用、监控和管理全面的数据安全。有关更多信息，请参阅 [Apache Ranger](#)。通过原生集成，您可以自带Apache Ranger来在亚马逊上实施精细的数据访问控制。EMR请参阅亚马逊EMR发布指南中的[将亚马逊EMR与 Apache Ranger 集成](#)。
- 亚马逊EMR版本 5.32.0 支持亚马逊EMR。EKS如需详细了解如何开始使用 EMR onEKS，请参阅 [Amaz EMR on 在做什么EKS](#)。
- 亚马逊EMR版本 5.32.0 支持亚马逊 EMR Studio (预览版)。有关 EMR Studio 入门的更多详细信息，请参阅 [Amazon EMR Studio \(预览版\)](#)。
- 限定范围的托管策略：为了与 AWS 最佳实践保持一致，Amazon 引入EMR了 v2 EMR 范围的默认托管策略，以取代即将弃用的策略。参见 [Amazon EMR 托管政策](#)。

## 已知问题

- 对于 Amazon EMR 6.3.0 和 6.2.0 私有子网集群，您无法访问 Ganglia 网页用户界面。您将收到“access denied (403)”错误。其他网络UIs，例如 Spark、Hue JupyterHub、Zeppelin、Livy 和 Tez，都正常运行。公有子网集群上的 Ganglia Web UI 访问也正常工作。要解决该问题，请在具有 `sudo systemctl restart httpd` 的主节点上重新启动 httpd 服务。此问题已在 Amazon EMR 6.4.0 中修复。
- 降低旧版本的“最大打开文件数”限制 AL2 [在新版本中已修复]。亚马逊EMR版本：  
emr-5.30.x、emr-5.31.0、emr-5.32.0、emr-6.0.0、emr-6.1.0 和 emr-6.2.0 基于较旧版本的 Linux ofAmazon 2 ()AL2，当使用默认版本创建亚马逊集群时，“最大打开文件数”的用户限制较低。EMR AMI亚马逊EMR发布的版本为5.30.1、5.30.2、5.31.1、5.32.1、6.0.1、6.1.1、6.2.1、5.33.0、6.3.0 及更高版本，包括具有更高的“最大打开文件数”设置的永久修复。如果使用打开文件数限制较低的发行版，会在提交 Spark 任务时导致“Too many open files”（打开的文件过多）错误。在受影响的版本中，亚马逊EMR默认AMI认“最大打开文件数”的ulimit设置为4096，低于Linux 2中的65536个文件限制。latestAmazon AMISpark 驱动程序和执行程序尝试打开超过 4096 个文件时，“打开的最大文件数”的较低 ulimit 设置会导致 Spark 任务失败。为了解决这个问题，Amazon EMR 提供了一个引导操作 (BA) 脚本，可以在创建集群时调整ulimit设置。

如果您使用的是无法永久修复此问题的旧EMR版 Amazon，则可以使用以下解决方法将实例控制器 ulimit 明确设置为最多 65536 个文件。

## 从命令行显式设置 ulimit

1. 编辑 `/etc/systemd/system/instance-controller.service` , 将以下参数添加到 Service (服务) 部分。

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. 重启 InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

## 使用引导操作 (BA) 设置 ulimit

您还可以在创建集群时使用引导操作 ( BA ) 脚本将实例控制器 ulimit 配置为 65536 个文件。

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

### Important

EMR运行 Amazon Linux 或 Amazon Linux 2 Amazon 机器映像 (AMIs) 的集群使用默认 Amazon Linux 行为 , 不会自动下载和安装需要重启的重要关键内核更新。这与运行默认

Amazon Linux 的其他亚马逊EC2实例的行为相同AMI。如果在 Amazon EMR 版本发布后需要重启的新 Amazon Linux 软件CUDA更新（例如内核和更新）可用，则默认运行的EMR集群实例AMI不会自动下载和安装这些更新。NVIDIA要获取内核更新，您可以[自定义您的亚马逊EMRAMI](#)以[使用最新的亚马逊 Linux AMI](#)。

- 该 GovCloud 区域目前不支持控制台支持创建指定 AWS Ranger 集成选项的安全配置。可以使用来完成安全配置CLI。请参阅 [《Amazon EMR 管理指南》中的创建EMR安全配置](#)。
- 在使用 Amazon EMR 5.31.0 AtRestEncryption 或 5.32.0 的集群上启用或HDFS加密时，Hive 查询会导致以下运行时异常。

```
TaskAttempt 3 failed, info=[Error: Error while running task ( failure ) :
attempt_1604112648850_0001_1_01_000000_3:java.lang.RuntimeException:
java.lang.RuntimeException: Hive Runtime Error while closing
operators: java.io.IOException: java.util.ServiceConfigurationError:
org.apache.hadoop.security.token.TokenIdentifier: Provider
org.apache.hadoop.hbase.security.token.AuthenticationTokenIdentifier not found
```

- 当你使用带有 Hive 分区位置格式的 Spark 来读取 Amazon S3 中的数据，并在亚马逊EMR版本 5.30.0 到 5.36.0 和 6.2.0 到 6.9.0 上运行 Spark 时，你可能会遇到一个问题，导致集群无法正确读取数据。如果您的分区具有以下所有特征，会发生这种情况：
  - 从同一个表扫描两个或多个分区。
  - 至少有一个分区目录路径是至少一个其他分区目录路径的前缀，例如，s3://bucket/table/p=a 是 s3://bucket/table/p=a b 的前缀。
  - 另一个分区目录中前缀之后的第一个字符的值为 UTF -8，小于该/字符 (U+002F)。例如，在 s3://bucket/table/p=a b 中，a 和 b 之间出现的空格字符 (U+0020) 就属于此类。请注意，还有其他 14 个非控制字符：!"#\$%&'()\*+,-。有关更多信息，请参阅 [UTF-8 编码表和 Unicode 字符](#)。

解决方法是在 spark-defaults 分类中将

spark.sql.sources.fastS3PartitionDiscovery.enabled 配置设置为 false。

## 5.32.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 emr 或 aws 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.4.1	亚马逊 SageMaker Spark SDK
emr-ddb	4.16.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.13.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.5.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-notebook-env	1.1.0	适用于 EMR Notebooks ( 可提供 jupyter 企业网关 ) 的 Conda env
emr-s3-dist-cp	2.17.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	1.6.0	EMRS3Select 连接器
emrfs	2.45.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.11.2	Apache Flink 命令行客户端脚本和应用程序。
flink-jobmanager-config	1.11.2	管理 Apache Flink EMR JobManager 节点上的资源。

组件	版本	描述
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.10.1-amzn-0	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.10.1-amzn-0	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.10.1-amzn-0	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.10.1-amzn-0	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	2.10.1-amzn-0	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	2.10.1-amzn-0	HTTPHDFS操作端点。
hadoop-kms-server	2.10.1-amzn-0	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.10.1-amzn-0	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。

组件	版本	描述
hadoop-yarn-nodemanager	2.10.1-amzn-0	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.10.1-amzn-0	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.10.1-amzn-0	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.13	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.13	为一个或多个HBase地区提供服务的服务。
hbase-client	1.4.13	HBase命令行客户端。
hbase-rest-server	1.4.13	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.13	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.3.7-amzn-3	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.7-amzn-3	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.7-amzn-3	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.7-amzn-3	Hive 命令行客户端。
hive-hbase	2.3.7-amzn-3	Hive-hbase 客户端。



组件	版本	描述
hive-metastore-server	2.3.7-amzn-3	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的 SQL 元数据。
hive-server2	2.3.7-amzn-3	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.6.0-amzn-0	增量处理框架，以支持低延迟和高效率的数据管道。
hudi-spark	0.6.0-amzn-0	用于运行 Spark 以及 Hudi 的捆绑库。
hudi-presto	0.6.0-amzn-0	用于运行 Presto 以及 Hudi 的捆绑库。
hue-server	4.8.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	1.1.0	Jupyter notebook 的多用户服务器
livy-server	0.7.0-incubating	REST 用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.7.0	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.68	我的 SQL 数据库服务器。

组件	版本	描述
nvidia-cuda	10.1.243	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.0	Oozie 命令行客户端。
oozie-server	5.2.0	用于接受 Oozie 工作流请求的服务。
opencv	4.4.0	开源计算机视觉库。
phoenix-library	4.14.3--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.14.3--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.240.1-amzn-0	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.240.1-amzn-0	用于执行查询的各个部分的服务。
presto-client	0.240.1-amzn-0	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
pig-client	0.17.0	Pig 命令行客户端。
r	3.4.3	用于统计计算的 R 项目
ranger-kms-server	1.2.0	Apache Ranger 密钥管理系统
spark-client	2.4.7-amzn-0	Spark 命令行客户端。
spark-history-server	2.4.7-amzn-0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。

组件	版本	描述
spark-on-yarn	2.4.7-amzn-0	的内存中执行引擎。YARN
spark-yarn-slave	2.4.7-amzn-0	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	2.3.1	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.2	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.8.2	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.14	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.14	ZooKeeper 命令行客户端。

### 5.32.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

为处于运行状态的集群中的实例组指定配置时，将发生重新配置操作。Amazon EMR 仅对您修改的分类启动重新配置操作。有关更多信息，请参阅 [在正在运行的集群中重新配置实例组](#)。

#### emr-5.32.0 分类

分类	描述	重新配置操作
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。	Restarts the ResourceM anager service.

分类	描述	重新配置操作
container-executor	更改 Hadoop YARN 的 container-executor.cfg 文件中的值。	Not available.
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。	Not available.
core-site	更改 Hadoop 的 core-site.xml 文件中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	更改 docker 相关设置。	Not available.

分类	描述	重新配置操作
emrfs-site	更改EMRFS设置。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	更改 flink-conf.yaml 设置。	Restarts Flink history server.
flink-log4j	更改 Flink log4j.properties 设置。	Restarts Flink history server.
flink-log4j-yarn-session	更改 Flink log4j-yarn-session.properties 设置。	Not available.
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。	Restarts Flink history server.

分类	描述	重新配置操作
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	更改 hadoop ssl 服务器配置	Not available.
hadoop-ssl-client	更改 hadoop ssl 客户端配置	Not available.
hbase	亚马逊EMR精心策划的 Apache 设置。HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.

分类	描述	重新配置操作
hbase-env	更改环境HBase中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	更改 hadoop-metrics2-hbase base.properties 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。	Not available.
hbase-site	更改HBase的 hbase-site.xml 文件中的值。	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	配置HDFS加密区域。	Should not be reconfigured.
hdfs-site	更改 hdfs-site.xml HDFS 中的值。	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	更改环境HCatalog中的值。	Restarts Hive HCatalog Server.
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。	Restarts Hive HCatalog Server.

分类	描述	重新配置操作
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat 中的值。	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。	Restarts Hive WebHCat Server.
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。	Not available.
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。	Not available.
hive-env	更改 Hive 环境中的值。	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	更改 Hive 的 hive-exec-log4j2.properties 文件中的值。	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log4j2.properties 文件中的值。	Not available.
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。	Not available.



分类	描述	重新配置操作
hive-site	更改 Hive 的 hive-site.xml 文件中的值	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值	Not available.
hue-ini	更改 Hue 的 ini 文件中的值	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	更改HTTPFS环境中的值。	Restarts Hadoop Httpfs service.
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。	Restarts Hadoop Httpfs service.
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。	Not available.
hadoop-kms-env	在 Hadoop KMS 环境中更改值。	Restarts Hadoop-KMS service.
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。	Not available.
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	更改 Hudi 环境中的值。	Not available.

分类	描述	重新配置操作
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。	Not available.
jupyter-hub-conf	更改 jupyterhub_config.py 文件 JupyterHubs 中的值。	Not available.
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。	Not available.
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。	Not available.
livy-conf	更改 Livy 的 livy.conf 文件中的值。	Restarts Livy Server.
livy-env	更改 Livy 环境中的值。	Restarts Livy Server.
livy-log4j	更改 Livy log4j.properties 设置。	Restarts Livy Server.
mapred-env	更改 MapReduce 应用程序环境中的值。	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	更改 Oozie 的环境中的值。	Restarts Oozie.
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。	Restarts Oozie.
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。	Restarts Oozie.
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。	Not available.

分类	描述	重新配置操作
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。	Not available.
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。	Restarts Phoenix-QueryServer.
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。	Not available.
pig-env	更改 Pig 环境中的值。	Not available.
pig-properties	更改 Pig 的 pig.properties 文件中的值。	Restarts Oozie.
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。	Not available.
presto-log	更改 Presto 的 log.properties 文件中的值。	Restarts Presto-Server.
presto-config	更改 Presto 的 config.properties 文件中的值。	Restarts Presto-Server.
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。	Not available.
presto-env	更改 Presto 的 presto-env.sh 文件中的值。	Restarts Presto-Server.
presto-node	更改 Presto 的 node.properties 文件中的值。	Not available.
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。	Not available.
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。	Restarts Presto-Server.
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。	Not available.
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。	Not available.
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。	Not available.
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。	Not available.
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。	Not available.
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。	Not available.
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。	Not available.
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。	Not available.
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。	Not available.
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。	Not available.
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。	Not available.

分类	描述	重新配置操作
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。	Not available.
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-site	更改 Ranger KMS 的 ranger-kms-site.xml 文件中的值。	Restarts Ranger KMS Server.
ranger-kms-env	在 Ranger KMS 环境中更改值。	Restarts Ranger KMS Server.
ranger-kms-log4j	更改 Ranger 的 kms-log4j.properties 文件中的值。KMS	Not available.
ranger-kms-db-ca	更改 S3 上 Ranger SQL SSL KMS 连接的 CA 文件值。	Not available.
recordserver-env	更改 EMR RecordServer 环境中的值。	Restarts EMR record server.
recordserver-conf	更改 server.properties EMR RecordServer 文件中的值。	Restarts EMR record server.
recordserver-log4j	更改 EMR RecordServer 的 log4j.properties 文件中的值。	Restarts EMR record server.
spark	亚马逊 EMR 精心策划的 Apache Spark 设置。	This property modifies spark-defaults. See actions there.
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-env	更改 Spark 环境中的值。	Restarts Spark history server and Spark thrift server.
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值	Not available.

分类	描述	重新配置操作
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。	Restarts Spark history server and Spark thrift server.
sqoop-env	更改 Sqoop 的环境中的值。	Not available.
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。	Not available.
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。	Not available.
tez-site	更改 Tez 的 tez-site.xml 文件中的值。	Restarts Oozie.
yarn-env	更改YARN环境中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	更改YARN的 yarn-site.xml 文件中的值。	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	更改 Zeppelin 环境中的值。	Restarts Zeppelin.
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。	Restarts Zookeeper server.

分类	描述	重新配置操作
zookeeper-log4j	更改 ZooKeeper 的 log4j.properties 文件中的值。	Restarts Zookeeper server.

## 亚马逊EMR版本 5.31.1

### 5.31.1 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Mahout](#) 和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-5.31.1	emr-5.31.0	emr-5.30.2	emr-5.30.1
适用于 Java 的 AWS SDK	1.11.852	1.11.852	1.11.759	1.11.759
Python	2.7、3.7	2.7、3.7	2.7、3.7	2.7、3.7
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-

	emr-5.31.1	emr-5.31.0	emr-5.30.2	emr-5.30.1
Flink	1.11.0	1.11.0	1.10.0	1.10.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13
HCatalog	2.3.7	2.3.7	2.3.6	2.3.6
Hadoop	2.10.0	2.10.0	2.8.5	2.8.5
Hive	2.3.7	2.3.7	2.3.6	2.3.6
Hudi	0.6.0-amzn-0	0.6.0-amzn-0	0.5.2-incubating	0.5.2-incubating
Hue	4.7.1	4.7.1	4.6.0	4.6.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.6.0	1.6.0	1.5.1	1.5.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.2.0	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.238.3	0.238.3	0.232	0.232
Spark	2.4.6	2.4.6	2.4.5	2.4.5
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7



	emr-5.31.1	emr-5.31.0	emr-5.30.2	emr-5.30.1
TensorFlow	2.1.0	2.1.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

### 5.31.1 发布说明

此版本旨在修复 Amazon EMR Scaling 无法成功扩展/缩小集群或导致应用程序故障时出现的问题。

#### 更改、增强和解决的问题

- 修复了 Amaz EMR on 集群上守护程序运行状况检查活动 ( 例如收集YARN节点状态和HDFS节点状态 ) 时，对利用率很高的大型集群的扩展请求失败的问题。之所以发生这种情况，是因为集群上的守护程序无法将节点的运行状况数据传送到内部 Ama EMR zon 组件。
- 改进了EMR集群上的守护程序，可在重复使用 IP 地址时正确跟踪节点状态，从而提高扩展操作期间的可靠性。
- [SPARK-29683](#)。修复了集群缩减期间出现任务失败的问题，因为 Spark 假定所有可用节点都被拒绝列出。
- [YARN-9011](#)。修复了集群尝试向上或向下扩展时，由于YARN停用过程中的争用条件而导致任务失败的问题。
- 通过确保 Amaz EMR on 集群上守护程序和 YARN /之间的节点状态始终保持一致，修复了集群扩展期间步骤或任务失败的问题。HDFS
- 修复了启用了 Kerberos 身份验证的 Amazon 集群的EMR集群操作 ( 例如缩减和步骤提交 ) 失败的问题。这是因为 Amaz EMR on 集群上守护程序没有续订 Kerberos 票证，而要与主节点上YARN运行的 HDFS /进行安全通信，则需要续订 Kerberos 票证。
- 较新的Amazon EMR 版本通过降低亚马逊较旧AL2版本的“最大打开文件数”限制来解决这个问题 EMR。亚马逊EMR发布的 5.30.1、5.30.2、5.31.1、5.32.1、5.32.1、6.0.1、6.1.1、6.2.1、5.33.0、6.3.0及更高版本现在包含一个永久修复程序，其中包含一个具有更高的“最大打开文件数”设置的永久修复。

- HTTPS现在，亚马逊 Linux 存储库已默认启用。如果您使用 Amazon S3 VPCE 策略来限制对特定存储桶的访问，则必须将新的 Amazon Linux 存储桶ARNarn:aws:s3:::amazonlinux-2-repos-\$region/\*添加到您的策略中（\$region替换为终端节点所在的区域）。有关更多信息，请在 AWS 讨论论坛中查看此主题。[公告：Amazon Linux 2 现在支持HTTPS在连接软件包存储库时使用。](#)

## 已知问题

- 当你使用带有 Hive 分区位置格式的 Spark 来读取 Amazon S3 中的数据，并在亚马逊EMR版本 5.30.0 到 5.36.0 和 6.2.0 到 6.9.0 上运行 Spark 时，你可能会遇到一个问题，导致集群无法正确读取数据。如果您的分区具有以下所有特征，会发生这种情况：
  - 从同一个表扫描两个或多个分区。
  - 至少有一个分区目录路径是至少一个其他分区目录路径的前缀，例如，s3://bucket/table/p=a 是 s3://bucket/table/p=a b 的前缀。
  - 另一个分区目录中前缀之后的第一个字符的值为 UTF -8，小于该/字符 (U+002F)。例如，在 s3://bucket/table/p=a b 中，a 和 b 之间出现的空格字符 (U+0020) 就属于此类。请注意，还有其他 14 个非控制字符：! "\$ % & ' ( ) \* + , -。有关更多信息，请参阅 [UTF-8 编码表和 Unicode 字符](#)。

解决方法是在 spark-defaults 分类中将

spark.sql.sources.fastS3PartitionDiscovery.enabled 配置设置为 false。

### 5.31.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 emr 或 aws 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.4.0	亚马逊 SageMaker Spark SDK
emr-ddb	4.15.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.13.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.5.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.15.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	1.6.0	EMRS3Select 连接器
emrfs	2.43.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.11.0	Apache Flink 命令行客户端脚本和应用程序。
flink-jobmanager-config	1.11.0	管理 Apache Flink EMR JobManager 节点上的资源。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。

组件	版本	描述
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.10.0-amzn-0.1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.10.0-amzn-0.1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.10.0-amzn-0.1	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.10.0-amzn-0.1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	2.10.0-amzn-0.1	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	2.10.0-amzn-0.1	HTTPHDFS操作端点。
hadoop-kms-server	2.10.0-amzn-0.1	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.10.0-amzn-0.1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.10.0-amzn-0.1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.10.0-amzn-0.1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.10.0-amzn-0.1	用于检索YARN应用程序的当前和历史信息的服务。

组件	版本	描述
hbase-hmaster	1.4.13	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.13	为一个或多个HBase地区提供服务的服务器。
hbase-client	1.4.13	HBase命令行客户端。
hbase-rest-server	1.4.13	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.13	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.3.7-amzn-1	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.7-amzn-1	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.7-amzn-1	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.7-amzn-1	Hive 命令行客户端。
hive-hbase	2.3.7-amzn-1	Hive-hbase 客户端。
hive-metastore-server	2.3.7-amzn-1	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.7-amzn-1	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.6.0-amzn-0	增量处理框架，以支持低延迟和高效率的数据管道。

组件	版本	描述
hudi-spark	0.6.0-amzn-0	用于运行 Spark 以及 Hudi 的捆绑库。
hudi-presto	0.6.0-amzn-0	用于运行 Presto 以及 Hudi 的捆绑库。
hue-server	4.7.1	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	1.1.0	Jupyter notebook 的多用户服务器
livy-server	0.7.0-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.6.0	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.64+	我的SQL数据库服务器。
nvidia-cuda	9.2.88	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.0	Oozie 命令行客户端。
oozie-server	5.2.0	用于接受 Oozie 工作流请求的服务。
opencv	4.3.0	开源计算机视觉库。
phoenix-library	4.14.3--1.4 HBase	服务器和客户端的 phoenix 库

组件	版本	描述
phoenix-query-server	4.14.3--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.238.3-amzn-0	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.238.3-amzn-0	用于执行查询的各个部分的服务。
presto-client	0.238.3-amzn-0	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
pig-client	0.17.0	Pig 命令行客户端。
r	3.4.3	用于统计计算的 R 项目
ranger-kms-server	1.2.0	Apache Ranger 密钥管理系统
spark-client	2.4.6-amzn-0.1	Spark 命令行客户端。
spark-history-server	2.4.6-amzn-0.1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.4.6-amzn-0.1	的内存中执行引擎。YARN
spark-yarn-slave	2.4.6-amzn-0.1	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	2.1.0	TensorFlow 用于高性能数值计算的开源软件库。

组件	版本	描述
tez-on-yarn	0.9.2	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.8.2	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.14	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.14	ZooKeeper 命令行客户端。

### 5.31.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.31.1 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。



分类	描述
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。

分类	描述
hcatalog-webhcat-site	更改 HCatalog WebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。

分类	描述
hudi-env	更改 Hudi 环境中的值。
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 jupyterhub_config.py 文件 JupyterHubs 中的值。
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。

分类	描述
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。

分类	描述
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。
ranger-kms-site	更改 Ran ranger-kms-site KMS ger .xml 文件中的值。
ranger-kms-env	在游侠KMS环境中更改值。
ranger-kms-log4j	更改 Ranger 的 kms-log4j.properties 文件中的值。KMS
ranger-kms-db-ca	更改 S3 上我与 Ranger SQL SSL KMS 连接的 CA 文件值。
recordserver-env	更改EMR RecordServer 环境中的值。
recordserver-conf	更改 server.proper EMR RecordServer ties 文件中的值。
recordserver-log4j	更改EMR RecordServer的 log4j.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。

分类	描述
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oracoop-site	更改 Sqoop OraOop 的 oracoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.31.0

### 5.31.0 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Mahout](#) 和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-5.31.0	emr-5.30.2	emr-5.30.1	emr-5.30.0
适用于 Java 的 AWS SDK	1.11.852	1.11.759	1.11.759	1.11.759
Python	2.7、3.7	2.7、3.7	2.7、3.7	2.7、3.7
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.11.0	1.10.0	1.10.0	1.10.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13
HCatalog	2.3.7	2.3.6	2.3.6	2.3.6
Hadoop	2.10.0	2.8.5	2.8.5	2.8.5
Hive	2.3.7	2.3.6	2.3.6	2.3.6
Hudi	0.6.0-amzn-0	0.5.2-incubating	0.5.2-incubating	0.5.2-incubating
Hue	4.7.1	4.6.0	4.6.0	4.6.0
Iceberg	-	-	-	-

	emr-5.31.0	emr-5.30.2	emr-5.30.1	emr-5.30.0
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.6.0	1.5.1	1.5.1	1.5.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.2.0	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.238.3	0.232	0.232	0.232
Spark	2.4.6	2.4.5	2.4.5	2.4.5
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.1.0	1.14.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 5.31.0 发布说明

以下发行说明包含有关亚马逊 5.31.0 EMR 版本的信息。更改与 5.30.1 有关。

首次发布日期：2020 年 10 月 9 日



上次更新日期：2020 年 10 月 15 日

## 升级

- 已将 Amazon Glue 连接器升级到版本 1.13.0
- 已将 Amazon SageMaker Spark 升级SDK到 1.4.0 版
- 已将 Amazon Kinesis 连接器升级到版本 3.5.9
- 已升级 AWS SDK for Java 到 1.11.852 版本
- 已将 Bigtop-tomcat 升级到版本 8.5.56
- 已EMR将 FS 升级到 2.43.0 版
- 已将EMR MetricsAndEventsApiGateway 客户端升级到 1.4.0 版
- 已EMR将 S3 Dist CP 升级到版本 2.15.0
- 已EMR将 S3 Select 升级到 1.6.0 版
- 已将 Flink 升级到版本 1.11.0
- 已将 Hadoop 升级到版本 2.10.0
- 已将 Hive 升级到版本 2.3.7
- 已将 Hudi 升级到版本 0.6.0
- 已将 Hue 升级到版本 4.7.1
- 已升级 JupyterHub 到 1.1.0 版
- 已将 Mxnet 升级到版本 1.6.0
- 已将 OpenCV 升级到版本 4.3.0
- 已将 Presto 升级到版本 0.238.3
- 已升级 TensorFlow 到 2.1.0 版

## 更改、增强功能和解决的问题

- 此版本旨在修复 Amazon EMR Scaling 无法成功扩展/缩小集群或导致应用程序故障时出现的问题。
- 修复了 Amaz EMR on 集群上守护程序运行状况检查活动（例如收集YARN节点状态和HDFS节点状态）时，对利用率很高的大型集群的扩展请求失败的问题。之所以发生这种情况，是因为集群上的守护程序无法将节点的运行状况数据传送到内部 Ama EMR zon 组件。
- 改进了EMR集群上的守护程序，可在重复使用 IP 地址时正确跟踪节点状态，从而提高扩展操作期间的可靠性。

- [SPARK-29683](#)。修复了集群缩减期间出现任务失败的问题，因为 Spark 假定所有可用节点都被拒绝列出。
- [YARN-9011](#)。修复了集群尝试向上或向下扩展时，由于YARN停用过程中的争用条件而导致任务失败的问题。
- 通过确保 Amaz EMR on 集群上守护程序和 YARN /之间的节点状态始终保持一致，修复了集群扩展期间步骤或任务失败的问题。HDFS
- 修复了启用了 Kerberos 身份验证的 Amazon 集群的EMR集群操作（例如缩减和步骤提交）失败的问题。这是因为 Amaz EMR on 集群上守护程序没有续订 Kerberos 票证，而要与主节点上YARN运行的 HDFS /进行安全通信，则需要续订 Kerberos 票证。
- 较新的Amazon EMR 版本通过降低亚马逊较旧AL2版本的“最大打开文件数”限制来解决这个问题 EMR。亚马逊EMR发布的 5.30.1、5.30.2、5.31.1、5.32.1、5.32.1、6.0.1、6.1.1、6.2.1、5.33.0、6.3.0及更高版本现在包含一个永久修复程序，其中包含一个具有更高的“最大打开文件数”设置的永久修复。
- 亚马逊 5.31.0 及更高EMR版本支持 [Hive 列统计信息](#)。
- 升级了组件版本。
- EMRFS亚马逊 EMR 5.31.0 中的 S3EC V2 Support。在 S3 Java SDK 版本 1.11.837 及更高版本中，引入了加密客户端版本 2 (S3EC V2)，其中包含各种安全增强功能。有关更多信息，请参阅下列内容：
  - S3 博客文章：[更新至 Amazon S3 加密客户端](#)。
  - AWS SDK for Java 开发者指南：[将加密和解密客户端迁移到 V2](#)。
  - EMR管理指南：[Amazon S3 客户端加密](#)。

加密客户端 V1 仍可在中使用，SDK以实现向后兼容。

## 新功能

- 降低旧版本的“最大打开文件数”限制 AL2 [在新版本中已修复]。亚马逊EMR版本：  
emr-5.30.x、emr-5.31.0、emr-5.32.0、emr-6.0.0、emr-6.1.0 和 emr-6.2.0 基于较旧版本的 Linux ofAmazon 2 ()AL2，当使用默认版本创建亚马逊集群时，“最大打开文件数”的用户限制较低。EMR AMI亚马逊EMR发布的版本为5.30.1、5.30.2、5.31.1、5.32.1、6.0.1、6.1.1、6.2.1、5.33.0、6.3.0及更高版本，包括具有更高的“最大打开文件数”设置的永久修复。如果使用打开文件数限制较低的发行版，会在提交 Spark 任务时导致“Too many open files”（打开的文件过多）错误。在受影响的版本中，亚马逊EMR默认AMI的“最大打开文件数”的ulimit设置为4096，低于Linux 2中的65536个文件限制。latestAmazon AMISpark 驱动程序和执行程序尝试打开超过 4096 个文件时，“打开的最大文件

数”的较低 ulimit 设置会导致 Spark 任务失败。为了解决这个问题，Amazon EMR 提供了一个引导操作 (BA) 脚本，可以在创建集群时调整 ulimit 设置。

如果您使用的是无法永久修复此问题的旧 EMR 版 Amazon，则可以使用以下解决方法将实例控制器 ulimit 明确设置为最多 65536 个文件。

从命令行显式设置 ulimit

1. 编辑 `/etc/systemd/system/instance-controller.service`，将以下参数添加到 Service (服务) 部分。

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. 重启 InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

使用引导操作 (BA) 设置 ulimit

您还可以在创建集群时使用引导操作 (BA) 脚本将实例控制器 ulimit 配置为 65536 个文件。

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
```

```
sudo systemctl daemon-reload
```

- 使用 Amazon EMR 5.31.0，您可以启动与 Lake Formation 集成的集群。这种集成为 Glue 数据目录中的数据库和表提供了精细的列级数据筛选。AWS 还支持从企业身份系统对 EMR 笔记本或 Apache Zeppelin 进行联合单点登录。有关更多信息，请参阅《[亚马逊EMR管理指南](#)》[AWS Lake Formation中的亚马逊EMR与集成](#)。

Amazon EMR with Lake Formation 目前在 16 个 AWS 地区上市：美国东部（俄亥俄州和弗吉尼亚北部）、美国西部（加利福尼亚北部和俄勒冈州）、亚太地区（孟买、首尔、新加坡、悉尼和东京）、加拿大（中部）、欧洲（法兰克福、爱尔兰、伦敦、巴黎和斯德哥尔摩）、南美洲（圣保罗）。

## 已知问题

- 具有多个主节点的集群和 Kerberos 身份验证中的已知问题

如果您在 Amazon 5.20.0 及更高 EMR 版本中运行具有多个主节点和 Kerberos 身份验证的集群，则在集群运行一段时间后，集群操作可能会遇到问题，例如缩小规模或步骤提交。具体时间段取决于您定义的 Kerberos 票证有效期。缩减问题会影响您提交的自动缩减和显式缩减请求。其它集群操作也可能受到影响。

解决办法：

- SSH 以 hadoop 用户身份访问具有多个主节点的 EMR 群集的主节点。
- 运行以下命令，为 hadoop 用户续订 Kerberos 票证。

```
kinit -kt <keytab_file> <principal>
```

通常情况下，keytab 文件位于 `/etc/hadoop.keytab`，而 principal 为 `hadoop/<hostname>@<REALM>` 格式。

### Note

此解决方法将在 Kerberos 票证有效期内生效。默认情况下，此持续时间为 10 个小时，但可以通过 Kerberos 设置进行配置。Kerberos 票证过期后，您必须重新运行上述命令。

- 在使用 Amazon EMR 5.31.0 AtRestEncryption 或 5.32.0 的集群上启用或 HDFS 加密时，Hive 查询会导致以下运行时异常。

```
TaskAttempt 3 failed, info=[Error: Error while running task ( failure ) :
attempt_1604112648850_0001_1_01_000000_3:java.lang.RuntimeException:
java.lang.RuntimeException: Hive Runtime Error while closing
operators: java.io.IOException: java.util.ServiceConfigurationError:
org.apache.hadoop.security.token.TokenIdentifier: Provider
org.apache.hadoop.hbase.security.token.AuthenticationTokenIdentifier not found
```

- 当你使用带有 Hive 分区位置格式的 Spark 来读取 Amazon S3 中的数据，并在亚马逊EMR版本 5.30.0 到 5.36.0 和 6.2.0 到 6.9.0 上运行 Spark 时，你可能会遇到一个问题，导致集群无法正确读取数据。如果您的分区具有以下所有特征，会发生这种情况：
  - 从同一个表扫描两个或多个分区。
  - 至少有一个分区目录路径是至少一个其他分区目录路径的前缀，例如，s3://bucket/table/p=a 是 s3://bucket/table/p=a b 的前缀。
  - 另一个分区目录中前缀之后的第一个字符的值为 UTF -8，小于该/字符 (U+002F)。例如，在 s3://bucket/table/p=a b 中，a 和 b 之间出现的空格字符 (U+0020) 就属于此类。请注意，还有其他 14 个非控制字符：!"#\$%&'()\*+,-。有关更多信息，请参阅 [UTF-8 编码表和 Unicode 字符](#)。

解决方法是在 spark-defaults 分类中将 spark.sql.sources.fastS3PartitionDiscovery.enabled 配置设置为 false。

### 5.31.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 emr 或 aws 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.4.0	亚马逊 SageMaker Spark SDK

组件	版本	描述
emr-ddb	4.15.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.13.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.5.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.15.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	1.6.0	EMRS3Select 连接器
emrfs	2.43.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.11.0	Apache Flink 命令行客户端脚本和应用程序。
flink-jobmanager-config	1.11.0	管理 Apache Flink EMR JobManager 节点上的资源。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。

组件	版本	描述
hadoop-client	2.10.0-amzn-0	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.10.0-amzn-0	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.10.0-amzn-0	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.10.0-amzn-0	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	2.10.0-amzn-0	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	2.10.0-amzn-0	HTTPHDFS操作端点。
hadoop-kms-server	2.10.0-amzn-0	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.10.0-amzn-0	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.10.0-amzn-0	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.10.0-amzn-0	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.10.0-amzn-0	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.13	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.13	为一个或多个HBase地区提供服务的服务。

组件	版本	描述
hbase-client	1.4.13	HBase命令行客户端。
hbase-rest-server	1.4.13	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.13	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.3.7-amzn-1	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.7-amzn-1	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.7-amzn-1	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.7-amzn-1	Hive 命令行客户端。
hive-hbase	2.3.7-amzn-1	Hive-hbase 客户端。
hive-metastore-server	2.3.7-amzn-1	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.7-amzn-1	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.6.0-amzn-0	增量处理框架，以支持低延迟和高效率的数据管道。
hudi-spark	0.6.0-amzn-0	用于运行 Spark 以及 Hudi 的捆绑库。
hudi-presto	0.6.0-amzn-0	用于运行 Presto 以及 Hudi 的捆绑库。



组件	版本	描述
hue-server	4.7.1	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	1.1.0	Jupyter notebook 的多用户服务器
livy-server	0.7.0-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.6.0	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.64	我的SQL数据库服务器。
nvidia-cuda	9.2.88	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.0	Oozie 命令行客户端。
oozie-server	5.2.0	用于接受 Oozie 工作流请求的服务。
opencv	4.3.0	开源计算机视觉库。
phoenix-library	4.14.3--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.14.3--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API

组件	版本	描述
presto-coordinator	0.238.3-amzn-0	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.238.3-amzn-0	用于执行查询的各个部分的服务。
presto-client	0.238.3-amzn-0	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
pig-client	0.17.0	Pig 命令行客户端。
r	3.4.3	用于统计计算的 R 项目
ranger-kms-server	1.2.0	Apache Ranger 密钥管理系统
spark-client	2.4.6-amzn-0	Spark 命令行客户端。
spark-history-server	2.4.6-amzn-0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.4.6-amzn-0	的内存中执行引擎。YARN
spark-yarn-slave	2.4.6-amzn-0	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	2.1.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.2	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。

组件	版本	描述
zeppelin-server	0.8.2	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.14	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.14	ZooKeeper 命令行客户端。

### 5.31.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.31.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。

分类	描述
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。

分类	描述
hcatalog-webhcat-site	更改 HCatalog WebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。

分类	描述
hudi-env	更改 Hudi 环境中的值。
jupyter-notebook-conf	更改 Jupyter notebook 的 <code>jupyter_notebook_config.py</code> 文件中的值。
jupyter-hub-conf	更改 <code>jupyterhub_config.py</code> 文件 JupyterHubs 中的值。
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 <code>config.json</code> 文件中的值。
livy-conf	更改 Livy 的 <code>livy.conf</code> 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy <code>log4j.properties</code> 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 <code>mapred-site.xml</code> 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 <code>oozie-log4j.properties</code> 文件中的值。
oozie-site	更改 Oozie 的 <code>oozie-site.xml</code> 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 <code>hadoop-metrics2-hbase.properties</code> 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 <code>hbase-site.xml</code> 文件中的值。
phoenix-log4j	更改 Phoenix 的 <code>log4j.properties</code> 文件中的值。
phoenix-metrics	更改 Phoenix 的 <code>hadoop-metrics2-phoenix.properties</code> 文件中的值。

分类	描述
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。

分类	描述
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。
ranger-kms-site	更改 Ran ranger-kms-site KMS ger .xml 文件中的值。
ranger-kms-env	在游侠KMS环境中更改值。
ranger-kms-log4j	更改 Ranger 的 kms-log4j.properties 文件中的值。KMS
ranger-kms-db-ca	更改 S3 上我与 Ranger SQL SSL KMS 连接的 CA 文件值。
recordserver-env	更改EMR RecordServer 环境中的值。
recordserver-conf	更改 server.proper EMR RecordServer ties 文件中的值。
recordserver-log4j	更改EMR RecordServer的 log4j.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。



分类	描述
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.30.2

### 5.30.2 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Mahout](#)和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-5.30.2	emr-5.30.1	emr-5.30.0	emr-5.29.0
适用于 Java 的 AWS SDK	1.11.759	1.11.759	1.11.759	1.11.682
Python	2.7、3.7	2.7、3.7	2.7、3.7	2.7、3.6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.10.0	1.10.0	1.10.0	1.9.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.10
HCatalog	2.3.6	2.3.6	2.3.6	2.3.6
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.6	2.3.6	2.3.6	2.3.6
Hudi	0.5.2-incubating	0.5.2-incubating	0.5.2-incubating	0.5.0-incubating
Hue	4.6.0	4.6.0	4.6.0	4.4.0
Iceberg	-	-	-	-

	emr-5.30.2	emr-5.30.1	emr-5.30.0	emr-5.29.0
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.1.0	1.0.0
Livy	0.7.0	0.7.0	0.7.0	0.6.0
MXNet	1.5.1	1.5.1	1.5.1	1.5.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.2.0	5.2.0	5.1.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.232	0.232	0.232	0.227
Spark	2.4.5	2.4.5	2.4.5	2.4.4
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.14.0	1.14.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 5.30.2 发布说明

此版本旨在修复 Amazon EMR Scaling 无法成功扩展/缩小集群或导致应用程序故障时出现的问题。

## 更改、增强和解决的问题

- 修复了 Amaz EMR on 集群上守护程序运行状况检查活动（例如收集YARN节点状态和HDFS节点状态）时，对利用率很高的大型集群的扩展请求失败的问题。之所以发生这种情况，是因为集群上的守护程序无法将节点的运行状况数据传送到内部 Ama EMR zon 组件。
- 改进了EMR集群上的守护程序，可在重复使用 IP 地址时正确跟踪节点状态，从而提高扩展操作期间的可靠性。
- [SPARK-29683](#)。修复了集群缩减期间出现任务失败的问题，因为 Spark 假定所有可用节点都被拒绝列出。
- [YARN-9011](#)。修复了集群尝试向上或向下扩展时，由于YARN停用过程中的争用条件而导致任务失败的问题。
- 通过确保 Amaz EMR on 集群上守护程序和 YARN /之间的节点状态始终保持一致，修复了集群扩展期间步骤或任务失败的问题。HDFS
- 修复了启用了 Kerberos 身份验证的 Amazon 集群的EMR集群操作（例如缩减和步骤提交）失败的问题。这是因为 Amaz EMR on 集群上守护程序没有续订 Kerberos 票证，而要与主节点上YARN运行的 HDFS /进行安全通信，则需要续订 Kerberos 票证。
- 较新的Amazon EMR 版本通过降低亚马逊较旧AL2版本的“最大打开文件数”限制来解决这个问题 EMR。亚马逊EMR发布的 5.30.1、5.30.2、5.31.1、5.32.1、5.32.1、6.0.1、6.1.1、6.2.1、5.33.0、6.3.0及更高版本现在包含一个永久修复程序，其中包含一个具有更高的“最大打开文件数”设置的永久修复。
- HTTPS现在，亚马逊 Linux 存储库已默认启用。如果您使用 Amazon S3 VPCE 策略来限制对特定存储桶的访问，则必须将新的 Amazon Linux 存储桶ARNarn:aws:s3:::amazonlinux-2-repos-\$region/\*添加到您的策略中（\$region替换为终端节点所在的区域）。有关更多信息，请在 AWS 讨论论坛中查看此主题。[公告：Amazon Linux 2 现在支持HTTPS在连接软件包存储库时使用。](#)

## 已知问题

- 当你使用带有 Hive 分区位置格式的 Spark 来读取 Amazon S3 中的数据，并在亚马逊EMR版本 5.30.0 到 5.36.0 和 6.2.0 到 6.9.0 上运行 Spark 时，你可能会遇到一个问题，导致集群无法正确读取数据。如果您的分区具有以下所有特征，会发生这种情况：
  - 从同一个表扫描两个或多个分区。
  - 至少有一个分区目录路径是至少一个其他分区目录路径的前缀，例如，s3://bucket/table/p=a 是 s3://bucket/table/p=a b 的前缀。

- 另一个分区目录中前缀之后的第一个字符的值为 UTF -8，小于该/字符 (U+002F)。例如，在 `s3://bucket/table/p=a b` 中，`a` 和 `b` 之间出现的空格字符 (U+0020) 就属于此类。请注意，还有其他 14 个非控制字符：`!"#$%&'()*+,-`。有关更多信息，请参阅 [UTF-8 编码表和 Unicode 字符](#)。

解决方法是在 `spark-defaults` 分类中将 `spark.sql.sources.fastS3PartitionDiscovery.enabled` 配置设置为 `false`。

### 5.30.2 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>aws-sagemaker-spark-sdk</code>	1.3.0	亚马逊 SageMaker Spark SDK
<code>emr-ddb</code>	4.14.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.13.0	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.5.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
<code>emr-s3-dist-cp</code>	<code>emrfs</code>	针对 Amazon S3 优化的分布式复制应用程序。

组件	版本	描述
emr-s3-select	1.5.0	EMRS3Select 连接器
emrfs	2.40.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.10.0	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.8.5-amzn-6.1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.8.5-amzn-6.1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.5-amzn-6.1	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.5-amzn-6.1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	2.8.5-amzn-6.1	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	2.8.5-amzn-6.1	HTTPHDFS操作端点。

组件	版本	描述
hadoop-kms-server	2.8.5-amzn-6.1	基于 Hadoop 的加密密钥管理服务。KeyProvider API
hadoop-mapred	2.8.5-amzn-6.1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.8.5-amzn-6.1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.8.5-amzn-6.1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.5-amzn-6.1	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.13	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.13	为一个或多个HBase地区提供服务的服务。
hbase-client	1.4.13	HBase命令行客户端。
hbase-rest-server	1.4.13	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.13	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	2.3.6-amzn-2	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.6-amzn-2	提供服务HCatalog，为分布式应用程序提供表和存储管理层。

组件	版本	描述
hcatalog-webhcat-server	2.3.6-amzn-2	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.6-amzn-2	Hive 命令行客户端。
hive-hbase	2.3.6-amzn-2	Hive-hbase 客户端。
hive-metastore-server	2.3.6-amzn-2	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.6-amzn-2	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.5.2-incubating	增量处理框架，以支持低延迟和高效率的数据管道。
hudi-presto	0.5.2-incubating	用于运行 Presto 以及 Hudi 的捆绑库。
hue-server	4.6.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	1.1.0	Jupyter notebook 的多用户服务器
livy-server	0.7.0-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.5.1	用于深度学习的灵活的、可扩展且高效的库。



组件	版本	描述
mariadb-server	5.5.64+	我的SQL数据库服务器。
nvidia-cuda	9.2.88	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.0	Oozie 命令行客户端。
oozie-server	5.2.0	用于接受 Oozie 工作流请求的服务。
opencv	3.4.0	开源计算机视觉库。
phoenix-library	4.14.3--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.14.3--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.232	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.232	用于执行查询的各个部分的服务。
presto-client	0.232	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
pig-client	0.17.0	Pig 命令行客户端。
r	3.4.3	用于统计计算的 R 项目
ranger-kms-server	1.2.0	Apache Ranger 密钥管理系统
spark-client	2.4.5-amzn-0.1	Spark 命令行客户端。

组件	版本	描述
spark-history-server	2.4.5-amzn-0.1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.4.5-amzn-0.1	的内存中执行引擎。YARN
spark-yarn-slave	2.4.5-amzn-0.1	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	1.14.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.2	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.8.2	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.14	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.14	ZooKeeper 命令行客户端。

### 5.30.2 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

## emr-5.30.2 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。

分类	描述
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。

分类	描述
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
hudi-env	更改 Hudi 环境中的值。
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 jupyterhub_config.py 文件 JupyterHubs 中的值。
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。

分类	描述
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。

分类	描述
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。
ranger-kms-site	更改 Ranger KMS 的 ranger-kms-site KMS ger .xml 文件中的值。
ranger-kms-env	在游侠KMS环境中更改值。

分类	描述
ranger-kms-log4j	更改 Ranger 的 kms-log4j.properties 文件中的值。KMS
ranger-kms-db-ca	更改 S3 上我与 Ranger SQL SSL KMS 连接的 CA 文件值。
recordserver-env	更改EMR RecordServer 环境中的值。
recordserver-conf	更改 server.proper EMR RecordServer ties 文件中的值。
recordserver-log4j	更改EMR RecordServer的 log4j.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。



分类	描述
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.30.1

### 5.30.1 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Mahout](#) 和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-5.30.1	emr-5.30.0	emr-5.29.0	emr-5.28.1
适用于 Java 的 AWS SDK	1.11.759	1.11.759	1.11.682	1.11.659
Python	2.7、3.7	2.7、3.7	2.7、3.6	2.7、3.6
Scala	2.11.12	2.11.12	2.11.12	2.11.12

	emr-5.30.1	emr-5.30.0	emr-5.29.0	emr-5.28.1
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.10.0	1.10.0	1.9.1	1.9.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.10	1.4.10
HCatalog	2.3.6	2.3.6	2.3.6	2.3.6
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.6	2.3.6	2.3.6	2.3.6
Hudi	0.5.2-incubating	0.5.2-incubating	0.5.0-incubating	0.5.0-incubating
Hue	4.6.0	4.6.0	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.0.0	1.0.0
Livy	0.7.0	0.7.0	0.6.0	0.6.0
MXNet	1.5.1	1.5.1	1.5.1	1.5.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.2.0	5.1.0	5.1.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0

	emr-5.30.1	emr-5.30.0	emr-5.29.0	emr-5.28.1
Presto	0.232	0.232	0.227	0.227
Spark	2.4.5	2.4.5	2.4.4	2.4.4
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.14.0	1.14.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

### 5.30.1 发布说明

以下发行说明包含有关亚马逊 5.30.1 EMR 版本的信息。更改与 5.30.0 有关。

首次发布日期：2020 年 6 月 30 日

上次更新时间：2020 年 8 月 24 日

#### 更改、增强功能和解决的问题

- 较新的 Amazon EMR 版本通过降低亚马逊较旧 AL2 版本的“最大打开文件数”限制来解决这个问题 EMR。亚马逊 EMR 发布的 5.30.1、5.30.2、5.31.1、5.32.1、5.32.1、6.0.1、6.1.1、6.2.1、5.33.0、6.3.0 及更高版本现在包含一个永久修复程序，其中包含一个具有更高的“最大打开文件数”设置的永久修复。
- 修复了实例控制器进程生成无限量进程的问题。
- 修复了以下问题：Hue 无法运行 Hive 查询并显示“database is locked ( 数据库已锁定 )”消息、阻止执行查询的问题。
- 修复了 Spark 问题，允许在 EMR 集群上同时运行更多任务。
- 修复了一个 Jupyter notebook 问题，该问题会导致 Jupyter 服务器中出现“too many files open error ( 打开过多文件错误 )”。

- 修复了集群启动时间的问题。

## 新功能

- Tez 用户界面和YARN时间轴服务器永久应用程序接口适用于亚马逊 6.x EMR 版本和 5.30.1 及EMR 更高版本。通过单击链接访问永久应用程序历史记录，无需通过SSH连接设置 Web 代理，即可快速访问作业历史记录。活动和已终止集群的日志将在应用程序结束后保留 30 天。有关更多信息，请参阅 Amazon EMR 管理指南中的[查看永久应用程序用户界面](#)。
- EMR笔记本执行APIs可以通过脚本或命令行执行EMR笔记本。无需 AWS 控制台即可启动、停止、列出和描述EMR笔记本执行的功能，使您能够以编程方式控制EMR笔记本电脑。借助参数化笔记本单元，您可以将不同的参数值传递给笔记本，而无需为每组新参数值创建笔记本副本。参见[EMR API 操作](#)。有关示例代码，请参阅[以编程方式执行 EMR Notebook 的示例命令](#)。

## 已知问题

- 降低旧版本的“最大打开文件数”限制 AL2 [在新版本中已修复]。亚马逊EMR版本：  
emr-5.30.x、emr-5.31.0、emr-5.32.0、emr-6.0.0、emr-6.1.0 和 emr-6.2.0 基于较旧版本的 Linux ofAmazon 2 ()AL2，当使用默认版本创建亚马逊集群时，“最大打开文件数”的用户限制较低。EMR AMI亚马逊EMR发布的版本为5.30.1、5.30.2、5.31.1、5.32.1、6.0.1、6.1.1、6.2.1、5.33.0、6.3.0 及更高版本，包括具有更高的“最大打开文件数”设置的永久修复。如果使用打开文件数限制较低的发行版，会在提交 Spark 任务时导致“Too many open files”（打开的文件过多）错误。在受影响的版本中，亚马逊EMR默认AMI的“最大打开文件数”的ulimit设置为4096，低于Linux 2中的65536个文件限制。latestAmazon AMISpark 驱动程序和执行程序尝试打开超过 4096 个文件时，“打开的最大文件数”的较低 ulimit 设置会导致 Spark 任务失败。为了解决这个问题，Amazon EMR 提供了一个引导操作 (BA) 脚本，可以在创建集群时调整ulimit设置。

如果您使用的是无法永久修复此问题的旧EMR版 Amazon，则可以使用以下解决方法将实例控制器 ulimit 明确设置为最多 65536 个文件。

### 从命令行显式设置 ulimit

1. 编辑 `/etc/systemd/system/instance-controller.service`，将以下参数添加到 Service (服务) 部分。

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. 重启 InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

## 使用引导操作 (BA) 设置 ulimit

您还可以在创建集群时使用引导操作 ( BA ) 脚本将实例控制器 ulimit 配置为 65536 个文件。

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- EMR笔记本电脑

在 5.30.1 EMR 版本中，允许您在集群主节点上安装内核和其他 Python 库的功能默认处于禁用状态。有关此功能的更多信息，请参阅[在集群主节点上安装内核和 Python 库](#)。

要启动此功能，请执行以下操作：

1. 确保附加到 Notebook 服务角色的权限EMR策略允许执行以下操作：

```
elasticmapreduce:ListSteps
```

有关更多信息，请参阅[EMR笔记本的服务角色](#)。

2. 使用在 AWS CLI 集群上运行设置 EMR Notebook 的步骤，如以下示例所示。Replace ( 替换 ) `us-east-1` 与您的集群所在的区域相同。有关更多信息，请参阅[使用 AWS CLI向集群中添加步骤](#)。

```
aws emr add-steps --cluster-id MyClusterID --steps
  Type=CUSTOM_JAR,Name=EMRNotebooksSetup,ActionOnFailure=CONTINUE,Jar=s3://us-
east-1.elasticmapreduce/libs/script-runner/script-runner.jar,Args=["s3://
  awssupportdatasvcs.com/bootstrap-actions/EMRNotebooksSetup/emr-notebooks-setup.sh"]
```

- 托管扩展

在未安装 Presto 的 5.30.0 和 5.30.1 的集群上进行托管扩展操作可能会导致应用程序故障或导致统一的实例组或实例集处于 ARRESTED 状态，尤其是在缩减操作之后快速执行扩展操作时。

解决方法是，即使您的任务不需要 Presto，也可以在创建包含亚马逊EMR版本 5.30.0 和 5.30.1 的集群时选择 Presto 作为要安装的应用程序。

- 具有多个主节点的集群和 Kerberos 身份验证中的已知问题


如果您在 Amazon 5.20.0 及更高EMR版本中运行具有多个主节点和 Kerberos 身份验证的集群，则在集群运行一段时间后，集群操作可能会遇到问题，例如缩小规模或步骤提交。具体时间段取决于您定义的 Kerberos 票证有效期。缩减问题会影响您提交的自动缩减和显式缩减请求。其它集群操作也可能受到影响。

解决办法：

- SSH以hadoop用户身份访问具有多个主节点的EMR群集的主节点。
- 运行以下命令，为 hadoop 用户续订 Kerberos 票证。

```
kinit -kt <keytab_file> <principal>
```

通常情况下，keytab 文件位于 `/etc/hadoop.keytab`，而 principal 为 `hadoop/<hostname>@<REALM>` 格式。

 Note

此解决方法将在 Kerberos 票证有效期内生效。默认情况下，此持续时间为 10 个小时，但可以通过 Kerberos 设置进行配置。Kerberos 票证过期后，您必须重新运行上述命令。

- 当你使用带有 Hive 分区位置格式的 Spark 来读取 Amazon S3 中的数据，并在亚马逊EMR版本 5.30.0 到 5.36.0 和 6.2.0 到 6.9.0 上运行 Spark 时，你可能会遇到一个问题，导致集群无法正确读取数据。如果您的分区具有以下所有特征，会发生这种情况：
  - 从同一个表扫描两个或多个分区。
  - 至少有一个分区目录路径是至少一个其他分区目录路径的前缀，例如，s3://bucket/table/p=a 是 s3://bucket/table/p=a b 的前缀。
  - 另一个分区目录中前缀之后的第一个字符的值为 UTF -8，小于该/字符 (U+002F)。例如，在 s3://bucket/table/p=a b 中，a 和 b 之间出现的空格字符 (U+0020) 就属于此类。请注意，还有其他 14 个非控制字符：!"#\$%&'()\*+,-。有关更多信息，请参阅 [UTF-8 编码表和 Unicode 字符](#)。

解决方法是在 spark-defaults 分类中将 `spark.sql.sources.fastS3PartitionDiscovery.enabled` 配置设置为 `false`。

### 5.30.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>aws-sagemaker-spark-sdk</code>	1.3.0	亚马逊 SageMaker Spark SDK
<code>emr-ddb</code>	4.14.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.13.0	适用于 Hadoop 生态系统的方便易用的库。

组件	版本	描述
emr-kinesis	3.5.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	emrfs	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	1.5.0	EMRS3Select 连接器
emrfs	2.40.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.10.0	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.8.5-amzn-6	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.8.5-amzn-6	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.5-amzn-6	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.5-amzn-6	HDFS用于跟踪文件名和区块位置的服务。



组件	版本	描述
hadoop-hdfs-journalnode	2.8.5-amzn-6	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httfs-server	2.8.5-amzn-6	HTTPHDFS操作端点。
hadoop-kms-server	2.8.5-amzn-6	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.8.5-amzn-6	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.8.5-amzn-6	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.8.5-amzn-6	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.5-amzn-6	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.13	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.13	为一个或多个HBase地区提供服务的服务。
hbase-client	1.4.13	HBase命令行客户端。
hbase-rest-server	1.4.13	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.13	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.3.6-amzn-2	用于操作 hcatalog-server 的“hcat”命令行客户端。

组件	版本	描述
hcatalog-server	2.3.6-amzn-2	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.6-amzn-2	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.6-amzn-2	Hive 命令行客户端。
hive-hbase	2.3.6-amzn-2	Hive-hbase 客户端。
hive-metastore-server	2.3.6-amzn-2	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.6-amzn-2	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.5.2-incubating	增量处理框架，以支持低延迟和高效率的数据管道。
hudi-presto	0.5.2-incubating	用于运行 Presto 以及 Hudi 的捆绑库。
hue-server	4.6.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	1.1.0	Jupyter notebook 的多用户服务器
livy-server	0.7.0-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器

组件	版本	描述
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.5.1	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.64	我的SQL数据库服务器。
nvidia-cuda	9.2.88	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.0	Oozie 命令行客户端。
oozie-server	5.2.0	用于接受 Oozie 工作流请求的服务。
opencv	3.4.0	开源计算机视觉库。
phoenix-library	4.14.3--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.14.3--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.232	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.232	用于执行查询的各个部分的服务。
presto-client	0.232	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
pig-client	0.17.0	Pig 命令行客户端。
r	3.4.3	用于统计计算的 R 项目

组件	版本	描述
ranger-kms-server	1.2.0	Apache Ranger 密钥管理系统
spark-client	2.4.5-amzn-0	Spark 命令行客户端。
spark-history-server	2.4.5-amzn-0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.4.5-amzn-0	的内存中执行引擎。YARN
spark-yarn-slave	2.4.5-amzn-0	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	1.14.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.2	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.8.2	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.14	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.14	ZooKeeper 命令行客户端。

### 5.30.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

## emr-5.30.1 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。

分类	描述
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。

分类	描述
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
hudi-env	更改 Hudi 环境中的值。
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 jupyterhub_config.py 文件 JupyterHubs 中的值。
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。

分类	描述
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。



分类	描述
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。
ranger-kms-site	更改 Ran ranger-kms-site KMS ger .xml 文件中的值。
ranger-kms-env	在游侠KMS环境中更改值。

分类	描述
ranger-kms-log4j	更改 Ranger 的 kms-log4j.properties 文件中的值。KMS
ranger-kms-db-ca	更改 S3 上我与 Ranger SQL SSL KMS 连接的 CA 文件值。
recordserver-env	更改EMR RecordServer 环境中的值。
recordserver-conf	更改 server.proper EMR RecordServer ties 文件中的值。
recordserver-log4j	更改EMR RecordServer的 log4j.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。

分类	描述
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.30.0

### 5.30.0 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Mahout](#) 和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-5.30.0	emr-5.29.0	emr-5.28.1	emr-5.28.0
适用于 Java 的 AWS SDK	1.11.759	1.11.682	1.11.659	1.11.659
Python	2.7、3.7	2.7、3.6	2.7、3.6	2.7、3.6
Scala	2.11.12	2.11.12	2.11.12	2.11.12

	emr-5.30.0	emr-5.29.0	emr-5.28.1	emr-5.28.0
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.10.0	1.9.1	1.9.0	1.9.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.10	1.4.10	1.4.10
HCatalog	2.3.6	2.3.6	2.3.6	2.3.6
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.6	2.3.6	2.3.6	2.3.6
Hudi	0.5.2-incubating	0.5.0-incubating	0.5.0-incubating	0.5.0-incubating
Hue	4.6.0	4.4.0	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.0.0	1.0.0	1.0.0
Livy	0.7.0	0.6.0	0.6.0	0.6.0
MXNet	1.5.1	1.5.1	1.5.1	1.5.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0

	emr-5.30.0	emr-5.29.0	emr-5.28.1	emr-5.28.0
Presto	0.232	0.227	0.227	0.227
Spark	2.4.5	2.4.4	2.4.4	2.4.4
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.14.0	1.14.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 5.30.0 发布说明

以下发行说明包含 Amazon EMR 版本 5.30.0 的信息。更改与 5.29.0 有关。

首次发布日期：2020 年 5 月 13 日

上次更新日期：2020 年 6 月 25 日

### 升级

- 已升级 AWS SDK for Java 到 1.11.759 版本
- 已将 Amazon SageMaker Spark 升级 SDK 到 1.3.0 版
- 已将 R EMR ecord Server 升级到 1.6.0 版
- 已将 Flink 升级到版本 1.10.0
- 已将 Ganglia 升级到版本 3.7.2
- 已升级 HBase 到 1.4.13 版
- 已将 Hudi 升级到版本 0.5.2-incubating
- 已将 Hue 升级到版本 4.6.0
- 已升级 JupyterHub 到 1.1.0 版

- 已将升级 Livy 到版本 0.7.0-incubating
- 已将 Oozie 升级到版本 5.2.0
- 已将 Presto 升级到版本 0.232
- 已将 Spark 升级到版本 2.4.5
- 升级后的连接器和驱动程序：亚马逊 Glue Connector 1.12.0；亚马逊 Kinesis Connector 3.5.0；DynamoDB 连接器 4.14.0 EMR

## 新功能

- EMR笔记本-与使用 5.30.0 创建的EMR集群一起使用时，EMR笔记本内核在集群上运行。这可以提高笔记本的性能，并允许您安装和自定义内核。您还可以在集群主节点上安装 Python 库。有关更多信息，请参阅《EMR管理指南》中的[安装和使用内核和库](#)。
- 托管扩展 — 在 Amazon 5.30.0 及更高EMR版本中，您可以启用EMR托管扩展，从而根据工作负载自动增加或减少集群中的实例或单元数量。Amazon 会EMR持续评估集群指标，以做出扩展决策，从而优化集群的成本和速度。有关更多信息，请参阅 Amazon EMR 管理指南中的[扩展集群资源](#)。
- 加密存储在 Amazon S3 中的日志文件 — 在 Amazon 5.30.0 及更高EMR版本中，您可以使用 AWS KMS 客户托管密钥加密存储在 Amazon S3 中的日志文件。有关更多信息，请参阅《亚马逊EMR管理指南》中的加密存储在 Amazon [S3 中的日志文件](#)。
- 亚马逊 Linux 2 支持 — 在 5.30.0 及更高EMR版本中，EMR uses Amazon Linux 2 操作系统。新的自定义AMIs ( Amazon 机器映像 ) 必须基于 the Amazon Linux 2 AMI。有关更多信息，请参阅[使用自定义AMI](#)。
- Presto Graceful Auto Scale — 使用 5.30.0 的EMR集群可以设置自动缩放超时时间，让 Presto 任务有时间在节点停用之前完成运行。有关更多信息，请参阅 [使用采用 Graceful Decommission 的 Presto 自动扩展配置](#)。
- 使用新的分配策略选项创建舰队实例 — 5.12.1 及更高EMR版本中提供了新的分配策略选项。它加快了集群预置、提高了 Spot 分配的准确性并减少了竞价型实例中断。需要更新非默认EMR服务角色。请查看[配置实例集](#)。
- sudo systemctl stop 和 sudo systemctl 启动命令 — 在 5.30.0 及更高EMR版本中，use Amazon Linux 2 操作系统使用和命令来重启服务。EMR sudo systemctl stop sudo systemctl start有关更多信息，请参阅[如何在 Amazon 中重启服务EMR？](#)。

## 更改、增强功能和解决的问题

- EMR默认情况下，版本 5.30.0 不安装 Ganglia。您可以在创建集群时明确选择 Ganglia 进行安装。

- Spark 性能优化。
- Presto 性能优化。
- Python 3 是亚马逊 5.30.0 及更高EMR版本的默认版本。
- 用于私有子网中服务访问的默认托管安全组已使用新规则进行更新。如果使用自定义安全组进行服务访问，则必须包含与默认托管安全组相同的规则。有关更多信息，请参阅 [EMRAmazon 服务访问托管安全组 \(私有子网\)](#)。如果您为 Amazon 使用自定义服务角色EMR，则必须向授予权限，`ec2:describeSecurityGroups`这样EMR才能验证安全组的创建是否正确。如果您使用 `EMR_DefaultRole`，则此权限已包含在默认托管式策略中。

## 已知问题

- 降低旧版本的“最大打开文件数”限制 AL2 [在新版本中已修复]。亚马逊EMR版本：`emr-5.30.x`、`emr-5.31.0`、`emr-5.32.0`、`emr-6.0.0`、`emr-6.1.0` 和 `emr-6.2.0` 基于较旧版本的 Linux ofAmazon 2 ()AL2，当使用默认版本创建亚马逊集群时，“最大打开文件数”的用户限制较低。EMR AMI亚马逊EMR发布的版本为5.30.1、5.30.2、5.31.1、5.32.1、6.0.1、6.1.1、6.2.1、5.33.0、6.3.0 及更高版本，包括具有更高的“最大打开文件数”设置的永久修复。如果使用打开文件数限制较低的发行版，会在提交 Spark 任务时导致“Too many open files”（打开的文件过多）错误。在受影响的版本中，亚马逊EMR默认AMI的“最大打开文件数”的ulimit设置为4096，低于Linux 2中的65536个文件限制。latestAmazon AMISpark 驱动程序和执行程序尝试打开超过 4096 个文件时，“打开的最大文件数”的较低 ulimit 设置会导致 Spark 任务失败。为了解决这个问题，Amazon EMR 提供了一个引导操作 (BA) 脚本，可以在创建集群时调整ulimit设置。

如果您使用的是无法永久修复此问题的旧EMR版 Amazon，则可以使用以下解决方法将实例控制器 ulimit 明确设置为最多 65536 个文件。

### 从命令行显式设置 ulimit

1. 编辑 `/etc/systemd/system/instance-controller.service`，将以下参数添加到 Service (服务) 部分。

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. 重启 InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

## 使用引导操作 (BA) 设置 ulimit

您还可以在创建集群时使用引导操作 ( BA ) 脚本将实例控制器 ulimit 配置为 65536 个文件。

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- 托管扩展

在未安装 Presto 的 5.30.0 和 5.30.1 的集群上进行托管扩展操作可能会导致应用程序故障或导致统一的实例组或实例集处于 ARRESTED 状态，尤其是在缩减操作之后快速执行扩展操作时。

解决方法是，即使您的任务不需要 Presto，也可以在创建包含亚马逊EMR版本 5.30.0 和 5.30.1 的集群时选择 Presto 作为要安装的应用程序。

- 具有多个主节点的集群和 Kerberos 身份验证中的已知问题

如果您在 Amazon 5.20.0 及更高EMR版本中运行具有多个主节点和 Kerberos 身份验证的集群，则在集群运行一段时间后，集群操作可能会遇到问题，例如缩小规模或步骤提交。具体时间段取决于您定义的 Kerberos 票证有效期。缩减问题会影响您提交的自动缩减和显式缩减请求。其它集群操作也可能受到影响。

解决办法：

- SSH以hadoop用户身份访问具有多个主节点的EMR群集的主节点。
- 运行以下命令，为 hadoop 用户续订 Kerberos 票证。



```
kinit -kt <keytab_file> <principal>
```

通常情况下，keytab 文件位于 `/etc/hadoop.keytab`，而 principal 为 `hadoop/<hostname>@<REALM>` 格式。

#### Note

此解决方法将在 Kerberos 票证有效期内生效。默认情况下，此持续时间为 10 个小时，但可以通过 Kerberos 设置进行配置。Kerberos 票证过期后，您必须重新运行上述命令。

- Hue 4.6.0 的默认数据库引擎是 SQLite，当您尝试将 Hue 与外部数据库配合使用时，这会导致出现问题。若要解决此问题，请在您的 `hue-ini` 配置分类中将 `engine` 设置为 `mysql`。此问题已在亚马逊 EMR 版本 5.30.1 中修复。
- 当你使用带有 Hive 分区位置格式的 Spark 来读取 Amazon S3 中的数据，并在亚马逊 EMR 版本 5.30.0 到 5.36.0 和 6.2.0 到 6.9.0 上运行 Spark 时，你可能会遇到一个问题，导致集群无法正确读取数据。如果您的分区具有以下所有特征，会发生这种情况：
  - 从同一个表扫描两个或多个分区。
  - 至少有一个分区目录路径是至少一个其他分区目录路径的前缀，例如，`s3://bucket/table/p=a` 是 `s3://bucket/table/p=a b` 的前缀。
  - 另一个分区目录中前缀之后的第一个字符的值为 UTF-8，小于该/字符 (U+002F)。例如，在 `s3://bucket/table/p=a b` 中，`a` 和 `b` 之间出现的空格字符 (U+0020) 就属于此类。请注意，还有其他 14 个非控制字符：`!"#$%&'()*+,-`。有关更多信息，请参阅 [UTF-8 编码表和 Unicode 字符](#)。

解决方法是在 `spark-defaults` 分类中将 `spark.sql.sources.fastS3PartitionDiscovery.enabled` 配置设置为 `false`。

## 5.30.0 组件版本

下面列出了 Amazon 随此版本 EMR 安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的 EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新 EMR 版本中的大数据应用程序包通常是社区中的最新版本。我们会 EMR 尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本 EMR 不同。这些组件具有以下形式的 `CommunityVersion-amzn-EmrVersion` 的发行版标注。`EmrVersion` 从 0 开始。例如，如果 `myapp-component` 以 2.2

版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.3.0	亚马逊 SageMaker Spark SDK
emr-ddb	4.14.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.13.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.5.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-notebook-env	1.0.0	适用于 emr notebook 的 Conda env
emr-s3-dist-cp	emrfs	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	1.5.0	EMRS3Select 连接器
emrfs	2.40.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.10.0	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。

组件	版本	描述
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.8.5-amzn-6	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.8.5-amzn-6	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.5-amzn-6	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.5-amzn-6	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	2.8.5-amzn-6	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	2.8.5-amzn-6	HTTPHDFS操作端点。
hadoop-kms-server	2.8.5-amzn-6	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.8.5-amzn-6	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.8.5-amzn-6	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.8.5-amzn-6	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.5-amzn-6	用于检索YARN应用程序的当前和历史信息的服务。

组件	版本	描述
hbase-hmaster	1.4.13	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.13	为一个或多个HBase地区提供服务的服务器。
hbase-client	1.4.13	HBase命令行客户端。
hbase-rest-server	1.4.13	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.13	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.3.6-amzn-2	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.6-amzn-2	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.6-amzn-2	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.6-amzn-2	Hive 命令行客户端。
hive-hbase	2.3.6-amzn-2	Hive-hbase 客户端。
hive-metastore-server	2.3.6-amzn-2	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.6-amzn-2	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.5.2-incubating	增量处理框架，以支持低延迟和高效率的数据管道。

组件	版本	描述
hudi-presto	0.5.2-incubating	用于运行 Presto 以及 Hudi 的捆绑库。
hue-server	4.6.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	1.1.0	Jupyter notebook 的多用户服务器
livy-server	0.7.0-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.5.1	用于深度学习的灵活的、可扩展且高效的库。
mariadb-server	5.5.64	我的SQL数据库服务器。
nvidia-cuda	9.2.88	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.2.0	Oozie 命令行客户端。
oozie-server	5.2.0	用于接受 Oozie 工作流请求的服务。
opencv	3.4.0	开源计算机视觉库。
phoenix-library	4.14.3--1.4 HBase	服务器和客户端的 phoenix 库

组件	版本	描述
phoenix-query-server	4.14.3--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.232	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.232	用于执行查询的各个部分的服务。
presto-client	0.232	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
pig-client	0.17.0	Pig 命令行客户端。
r	3.4.3	用于统计计算的 R 项目
ranger-kms-server	1.2.0	Apache Ranger 密钥管理系统
spark-client	2.4.5-amzn-0	Spark 命令行客户端。
spark-history-server	2.4.5-amzn-0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.4.5-amzn-0	的内存中执行引擎。YARN
spark-yarn-slave	2.4.5-amzn-0	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	1.14.0	TensorFlow 用于高性能数值计算的开源软件库。

组件	版本	描述
tez-on-yarn	0.9.2	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.8.2	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.14	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.14	ZooKeeper 命令行客户端。

### 5.30.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.30.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。

分类	描述
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。



分类	描述
hcatalog-webhcat-site	更改 HCatalog WebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。

分类	描述
hudi-env	更改 Hudi 环境中的值。
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 jupyterhub_config.py 文件 JupyterHubs 中的值。
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。

分类	描述
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。

分类	描述
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。
ranger-kms-site	更改 Ran ranger-kms-site KMS ger .xml 文件中的值。
ranger-kms-env	在游侠KMS环境中更改值。
ranger-kms-log4j	更改 Ranger 的 kms-log4j.properties 文件中的值。KMS
ranger-kms-db-ca	更改 S3 上我与 Ranger SQL SSL KMS 连接的 CA 文件值。
recordserver-env	更改EMR RecordServer 环境中的值。
recordserver-conf	更改 server.proper EMR RecordServer ties 文件中的值。
recordserver-log4j	更改EMR RecordServer的 log4j.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。

分类	描述
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.29.0

### 5.29.0 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Mahout](#) 和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-5.29.0	emr-5.28.1	emr-5.28.0	emr-5.27.1
适用于 Java 的 AWS SDK	1.11.682	1.11.659	1.11.659	1.11.615
Python	2.7、3.6	2.7、3.6	2.7、3.6	2.7、3.6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.9.1	1.9.0	1.9.0	1.8.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.10	1.4.10	1.4.10	1.4.10
HCatalog	2.3.6	2.3.6	2.3.6	2.3.5
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.6	2.3.6	2.3.6	2.3.5
Hudi	0.5.0-incubating	0.5.0-incubating	0.5.0-incubating	-
Hue	4.4.0	4.4.0	4.4.0	4.4.0
Iceberg	-	-	-	-

	emr-5.29.0	emr-5.28.1	emr-5.28.0	emr-5.27.1
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.0.0	1.0.0	1.0.0	1.0.0
Livy	0.6.0	0.6.0	0.6.0	0.6.0
MXNet	1.5.1	1.5.1	1.5.1	1.4.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.227	0.227	0.227	0.224
Spark	2.4.4	2.4.4	2.4.4	2.4.4
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.14.0	1.14.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.2	0.8.1
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 5.29.0 发布说明

以下发行说明包含有关亚马逊 5.29. EMR 0 版本的信息。更改与 5.28.1 有关。

首次发布日期：2020 年 1 月 17 日

## 升级

- 已升级 AWS SDK for Java 到 1.11.682 版本
- 已将 Hive 升级到版本 2.3.6
- 已将 Flink 升级到版本 1.9.1
- 已将 EMRFS 升级到版本 2.38.0
- 已将 EMR DynamoDB 连接器升级到版本 4.13.0

## 更改、增强功能和解决的问题

- Spark
  - Spark 性能优化。
- EMRFS
  - 将管理指南更新为 `emrfs-site.xml` 默认设置以实现了一致视图。

## 已知问题

- 具有多个主节点的集群和 Kerberos 身份验证中的已知问题

如果您在 Amazon 5.20.0 及更高EMR版本中运行具有多个主节点和 Kerberos 身份验证的集群，则在集群运行一段时间后，集群操作可能会遇到问题，例如缩小规模或步骤提交。具体时间段取决于您定义的 Kerberos 票证有效期。缩减问题会影响您提交的自动缩减和显式缩减请求。其它集群操作也可能会受到影响。

### 解决办法：

- SSH以hadoop用户身份访问具有多个主节点的EMR群集的主节点。
- 运行以下命令，为 hadoop 用户续订 Kerberos 票证。

```
kinit -kt <keytab_file> <principal>
```

通常情况下，keytab 文件位于 `/etc/hadoop.keytab`，而 principal 为 `hadoop/<hostname>@<REALM>` 格式。



**Note**

此解决方法将在 Kerberos 票证有效期内生效。默认情况下，此持续时间为 10 个小时，但可以通过 Kerberos 设置进行配置。Kerberos 票证过期后，您必须重新运行上述命令。

## 5.29.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>aws-sagemaker-spark-sdk</code>	1.2.6	亚马逊 SageMaker Spark SDK
<code>emr-ddb</code>	4.13.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.12.0	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
<code>emr-s3-dist-cp</code>	2.13.0	针对 Amazon S3 优化的分布式复制应用程序。
<code>emr-s3-select</code>	1.4.0	EMRS3Select 连接器

组件	版本	描述
emrfs	2.38.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.9.1	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.8.5-amzn-5	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.8.5-amzn-5	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.5-amzn-5	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.5-amzn-5	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	2.8.5-amzn-5	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	2.8.5-amzn-5	HTTPHDFS操作端点。
hadoop-kms-server	2.8.5-amzn-5	基于 Hadoop 的加密密钥管理服务。 KeyProvider API

组件	版本	描述
hadoop-mapred	2.8.5-amzn-5	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.8.5-amzn-5	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.8.5-amzn-5	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.5-amzn-5	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.10	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.10	为一个或多个HBase地区提供服务的服务。
hbase-client	1.4.10	HBase命令行客户端。
hbase-rest-server	1.4.10	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.10	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.3.6-amzn-1	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.6-amzn-1	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.6-amzn-1	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.6-amzn-1	Hive 命令行客户端。

组件	版本	描述
hive-hbase	2.3.6-amzn-1	Hive-hbase 客户端。
hive-metastore-server	2.3.6-amzn-1	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的 SQL 元数据。
hive-server2	2.3.6-amzn-1	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.5.0-incubating	增量处理框架，以支持低延迟和高效率的数据管道。
hudi-presto	0.5.0-incubating	用于运行 Presto 以及 Hudi 的捆绑库。
hue-server	4.4.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	1.0.0	Jupyter notebook 的多用户服务器
livy-server	0.6.0-incubating	REST 用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.5.1	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的 SQL 数据库服务器。
nvidia-cuda	9.2.88	Nvidia 驱动程序和 Cuda 工具包

组件	版本	描述
oozie-client	5.1.0	Oozie 命令行客户端。
oozie-server	5.1.0	用于接受 Oozie 工作流请求的服务。
opencv	3.4.0	开源计算机视觉库。
phoenix-library	4.14.3--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.14.3--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.227	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.227	用于执行查询的各个部分的服务。
presto-client	0.227	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
pig-client	0.17.0	Pig 命令行客户端。
r	3.4.1	用于统计计算的 R 项目
spark-client	2.4.4	Spark 命令行客户端。
spark-history-server	2.4.4	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.4.4	的内存中执行引擎。YARN
spark-yarn-slave	2.4.4	YARN从属服务器需要的 Apache Spark 库。

组件	版本	描述
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	1.14.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.2	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.8.2	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.14	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.14	ZooKeeper 命令行客户端。

## 5.29.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

### emr-5.29.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。

分类	描述
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。

分类	描述
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。



分类	描述
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 jupyterhub_config.py 文件 JupyterHubs 中的值。
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。

分类	描述
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。

分类	描述
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。
ranger-kms-site	更改 Ran ranger-kms-site KMS ger .xml 文件中的值。
ranger-kms-env	在游侠KMS环境中更改值。
ranger-kms-log4j	更改 Ranger 的 kms-log4j.properties 文件中的值。KMS
ranger-kms-db-ca	更改 S3 上我与 Ranger SQL SSL KMS 连接的 CA 文件值。
recordserver-env	更改EMR RecordServer 环境中的值。
recordserver-conf	更改 server.proper EMR RecordServer ties 文件中的值。
recordserver-log4j	更改EMR RecordServer的 log4j.properties 文件中的值。

分类	描述
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.28.1

### 5.28.1 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Mahout](#)和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-5.28.1	emr-5.28.0	emr-5.27.1	emr-5.27.0
适用于 Java 的 AWS SDK	1.11.659	1.11.659	1.11.615	1.11.615
Python	2.7、3.6	2.7、3.6	2.7、3.6	2.7、3.6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.9.0	1.9.0	1.8.1	1.8.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.10	1.4.10	1.4.10	1.4.10

	emr-5.28.1	emr-5.28.0	emr-5.27.1	emr-5.27.0
HCatalog	2.3.6	2.3.6	2.3.5	2.3.5
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.6	2.3.6	2.3.5	2.3.5
Hudi	0.5.0-incubating	0.5.0-incubating	-	-
Hue	4.4.0	4.4.0	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.0.0	1.0.0	1.0.0	1.0.0
Livy	0.6.0	0.6.0	0.6.0	0.6.0
MXNet	1.5.1	1.5.1	1.4.0	1.4.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.3	4.14.3	4.14.2	4.14.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.227	0.227	0.224	0.224
Spark	2.4.4	2.4.4	2.4.4	2.4.4
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.14.0	1.14.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2

	emr-5.28.1	emr-5.28.0	emr-5.27.1	emr-5.27.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.1	0.8.1
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 5.28.1 发布说明

以下发行说明包含有关亚马逊 5.28.1 EMR 版本的信息。更改与 5.28.0 有关。

首次发布日期：2020 年 1 月 10 日

更改、增强功能和解决的问题

- Spark
  - 修复了 Spark 兼容性问题。
- CloudWatch 指标
  - 修复了在具有多个主节点的EMR集群上发布亚马逊 CloudWatch 指标的问题。
- 已禁用日志消息
  - 已禁用假日志消息“...using old version (<4.5.8) of Apache http client” (使用低于版本 4.5.8 的 Apache http 客户端)。

已知问题

- 具有多个主节点的集群和 Kerberos 身份验证中的已知问题

如果您在 Amazon 5.20.0 及更高EMR版本中运行具有多个主节点和 Kerberos 身份验证的集群，则在集群运行一段时间后，集群操作可能会遇到问题，例如缩小规模或步骤提交。具体时间段取决于您定义的 Kerberos 票证有效期。缩减问题会影响您提交的自动缩减和显式缩减请求。其它集群操作也可能受到影响。

解决办法：

- SSH以hadoop用户身份访问具有多个主节点的EMR群集的主节点。
- 运行以下命令，为 hadoop 用户续订 Kerberos 票证。

```
kinit -kt <keytab_file> <principal>
```

通常情况下，keytab 文件位于 `/etc/hadoop.keytab`，而 principal 为 `hadoop/<hostname>@<REALM>` 格式。

#### Note

此解决方法将在 Kerberos 票证有效期内生效。默认情况下，此持续时间为 10 个小时，但可以通过 Kerberos 设置进行配置。Kerberos 票证过期后，您必须重新运行上述命令。

### 5.28.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>aws-sagemaker-spark-sdk</code>	1.2.6	亚马逊 SageMaker Spark SDK
<code>emr-ddb</code>	4.12.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.11.0	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。



组件	版本	描述
emr-s3-dist-cp	2.13.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	1.3.0	EMRS3Select 连接器
emrfs	2.37.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.9.0	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.8.5-amzn-5	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.8.5-amzn-5	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.5-amzn-5	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.5-amzn-5	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	2.8.5-amzn-5	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。

组件	版本	描述
hadoop-httfs-server	2.8.5-amzn-5	HTTPHDFS操作端点。
hadoop-kms-server	2.8.5-amzn-5	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.8.5-amzn-5	MapReduce 用于运行 MapReduce 应用程序的执行引擎。
hadoop-yarn-nodemanager	2.8.5-amzn-5	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.8.5-amzn-5	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.5-amzn-5	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.10	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.10	为一个或多个HBase地区提供服务的服务。
hbase-client	1.4.10	HBase命令行客户端。
hbase-rest-server	1.4.10	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.10	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.3.6-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.6-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。

组件	版本	描述
hcatalog-webhcat-server	2.3.6-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.6-amzn-0	Hive 命令行客户端。
hive-hbase	2.3.6-amzn-0	Hive-hbase 客户端。
hive-metastore-server	2.3.6-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.6-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.5.0-incubating	增量处理框架，以支持低延迟和高效率的数据管道。
hudi-presto	0.5.0-incubating	用于运行 Presto 以及 Hudi 的捆绑库。
hue-server	4.4.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	1.0.0	Jupyter notebook 的多用户服务器
livy-server	0.6.0-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.5.1	用于深度学习的灵活的、可扩展且高效的库。

组件	版本	描述
mysql-server	5.5.54+	我的SQL数据库服务器。
nvidia-cuda	9.2.88	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.1.0	Oozie 命令行客户端。
oozie-server	5.1.0	用于接受 Oozie 工作流请求的服务。
opencv	3.4.0	开源计算机视觉库。
phoenix-library	4.14.3--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.14.3--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.227	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.227	用于执行查询的各个部分的服务。
presto-client	0.227	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
pig-client	0.17.0	Pig 命令行客户端。
r	3.4.1	用于统计计算的 R 项目
spark-client	2.4.4	Spark 命令行客户端。
spark-history-server	2.4.4	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。

组件	版本	描述
spark-on-yarn	2.4.4	的内存中执行引擎。YARN
spark-yarn-slave	2.4.4	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	1.14.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.2	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.8.2	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.14	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.14	ZooKeeper 命令行客户端。

### 5.28.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.28.1 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。

分类	描述
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。

分类	描述
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。

分类	描述
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 jupyterhub_config.py 文件 JupyterHubs 中的值。
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。



分类	描述
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。

分类	描述
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。
ranger-kms-site	更改 Ranger KMS 的 ranger-kms-site KMS ger .xml 文件中的值。
ranger-kms-env	在游侠KMS环境中更改值。
ranger-kms-log4j	更改 Ranger 的 kms-log4j.properties 文件中的值。KMS
ranger-kms-db-ca	更改 S3 上我与 Ranger SQL SSL KMS 连接的 CA 文件值。
recordserver-env	更改EMR RecordServer 环境中的值。

分类	描述
recordserver-conf	更改 server.proper EMR RecordServer ties 文件中的值。
recordserver-log4j	更改EMR RecordServer的 log4j.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.28.0

### 5.28.0 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hudi](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Mahout](#)和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-5.28.0	emr-5.27.1	emr-5.27.0	emr-5.26.0
适用于 Java 的 AWS SDK	1.11.659	1.11.615	1.11.615	1.11.595
Python	2.7、3.6	2.7、3.6	2.7、3.6	2.7、3.6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.9.0	1.8.1	1.8.1	1.8.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.10	1.4.10	1.4.10	1.4.10

	emr-5.28.0	emr-5.27.1	emr-5.27.0	emr-5.26.0
HCatalog	2.3.6	2.3.5	2.3.5	2.3.5
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.6	2.3.5	2.3.5	2.3.5
Hudi	0.5.0-incubating	-	-	-
Hue	4.4.0	4.4.0	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.0.0	1.0.0	1.0.0	0.9.6
Livy	0.6.0	0.6.0	0.6.0	0.6.0
MXNet	1.5.1	1.4.0	1.4.0	1.4.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.3	4.14.2	4.14.2	4.14.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.227	0.224	0.224	0.220
Spark	2.4.4	2.4.4	2.4.4	2.4.3
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.14.0	1.14.0	1.14.0	1.13.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2

	emr-5.28.0	emr-5.27.1	emr-5.27.0	emr-5.26.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.1	0.8.1	0.8.1
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 5.28.0 发布说明

以下发行说明包含有关亚马逊 5.28. EMR 0 版本的信息。更改与 5.27.0 有关。

首次发布日期：2019 年 11 月 12 日

### 升级

- 已将 Flink 升级到版本 1.9.0
- 已将 Hive 升级到版本 2.3.6
- 已升级MXNet到 1.5.1 版
- 已将 Phoenix 升级到版本 4.14.3
- 已将 Presto 升级到版本 0.227
- 已将 Zeppelin 升级到版本 0.8.2

### 新功能

- [Apache Hudi](#) 现在EMR可供亚马逊在你创建集群时安装。有关更多信息，请参阅 [Hudi](#)。
- ( 2019 年 11 月 25 日 ) 您现在可以选择并行运行多个步骤以提高集群利用率并节省成本。您还可以取消待处理和正在运行的步骤。有关更多信息，请参阅[使用 AWS CLI 和控制台处理步骤](#)。
- ( 2019 年 12 月 3 日 ) 您现在可以在上创建和运行EMR集群 AWS Outposts。AWS Outposts 在本地设施中启用原生 AWS 服务、基础设施和运营模式。在 AWS Outposts 环境中，您可以使用与 AWS 云端相同的 AWS APIs工具和基础架构。有关更多信息，请参阅[上的EMR集群 AWS Outposts](#)。
- ( 2020 年 3 月 11 日 ) 从 Amazon EMR 版本 5.28.0 开始，您可以在本地区域子网上创建和运行亚马逊EMR集群，作为支持 AWS 本地区域的 AWS 区域的逻辑扩展。本地区域使得 Amazon EMR 功

能和部分 AWS 服务（例如计算和存储服务）可以位于离用户更远的地方，从而为本地运行的应用程序提供极低的延迟访问。有关可用的 Local Zones 列表，请参阅 [AWS Local Zones](#)。有关访问可用 AWS 本地区域的信息，请参阅 [区域、可用区和本地区域](#)。

Local Zones 目前不支持亚马逊EMR笔记本电脑，也不支持EMR使用接口VPC终端节点 (AWS PrivateLink) 直接连接到亚马逊。

## 更改、增强功能和解决的问题

- 扩展了对高可用性集群的应用程序支持
  - 有关更多信息，请参阅 Amazon EMR 管理指南 [中的具有多个主节点的EMR集群中支持的应用程序](#)。
- Spark
  - 性能优化
- Hive
  - 性能优化
- Presto
  - 性能优化

## 已知问题

- 具有多个主节点的集群和 Kerberos 身份验证中的已知问题

如果您在 Amazon 5.20.0 及更高EMR版本中运行具有多个主节点和 Kerberos 身份验证的集群，则在集群运行一段时间后，集群操作可能会遇到问题，例如缩小规模或步骤提交。具体时间段取决于您定义的 Kerberos 票证有效期。缩减问题会影响您提交的自动缩减和显式缩减请求。其它集群操作也可能受到影响。

### 解决办法：

- SSH以hadoop用户身份访问具有多个主节点的EMR群集的主节点。
- 运行以下命令，为 hadoop 用户续订 Kerberos 票证。

```
kinit -kt <keytab_file> <principal>
```

通常情况下，keytab 文件位于 `/etc/hadoop.keytab`，而 principal 为 `hadoop/<hostname>@<REALM>` 格式。

**Note**

此解决方法将在 Kerberos 票证有效期内生效。默认情况下，此持续时间为 10 个小时，但可以通过 Kerberos 设置进行配置。Kerberos 票证过期后，您必须重新运行上述命令。

## 5.28.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>aws-sagemaker-spark-sdk</code>	1.2.6	亚马逊 SageMaker Spark SDK
<code>emr-ddb</code>	4.12.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.11.0	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
<code>emr-s3-dist-cp</code>	2.13.0	针对 Amazon S3 优化的分布式复制应用程序。
<code>emr-s3-select</code>	1.3.0	EMRS3Select 连接器



组件	版本	描述
emrfs	2.37.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.9.0	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.8.5-amzn-5	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.8.5-amzn-5	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.5-amzn-5	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.5-amzn-5	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	2.8.5-amzn-5	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	2.8.5-amzn-5	HTTPHDFS操作端点。
hadoop-kms-server	2.8.5-amzn-5	基于 Hadoop 的加密密钥管理服务。 KeyProvider API

组件	版本	描述
hadoop-mapred	2.8.5-amzn-5	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.8.5-amzn-5	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.8.5-amzn-5	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.5-amzn-5	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.10	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.10	为一个或多个HBase地区提供服务的服务。
hbase-client	1.4.10	HBase命令行客户端。
hbase-rest-server	1.4.10	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.10	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.3.6-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.6-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.6-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.6-amzn-0	Hive 命令行客户端。

组件	版本	描述
hive-hbase	2.3.6-amzn-0	Hive-hbase 客户端。
hive-metastore-server	2.3.6-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的 SQL 元数据。
hive-server2	2.3.6-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hudi	0.5.0-incubating	增量处理框架，以支持低延迟和高效率的数据管道。
hudi-presto	0.5.0-incubating	用于运行 Presto 以及 Hudi 的捆绑库。
hue-server	4.4.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	1.0.0	Jupyter notebook 的多用户服务器
livy-server	0.6.0-incubating	REST 用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.5.1	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的 SQL 数据库服务器。
nvidia-cuda	9.2.88	Nvidia 驱动程序和 Cuda 工具包

组件	版本	描述
oozie-client	5.1.0	Oozie 命令行客户端。
oozie-server	5.1.0	用于接受 Oozie 工作流请求的服务。
opencv	3.4.0	开源计算机视觉库。
phoenix-library	4.14.3--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.14.3--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.227	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.227	用于执行查询的各个部分的服务。
presto-client	0.227	Presto 命令行客户端，安装在 HA 集群的备用主节点（未启动 Presto 服务器）上。
pig-client	0.17.0	Pig 命令行客户端。
r	3.4.1	用于统计计算的 R 项目
spark-client	2.4.4	Spark 命令行客户端。
spark-history-server	2.4.4	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.4.4	的内存中执行引擎。YARN
spark-yarn-slave	2.4.4	YARN从属服务器需要的 Apache Spark 库。

组件	版本	描述
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	1.14.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.2	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.8.2	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.14	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.14	ZooKeeper 命令行客户端。

## 5.28.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

### emr-5.28.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。

分类	描述
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。

分类	描述
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。

分类	描述
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 jupyterhub_config.py 文件 JupyterHubs 中的值。
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。



分类	描述
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。

分类	描述
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。
ranger-kms-site	更改 Ran ranger-kms-site KMS ger .xml 文件中的值。
ranger-kms-env	在游侠KMS环境中更改值。
ranger-kms-log4j	更改 Ranger 的 kms-log4j.properties 文件中的值。KMS
ranger-kms-db-ca	更改 S3 上我与 Ranger SQL SSL KMS 连接的 CA 文件值。
recordserver-env	更改EMR RecordServer 环境中的值。
recordserver-conf	更改 server.proper EMR RecordServer ties 文件中的值。
recordserver-log4j	更改EMR RecordServer的 log4j.properties 文件中的值。

分类	描述
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.27.1

### 5.27.1 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#)和[ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-5.27.1	emr-5.27.0	emr-5.26.0	emr-5.25.0
适用于 Java 的 AWS SDK	1.11.615	1.11.615	1.11.595	1.11.566
Python	2.7、3.6	2.7、3.6	2.7、3.6	2.7、3.6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.8.1	1.8.1	1.8.0	1.8.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.10	1.4.10	1.4.10	1.4.9

	emr-5.27.1	emr-5.27.0	emr-5.26.0	emr-5.25.0
HCatalog	2.3.5	2.3.5	2.3.5	2.3.5
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.5	2.3.5	2.3.5	2.3.5
Hudi	-	-	-	-
Hue	4.4.0	4.4.0	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.0.0	1.0.0	0.9.6	0.9.6
Livy	0.6.0	0.6.0	0.6.0	0.6.0
MXNet	1.4.0	1.4.0	1.4.0	1.4.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.2	4.14.2	4.14.2	4.14.1
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.224	0.224	0.220	0.220
Spark	2.4.4	2.4.4	2.4.3	2.4.3
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.14.0	1.14.0	1.13.1	1.13.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2

	emr-5.27.1	emr-5.27.0	emr-5.26.0	emr-5.25.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.1	0.8.1	0.8.1
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

## 5.27.1 发布说明

这是补丁版本。所有应用程序和组件都与之前的 Amazon EMR 版本相同。

实例元数据服务 (IMDS) V2 支持状态：Amazon EMR 5.23.1、5.27.1 和 5.32 或更高版本的组件用于所有调用。IMDSv2 IMDS对于应用程序代码中的IMDS调用，您可以同时使用IMDSv1和IMDSv2，或者将配置IMDS为仅IMDSv2用于增强安全性。对于其他 5.x EMR 版本，禁用IMDSv1会导致集群启动失败。

## 5.27.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 emr 或 aws 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.2.4	亚马逊 SageMaker Spark SDK
emr-ddb	4.12.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.11.0	适用于 Hadoop 生态系统的方便易用的库。

组件	版本	描述
emr-kinesis	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.13.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	1.3.0	EMRS3Select 连接器
emrfs	2.36.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.8.1	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.8.5-amzn-4	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.8.5-amzn-4	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.5-amzn-4	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.5-amzn-4	HDFS用于跟踪文件名和区块位置的服务。

组件	版本	描述
hadoop-hdfs-journalnode	2.8.5-amzn-4	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httfs-server	2.8.5-amzn-4	HTTPHDFS操作端点。
hadoop-kms-server	2.8.5-amzn-4	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.8.5-amzn-4	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.8.5-amzn-4	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.8.5-amzn-4	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.5-amzn-4	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.10	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.10	为一个或多个HBase地区提供服务的服务。
hbase-client	1.4.10	HBase命令行客户端。
hbase-rest-server	1.4.10	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.10	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.3.5-amzn-1	用于操作 hcatalog-server 的“hcat”命令行客户端。



组件	版本	描述
hcatalog-server	2.3.5-amzn-1	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.5-amzn-1	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.5-amzn-1	Hive 命令行客户端。
hive-hbase	2.3.5-amzn-1	Hive-hbase 客户端。
hive-metastore-server	2.3.5-amzn-1	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.5-amzn-1	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	4.4.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	1.0.0	Jupyter notebook 的多用户服务器
livy-server	0.6.0-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.4.0	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的SQL数据库服务器。

组件	版本	描述
nvidia-cuda	9.2.88	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.1.0	Oozie 命令行客户端。
oozie-server	5.1.0	用于接受 Oozie 工作流请求的服务。
opencv	3.4.0	开源计算机视觉库。
phoenix-library	4.14.2--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.14.2--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.224	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.224	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。
r	3.4.1	用于统计计算的 R 项目
spark-client	2.4.4	Spark 命令行客户端。
spark-history-server	2.4.4	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.4.4	的内存中执行引擎。YARN
spark-yarn-slave	2.4.4	YARN从属服务器需要的 Apache Spark 库。

组件	版本	描述
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	1.14.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.2	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.8.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.14	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.14	ZooKeeper 命令行客户端。

### 5.27.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.27.1 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。

分类	描述
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。

分类	描述
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。

分类	描述
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 jupyterhub_config.py 文件 JupyterHubs 中的值。
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。

分类	描述
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。

分类	描述
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。
ranger-kms-site	更改 Ran ranger-kms-site KMS ger .xml 文件中的值。
ranger-kms-env	在游侠KMS环境中更改值。
ranger-kms-log4j	更改 Ranger 的 kms-log4j.properties 文件中的值。KMS
ranger-kms-db-ca	更改 S3 上我与 Ranger SQL SSL KMS 连接的 CA 文件值。
recordserver-env	更改EMR RecordServer 环境中的值。
recordserver-conf	更改 server.proper EMR RecordServer ties 文件中的值。
recordserver-log4j	更改EMR RecordServer的 log4j.properties 文件中的值。



分类	描述
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.27.0

### 5.27.0 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#)和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-5.27.0	emr-5.26.0	emr-5.25.0	emr-5.24.1
适用于 Java 的 AWS SDK	1.11.615	1.11.595	1.11.566	1.11.546
Python	2.7、3.6	2.7、3.6	2.7、3.6	2.7、3.6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.8.1	1.8.0	1.8.0	1.8.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.10	1.4.10	1.4.9	1.4.9

	emr-5.27.0	emr-5.26.0	emr-5.25.0	emr-5.24.1
HCatalog	2.3.5	2.3.5	2.3.5	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.5	2.3.5	2.3.5	2.3.4
Hudi	-	-	-	-
Hue	4.4.0	4.4.0	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.0.0	0.9.6	0.9.6	0.9.6
Livy	0.6.0	0.6.0	0.6.0	0.6.0
MXNet	1.4.0	1.4.0	1.4.0	1.4.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.2	4.14.2	4.14.1	4.14.1
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.224	0.220	0.220	0.219
Spark	2.4.4	2.4.3	2.4.3	2.4.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.14.0	1.13.1	1.13.1	1.12.0
Tez	0.9.2	0.9.2	0.9.2	0.9.1

	emr-5.27.0	emr-5.26.0	emr-5.25.0	emr-5.24.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.1	0.8.1	0.8.1
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.13

## 5.27.0 发布说明

以下发行说明包含有关亚马逊 5.27. EMR 0 版本的信息。更改与 5.26.0 有关。

首次发布日期：2019 年 9 月 23 日

### 升级

- AWS SDK for Java 1.11.615
- Flink 1.8.1
- JupyterHub 1.0.0
- Spark 2.4.4
- Tensorflow 1.14.0
- 连接器和驱动程序：
  - DynamoDB 连接器 4.12.0

### 新功能

- ( 2019 年 10 月 24 日 ) 所有亚马逊EMR版本均提供EMR笔记本电脑的以下新功能。
  - 现在，您可以将 Git 存储库与EMR笔记本关联，将笔记本存储在版本控制的环境中。您可以通过远程 Git 存储库与同行共享代码，并重复使用现有的 Jupyter notebook。有关更多信息，请参阅《亚马逊EMR管理指南》中的“[将 Git 存储库与亚马逊EMR笔记本关联](#)”。
  - [nbdime 实用程序](#)现已在EMR笔记本中提供，用于简化比较和合并笔记本电脑。
  - EMR现在支持笔记本电脑 JupyterLab。 JupyterLab 是一个基于 Web 的交互式开发环境，与 Jupyter 笔记本电脑完全兼容。现在，你可以选择在 Jupyter 笔记本编辑器 JupyterLab 或 Jupyter 笔记本编辑器中打开笔记本。

- (2019年10月30日) 在 Amazon 5.25.0 及更高EMR版本中，您可以从集群摘要页面或控制台的“应用程序历史记录”选项卡连接到 Spark 历史服务器用户界面。您无需通过SSH连接设置 Web 代理，而是可以快速访问 Spark 历史服务器用户界面来查看应用程序指标并访问活动集群和已终止集群的相关日志文件。有关更多信息，请参阅 Amazon EMR 管理指南中的[集群外访问永久应用程序用户界面](#)。

## 更改、增强功能和解决的问题

- 带有多个主节点的 Amazon EMR 集群
  - 您可以在具有多个主节点的 Amazon EMR 集群上安装和运行 Flink。有关更多信息，请参阅[支持的应用程序和功能](#)。
  - 您可以在具有多个主节点的 Amazon EMR 集群上配置HDFS透明加密。有关更多信息，请参阅在[具有多个主节点的EMR集群上进行HDFS透明加密](#)。
  - 现在，您可以修改在具有多个主节点的 Amazon EMR 集群上运行的应用程序的配置。有关更多信息，请参阅[为运行的集群中的实例组提供配置](#)。
- Amazon EMR-DynamoDB 连接器
  - Amazon EMR-DynamoDB Connector 现在支持以下 DynamoDB 数据类型：布尔值、列表、地图、项目、空。有关更多信息，请参阅[设置 Hive 表以运行 Hive 命令](#)。

## 已知问题

- 具有多个主节点的集群和 Kerberos 身份验证中的已知问题

如果您在 Amazon 5.20.0 及更高EMR版本中运行具有多个主节点和 Kerberos 身份验证的集群，则在集群运行一段时间后，集群操作可能会遇到问题，例如缩小规模或步骤提交。具体时间段取决于您定义的 Kerberos 票证有效期。缩减问题会影响您提交的自动缩减和显式缩减请求。其它集群操作也可能受到影响。

### 解决办法：

- SSH以hadoop用户身份访问具有多个主节点的EMR群集的主节点。
- 运行以下命令，为 hadoop 用户续订 Kerberos 票证。

```
kinit -kt <keytab_file> <principal>
```

通常情况下，keytab 文件位于 /etc/hadoop.keytab，而 principal 为 hadoop/  
<hostname>@<REALM> 格式。

**Note**

此解决方法将在 Kerberos 票证有效期内生效。默认情况下，此持续时间为 10 个小时，但可以通过 Kerberos 设置进行配置。Kerberos 票证过期后，您必须重新运行上述命令。

## 5.27.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>aws-sagemaker-spark-sdk</code>	1.2.4	亚马逊 SageMaker Spark SDK
<code>emr-ddb</code>	4.12.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.11.0	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
<code>emr-s3-dist-cp</code>	2.13.0	针对 Amazon S3 优化的分布式复制应用程序。
<code>emr-s3-select</code>	1.3.0	EMRS3Select 连接器

组件	版本	描述
emrfs	2.36.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.8.1	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.8.5-amzn-4	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.8.5-amzn-4	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.5-amzn-4	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.5-amzn-4	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	2.8.5-amzn-4	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	2.8.5-amzn-4	HTTPHDFS操作端点。
hadoop-kms-server	2.8.5-amzn-4	基于 Hadoop 的加密密钥管理服务。 KeyProvider API

组件	版本	描述
hadoop-mapred	2.8.5-amzn-4	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.8.5-amzn-4	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.8.5-amzn-4	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.5-amzn-4	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.10	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.10	为一个或多个HBase地区提供服务的服务。
hbase-client	1.4.10	HBase命令行客户端。
hbase-rest-server	1.4.10	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.10	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.3.5-amzn-1	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.5-amzn-1	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.5-amzn-1	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.5-amzn-1	Hive 命令行客户端。



组件	版本	描述
hive-hbase	2.3.5-amzn-1	Hive-hbase 客户端。
hive-metastore-server	2.3.5-amzn-1	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的 SQL 元数据。
hive-server2	2.3.5-amzn-1	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	4.4.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	1.0.0	Jupyter notebook 的多用户服务器
livy-server	0.6.0-incubating	REST 用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.4.0	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的 SQL 数据库服务器。
nvidia-cuda	9.2.88	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.1.0	Oozie 命令行客户端。
oozie-server	5.1.0	用于接受 Oozie 工作流请求的服务。

组件	版本	描述
opencv	3.4.0	开源计算机视觉库。
phoenix-library	4.14.2--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.14.2--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.224	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.224	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。
r	3.4.1	用于统计计算的 R 项目
spark-client	2.4.4	Spark 命令行客户端。
spark-history-server	2.4.4	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.4.4	的内存中执行引擎。YARN
spark-yarn-slave	2.4.4	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	1.14.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.2	tez YARN 应用程序和库。

组件	版本	描述
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.8.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.14	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.14	ZooKeeper 命令行客户端。

### 5.27.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.27.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。

分类	描述
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。

分类	描述
hcatalog-webhcat-site	更改 HCatalog WebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。

分类	描述
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 jupyterhub_config.py 文件 JupyterHubs 中的值。
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。

分类	描述
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。

分类	描述
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。
ranger-kms-site	更改 Ran ranger-kms-site KMS ger .xml 文件中的值。
ranger-kms-env	在游侠KMS环境中更改值。
ranger-kms-log4j	更改 Ranger 的 kms-log4j.properties 文件中的值。KMS
ranger-kms-db-ca	更改 S3 上我与 Ranger SQL SSL KMS 连接的 CA 文件值。
recordserver-env	更改EMR RecordServer 环境中的值。
recordserver-conf	更改 server.proper EMR RecordServer ties 文件中的值。
recordserver-log4j	更改EMR RecordServer的 log4j.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值



分类	描述
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.26.0

### 5.26.0 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#)和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)

- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

## 应用程序版本信息

	emr-5.26.0	emr-5.25.0	emr-5.24.1	emr-5.24.0
适用于 Java 的 AWS SDK	1.11.595	1.11.566	1.11.546	1.11.546
Python	2.7、3.6	2.7、3.6	2.7、3.6	2.7、3.6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.8.0	1.8.0	1.8.0	1.8.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.10	1.4.9	1.4.9	1.4.9
HCatalog	2.3.5	2.3.5	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.5	2.3.5	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.4.0	4.4.0	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-

	emr-5.26.0	emr-5.25.0	emr-5.24.1	emr-5.24.0
JupyterHub	0.9.6	0.9.6	0.9.6	0.9.6
Livy	0.6.0	0.6.0	0.6.0	0.6.0
MXNet	1.4.0	1.4.0	1.4.0	1.4.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.2	4.14.1	4.14.1	4.14.1
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.220	0.220	0.219	0.219
Spark	2.4.3	2.4.3	2.4.2	2.4.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.13.1	1.13.1	1.12.0	1.12.0
Tez	0.9.2	0.9.2	0.9.1	0.9.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.1	0.8.1	0.8.1
ZooKeeper	3.4.14	3.4.14	3.4.13	3.4.13

## 5.26.0 发布说明

以下发行说明包含有关亚马逊 5.26. EMR 0 版本的信息。更改与 5.25.0 有关。

首次发布日期：2019 年 8 月 8 日

上次更新日期：2019 年 8 月 19 日

## 升级

- AWS SDK for Java 1.11.595
- HBase1.4.10
- Phoenix 4.14.2
- 连接器和驱动程序：
  - DynamoDB 连接器 4.11.0
  - MariaDB 连接器 2.4.2
  - 亚马逊 Redshift JDBC Driver 1.2.32.1056

## 新功能

- (测试版) 使用亚马逊 EMR 5.26.0，您可以启动与 Lake Formation 集成的集群。这种集成提供了对 Glue 数据目录中的数据库和表的精细列级访问。AWS 它还支持从企业身份系统对 EMR 笔记本或 Apache Zeppelin 进行联合单点登录。有关更多信息，请参阅[将 Amazon EMR 与 AWS Lake Formation \(测试版\) 集成](#)。
- (2019 年 8 月 19 日) 所有支持安全组的亚马逊 EMR 版本现已推出亚马逊 EMR 封锁公开访问权限。屏蔽公共访问是适用于每个 AWS 地区的账户范围的设置。除非将端口指定为例外，否则当任何与集群关联的安全组都有允许某个端口上来自 IPv4 0.0.0.0/0 或 IPv6:: /0 (公共访问) 的入站流量的规则时，阻止公共访问会阻止集群启动。默认情况下，端口 22 是一个例外。有关更多信息，请参阅《[亚马逊 EMR 管理指南](#)》中的“[使用 Amazon EMR 阻止公共访问](#)”。

## 更改、增强功能和解决的问题

- EMR 笔记本
  - 在 EMR 5.26.0 及更高版本中，EMR 笔记本除了默认的 Python 库外，还支持笔记本电脑范围的 Python 库。无需重新创建集群或重新将笔记本附加到集群，您即可从笔记本编辑器中安装笔记本范围的库。笔记本范围的库是在 Python 虚拟环境中创建的，因此适用于当前笔记本会话。这使得您可以隔离笔记本依赖项。有关更多信息，请参阅 Amazon EMR 管理指南中的[使用笔记本范围的库](#)。
- EMRFS
  - 您可以通过将设置 `fs.s3.consistent.metadata.etag.validation.enabled` 为来启用 ETag 验证功能 (测试版) `true`。借助此功能，ETags 可以 EMRFS 使用 Amazon S3 验证正在读取的对象是否为最新可用版本。此功能对于 Amazon S3 上的文件被覆盖，同时保留相同名称的

read-after-update 用例非常有用。此ETag验证功能目前不适用于 S3 Select。有关更多信息，请参阅[配置统一视图](#)。

- Spark

- 现在默认启用了以下优化：DISTINCT之前的动态分区修剪、JOIN后续查询的SQL计划统计推断的改进INTERSECT、标量子DISTINCT查询的扁平化、优化的联接重新排序以及布隆过滤器联接。有关更多信息，请参阅[优化 Spark 性能](#)。
- 改进了排序合并连接的整个阶段代码生成。
- 改进了查询片段和子查询重用。
- 改进了 Spark 启动时的预分配执行程序。
- 连接的较小侧包含广播提示时，不再应用 Bloom 筛选条件连接。

- Tez

- 已解决 Tez 中存在的问题。Tez UI 现在可以在具有多个主节点的 Amazon EMR 集群上运行。

## 已知问题

- 改进的“排序合并连接的整个阶段代码生成”功能在启用后会增加内存压力。此优化可提高性能，但如果 `spark.yarn.executor.memoryOverheadFactor` 未调整，不能提供足够的内存，则会导致任务重试或失败。要禁用此功能，请将 `spark.sql.sortMergeJoinExec.extendedCodegen.enabled` 设置为 `false`。
- 具有多个主节点的集群和 Kerberos 身份验证中的已知问题

如果您在 Amazon 5.20.0 及更高EMR版本中运行具有多个主节点和 Kerberos 身份验证的集群，则在集群运行一段时间后，集群操作可能会遇到问题，例如缩小规模或步骤提交。具体时间段取决于您定义的 Kerberos 票证有效期。缩减问题会影响您提交的自动缩减和显式缩减请求。其它集群操作也可能受到影响。

解决办法：

- SSH以hadoop用户身份访问具有多个主节点的EMR群集的主节点。
- 运行以下命令，为 hadoop 用户续订 Kerberos 票证。

```
kinit -kt <keytab_file> <principal>
```

通常情况下，keytab 文件位于 `/etc/hadoop.keytab`，而 principal 为 `hadoop/<hostname>@<REALM>` 格式。

**Note**

此解决方法将在 Kerberos 票证有效期内生效。默认情况下，此持续时间为 10 个小时，但可以通过 Kerberos 设置进行配置。Kerberos 票证过期后，您必须重新运行上述命令。

## 5.26.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>aws-sagemaker-spark-sdk</code>	1.2.4	亚马逊 SageMaker Spark SDK
<code>emr-ddb</code>	4.11.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.10.0	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
<code>emr-s3-dist-cp</code>	2.12.0	针对 Amazon S3 优化的分布式复制应用程序。
<code>emr-s3-select</code>	1.3.0	EMRS3Select 连接器

组件	版本	描述
emrfs	2.35.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.8.0	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.8.5-amzn-4	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.8.5-amzn-4	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.5-amzn-4	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.5-amzn-4	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	2.8.5-amzn-4	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	2.8.5-amzn-4	HTTPHDFS操作端点。
hadoop-kms-server	2.8.5-amzn-4	基于 Hadoop 的加密密钥管理服务。 KeyProvider API

组件	版本	描述
hadoop-mapred	2.8.5-amzn-4	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.8.5-amzn-4	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.8.5-amzn-4	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.5-amzn-4	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.10	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.10	为一个或多个HBase地区提供服务的服务。
hbase-client	1.4.10	HBase命令行客户端。
hbase-rest-server	1.4.10	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.10	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.3.5-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.5-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.5-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.5-amzn-0	Hive 命令行客户端。



组件	版本	描述
hive-hbase	2.3.5-amzn-0	Hive-hbase 客户端。
hive-metastore-server	2.3.5-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的 SQL 元数据。
hive-server2	2.3.5-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	4.4.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	0.9.6	Jupyter notebook 的多用户服务器
livy-server	0.6.0-incubating	REST 用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.4.0	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的 SQL 数据库服务器。
nvidia-cuda	9.2.88	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.1.0	Oozie 命令行客户端。
oozie-server	5.1.0	用于接受 Oozie 工作流请求的服务。

组件	版本	描述
opencv	3.4.0	开源计算机视觉库。
phoenix-library	4.14.2--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.14.2--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.220	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.220	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。
r	3.4.1	用于统计计算的 R 项目
spark-client	2.4.3	Spark 命令行客户端。
spark-history-server	2.4.3	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.4.3	的内存中执行引擎。YARN
spark-yarn-slave	2.4.3	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	1.13.1	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.2	tez YARN 应用程序和库。

组件	版本	描述
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.8.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.14	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.14	ZooKeeper 命令行客户端。

### 5.26.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.26.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。

分类	描述
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。

分类	描述
hcatalog-webhcat-site	更改 HCatalog WebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。

分类	描述
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 jupyterhub_config.py 文件 JupyterHubs 中的值。
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。

分类	描述
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。

分类	描述
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。
recordserver-env	更改EMR RecordServer 环境中的值。
recordserver-conf	更改 server.proper EMR RecordServer ties 文件中的值。
recordserver-log4j	更改EMR RecordServer的 log4j.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。



分类	描述
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.25.0

### 5.25.0 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#)和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-5.25.0	emr-5.24.1	emr-5.24.0	emr-5.23.1
适用于 Java 的 AWS SDK	1.11.566	1.11.546	1.11.546	1.11.519
Python	2.7、3.6	2.7、3.6	2.7、3.6	2.7、3.6

	emr-5.25.0	emr-5.24.1	emr-5.24.0	emr-5.23.1
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.8.0	1.8.0	1.8.0	1.7.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.9	1.4.9	1.4.9	1.4.9
HCatalog	2.3.5	2.3.4	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.5	2.3.4	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.4.0	4.4.0	4.4.0	4.3.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.6	0.9.6	0.9.6	0.9.4
Livy	0.6.0	0.6.0	0.6.0	0.5.0
MXNet	1.4.0	1.4.0	1.4.0	1.3.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.1	4.14.1	4.14.1	4.14.1

	emr-5.25.0	emr-5.24.1	emr-5.24.0	emr-5.23.1
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.220	0.219	0.219	0.215
Spark	2.4.3	2.4.2	2.4.2	2.4.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.13.1	1.12.0	1.12.0	1.12.0
Tez	0.9.2	0.9.1	0.9.1	0.9.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.1	0.8.1	0.8.1
ZooKeeper	3.4.14	3.4.13	3.4.13	3.4.13

## 5.25.0 发布说明

以下发行说明包含 Amazon EMR 版本 5.25.0 的信息。更改与 5.24.1 有关。

首次发布日期：2019 年 7 月 17 日

上次更新日期：2019 年 10 月 30 日

亚马逊 EMR 5.25.0

### 升级

- AWS SDK for Java 1.11.566
- Hive 2.3.5
- Presto 0.220
- Spark 2.4.3
- TensorFlow 1.13.1
- Tez 0.9.2

- Zookeeper 3.4.14

## 新功能

- (2019年10月30日)从Amazon EMR版本5.25.0开始，您可以从集群摘要页面或控制台的“应用程序历史记录”选项卡连接到Spark历史服务器用户界面。您无需通过SSH连接设置Web代理，而是可以快速访问Spark历史服务器用户界面来查看应用程序指标并访问活动集群和已终止集群的相关日志文件。有关更多信息，请参阅Amazon EMR管理指南中的[集群外访问永久应用程序用户界面](#)。

## 更改、增强功能和解决的问题

- Spark
  - 通过使用Bloom筛选条件预筛选输入，提高了某些连接的性能。默认情况下，优化处于禁用状态，但可以通过以下方式启用：将Spark配置参数`spark.sql.bloomFilterJoin.enabled`设置为`true`。
  - 改进了按字符串类型列分组的性能。
  - 改进了未HBase安装的集群的R4实例类型的默认Spark执行器内存和内核配置。
  - 解决了动态分区修剪功能之前存在的一个问题，即修剪的表必须位于联接的左侧。
  - DISTINCT在INTERSECT优化之前进行了改进，适用于其他涉及别名的情况。
  - 改进了JOIN后续DISTINCT查询的SQL计划统计数据推断。默认情况下，该改进处于禁用状态，但可以通过以下方式启用：将Spark配置参数`spark.sql.statsImprovements.enabled`设置为`true`。此优化是“Distinct before Intersect”功能所需的，将`spark.sql.optimizer.distinctBeforeIntersect.enabled`设置为`true`时将自动启用。
  - 根据表格大小和筛选条件优化了联接顺序。默认情况下，该优化处于禁用状态，但可以通过以下方式启用：将Spark配置参数`spark.sql.optimizer.sizeBasedJoinReorder.enabled`设置为`true`。

有关更多信息，请参阅[优化 Spark 性能](#)。

- EMRFS
  - 现在，默认情况下`fs.s3.buckets.create.enabled`，该EMRFS设置处于禁用状态。通过测试，我们发现禁用此设置可提高性能并可防止意外创建S3存储桶。如果您的应用程序需使用此功能，则可以通过以下方式启用：将`emrfs-site`配置分类中的

`fs.s3.buckets.create.enabled` 设置为 `true`。有关更多信息，请参阅[在创建集群时提供配置](#)。

- 安全配置中的本地磁盘加密和 S3 加密改进 (2019 年 8 月 5 日)
  - 在安全配置设置中将 Amazon S3 加密设置与本地磁盘加密设置分开。
  - 添加了在 5.24.0 及更高版本中启用 EBS 加密的选项。选择此选项后，除了存储卷之外，还会加密根设备卷。以前的版本需要使用自定义 AMI 来加密根设备卷。
  - 有关更多信息，请参阅 Amazon EMR 管理指南中的[加密选项](#)。

## 已知问题

- 具有多个主节点的集群和 Kerberos 身份验证中的已知问题

如果您在 Amazon 5.20.0 及更高 EMR 版本中运行具有多个主节点和 Kerberos 身份验证的集群，则在集群运行一段时间后，集群操作可能会遇到问题，例如缩小规模或步骤提交。具体时间段取决于您定义的 Kerberos 票证有效期。缩减问题会影响您提交的自动缩减和显式缩减请求。其它集群操作也可能受到影响。

解决办法：

- SSH 以 hadoop 用户身份访问具有多个主节点的 EMR 群集的主节点。
- 运行以下命令，为 hadoop 用户续订 Kerberos 票证。

```
kinit -kt <keytab_file> <principal>
```

通常情况下，keytab 文件位于 `/etc/hadoop.keytab`，而 principal 为 `hadoop/<hostname>@<REALM>` 格式。

### Note

此解决方法将在 Kerberos 票证有效期内生效。默认情况下，此持续时间为 10 个小时，但可以通过 Kerberos 设置进行配置。Kerberos 票证过期后，您必须重新运行上述命令。

## 5.25.0 组件版本

下面列出了 Amazon 随此版本 EMR 安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的 EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon

最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.2.4	亚马逊 SageMaker Spark SDK
emr-ddb	4.10.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.9.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.11.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	1.3.0	EMRS3Select 连接器
emrfs	2.34.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.8.0	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。

组件	版本	描述
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.8.5-amzn-4	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.8.5-amzn-4	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.5-amzn-4	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.5-amzn-4	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	2.8.5-amzn-4	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	2.8.5-amzn-4	HTTPHDFS操作端点。
hadoop-kms-server	2.8.5-amzn-4	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.8.5-amzn-4	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.8.5-amzn-4	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.8.5-amzn-4	YARN用于分配和管理群集资源和分布式应用程序的服务。

组件	版本	描述
hadoop-yarn-timeline-server	2.8.5-amzn-4	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.9	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.9	为一个或多个HBase地区提供服务的服务。
hbase-client	1.4.9	HBase命令行客户端。
hbase-rest-server	1.4.9	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.9	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.3.5-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.5-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.5-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.5-amzn-0	Hive 命令行客户端。
hive-hbase	2.3.5-amzn-0	Hive-hbase 客户端。
hive-metastore-server	2.3.5-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.5-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。



组件	版本	描述
hue-server	4.4.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	0.9.6	Jupyter notebook 的多用户服务器
livy-server	0.6.0-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.4.0	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
nvidia-cuda	9.2.88	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.1.0	Oozie 命令行客户端。
oozie-server	5.1.0	用于接受 Oozie 工作流请求的服务。
opencv	3.4.0	开源计算机视觉库。
phoenix-library	4.14.1--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.14.1--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API

组件	版本	描述
presto-coordinator	0.220	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.220	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。
r	3.4.1	用于统计计算的 R 项目
spark-client	2.4.3	Spark 命令行客户端。
spark-history-server	2.4.3	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.4.3	的内存中执行引擎。YARN
spark-yarn-slave	2.4.3	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	1.13.1	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.2	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.8.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.14	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。

组件	版本	描述
zookeeper-client	3.4.14	ZooKeeper 命令行客户端。

### 5.25.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.25.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置

分类	描述
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。

分类	描述
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 JupyterHubs的 jupyterhub_config.py 文件中的值。
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。

分类	描述
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。

分类	描述
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。

分类	描述
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。
recordserver-env	更改EMR RecordServer 环境中的值。
recordserver-conf	更改 server.proper EMR RecordServer ties 文件中的值。
recordserver-log4j	更改EMR RecordServer的 log4j.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。



分类	描述
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.24.1

### 5.24.1 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#)和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-5.24.1	emr-5.24.0	emr-5.23.1	emr-5.23.0
适用于 Java 的 AWS SDK	1.11.546	1.11.546	1.11.519	1.11.519
Python	2.7、3.6	2.7、3.6	2.7、3.6	2.7、3.6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-

	emr-5.24.1	emr-5.24.0	emr-5.23.1	emr-5.23.0
Flink	1.8.0	1.8.0	1.7.1	1.7.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.9	1.4.9	1.4.9	1.4.9
HCatalog	2.3.4	2.3.4	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.4.0	4.4.0	4.3.0	4.3.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.6	0.9.6	0.9.4	0.9.4
Livy	0.6.0	0.6.0	0.5.0	0.5.0
MXNet	1.4.0	1.4.0	1.3.1	1.3.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.1	4.14.1	4.14.1	4.14.1
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.219	0.219	0.215	0.215
Spark	2.4.2	2.4.2	2.4.0	2.4.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7

	emr-5.24.1	emr-5.24.0	emr-5.23.1	emr-5.23.0
TensorFlow	1.12.0	1.12.0	1.12.0	1.12.0
Tez	0.9.1	0.9.1	0.9.1	0.9.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.1	0.8.1	0.8.1
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

### 5.24.1 发布说明

以下发行说明包含有关亚马逊 5.24.1 EMR 版本的信息。更改与 5.24.0 有关。

首次发布日期：2019 年 6 月 26 日

#### 更改、增强功能和解决的问题

- 更新了亚马逊的默认 Amazon Linux AMI，EMR使其包含重要的 Linux 内核安全更新，包括 TCPSACK拒绝服务问题 ([AWS-2019-005](#))。

#### 已知问题

- 具有多个主节点的集群和 Kerberos 身份验证中的已知问题

如果您在 Amazon 5.20.0 及更高EMR版本中运行具有多个主节点和 Kerberos 身份验证的集群，则在集群运行一段时间后，集群操作可能会遇到问题，例如缩小规模或步骤提交。具体时间段取决于您定义的 Kerberos 票证有效期。缩减问题会影响您提交的自动缩减和显式缩减请求。其它集群操作也可能受到影响。

#### 解决办法：

- SSH以hadoop用户身份访问具有多个主节点的EMR群集的主节点。
- 运行以下命令，为 hadoop 用户续订 Kerberos 票证。

```
kinit -kt <keytab_file> <principal>
```

通常情况下，keytab 文件位于 `/etc/hadoop.keytab`，而 principal 为 `hadoop/<hostname>@<REALM>` 格式。

#### Note

此解决方法将在 Kerberos 票证有效期内生效。默认情况下，此持续时间为 10 个小时，但可以通过 Kerberos 设置进行配置。Kerberos 票证过期后，您必须重新运行上述命令。

### 5.24.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>aws-sagemaker-spark-sdk</code>	1.2.1	亚马逊 SageMaker Spark SDK
<code>emr-ddb</code>	4.9.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.8.0	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
<code>emr-s3-dist-cp</code>	2.11.0	针对 Amazon S3 优化的分布式复制应用程序。

组件	版本	描述
emr-s3-select	1.3.0	EMRS3Select 连接器
emrfs	2.33.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.8.0	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.8.5-amzn-4	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.8.5-amzn-4	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.5-amzn-4	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.5-amzn-4	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	2.8.5-amzn-4	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	2.8.5-amzn-4	HTTPHDFS操作端点。

组件	版本	描述
hadoop-kms-server	2.8.5-amzn-4	基于 Hadoop 的加密密钥管理服务。KeyProvider API
hadoop-mapred	2.8.5-amzn-4	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.8.5-amzn-4	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.8.5-amzn-4	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.5-amzn-4	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.9	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.9	为一个或多个HBase地区提供服务的服务。
hbase-client	1.4.9	HBase命令行客户端。
hbase-rest-server	1.4.9	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.9	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	2.3.4-amzn-2	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.4-amzn-2	提供服务HCatalog，为分布式应用程序提供表和存储管理层。

组件	版本	描述
hcatalog-webhcat-server	2.3.4-amzn-2	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.4-amzn-2	Hive 命令行客户端。
hive-hbase	2.3.4-amzn-2	Hive-hbase 客户端。
hive-metastore-server	2.3.4-amzn-2	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.4-amzn-2	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	4.4.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	0.9.6	Jupyter notebook 的多用户服务器
livy-server	0.6.0-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.4.0	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
nvidia-cuda	9.2.88	Nvidia 驱动程序和 Cuda 工具包

组件	版本	描述
oozie-client	5.1.0	Oozie 命令行客户端。
oozie-server	5.1.0	用于接受 Oozie 工作流请求的服务。
opencv	3.4.0	开源计算机视觉库。
phoenix-library	4.14.1--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.14.1--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.219	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.219	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。
r	3.4.1	用于统计计算的 R 项目
spark-client	2.4.2	Spark 命令行客户端。
spark-history-server	2.4.2	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.4.2	的内存中执行引擎。YARN
spark-yarn-slave	2.4.2	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。



组件	版本	描述
tensorflow	1.12.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.1	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.8.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.13	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.13	ZooKeeper 命令行客户端。

### 5.24.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.24.1 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。

分类	描述
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。

分类	描述
hcatalog-webhcat-log4j2	更改 HCatalog WebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog WebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改 HTTPFS 环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。

分类	描述
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 JupyterHubs的 jupyterhub_config.py 文件中的值。
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。

分类	描述
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。

分类	描述
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。

分类	描述
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper 中的值。
zookeeper-log4j	更改 ZooKeeper 的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.24.0

### 5.24.0 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#) 和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-5.24.0	emr-5.23.1	emr-5.23.0	emr-5.22.0
适用于 Java 的 AWS SDK	1.11.546	1.11.519	1.11.519	1.11.510
Python	2.7、3.6	2.7、3.6	2.7、3.6	2.7、3.6
Scala	2.11.12	2.11.12	2.11.12	2.11.12

	emr-5.24.0	emr-5.23.1	emr-5.23.0	emr-5.22.0
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.8.0	1.7.1	1.7.1	1.7.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.9	1.4.9	1.4.9	1.4.9
HCatalog	2.3.4	2.3.4	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.4.0	4.3.0	4.3.0	4.3.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.6	0.9.4	0.9.4	0.9.4
Livy	0.6.0	0.5.0	0.5.0	0.5.0
MXNet	1.4.0	1.3.1	1.3.1	1.3.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.1	4.14.1	4.14.1	4.14.1
Pig	0.17.0	0.17.0	0.17.0	0.17.0



	emr-5.24.0	emr-5.23.1	emr-5.23.0	emr-5.22.0
Presto	0.219	0.215	0.215	0.215
Spark	2.4.2	2.4.0	2.4.0	2.4.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.12.0	1.12.0
Tez	0.9.1	0.9.1	0.9.1	0.9.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.1	0.8.1	0.8.1
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

## 5.24.0 发布说明

以下发行说明包含有关亚马逊 5.24. EMR 0 版本的信息。更改与 5.23.0 有关。

首次发布日期：2019 年 6 月 11 日

上次更新时间：2019 年 8 月 5 日

### 升级

- Flink 1.8.0
- Hue 4.4.0
- JupyterHub 0.9.6
- Livy 0.6.0
- MxNet 1.4.0
- Presto 0.219
- Spark 2.4.2
- AWS SDK for Java 1.11.546
- 连接器和驱动程序：

- DynamoDB 连接器 4.9.0
- MariaDB 连接器 2.4.1
- 亚马逊 Redshift JDBC Driver 1.2.27.1051

## 更改、增强功能和解决的问题

- Spark
  - 添加了对动态修剪分区的优化。默认情况下禁用优化。要启用该优化，请将 Spark 参数 `spark.sql.dynamicPartitionPruning.enabled` 设置为 `true`。
  - 改进了 INTERSECT 查询的性能。默认情况下禁用此优化。要启用该优化，请将 Spark 参数 `spark.sql.optimizer.distinctBeforeIntersect.enabled` 设置为 `true`。
  - 添加了对展平标量子查询的优化，可使用相同关系进行聚合。默认情况下禁用优化。要启用该优化，请将 Spark 参数 `spark.sql.optimizer.flattenScalarSubqueriesWithAggregates.enabled` 设置为 `true`。
  - 改进了整个阶段代码生成。

有关更多信息，请参阅[优化 Spark 性能](#)。

- 安全配置中的本地磁盘加密和 S3 加密改进 ( 2019 年 8 月 5 日 )
  - 在安全配置设置中将 Amazon S3 加密设置与本地磁盘加密设置分开。
  - 添加了启用EBS加密的选项。选择此选项后，除了存储卷之外，还会加密根设备卷。以前的版本需要使用自定义AMI来加密根设备卷。
  - 有关更多信息，请参阅 Amazon EMR 管理指南中的[加密选项](#)。

## 已知问题

- 具有多个主节点的集群和 Kerberos 身份验证中的已知问题

如果您在 Amazon 5.20.0 及更高EMR版本中运行具有多个主节点和 Kerberos 身份验证的集群，则在集群运行一段时间后，集群操作可能会遇到问题，例如缩小规模或步骤提交。具体时间段取决于您定义的 Kerberos 票证有效期。缩减问题会影响您提交的自动缩减和显式缩减请求。其它集群操作也可能受到受影响。

解决办法：

- SSH以hadoop用户身份访问具有多个主节点的EMR群集的主节点。

- 运行以下命令，为 hadoop 用户续订 Kerberos 票证。

```
kinit -kt <keytab_file> <principal>
```

通常情况下，keytab 文件位于 `/etc/hadoop.keytab`，而 principal 为 `hadoop/<hostname>@<REALM>` 格式。

#### Note

此解决方法将在 Kerberos 票证有效期内生效。默认情况下，此持续时间为 10 个小时，但可以通过 Kerberos 设置进行配置。Kerberos 票证过期后，您必须重新运行上述命令。

## 5.24.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>aws-sagemaker-spark-sdk</code>	1.2.1	亚马逊 SageMaker Spark SDK
<code>emr-ddb</code>	4.9.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.8.0	适用于 Hadoop 生态系统的方便易用的库。

组件	版本	描述
emr-kinesis	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.11.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	1.3.0	EMRS3Select 连接器
emrfs	2.33.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.8.0	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.8.5-amzn-4	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.8.5-amzn-4	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.5-amzn-4	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.5-amzn-4	HDFS用于跟踪文件名和区块位置的服务。

组件	版本	描述
hadoop-hdfs-journalnode	2.8.5-amzn-4	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httfs-server	2.8.5-amzn-4	HTTPHDFS操作端点。
hadoop-kms-server	2.8.5-amzn-4	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.8.5-amzn-4	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.8.5-amzn-4	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.8.5-amzn-4	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.5-amzn-4	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.9	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.9	为一个或多个HBase地区提供服务的服务。
hbase-client	1.4.9	HBase命令行客户端。
hbase-rest-server	1.4.9	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.9	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.3.4-amzn-2	用于操作 hcatalog-server 的“hcat”命令行客户端。

组件	版本	描述
hcatalog-server	2.3.4-amzn-2	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.4-amzn-2	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.4-amzn-2	Hive 命令行客户端。
hive-hbase	2.3.4-amzn-2	Hive-hbase 客户端。
hive-metastore-server	2.3.4-amzn-2	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.4-amzn-2	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	4.4.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	0.9.6	Jupyter notebook 的多用户服务器
livy-server	0.6.0-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.4.0	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的SQL数据库服务器。

组件	版本	描述
nvidia-cuda	9.2.88	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.1.0	Oozie 命令行客户端。
oozie-server	5.1.0	用于接受 Oozie 工作流请求的服务。
opencv	3.4.0	开源计算机视觉库。
phoenix-library	4.14.1--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.14.1--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.219	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.219	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。
r	3.4.1	用于统计计算的 R 项目
spark-client	2.4.2	Spark 命令行客户端。
spark-history-server	2.4.2	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.4.2	的内存中执行引擎。YARN
spark-yarn-slave	2.4.2	YARN从属服务器需要的 Apache Spark 库。

组件	版本	描述
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	1.12.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.1	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.8.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.13	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.13	ZooKeeper 命令行客户端。

## 5.24.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

### emr-5.24.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。



分类	描述
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。

分类	描述
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。

分类	描述
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 JupyterHubs 的 jupyterhub_config.py 文件中的值。
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。

分类	描述
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。

分类	描述
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。

分类	描述
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.23.1

### 5.23.1 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#)和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-5.23.1	emr-5.23.0	emr-5.22.0	emr-5.21.2
适用于 Java 的 AWS SDK	1.11.519	1.11.519	1.11.510	1.11.479
Python	2.7、3.6	2.7、3.6	2.7、3.6	2.7、3.6

	emr-5.23.1	emr-5.23.0	emr-5.22.0	emr-5.21.2
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.7.1	1.7.1	1.7.1	1.7.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.9	1.4.9	1.4.9	1.4.8
HCatalog	2.3.4	2.3.4	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.3.0	4.3.0	4.3.0	4.3.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.9.4	0.9.4
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.1	1.3.1	1.3.1	1.3.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.0.0
Phoenix	4.14.1	4.14.1	4.14.1	4.14.0

	emr-5.23.1	emr-5.23.0	emr-5.22.0	emr-5.21.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.215	0.215	0.215	0.215
Spark	2.4.0	2.4.0	2.4.0	2.4.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.12.0	1.12.0
Tez	0.9.1	0.9.1	0.9.1	0.9.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.1	0.8.1	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

### 5.23.1 发布说明

这是补丁版本。所有应用程序和组件都与之前的 Amazon EMR 版本相同。

实例元数据服务 (IMDS) V2 支持状态：Amazon EMR 5.23.1、5.27.1 和 5.32 或更高版本的组件用于所有调用。IMDSv2 对于应用程序代码中的 IMDS 调用，您可以同时使用 IMDSv1 和 IMDSv2，或者将配置 IMDS 为仅 IMDSv2 用于增强安全性。对于其他 5.x EMR 版本，禁用 IMDSv1 会导致集群启动失败。

### 5.23.1 组件版本

下面列出了 Amazon 随此版本 EMR 安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的 EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新 EMR 版本中的大数据应用程序包通常是社区中的最新版本。我们会 EMR 尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本 EMR 不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。



组件	版本	描述
aws-sagemaker-spark-sdk	1.2.1	亚马逊 SageMaker Spark SDK
emr-ddb	4.8.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.7.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.11.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	1.2.0	EMRS3Select 连接器
emrfs	2.32.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.7.1	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。

组件	版本	描述
hadoop-client	2.8.5-amzn-3	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.8.5-amzn-3	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.5-amzn-3	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.5-amzn-3	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	2.8.5-amzn-3	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	2.8.5-amzn-3	HTTPHDFS操作端点。
hadoop-kms-server	2.8.5-amzn-3	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.8.5-amzn-3	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.8.5-amzn-3	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.8.5-amzn-3	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.5-amzn-3	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.9	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.9	为一个或多个HBase地区提供服务的服务。

组件	版本	描述
hbase-client	1.4.9	HBase命令行客户端。
hbase-rest-server	1.4.9	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.9	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	2.3.4-amzn-1	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.4-amzn-1	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.4-amzn-1	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.4-amzn-1	Hive 命令行客户端。
hive-hbase	2.3.4-amzn-1	Hive-hbase 客户端。
hive-metastore-server	2.3.4-amzn-1	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.4-amzn-1	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	4.3.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	0.9.4	Jupyter notebook 的多用户服务器

组件	版本	描述
livy-server	0.5.0-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.3.1	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
nvidia-cuda	9.2.88	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.1.0	Oozie 命令行客户端。
oozie-server	5.1.0	用于接受 Oozie 工作流请求的服务。
opencv	3.4.0	开源计算机视觉库。
phoenix-library	4.14.1--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.14.1--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.215	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.215	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。

组件	版本	描述
r	3.4.1	用于统计计算的 R 项目
spark-client	2.4.0	Spark 命令行客户端。
spark-history-server	2.4.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.4.0	的内存中执行引擎。YARN
spark-yarn-slave	2.4.0	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	1.12.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.1	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.8.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.13	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.13	ZooKeeper 命令行客户端。

### 5.23.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

## emr-5.23.1 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。

分类	描述
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。

分类	描述
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 JupyterHubs的 jupyterhub_config.py 文件中的值。
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。



分类	描述
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。

分类	描述
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值

分类	描述
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR发布 5.23.0

### 5.23.0 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#) 和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)

- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

## 应用程序版本信息

	emr-5.23.0	emr-5.22.0	emr-5.21.2	emr-5.21.1
适用于 Java 的 AWS SDK	1.11.519	1.11.510	1.11.479	1.11.479
Python	2.7、3.6	2.7、3.6	2.7、3.6	2.7、3.6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.7.1	1.7.1	1.7.0	1.7.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.9	1.4.9	1.4.8	1.4.8
HCatalog	2.3.4	2.3.4	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.3.0	4.3.0	4.3.0	4.3.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-

	emr-5.23.0	emr-5.22.0	emr-5.21.2	emr-5.21.1
JupyterHub	0.9.4	0.9.4	0.9.4	0.9.4
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.1	1.3.1	1.3.1	1.3.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.0.0	5.0.0
Phoenix	4.14.1	4.14.1	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.215	0.215	0.215	0.215
Spark	2.4.0	2.4.0	2.4.0	2.4.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.12.0	1.12.0
Tez	0.9.1	0.9.1	0.9.1	0.9.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.1	0.8.0	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

## 5.23.0 发布说明

以下发行说明包含有关亚马逊 5.23. EMR 0 版本的信息。更改与 5.22.0 有关。

首次发布日期：2019 年 4 月 1 日

上次更新时间：2019 年 4 月 30 日

## 升级

- AWS SDK for Java 1.11.519

## 新功能

- (2019年4月30日) 在 Amazon EMR 5.23.0 及更高版本中，您可以启动包含三个主节点的集群，以支持YARN资源管理器、Spark HDFS NameNode、Hive 和 Ganglia 等应用程序的高可用性。使用此功能，主节点不再发生潜在的单点故障。如果其中一个主节点出现故障，Amazon EMR 会自动故障转移到备用主节点，并将故障主节点替换为具有相同配置和引导操作的新主节点。有关更多信息，请参阅[计划和配置主节点](#)。

## 已知问题

- Tez 用户界面 (已在亚马逊EMR版本 5.26.0 中修复)

Tez UI 不适用于具有多个主节点的EMR集群。

- Hue (已在亚马逊 5.24. EMR 0 版本中修复)

- 在亚马逊上运行的 Hue EMR 不支持 Solr。从 Amazon 5.20.0 EMR 版本开始，配置错误问题会导致 Solr 被启用，并显示类似于以下内容的无害错误消息：

```
Solr server could not be contacted properly:
HTTPConnectionPool('host=ip-xx-xx-xx-xx.ec2.internal',
port=1978): Max retries exceeded with url: /solr/admin/info/
system?user.name=hue&doAs=administrator&wt=json (Caused by
NewConnectionError(': Failed to establish a new connection: [Errno 111]
Connection refused',))
```

要防止显示 Solr 错误消息：

1. 使用连接到主节点命令行SSH。
2. 使用文本编辑器打开 hue.ini 文件。例如：

```
sudo vim /etc/hue/conf/hue.ini
```

3. 搜索术语 appblacklist，并将该行修改为以下内容：

```
appblacklist = search
```

4. 保存更改并重新启动 Hue，如以下示例所示：

```
sudo stop hue; sudo start hue
```

- 具有多个主节点的集群和 Kerberos 身份验证中的已知问题

如果您在 Amazon 5.20.0 及更高EMR版本中运行具有多个主节点和 Kerberos 身份验证的集群，则在集群运行一段时间后，集群操作可能会遇到问题，例如缩小规模或步骤提交。具体时间段取决于您定义的 Kerberos 票证有效期。缩减问题会影响您提交的自动缩减和显式缩减请求。其它集群操作也可能受到影响。

解决办法：

- SSH以hadoop用户身份访问具有多个主节点的EMR群集的主节点。
- 运行以下命令，为 hadoop 用户续订 Kerberos 票证。

```
kinit -kt <keytab_file> <principal>
```

通常情况下，keytab 文件位于 `/etc/hadoop.keytab`，而 principal 为 `hadoop/<hostname>@<REALM>` 格式。

#### Note

此解决方法将在 Kerberos 票证有效期内生效。默认情况下，此持续时间为 10 个小时，但可以通过 Kerberos 设置进行配置。Kerberos 票证过期后，您必须重新运行上述命令。

## 5.23.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
aws-sagemaker-spark-sdk	1.2.1	亚马逊 SageMaker Spark SDK
emr-ddb	4.8.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.7.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.11.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	1.2.0	EMRS3Select 连接器
emrfs	2.32.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.7.1	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。



组件	版本	描述
hadoop-client	2.8.5-amzn-3	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.8.5-amzn-3	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.5-amzn-3	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.5-amzn-3	HDFS用于跟踪文件名和区块位置的服务。
hadoop-hdfs-journalnode	2.8.5-amzn-3	HDFS用于管理 HA 集群上的 Hadoop 文件系统日志的服务。
hadoop-httpfs-server	2.8.5-amzn-3	HTTPHDFS操作端点。
hadoop-kms-server	2.8.5-amzn-3	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.8.5-amzn-3	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.8.5-amzn-3	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.8.5-amzn-3	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.5-amzn-3	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.9	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.9	为一个或多个HBase地区提供服务的服务。

组件	版本	描述
hbase-client	1.4.9	HBase命令行客户端。
hbase-rest-server	1.4.9	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.9	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	2.3.4-amzn-1	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.4-amzn-1	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.4-amzn-1	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.4-amzn-1	Hive 命令行客户端。
hive-hbase	2.3.4-amzn-1	Hive-hbase 客户端。
hive-metastore-server	2.3.4-amzn-1	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.4-amzn-1	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	4.3.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	0.9.4	Jupyter notebook 的多用户服务器

组件	版本	描述
livy-server	0.5.0-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.3.1	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
nvidia-cuda	9.2.88	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.1.0	Oozie 命令行客户端。
oozie-server	5.1.0	用于接受 Oozie 工作流请求的服务。
opencv	3.4.0	开源计算机视觉库。
phoenix-library	4.14.1--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.14.1--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.215	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.215	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。

组件	版本	描述
r	3.4.1	用于统计计算的 R 项目
spark-client	2.4.0	Spark 命令行客户端。
spark-history-server	2.4.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.4.0	的内存中执行引擎。YARN
spark-yarn-slave	2.4.0	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	1.12.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.1	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.8.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.13	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.13	ZooKeeper 命令行客户端。

### 5.23.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

## emr-5.23.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。

分类	描述
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。

分类	描述
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 JupyterHubs的 jupyterhub_config.py 文件中的值。
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。

分类	描述
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。



分类	描述
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值

分类	描述
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.22.0

### 5.22.0 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#) 和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)

- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

## 应用程序版本信息

	emr-5.22.0	emr-5.21.2	emr-5.21.1	emr-5.21.0
适用于 Java 的 AWS SDK	1.11.510	1.11.479	1.11.479	1.11.479
Python	2.7、3.6	2.7、3.6	2.7、3.6	2.7、3.6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.7.1	1.7.0	1.7.0	1.7.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.9	1.4.8	1.4.8	1.4.8
HCatalog	2.3.4	2.3.4	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.3.0	4.3.0	4.3.0	4.3.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-

	emr-5.22.0	emr-5.21.2	emr-5.21.1	emr-5.21.0
JupyterHub	0.9.4	0.9.4	0.9.4	0.9.4
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.1	1.3.1	1.3.1	1.3.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.1	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.215	0.215	0.215	0.215
Spark	2.4.0	2.4.0	2.4.0	2.4.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.12.0	1.12.0
Tez	0.9.1	0.9.1	0.9.1	0.9.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.0	0.8.0	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

## 5.22.0 发布说明

以下发行说明包含有关亚马逊 5.22. EMR 0 版本的信息。更改与 5.21.0 有关。

### Important

从亚马逊EMR版本 5.22.0 开始，亚马逊仅EMR使用 AWS 签名版本 4 来验证向亚马逊 S3 发出的请求。早期的 Amazon EMR 版本在某些情况下使用 AWS 签名版本 2，除非发行说明中注明

仅使用签名版本 4。有关更多信息，请参阅《[亚马逊简单存储服务开发者AWS 指南](#)》中的[对请求进行身份验证 \(AWS 签名版本 4\)](#) 和[对请求进行身份验证 \(签名版本 2\)](#)。

首次发布日期：2019 年 3 月 20 日

## 升级

- Flink 1.7.1
- HBase1.4.9
- Oozie 5.1.0
- Phoenix 4.14.1
- Zeppelin 0.8.1
- 连接器和驱动程序：
  - DynamoDB 连接器 4.8.0
  - MariaDB 连接器 2.2.6
  - 亚马逊 Redshift JDBC Driver 1.2.20.1043

## 新功能

- 修改了EBS仅限存储空间的EC2实例类型的默认EBS配置。当您使用 Amazon EMR 版本 5.22.0 及更高版本创建集群时，默认EBS存储量会根据实例的大小而增加。此外，我们将增加的存储空间分成多个卷，从而提高了IOPS性能。如果要使用不同的EBS实例存储配置，可以在创建EMR集群或向现有集群添加节点时指定该配置。有关默认为每种实例类型分配的存储量和卷数量的更多信息，请参阅 Amazon EMR 管理指南中的[实例默认EBS存储](#)。

## 更改、增强功能和解决的问题

- Spark
  - 在 Spark 上YARN引入了一个新的配置属性 `spark.yarn.executor.memoryOverheadFactor`。此属性的值是一个缩放系数，它将内存开销值设置为执行程序内存的百分比，最小为 384 MB。如果内存开销设置为使用 `spark.yarn.executor.memoryOverhead`，则此属性不发挥任何作用。默认值为 0.1875，表示 18.75%。与Spark内部设置的10%默认值相比，Amazon的默认值在YARN容器中为执行程序的内存开销EMR留出了更多的空间。根据经验，Amazon的EMR默认值为18.75%，表明-DS基准测试中与内存相关的故障较少。TPC

- 向后移植了 [SPARK-26316](#) 以提高性能。
- 在亚马逊EMR版本 5.19.0、5.20.0 和 5.21.0 中，YARN节点标签存储在目录中。HDFS在某些情况下，这会导致核心节点启动延迟，然后导致集群超时和启动失败。从 Amazon EMR 5.22.0 开始，此问题已得到解决。YARN节点标签存储在每个群集节点的本地磁盘上，避免依赖于HDFS。

## 已知问题

- Hue ( 已在亚马逊 5.24. EMR 0 版本中修复 )
- 在亚马逊上运行的 Hue EMR 不支持 Solr。从 Amazon 5.20.0 EMR 版本开始，配置错误问题会导致 Solr 被启用，并显示类似于以下内容的无害错误消息：

```
Solr server could not be contacted properly:
HTTPConnectionPool('host=ip-xx-xx-xx-xx.ec2.internal',
port=1978): Max retries exceeded with url: /solr/admin/info/
system?user.name=hue&doAs=administrator&wt=json (Caused by
NewConnectionError(': Failed to establish a new connection: [Errno 111]
Connection refused',))
```

要防止显示 Solr 错误消息：

1. 使用连接到主节点命令行SSH。
2. 使用文本编辑器打开 hue.ini 文件。例如：

```
sudo vim /etc/hue/conf/hue.ini
```

3. 搜索术语 appblacklist，并将该行修改为以下内容：

```
appblacklist = search
```

4. 保存更改并重新启动 Hue，如以下示例所示：

```
sudo stop hue; sudo start hue
```

- 具有多个主节点的集群和 Kerberos 身份验证中的已知问题

如果您在 Amazon 5.20.0 及更高EMR版本中运行具有多个主节点和 Kerberos 身份验证的集群，则在集群运行一段时间后，集群操作可能会遇到问题，例如缩小规模或步骤提交。具体时间段取决于您定义的 Kerberos 票证有效期。缩减问题会影响您提交的自动缩减和显式缩减请求。其它集群操作也可能受到影响。

解决办法：

- SSH以hadoop用户身份访问具有多个主节点的EMR群集的主节点。
- 运行以下命令，为 hadoop 用户续订 Kerberos 票证。

```
kinit -kt <keytab_file> <principal>
```

通常情况下，keytab 文件位于 `/etc/hadoop.keytab`，而 principal 为 `hadoop/<hostname>@<REALM>` 格式。

#### Note

此解决方法将在 Kerberos 票证有效期内生效。默认情况下，此持续时间为 10 个小时，但可以通过 Kerberos 设置进行配置。Kerberos 票证过期后，您必须重新运行上述命令。

## 5.22.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>aws-sagemaker-spark-sdk</code>	1.2.1	亚马逊 SageMaker Spark SDK
<code>emr-ddb</code>	4.8.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.6.0	适用于 Hadoop 生态系统的方便易用的库。

组件	版本	描述
emr-kinesis	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.11.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	1.2.0	EMRS3Select 连接器
emrfs	2.31.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.7.1	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.8.5-amzn-2	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.8.5-amzn-2	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.5-amzn-2	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.5-amzn-2	HDFS用于跟踪文件名和区块位置的服务。



组件	版本	描述
hadoop-httfs-server	2.8.5-amzn-2	HTTPHDFS操作端点。
hadoop-kms-server	2.8.5-amzn-2	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.8.5-amzn-2	MapReduce 用于运行 MapReduce 应用程序的执行引擎。
hadoop-yarn-nodemanager	2.8.5-amzn-2	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.8.5-amzn-2	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.5-amzn-2	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.9	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.9	为一个或多个HBase地区提供服务的服务。
hbase-client	1.4.9	HBase命令行客户端。
hbase-rest-server	1.4.9	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.9	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.3.4-amzn-1	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.4-amzn-1	提供服务HCatalog，为分布式应用程序提供表和存储管理层。

组件	版本	描述
hcatalog-webhcat-server	2.3.4-amzn-1	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.4-amzn-1	Hive 命令行客户端。
hive-hbase	2.3.4-amzn-1	Hive-hbase 客户端。
hive-metastore-server	2.3.4-amzn-1	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.4-amzn-1	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	4.3.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	0.9.4	Jupyter notebook 的多用户服务器
livy-server	0.5.0-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.3.1	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
nvidia-cuda	9.2.88	Nvidia 驱动程序和 Cuda 工具包

组件	版本	描述
oozie-client	5.1.0	Oozie 命令行客户端。
oozie-server	5.1.0	用于接受 Oozie 工作流请求的服务。
opencv	3.4.0	开源计算机视觉库。
phoenix-library	4.14.1--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.14.1--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.215	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.215	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。
r	3.4.1	用于统计计算的 R 项目
spark-client	2.4.0	Spark 命令行客户端。
spark-history-server	2.4.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.4.0	的内存中执行引擎。YARN
spark-yarn-slave	2.4.0	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。

组件	版本	描述
tensorflow	1.12.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.1	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.8.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.13	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.13	ZooKeeper 命令行客户端。

## 5.22.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

### emr-5.22.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。

分类	描述
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。

分类	描述
hcatalog-webhcat-log4j2	更改 HCatalog WebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog WebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改 HTTPFS 环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。

分类	描述
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 JupyterHubs 的 jupyterhub_config.py 文件中的值。
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。

分类	描述
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。



分类	描述
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。

分类	描述
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper 中的值。
zookeeper-log4j	更改 ZooKeeper 的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.21.2

### 5.21.2 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#) 和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-5.21.2	emr-5.21.1	emr-5.21.0	emr-5.20.1
适用于 Java 的 AWS SDK	1.11.479	1.11.479	1.11.479	1.11.461
Python	2.7、3.6	2.7、3.6	2.7、3.6	2.7、3.6
Scala	2.11.12	2.11.12	2.11.12	2.11.12

	emr-5.21.2	emr-5.21.1	emr-5.21.0	emr-5.20.1
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.7.0	1.7.0	1.7.0	1.6.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.8	1.4.8	1.4.8	1.4.8
HCatalog	2.3.4	2.3.4	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.3.0	4.3.0	4.3.0	4.3.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.9.4	0.9.4
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.1	1.3.1	1.3.1	1.3.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0

	emr-5.21.2	emr-5.21.1	emr-5.21.0	emr-5.20.1
Presto	0.215	0.215	0.215	0.214
Spark	2.4.0	2.4.0	2.4.0	2.4.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.12.0	1.12.0
Tez	0.9.1	0.9.1	0.9.1	0.9.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.8.0	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

## 5.21.2 发布说明

这是一个补丁版本，用于为向 Amazon S3 发出的请求添加 AWS 签名版本 4 身份验证。所有应用程序和组件都与之前的 Amazon EMR 版本相同。

### Important

在此版本中，亚马逊仅EMR使用 AWS 签名版本 4 来验证向 Amazon S3 发出的请求。有关更多信息，请参阅[新功能](#)。

## 5.21.2 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 `CommunityVersion-amzn-EmrVersion` 的发行版标注。`EmrVersion` 从 0 开始。例如，如果 `myapp-component` 以 2.2

版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.2.1	亚马逊 SageMaker Spark SDK
emr-ddb	4.7.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.5.1	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.11.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	1.2.0	EMRS3Select 连接器
emrfs	2.30.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.7.0	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。

组件	版本	描述
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.8.5-amzn-1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.8.5-amzn-1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.5-amzn-1	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.5-amzn-1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.8.5-amzn-1	HTTPHDFS操作端点。
hadoop-kms-server	2.8.5-amzn-1	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.8.5-amzn-1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.8.5-amzn-1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.8.5-amzn-1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.5-amzn-1	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.8	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.8	为一个或多个HBase地区提供服务的服务。

组件	版本	描述
hbase-client	1.4.8	HBase命令行客户端。
hbase-rest-server	1.4.8	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.8	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	2.3.4-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.4-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.4-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.4-amzn-0	Hive 命令行客户端。
hive-hbase	2.3.4-amzn-0	Hive-hbase 客户端。
hive-metastore-server	2.3.4-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.4-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	4.3.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	0.9.4	Jupyter notebook 的多用户服务器

组件	版本	描述
livy-server	0.5.0-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.3.1	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
nvidia-cuda	9.2.88	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.0.0	Oozie 命令行客户端。
oozie-server	5.0.0	用于接受 Oozie 工作流请求的服务。
opencv	3.4.0	开源计算机视觉库。
phoenix-library	4.14.0--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.14.0--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.215	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.215	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。



组件	版本	描述
r	3.4.1	用于统计计算的 R 项目
spark-client	2.4.0	Spark 命令行客户端。
spark-history-server	2.4.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.4.0	的内存中执行引擎。YARN
spark-yarn-slave	2.4.0	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	1.12.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.1	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.8.0	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.13	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.13	ZooKeeper 命令行客户端。

### 5.21.2 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

## emr-5.21.2 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。

分类	描述
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。

分类	描述
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 JupyterHubs的 jupyterhub_config.py 文件中的值。
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。

分类	描述
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。

分类	描述
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值

分类	描述
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.21.1

### 5.21.1 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#)和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)

- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

## 应用程序版本信息

	emr-5.21.1	emr-5.21.0	emr-5.20.1	emr-5.20.0
适用于 Java 的 AWS SDK	1.11.479	1.11.479	1.11.461	1.11.461
Python	2.7、3.6	2.7、3.6	2.7、3.6	2.7、3.6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.7.0	1.7.0	1.6.2	1.6.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.8	1.4.8	1.4.8	1.4.8
HCatalog	2.3.4	2.3.4	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.3.0	4.3.0	4.3.0	4.3.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-



	emr-5.21.1	emr-5.21.0	emr-5.20.1	emr-5.20.0
JupyterHub	0.9.4	0.9.4	0.9.4	0.9.4
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.1	1.3.1	1.3.1	1.3.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.215	0.215	0.214	0.214
Spark	2.4.0	2.4.0	2.4.0	2.4.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.12.0	1.12.0
Tez	0.9.1	0.9.1	0.9.1	0.9.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.8.0	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

## 5.21.1 发布说明

以下发行说明包含有关亚马逊 5.21.1 EMR 版本的信息。更改与 5.21.0 有关。

首次发布日期：2019 年 7 月 18 日

## 更改、增强功能和解决的问题

- 更新了亚马逊的默认 Amazon L AMI linux，EMR使其包含重要的 Linux 内核安全更新，包括 TCPSACK拒绝服务问题 ([AWS-2019-005](#))。

## 已知问题

- 具有多个主节点的集群和 Kerberos 身份验证中的已知问题

如果您在 Amazon 5.20.0 及更高EMR版本中运行具有多个主节点和 Kerberos 身份验证的集群，则在集群运行一段时间后，集群操作可能会遇到问题，例如缩小规模或步骤提交。具体时间段取决于您定义的 Kerberos 票证有效期。缩减问题会影响您提交的自动缩减和显式缩减请求。其它集群操作也可能受到影响。

### 解决办法：

- SSH以hadoop用户身份访问具有多个主节点的EMR群集的主主节点。
- 运行以下命令，为 hadoop 用户续订 Kerberos 票证。

```
kinit -kt <keytab_file> <principal>
```

通常情况下，keytab 文件位于 `/etc/hadoop.keytab`，而 principal 为 `hadoop/<hostname>@<REALM>` 格式。

### Note

此解决方法将在 Kerberos 票证有效期内生效。默认情况下，此持续时间为 10 个小时，但可以通过 Kerberos 设置进行配置。Kerberos 票证过期后，您必须重新运行上述命令。

## 5.21.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 `CommunityVersion-amzn-EmrVersion` 的发行版标注。`EmrVersion` 从 0 开始。例如，如果 `myapp-component` 以 2.2

版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.2.1	亚马逊 SageMaker Spark SDK
emr-ddb	4.7.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.5.1	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.11.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	1.2.0	EMRS3Select 连接器
emrfs	2.30.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.7.0	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。

组件	版本	描述
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.8.5-amzn-1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.8.5-amzn-1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.5-amzn-1	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.5-amzn-1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.8.5-amzn-1	HTTPHDFS操作端点。
hadoop-kms-server	2.8.5-amzn-1	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.8.5-amzn-1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.8.5-amzn-1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.8.5-amzn-1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.5-amzn-1	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.8	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.8	为一个或多个HBase地区提供服务的服务。

组件	版本	描述
hbase-client	1.4.8	HBase 命令行客户端。
hbase-rest-server	1.4.8	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.8	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	2.3.4-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.4-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.4-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.4-amzn-0	Hive 命令行客户端。
hive-hbase	2.3.4-amzn-0	Hive-hbase 客户端。
hive-metastore-server	2.3.4-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.4-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	4.3.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	0.9.4	Jupyter notebook 的多用户服务器

组件	版本	描述
livy-server	0.5.0-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.3.1	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
nvidia-cuda	9.2.88	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.0.0	Oozie 命令行客户端。
oozie-server	5.0.0	用于接受 Oozie 工作流请求的服务。
opencv	3.4.0	开源计算机视觉库。
phoenix-library	4.14.0--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.14.0--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.215	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.215	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。

组件	版本	描述
r	3.4.1	用于统计计算的 R 项目
spark-client	2.4.0	Spark 命令行客户端。
spark-history-server	2.4.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.4.0	的内存中执行引擎。YARN
spark-yarn-slave	2.4.0	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	1.12.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.1	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.8.0	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.13	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.13	ZooKeeper 命令行客户端。

### 5.21.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

## emr-5.21.1 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。



分类	描述
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。

分类	描述
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 jupyterhub_config.py 文件 JupyterHubs 中的值。
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。

分类	描述
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。

分类	描述
presto-connector-cassandra	更改 Presto 的 <code>cassandra.properties</code> 文件中的值。
presto-connector-hive	更改 Presto 的 <code>hive.properties</code> 文件中的值。
presto-connector-jmx	更改 Presto 的 <code>jmx.properties</code> 文件中的值。
presto-connector-kafka	更改 Presto 的 <code>kafka.properties</code> 文件中的值。
presto-connector-localfile	更改 Presto 的 <code>localfile.properties</code> 文件中的值。
presto-connector-memory	更改 Presto 的 <code>memory.properties</code> 文件中的值。
presto-connector-mongodb	更改 Presto 的 <code>mongodb.properties</code> 文件中的值。
presto-connector-mysql	更改 Presto 的 <code>mysql.properties</code> 文件中的值。
presto-connector-postgresql	更改 Presto 的 <code>postgresql.properties</code> 文件中的值。
presto-connector-raptor	更改 Presto 的 <code>raptor.properties</code> 文件中的值。
presto-connector-redis	更改 Presto 的 <code>redis.properties</code> 文件中的值。
presto-connector-redshift	更改 Presto 的 <code>redshift.properties</code> 文件中的值。
presto-connector-tpch	更改 Presto 的 <code>tpch.properties</code> 文件中的值。
presto-connector-tpcds	更改 Presto 的 <code>tpcds.properties</code> 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 <code>spark-defaults.conf</code> 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 <code>hive-site.xml</code> 文件中的值

分类	描述
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.21.0

### 5.21.0 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#)和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)

- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

## 应用程序版本信息

	emr-5.21.0	emr-5.20.1	emr-5.20.0	emr-5.19.1
适用于 Java 的 AWS SDK	1.11.479	1.11.461	1.11.461	1.11.433
Python	2.7、3.6	2.7、3.6	2.7、3.6	2.7、3.4
Scala	2.11.12	2.11.12	2.11.12	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.7.0	1.6.2	1.6.2	1.6.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.8	1.4.8	1.4.8	1.4.7
HCatalog	2.3.4	2.3.4	2.3.4	2.3.3
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.3
Hudi	-	-	-	-
Hue	4.3.0	4.3.0	4.3.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-

	emr-5.21.0	emr-5.20.1	emr-5.20.0	emr-5.19.1
JupyterHub	0.9.4	0.9.4	0.9.4	0.9.4
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.1	1.3.1	1.3.1	1.3.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.215	0.214	0.214	0.212
Spark	2.4.0	2.4.0	2.4.0	2.3.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.12.0	1.11.0
Tez	0.9.1	0.9.1	0.9.1	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.8.0	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

## 5.21.0 发布说明

以下发行说明包含有关亚马逊 5.21. EMR 0 版本的信息。更改与 5.20.0 有关。

首次发布日期：2019 年 2 月 18 日

上次更新时间：2019 年 4 月 3 日

## 升级

- Flink 1.7.0
- Presto 0.215
- AWS SDK for Java 1.11.479

## 新功能

- ( 2019 年 4 月 3 日 ) 在 Amazon 5.21.0 及更高EMR版本中，您可以覆盖集群配置，并为正在运行的集群中的每个实例组指定其他配置分类。为此，您可以使用 Amazon EMR 控制台、AWS Command Line Interface (AWS CLI) 或 AWS SDK。有关更多信息，请参阅[为运行的集群中的实例组提供配置](#)。

## 更改、增强功能和解决的问题

- Zeppelin
  - 向后移植 [ZEPPELIN-3878](#)。

## 已知问题

- Hue ( 已在亚马逊 5.24. EMR 0 版本中修复 )
  - 在亚马逊上运行的 Hue EMR 不支持 Solr。从 Amazon 5.20.0 EMR 版本开始，配置错误问题会导致 Solr 被启用，并显示类似于以下内容的无害错误消息：

```
Solr server could not be contacted properly:
HTTPConnectionPool('host=ip-xx-xx-xx-xx.ec2.internal',
port=1978): Max retries exceeded with url: /solr/admin/info/
system?user.name=hue&doAs=administrator&wt=json (Caused by
NewConnectionError(': Failed to establish a new connection: [Errno 111]
Connection refused',))
```

要防止显示 Solr 错误消息:

1. 使用连接到主节点命令行SSH。
2. 使用文本编辑器打开 hue.ini 文件。例如：

```
sudo vim /etc/hue/conf/hue.ini
```



3. 搜索术语 `appblacklist`，并将该行修改为以下内容：

```
appblacklist = search
```

4. 保存更改并重新启动 Hue，如以下示例所示：

```
sudo stop hue; sudo start hue
```

- Tez

- 此问题已在亚马逊 EMR 5.22.0 中修复。

当你通过 `http://` 连接到 Tez 用户界面时 `MasterDNS:8080/tez-ui` 通过 SSH 连接到群集主节点，出现错误“适配器操作失败-时间轴服务器 () ATS 遥不可及。要么已关闭，要么 CORS 出现“未启用”，要么任务意外显示 N/A。

这是由于 Tez UI 使用 `localhost` 而不是主节点的主机名向 YARN 时间轴服务器发出请求所致。解决方法：将脚本作为引导操作或步骤运行。脚本更新 Tez `configs.env` 文件中的主机名。有关更多信息以及脚本的位置信息，请参阅[引导说明](#)。

- 在亚马逊 EMR 版本 5.19.0、5.20.0 和 5.21.0 中，YARN 节点标签存储在目录中。HDFS 在某些情况下，这会导致核心节点启动延迟，然后导致集群超时和启动失败。从 Amazon EMR 5.22.0 开始，此问题已得到解决。YARN 节点标签存储在每个群集节点的本地磁盘上，从而避免依赖于 HDFS。
- 具有多个主节点的集群和 Kerberos 身份验证中的已知问题

如果您在 Amazon 5.20.0 及更高 EMR 版本中运行具有多个主节点和 Kerberos 身份验证的集群，则在集群运行一段时间后，集群操作可能会遇到问题，例如缩小规模或步骤提交。具体时间段取决于您定义的 Kerberos 票证有效期。缩减问题会影响您提交的自动缩减和显式缩减请求。其它集群操作也可能受到影响。

解决办法：

- SSH 以 `hadoop` 用户身份访问具有多个主节点的 EMR 群集的主节点。
- 运行以下命令，为 `hadoop` 用户续订 Kerberos 票证。

```
kinit -kt <keytab_file> <principal>
```

通常情况下，`keytab` 文件位于 `/etc/hadoop.keytab`，而 `principal` 为 `hadoop/<hostname>@<REALM>` 格式。

**Note**

此解决方法将在 Kerberos 票证有效期内生效。默认情况下，此持续时间为 10 个小时，但可以通过 Kerberos 设置进行配置。Kerberos 票证过期后，您必须重新运行上述命令。

### 5.21.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>aws-sagemaker-spark-sdk</code>	1.2.1	亚马逊 SageMaker Spark SDK
<code>emr-ddb</code>	4.7.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.5.1	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
<code>emr-s3-dist-cp</code>	2.11.0	针对 Amazon S3 优化的分布式复制应用程序。
<code>emr-s3-select</code>	1.2.0	EMRS3Select 连接器

组件	版本	描述
emrfs	2.30.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.7.0	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.8.5-amzn-1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.8.5-amzn-1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.5-amzn-1	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.5-amzn-1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.8.5-amzn-1	HTTPHDFS操作端点。
hadoop-kms-server	2.8.5-amzn-1	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.8.5-amzn-1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。

组件	版本	描述
hadoop-yarn-nodemanager	2.8.5-amzn-1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.8.5-amzn-1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.5-amzn-1	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.8	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.8	为一个或多个HBase地区提供服务的服务。
hbase-client	1.4.8	HBase命令行客户端。
hbase-rest-server	1.4.8	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.8	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.3.4-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.4-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.4-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.4-amzn-0	Hive 命令行客户端。
hive-hbase	2.3.4-amzn-0	Hive-hbase 客户端。

组件	版本	描述
hive-metastore-server	2.3.4-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的 SQL 元数据。
hive-server2	2.3.4-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	4.3.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	0.9.4	Jupyter notebook 的多用户服务器
livy-server	0.5.0-incubating	REST 用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.3.1	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的 SQL 数据库服务器。
nvidia-cuda	9.2.88	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.0.0	Oozie 命令行客户端。
oozie-server	5.0.0	用于接受 Oozie 工作流请求的服务。
opencv	3.4.0	开源计算机视觉库。

组件	版本	描述
phoenix-library	4.14.0--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.14.0--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.215	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.215	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。
r	3.4.1	用于统计计算的 R 项目
spark-client	2.4.0	Spark 命令行客户端。
spark-history-server	2.4.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.4.0	的内存中执行引擎。YARN
spark-yarn-slave	2.4.0	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	1.12.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.1	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。

组件	版本	描述
zeppelin-server	0.8.0	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.13	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.13	ZooKeeper 命令行客户端。

## 5.21.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

### emr-5.21.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。

分类	描述
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。



分类	描述
hcatalog-webhcat-site	更改 HCatalog WebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。

分类	描述
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 jupyterhub_config.py 文件 JupyterHubs 中的值。
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。

分类	描述
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。

分类	描述
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。

分类	描述
zookeeper-log4j	更改 ZooKeeper 的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.20.1

### 5.20.1 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#) 和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-5.20.1	emr-5.20.0	emr-5.19.1	emr-5.19.0
适用于 Java 的 AWS SDK	1.11.461	1.11.461	1.11.433	1.11.433
Python	2.7、3.6	2.7、3.6	2.7、3.4	2.7、3.4
Scala	2.11.12	2.11.12	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-

	emr-5.20.1	emr-5.20.0	emr-5.19.1	emr-5.19.0
Flink	1.6.2	1.6.2	1.6.1	1.6.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.8	1.4.8	1.4.7	1.4.7
HCatalog	2.3.4	2.3.4	2.3.3	2.3.3
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.3.0	4.3.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.9.4	0.9.4
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.1	1.3.1	1.3.0	1.3.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.214	0.214	0.212	0.212
Spark	2.4.0	2.4.0	2.3.2	2.3.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7

	emr-5.20.1	emr-5.20.0	emr-5.19.1	emr-5.19.0
TensorFlow	1.12.0	1.12.0	1.11.0	1.11.0
Tez	0.9.1	0.9.1	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.8.0	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

### 5.20.1 发布说明

这是一个补丁版本，用于为向 Amazon S3 发出的请求添加 AWS 签名版本 4 身份验证。所有应用程序和组件都与之前的 Amazon EMR 版本相同。

#### Important

在此版本中，亚马逊仅EMR使用 AWS 签名版本 4 来验证向 Amazon S3 发出的请求。有关更多信息，请参阅[新功能](#)。

### 5.20.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.2.1	亚马逊 SageMaker Spark SDK
emr-ddb	4.7.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.5.1	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.10.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	1.2.0	EMRS3Select 连接器
emrfs	2.29.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.6.2	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。



组件	版本	描述
hadoop-client	2.8.5-amzn-1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.8.5-amzn-1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.5-amzn-1	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.5-amzn-1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-https-server	2.8.5-amzn-1	HTTPHDFS操作端点。
hadoop-kms-server	2.8.5-amzn-1	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.8.5-amzn-1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.8.5-amzn-1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.8.5-amzn-1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.5-amzn-1	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.8	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.8	为一个或多个HBase地区提供服务的服务。
hbase-client	1.4.8	HBase命令行客户端。

组件	版本	描述
hbase-rest-server	1.4.8	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.8	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	2.3.4-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.4-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.4-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.4-amzn-0	Hive 命令行客户端。
hive-hbase	2.3.4-amzn-0	Hive-hbase 客户端。
hive-metastore-server	2.3.4-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.4-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	4.3.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	0.9.4	Jupyter notebook 的多用户服务器
livy-server	0.5.0-incubating	REST用于与 Apache Spark 交互的接口

组件	版本	描述
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.3.1	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
nvidia-cuda	9.2.88	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.0.0	Oozie 命令行客户端。
oozie-server	5.0.0	用于接受 Oozie 工作流请求的服务。
opencv	3.4.0	开源计算机视觉库。
phoenix-library	4.14.0--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.14.0--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.214	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.214	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。
r	3.4.1	用于统计计算的 R 项目
spark-client	2.4.0	Spark 命令行客户端。

组件	版本	描述
spark-history-server	2.4.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.4.0	的内存中执行引擎。YARN
spark-yarn-slave	2.4.0	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	1.12.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.1	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.8.0	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.13	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.13	ZooKeeper 命令行客户端。

### 5.20.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

## emr-5.20.1 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。

分类	描述
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。

分类	描述
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 jupyterhub_config.py 文件 JupyterHubs 中的值。
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。

分类	描述
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。



分类	描述
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值

分类	描述
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.20.0

### 5.20.0 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#) 和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)

- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

## 应用程序版本信息

	emr-5.20.0	emr-5.19.1	emr-5.19.0	emr-5.18.1
适用于 Java 的 AWS SDK	1.11.461	1.11.433	1.11.433	1.11.393
Python	2.7、3.6	2.7、3.4	2.7、3.4	2.7、3.4
Scala	2.11.12	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.6.2	1.6.1	1.6.1	1.6.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.8	1.4.7	1.4.7	1.4.7
HCatalog	2.3.4	2.3.3	2.3.3	2.3.3
Hadoop	2.8.5	2.8.5	2.8.5	2.8.4
Hive	2.3.4	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.3.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-

	emr-5.20.0	emr-5.19.1	emr-5.19.0	emr-5.18.1
JupyterHub	0.9.4	0.9.4	0.9.4	0.8.1
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.1	1.3.0	1.3.0	1.2.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.214	0.212	0.212	0.210
Spark	2.4.0	2.3.2	2.3.2	2.3.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.11.0	1.11.0	1.9.0
Tez	0.9.1	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.8.0	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.12

## 5.20.0 发布说明

以下发行说明包含有关亚马逊 5.20. EMR 0 版本的信息。更改与 5.19.0 有关。

首次发布日期：2018 年 12 月 18 日

上次更新时间：2019 年 1 月 22 日

## 升级

- Flink 1.6.2
- HBase1.4.8
- Hive 2.3.4
- Hue 4.3.0
- MXNet1.3.1
- Presto 0.214
- Spark 2.4.0
- TensorFlow 1.12.0
- Tez 0.9.1
- AWS SDK for Java 1.11.461

## 新功能

- (2019年1月22日) 亚马逊中的 Kerberos EMR 已得到改进，支持从外部对委托人进行身份验证。KDC这可以集中管理委托人，因为多个集群可以共享一个外部KDC集群。此外，外部KDC用户可以与 Active Directory 域建立跨领域信任。这使得所有集群可以从 Active Directory 对委托人进行身份验证。有关更多信息，请参阅《亚马逊EMR管理指南》中的“[使用 Kerberos 身份验证](#)”。

## 更改、增强功能和解决的问题

- 亚马逊的默认 Amazon AMI Linux EMR
  - Python 3 软件包已从 Python 3.4 升级到 3.6。
- EMRFSS3 优化的提交者
  - EMRFSS3 优化的提交器现在默认处于启用状态，这提高了写入性能。有关更多信息，请参阅 [使用 EMRFS S3 优化的提交器](#)。
- Hive
  - 向后移植 [HIVE-16686](#)。
- 集成 Spark 和 Hive 的 Glue
  - 在 EMR 5.20.0 或更高版本中，当使用 G AWS Iue Data Catalog 作为元数据存储时，Spark 和 Hive 会自动启用并行分区修剪。此更改通过并行执行多个请求来检索分区，显著缩短查询计划时间。可同时执行的分段总数介于 1 到 10 之间。默认值为 5，这是建议的设置。您可以通过以下方式更改该值：指定 hive-site 配置分类中的属性 `aws.glue.partition.num.segments`。如

果发生节流，则可以通过将值更改为 1 来关闭此功能。有关更多信息，请参阅 [AWS Glue 分段结构](#)。

## 已知问题

- Hue (已在亚马逊 5.24. EMR 0 版本中修复)

- 在亚马逊上运行的 Hue EMR 不支持 Solr。从 Amazon 5.20.0 EMR 版本开始，配置错误问题会导致 Solr 被启用，并显示类似于以下内容的无害错误消息：

```
Solr server could not be contacted properly:
HTTPConnectionPool('host=ip-xx-xx-xx-xx.ec2.internal',
port=1978): Max retries exceeded with url: /solr/admin/info/
system?user.name=hue&doAs=administrator&wt=json (Caused by
NewConnectionError(': Failed to establish a new connection: [Errno 111]
Connection refused',))
```

要防止显示 Solr 错误消息：

1. 使用连接到主节点命令行SSH。
2. 使用文本编辑器打开 hue.ini 文件。例如：

```
sudo vim /etc/hue/conf/hue.ini
```

3. 搜索术语 appblacklist，并将该行修改为以下内容：

```
appblacklist = search
```

4. 保存更改并重新启动 Hue，如以下示例所示：

```
sudo stop hue; sudo start hue
```

- Tez

- 此问题已在亚马逊 EMR 5.22.0 中修复。

当你通过 http:// 连接到 Tez 用户界面时 *MasterDNS*:8080/tez-ui 通过SSH连接到群集主节点，出现错误“适配器操作失败-时间轴服务器 () ATS 遥不可及。要么已关闭，要么CORS出现“未启用”，要么任务意外显示 N/A。

这是由于 Tez UI 使用localhost而不是主节点的主机名向YARN时间轴服务器发出请求所致。解决方法：将脚本作为引导操作或步骤运行。脚本更新 Tez configs.env 文件中的主机名。有关更多信息以及脚本的位置信息，请参阅[引导说明](#)。

- 在亚马逊EMR版本 5.19.0、5.20.0 和 5.21.0 中，YARN节点标签存储在目录中。HDFS在某些情况下，这会导致核心节点启动延迟，然后导致集群超时和启动失败。从 Amazon EMR 5.22.0 开始，此问题已得到解决。YARN节点标签存储在每个群集节点的本地磁盘上，从而避免依赖于HDFS。
- 具有多个主节点的集群和 Kerberos 身份验证中的已知问题

如果您在 Amazon 5.20.0 及更高EMR版本中运行具有多个主节点和 Kerberos 身份验证的集群，则在集群运行一段时间后，集群操作可能会遇到问题，例如缩小规模或步骤提交。具体时间段取决于您定义的 Kerberos 票证有效期。缩减问题会影响您提交的自动缩减和显式缩减请求。其它集群操作也可能受到影响。

解决办法：

- SSH以hadoop用户身份访问具有多个主节点的EMR群集的主节点。
- 运行以下命令，为 hadoop 用户续订 Kerberos 票证。

```
kinit -kt <keytab_file> <principal>
```

通常情况下，keytab 文件位于 /etc/hadoop.keytab，而 principal 为 hadoop/<hostname>@<REALM> 格式。

#### Note

此解决方法将在 Kerberos 票证有效期内生效。默认情况下，此持续时间为 10 个小时，但可以通过 Kerberos 设置进行配置。Kerberos 票证过期后，您必须重新运行上述命令。

## 5.20.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 emr 或 aws 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2

版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.2.1	亚马逊 SageMaker Spark SDK
emr-ddb	4.7.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.5.1	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.10.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	1.2.0	EMRS3Select 连接器
emrfs	2.29.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.6.2	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。



组件	版本	描述
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.8.5-amzn-1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.8.5-amzn-1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.5-amzn-1	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.5-amzn-1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.8.5-amzn-1	HTTPHDFS操作端点。
hadoop-kms-server	2.8.5-amzn-1	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.8.5-amzn-1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.8.5-amzn-1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.8.5-amzn-1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.5-amzn-1	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.8	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.8	为一个或多个HBase地区提供服务的服务。

组件	版本	描述
hbase-client	1.4.8	HBase命令行客户端。
hbase-rest-server	1.4.8	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.8	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.3.4-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.4-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.4-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.4-amzn-0	Hive 命令行客户端。
hive-hbase	2.3.4-amzn-0	Hive-hbase 客户端。
hive-metastore-server	2.3.4-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.4-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	4.3.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	0.9.4	Jupyter notebook 的多用户服务器

组件	版本	描述
livy-server	0.5.0-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.3.1	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
nvidia-cuda	9.2.88	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.0.0	Oozie 命令行客户端。
oozie-server	5.0.0	用于接受 Oozie 工作流请求的服务。
opencv	3.4.0	开源计算机视觉库。
phoenix-library	4.14.0--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.14.0--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.214	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.214	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。

组件	版本	描述
r	3.4.1	用于统计计算的 R 项目
spark-client	2.4.0	Spark 命令行客户端。
spark-history-server	2.4.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.4.0	的内存中执行引擎。YARN
spark-yarn-slave	2.4.0	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	1.12.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.9.1	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.8.0	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.13	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.13	ZooKeeper 命令行客户端。

### 5.20.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

## emr-5.20.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。

分类	描述
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。

分类	描述
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 jupyterhub_config.py 文件 JupyterHubs 中的值。
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。

分类	描述
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。



分类	描述
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值

分类	描述
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.19.1

### 5.19.1 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#)和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)

- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

## 应用程序版本信息

	emr-5.19.1	emr-5.19.0	emr-5.18.1	emr-5.18.0
适用于 Java 的 AWS SDK	1.11.433	1.11.433	1.11.393	1.11.393
Python	2.7、3.4	2.7、3.4	2.7、3.4	2.7、3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.6.1	1.6.1	1.6.0	1.6.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.7	1.4.7	1.4.7	1.4.7
HCatalog	2.3.3	2.3.3	2.3.3	2.3.3
Hadoop	2.8.5	2.8.5	2.8.4	2.8.4
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-

	emr-5.19.1	emr-5.19.0	emr-5.18.1	emr-5.18.0
JupyterHub	0.9.4	0.9.4	0.8.1	0.8.1
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.0	1.3.0	1.2.0	1.2.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.212	0.212	0.210	0.210
Spark	2.3.2	2.3.2	2.3.2	2.3.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.11.0	1.11.0	1.9.0	1.9.0
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.8.0	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.12	3.4.12

### 5.19.1 发布说明

这是一个补丁版本，用于为向 Amazon S3 发出的请求添加 AWS 签名版本 4 身份验证。所有应用程序和组件都与之前的 Amazon EMR 版本相同。

**⚠ Important**

在此版本中，亚马逊仅EMR使用 AWS 签名版本 4 来验证向 Amazon S3 发出的请求。有关更多信息，请参阅[新功能](#)。

### 5.19.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>aws-sagemaker-spark-sdk</code>	1.2.0	亚马逊 SageMaker Spark SDK
<code>emr-ddb</code>	4.7.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.5.1	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
<code>emr-s3-dist-cp</code>	2.10.0	针对 Amazon S3 优化的分布式复制应用程序。
<code>emr-s3-select</code>	1.1.0	EMRS3Select 连接器

组件	版本	描述
emrfs	2.28.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.6.1	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.8.5-amzn-0	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.8.5-amzn-0	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.5-amzn-0	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.5-amzn-0	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.8.5-amzn-0	HTTPHDFS操作端点。
hadoop-kms-server	2.8.5-amzn-0	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.8.5-amzn-0	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。

组件	版本	描述
hadoop-yarn-nodemanager	2.8.5-amzn-0	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.8.5-amzn-0	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.5-amzn-0	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.7	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.7	为一个或多个HBase地区提供服务的服务。
hbase-client	1.4.7	HBase命令行客户端。
hbase-rest-server	1.4.7	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.7	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.3.3-amzn-2	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.3-amzn-2	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.3-amzn-2	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.3-amzn-2	Hive 命令行客户端。
hive-hbase	2.3.3-amzn-2	Hive-hbase 客户端。

组件	版本	描述
hive-metastore-server	2.3.3-amzn-2	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的 SQL 元数据。
hive-server2	2.3.3-amzn-2	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	4.2.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	0.9.4	Jupyter notebook 的多用户服务器
livy-server	0.5.0-incubating	REST 用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.3.0	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的 SQL 数据库服务器。
nvidia-cuda	9.2.88	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.0.0	Oozie 命令行客户端。
oozie-server	5.0.0	用于接受 Oozie 工作流请求的服务。
opencv	3.4.0	开源计算机视觉库。



组件	版本	描述
phoenix-library	4.14.0--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.14.0--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.212	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.212	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。
r	3.4.1	用于统计计算的 R 项目
spark-client	2.3.2	Spark 命令行客户端。
spark-history-server	2.3.2	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.3.2	的内存中执行引擎。YARN
spark-yarn-slave	2.3.2	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	1.11.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。

组件	版本	描述
zeppelin-server	0.8.0	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.13	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.13	ZooKeeper 命令行客户端。

### 5.19.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.19.1 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。

分类	描述
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。

分类	描述
hcatalog-webhcat-site	更改 HCatalog WebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。

分类	描述
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 jupyterhub_config.py 文件 JupyterHubs 中的值。
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。

分类	描述
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。

分类	描述
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。

分类	描述
zookeeper-log4j	更改 ZooKeeper 的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.19.0

### 5.19.0 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#) 和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-5.19.0	emr-5.18.1	emr-5.18.0	emr-5.17.2
适用于 Java 的 AWS SDK	1.11.433	1.11.393	1.11.393	1.11.336
Python	2.7、3.4	2.7、3.4	2.7、3.4	2.7、3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-



	emr-5.19.0	emr-5.18.1	emr-5.18.0	emr-5.17.2
Flink	1.6.1	1.6.0	1.6.0	1.5.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.7	1.4.7	1.4.7	1.4.6
HCatalog	2.3.3	2.3.3	2.3.3	2.3.3
Hadoop	2.8.5	2.8.4	2.8.4	2.8.4
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.8.1	0.8.1	0.8.1
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.3.0	1.2.0	1.2.0	1.2.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.212	0.210	0.210	0.206
Spark	2.3.2	2.3.2	2.3.2	2.3.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7

	emr-5.19.0	emr-5.18.1	emr-5.18.0	emr-5.17.2
TensorFlow	1.11.0	1.9.0	1.9.0	1.9.0
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.8.0	0.7.3
ZooKeeper	3.4.13	3.4.12	3.4.12	3.4.12

## 5.19.0 发布说明

以下发行说明包含有关亚马逊 5.19.0 EMR 版本的信息。更改与 5.18.0 有关。

首次发布日期：2018 年 11 月 7 日

上次更新时间：2018 年 11 月 19 日

### 升级

- Hadoop 2.8.5
- Flink 1.6.1
- JupyterHub 0.9.4
- MXNet1.3.0
- Presto 0.212
- TensorFlow 1.11.0
- Zookeeper 3.4.13
- AWS SDK for Java 1.11.433

### 新功能

- (2018 年 11 月 19 日) EMR 笔记本是一个基于 Jupyter Notebook 的托管环境。它支持 Spark PySpark、Spark R 和 Scala 的 Spark 魔法内核。SQLEMR 笔记本可以与使用 Amazon EMR 版本 5.18.0 及更高版本创建的集群一起使用。有关更多信息，请参阅《Amazon EMR 管理指南》中的[“使用EMR笔记本”](#)。

- 使用 Spark 和 Parquet 文件写入 Parquet 文件时，可以使用 EMRFS S3 优化的提交器。EMRFS 此提交程序改进了写入性能。有关更多信息，请参阅[使用 EMRFS S3 优化的提交器](#)。

## 更改、增强功能和解决的问题

- YARN
  - 修改了限制应用程序主进程在核心节点上运行的逻辑。此功能现在使用 YARN 节点标签功能和 `yarn-site` 和 `capacity-scheduler` 配置分类中的属性。有关信息，请参阅。<https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-plan-instances-guidelines.html#emr-plan-spot-YARN>。
- 亚马逊的默认 Amazon AMI Linux EMR
  - 默认情况下，不再安装 `ruby18`、`php56` 和 `gcc48`。如果需要，可以使用 `yum` 安装它们。
  - 默认情况下，不再安装 `aws-sdk ruby gem`。如果需要，可以使用 `gem install aws-sdk` 进行安装。此外，还可以安装特定组件。例如，`gem install aws-sdk-s3`。

## 已知问题

- EMR 笔记本-在某些情况下，当多个笔记本编辑器处于打开状态时，笔记本编辑器可能无法连接到集群。如果发生这种情况，请清除浏览器 Cookie，然后重新打开笔记本编辑器。
- CloudWatch ContainerPending 指标和自动缩放 — (在 5.20.0 中已修复) Amazon EMR 可能会发出负值。ContainerPending 如果在自动伸缩规则中使用 ContainerPending，自动伸缩的行为方式可能会不符合预期。请避免在自动伸缩中使用 ContainerPending。
- 在亚马逊 EMR 版本 5.19.0、5.20.0 和 5.21.0 中，YARN 节点标签存储在目录中。HDFS 在某些情况下，这会导致核心节点启动延迟，然后导致集群超时和启动失败。从 Amazon EMR 5.22.0 开始，此问题已得到解决。YARN 节点标签存储在每个群集节点的本地磁盘上，从而避免依赖于 HDFS。

## 5.19.0 组件版本

下面列出了 Amazon 随此版本 EMR 安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的 EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新 EMR 版本中的大数据应用程序包通常是社区中的最新版本。我们会 EMR 尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本 EMR 不同。这些组件具有以下形式的 `CommunityVersion-amzn-EmrVersion` 的发行版标注。`EmrVersion` 从 0 开始。例如，如果 `myapp-component` 以 2.2

版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.2.0	亚马逊 SageMaker Spark SDK
emr-ddb	4.7.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.5.1	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.10.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	1.1.0	EMRS3Select 连接器
emrfs	2.28.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.6.1	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。

组件	版本	描述
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.8.5-amzn-0	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.8.5-amzn-0	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.5-amzn-0	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.5-amzn-0	HDFS用于跟踪文件名和区块位置的服务。
hadoop-https-server	2.8.5-amzn-0	HTTPHDFS操作端点。
hadoop-kms-server	2.8.5-amzn-0	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.8.5-amzn-0	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.8.5-amzn-0	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.8.5-amzn-0	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.5-amzn-0	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.7	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.7	为一个或多个HBase地区提供服务的服务。

组件	版本	描述
hbase-client	1.4.7	HBase命令行客户端。
hbase-rest-server	1.4.7	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.7	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.3.3-amzn-2	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.3-amzn-2	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.3-amzn-2	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.3-amzn-2	Hive 命令行客户端。
hive-hbase	2.3.3-amzn-2	Hive-hbase 客户端。
hive-metastore-server	2.3.3-amzn-2	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.3-amzn-2	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	4.2.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	0.9.4	Jupyter notebook 的多用户服务器

组件	版本	描述
livy-server	0.5.0-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.3.0	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
nvidia-cuda	9.2.88	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.0.0	Oozie 命令行客户端。
oozie-server	5.0.0	用于接受 Oozie 工作流请求的服务。
opencv	3.4.0	开源计算机视觉库。
phoenix-library	4.14.0--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.14.0--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.212	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.212	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。

组件	版本	描述
r	3.4.1	用于统计计算的 R 项目
spark-client	2.3.2	Spark 命令行客户端。
spark-history-server	2.3.2	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.3.2	的内存中执行引擎。YARN
spark-yarn-slave	2.3.2	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	1.11.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.8.0	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.13	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.13	ZooKeeper 命令行客户端。

### 5.19.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。



## emr-5.19.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。

分类	描述
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。

分类	描述
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 jupyterhub_config.py 文件 JupyterHubs 中的值。
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。

分类	描述
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。

分类	描述
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-memory	更改 Presto 的 memory.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
presto-connector-tpcds	更改 Presto 的 tpcds.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值

分类	描述
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.18.1

### 5.18.1 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#)和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)

- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

## 应用程序版本信息

	emr-5.18.1	emr-5.18.0	emr-5.17.2	emr-5.17.1
适用于 Java 的 AWS SDK	1.11.393	1.11.393	1.11.336	1.11.336
Python	2.7、3.4	2.7、3.4	2.7、3.4	2.7、3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.6.0	1.6.0	1.5.2	1.5.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.7	1.4.7	1.4.6	1.4.6
HCatalog	2.3.3	2.3.3	2.3.3	2.3.3
Hadoop	2.8.4	2.8.4	2.8.4	2.8.4
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-

	emr-5.18.1	emr-5.18.0	emr-5.17.2	emr-5.17.1
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.2.0	1.2.0	1.2.0	1.2.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.210	0.210	0.206	0.206
Spark	2.3.2	2.3.2	2.3.1	2.3.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.9.0	1.9.0	1.9.0	1.9.0
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.12	3.4.12	3.4.12

### 5.18.1 发布说明

这是一个补丁版本，用于为向 Amazon S3 发出的请求添加 AWS 签名版本 4 身份验证。所有应用程序和组件都与之前的 Amazon EMR 版本相同。



**⚠ Important**

在此版本中，亚马逊仅EMR使用 AWS 签名版本 4 来验证向 Amazon S3 发出的请求。有关更多信息，请参阅[新功能](#)。

### 5.18.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>aws-sagemaker-spark-sdk</code>	1.1.3	亚马逊 SageMaker Spark SDK
<code>emr-ddb</code>	4.6.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.5.0	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
<code>emr-s3-dist-cp</code>	2.10.0	针对 Amazon S3 优化的分布式复制应用程序。
<code>emr-s3-select</code>	1.1.0	EMRS3Select 连接器

组件	版本	描述
emrfs	2.27.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.6.0	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.8.4-amzn-1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.8.4-amzn-1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.4-amzn-1	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.4-amzn-1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.8.4-amzn-1	HTTPHDFS操作端点。
hadoop-kms-server	2.8.4-amzn-1	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.8.4-amzn-1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。

组件	版本	描述
hadoop-yarn-nodemanager	2.8.4-amzn-1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.8.4-amzn-1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.4-amzn-1	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.7	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.7	为一个或多个HBase地区提供服务的服务。
hbase-client	1.4.7	HBase命令行客户端。
hbase-rest-server	1.4.7	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.7	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.3.3-amzn-2	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.3-amzn-2	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.3-amzn-2	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.3-amzn-2	Hive 命令行客户端。
hive-hbase	2.3.3-amzn-2	Hive-hbase 客户端。

组件	版本	描述
hive-metastore-server	2.3.3-amzn-2	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的 SQL 元数据。
hive-server2	2.3.3-amzn-2	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	4.2.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	0.8.1	Jupyter notebook 的多用户服务器
livy-server	0.5.0-incubating	REST 用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.2.0	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的 SQL 数据库服务器。
nvidia-cuda	9.2.88	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.0.0	Oozie 命令行客户端。
oozie-server	5.0.0	用于接受 Oozie 工作流请求的服务。
opencv	3.4.0	开源计算机视觉库。

组件	版本	描述
phoenix-library	4.14.0--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.14.0--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.210	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.210	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。
r	3.4.1	用于统计计算的 R 项目
spark-client	2.3.2	Spark 命令行客户端。
spark-history-server	2.3.2	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.3.2	的内存中执行引擎。YARN
spark-yarn-slave	2.3.2	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	1.9.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。

组件	版本	描述
zeppelin-server	0.8.0	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.12	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.12	ZooKeeper 命令行客户端。

### 5.18.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.18.1 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。

分类	描述
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。

分类	描述
hcatalog-webhcat-site	更改 HCatalog WebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。



分类	描述
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 jupyterhub_config.py 文件 JupyterHubs 中的值。
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。

分类	描述
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。

分类	描述
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.18.0

### 5.18.0 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#)和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-5.18.0	emr-5.17.2	emr-5.17.1	emr-5.17.0
适用于 Java 的 AWS SDK	1.11.393	1.11.336	1.11.336	1.11.336
Python	2.7、3.4	2.7、3.4	2.7、3.4	2.7、3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.6.0	1.5.2	1.5.2	1.5.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.7	1.4.6	1.4.6	1.4.6

	emr-5.18.0	emr-5.17.2	emr-5.17.1	emr-5.17.0
HCatalog	2.3.3	2.3.3	2.3.3	2.3.3
Hadoop	2.8.4	2.8.4	2.8.4	2.8.4
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.2.0	1.2.0	1.2.0	1.2.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.210	0.206	0.206	0.206
Spark	2.3.2	2.3.1	2.3.1	2.3.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.9.0	1.9.0	1.9.0	1.9.0
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.18.0	emr-5.17.2	emr-5.17.1	emr-5.17.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.12	3.4.12	3.4.12

## 5.18.0 发布说明

以下发行说明包含有关亚马逊 5.18. EMR 0 版本的信息。更改与 5.17.0 有关。

首次发布日期：2018 年 10 月 24 日

### 升级

- Flink 1.6.0
- HBase1.4.7
- Presto 0.210
- Spark 2.3.2
- Zeppelin 0.8.0

### 新功能

- 从 Amazon EMR 5.18.0 开始，您可以使用 Amazon 构EMR件存储库针对特定亚马逊版本中提供的库和依赖项的确切版本构建任务代码。EMR有关更多信息，请参阅 [使用 Amazon EMR 项目存储库检查依赖项](#)。

### 更改、增强功能和解决的问题

- Hive
  - 添加了对 S3 Select 的支持。有关更多信息，请参阅[将 S3 Select 与 Hive 结合使用以提高查询性能](#)。
- Presto
  - 添加了对 [S3 Select](#) Pushdown 的支持。有关更多信息，请参阅[使用 S3 Select Pushdown 搭配 Presto 提高性能](#)。

- Spark
  - Spark 的默认 log4j 配置已更改为 Spark Streaming 任务每小时的滚动容器日志。这有助于防止删除长时间运行的 Spark Streaming 任务的日志。

### 5.18.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>aws-sagemaker-spark-sdk</code>	1.1.3	亚马逊 SageMaker Spark SDK
<code>emr-ddb</code>	4.6.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.5.0	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
<code>emr-s3-dist-cp</code>	2.10.0	针对 Amazon S3 优化的分布式复制应用程序。
<code>emr-s3-select</code>	1.1.0	EMRS3Select 连接器
<code>emrfs</code>	2.27.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。

组件	版本	描述
flink-client	1.6.0	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.8.4-amzn-1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.8.4-amzn-1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.4-amzn-1	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.4-amzn-1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.8.4-amzn-1	HTTPHDFS操作端点。
hadoop-kms-server	2.8.4-amzn-1	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.8.4-amzn-1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.8.4-amzn-1	YARN用于管理单个节点上的容器的服务。



组件	版本	描述
hadoop-yarn-resourcemanager	2.8.4-amzn-1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.4-amzn-1	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.7	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.7	为一个或多个HBase地区提供服务的服务。
hbase-client	1.4.7	HBase命令行客户端。
hbase-rest-server	1.4.7	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.7	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.3.3-amzn-2	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.3-amzn-2	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.3-amzn-2	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.3-amzn-2	Hive 命令行客户端。
hive-hbase	2.3.3-amzn-2	Hive-hbase 客户端。
hive-metastore-server	2.3.3-amzn-2	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。

组件	版本	描述
hive-server2	2.3.3-amzn-2	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	4.2.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	0.8.1	Jupyter notebook 的多用户服务器
livy-server	0.5.0-incubating	REST用于与 Apache Spark 交互的接口
nginx	1.12.1	nginx [engine x] HTTP 是一个反向代理服务器
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.2.0	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
nvidia-cuda	9.2.88	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.0.0	Oozie 命令行客户端。
oozie-server	5.0.0	用于接受 Oozie 工作流请求的服务。
opencv	3.4.0	开源计算机视觉库。
phoenix-library	4.14.0--1.4 HBase	服务器和客户端的 phoenix 库

组件	版本	描述
phoenix-query-server	4.14.0--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.210	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.210	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。
r	3.4.1	用于统计计算的 R 项目
spark-client	2.3.2	Spark 命令行客户端。
spark-history-server	2.3.2	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.3.2	的内存中执行引擎。YARN
spark-yarn-slave	2.3.2	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	1.9.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.8.0	支持交互式数据分析的基于 Web 的笔记本电脑。

组件	版本	描述
zookeeper-server	3.4.12	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.12	ZooKeeper 命令行客户端。

## 5.18.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

### emr-5.18.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。

分类	描述
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。

分类	描述
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 jupyterhub_config.py 文件 JupyterHubs 中的值。

分类	描述
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。

分类	描述
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。



分类	描述
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.17.2

### 5.17.2 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#)和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-5.17.2	emr-5.17.1	emr-5.17.0	emr-5.16.1
适用于 Java 的 AWS SDK	1.11.336	1.11.336	1.11.336	1.11.336
Python	2.7、3.4	2.7、3.4	2.7、3.4	2.7、3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.5.2	1.5.2	1.5.2	1.5.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.6	1.4.6	1.4.6	1.4.4

	emr-5.17.2	emr-5.17.1	emr-5.17.0	emr-5.16.1
HCatalog	2.3.3	2.3.3	2.3.3	2.3.3
Hadoop	2.8.4	2.8.4	2.8.4	2.8.4
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.2.0	1.2.0	1.2.0	1.2.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.206	0.206	0.206	0.203
Spark	2.3.1	2.3.1	2.3.1	2.3.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.9.0	1.9.0	1.9.0	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.17.2	emr-5.17.1	emr-5.17.0	emr-5.16.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.12	3.4.12	3.4.12

## 5.17.2 发布说明

这是一个补丁版本，用于为向 Amazon S3 发出的请求添加 AWS 签名版本 4 身份验证。所有应用程序和组件都与之前的 Amazon EMR 版本相同。

### Important

在此版本中，亚马逊仅EMR使用 AWS 签名版本 4 来验证向 Amazon S3 发出的请求。有关更多信息，请参阅[新功能](#)。

## 5.17.2 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>aws-sagemaker-spark-sdk</code>	1.1.3	亚马逊 SageMaker Spark SDK

组件	版本	描述
emr-ddb	4.6.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.5.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.10.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	1.0.0	EMRS3Select 连接器
emrfs	2.26.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.5.2	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.8.4-amzn-1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。

组件	版本	描述
hadoop-hdfs-datanode	2.8.4-amzn-1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.4-amzn-1	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.4-amzn-1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.8.4-amzn-1	HTTPHDFS操作端点。
hadoop-kms-server	2.8.4-amzn-1	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.8.4-amzn-1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.8.4-amzn-1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.8.4-amzn-1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.4-amzn-1	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.6	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.6	为一个或多个HBase地区提供服务的服务。
hbase-client	1.4.6	HBase命令行客户端。
hbase-rest-server	1.4.6	为提供RESTfulHTTP终端节点的服务HBase。

组件	版本	描述
hbase-thrift-server	1.4.6	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	2.3.3-amzn-1	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.3-amzn-1	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.3-amzn-1	HTTP端点为其提供REST接口 HCatalog。
hive-client	2.3.3-amzn-1	Hive 命令行客户端。
hive-hbase	2.3.3-amzn-1	Hive-hbase 客户端。
hive-metastore-server	2.3.3-amzn-1	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.3-amzn-1	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	4.2.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	0.8.1	Jupyter notebook 的多用户服务器
livy-server	0.5.0-incubating	REST用于与 Apache Spark 交互的接口
mahout-client	0.13.0	用于机器学习的库。

组件	版本	描述
mxnet	1.2.0	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
nvidia-cuda	9.2.88	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.0.0	Oozie 命令行客户端。
oozie-server	5.0.0	用于接受 Oozie 工作流请求的服务。
opencv	3.4.0	开源计算机视觉库。
phoenix-library	4.14.0--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.14.0--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.206	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.206	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。
r	3.4.1	用于统计计算的 R 项目
spark-client	2.3.1	Spark 命令行客户端。
spark-history-server	2.3.1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。



组件	版本	描述
spark-on-yarn	2.3.1	的内存中执行引擎。YARN
spark-yarn-slave	2.3.1	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	1.9.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.7.3	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.12	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.12	ZooKeeper 命令行客户端。

## 5.17.2 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

### emr-5.17.2 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。

分类	描述
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。

分类	描述
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。

分类	描述
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 jupyterhub_config.py 文件 JupyterHubs 中的值。
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。

分类	描述
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。

分类	描述
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。

分类	描述
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.17.1

### 5.17.1 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#)和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-5.17.1	emr-5.17.0	emr-5.16.1	emr-5.16.0
适用于 Java 的 AWS SDK	1.11.336	1.11.336	1.11.336	1.11.336
Python	2.7、3.4	2.7、3.4	2.7、3.4	2.7、3.4

	emr-5.17.1	emr-5.17.0	emr-5.16.1	emr-5.16.0
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.5.2	1.5.2	1.5.0	1.5.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.6	1.4.6	1.4.4	1.4.4
HCatalog	2.3.3	2.3.3	2.3.3	2.3.3
Hadoop	2.8.4	2.8.4	2.8.4	2.8.4
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0.5.0	0.5.0	0.5.0	0.5.0
MXNet	1.2.0	1.2.0	1.2.0	1.2.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0



	emr-5.17.1	emr-5.17.0	emr-5.16.1	emr-5.16.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.206	0.206	0.203	0.203
Spark	2.3.1	2.3.1	2.3.1	2.3.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.9.0	1.9.0	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.12	3.4.12	3.4.12

### 5.17.1 发布说明

以下发行说明包含有关亚马逊 5.17.1 EMR 版本的信息。更改与 5.17.0 有关。

首次发布日期：2019 年 7 月 18 日

更改、增强功能和解决的问题

- 更新了亚马逊的默认 Amazon Linux AMI，EMR使其包含重要的 Linux 内核安全更新，包括 TCPSACK拒绝服务问题 ([AWS-2019-005](#))。

### 5.17.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 emr 或 aws 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.1.3	亚马逊 SageMaker Spark SDK
emr-ddb	4.6.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.5.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.10.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	1.0.0	EMRS3Select 连接器
emrfs	2.26.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.5.2	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。

组件	版本	描述
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.8.4-amzn-1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.8.4-amzn-1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.4-amzn-1	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.4-amzn-1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.8.4-amzn-1	HTTPHDFS操作端点。
hadoop-kms-server	2.8.4-amzn-1	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.8.4-amzn-1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.8.4-amzn-1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.8.4-amzn-1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.4-amzn-1	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.6	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.6	为一个或多个HBase地区提供服务的服务。

组件	版本	描述
hbase-client	1.4.6	HBase 命令行客户端。
hbase-rest-server	1.4.6	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.6	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	2.3.3-amzn-1	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.3-amzn-1	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.3-amzn-1	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.3-amzn-1	Hive 命令行客户端。
hive-hbase	2.3.3-amzn-1	Hive-hbase 客户端。
hive-metastore-server	2.3.3-amzn-1	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.3-amzn-1	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	4.2.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	0.8.1	Jupyter notebook 的多用户服务器

组件	版本	描述
livy-server	0.5.0-incubating	REST用于与 Apache Spark 交互的接口
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.2.0	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
nvidia-cuda	9.2.88	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.0.0	Oozie 命令行客户端。
oozie-server	5.0.0	用于接受 Oozie 工作流请求的服务。
opencv	3.4.0	开源计算机视觉库。
phoenix-library	4.14.0--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.14.0--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.206	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.206	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。
r	3.4.1	用于统计计算的 R 项目
spark-client	2.3.1	Spark 命令行客户端。

组件	版本	描述
spark-history-server	2.3.1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.3.1	的内存中执行引擎。YARN
spark-yarn-slave	2.3.1	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	1.9.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.7.3	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.12	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.12	ZooKeeper 命令行客户端。

### 5.17.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

## emr-5.17.1 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。

分类	描述
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。



分类	描述
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 jupyterhub_config.py 文件 JupyterHubs 中的值。
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。

分类	描述
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。

分类	描述
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。

分类	描述
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.17.0

### 5.17.0 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#)和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

## 应用程序版本信息

	emr-5.17.0	emr-5.16.1	emr-5.16.0	emr-5.15.1
适用于 Java 的 AWS SDK	1.11.336	1.11.336	1.11.336	1.11.333
Python	2.7、3.4	2.7、3.4	2.7、3.4	2.7、3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.5.2	1.5.0	1.5.0	1.4.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.6	1.4.4	1.4.4	1.4.4
HCatalog	2.3.3	2.3.3	2.3.3	2.3.3
Hadoop	2.8.4	2.8.4	2.8.4	2.8.3
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0.5.0	0.5.0	0.5.0	0.4.0
MXNet	1.2.0	1.2.0	1.2.0	1.1.0

	emr-5.17.0	emr-5.16.1	emr-5.16.0	emr-5.15.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.206	0.203	0.203	0.194
Spark	2.3.1	2.3.1	2.3.1	2.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.9.0	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.12	3.4.12	3.4.12

## 5.17.0 发布说明

以下发行说明包含有关亚马逊 5.17. EMR 0 版本的信息。更改与 5.16.0 有关。

首次发布日期：2018 年 8 月 30 日

### 升级

- Flink 1.5.2
- HBase1.4.6
- Presto 0.206

## 新功能

- 添加了对 Tensorflow 的支持。有关更多信息，请参阅[TensorFlow](#)。

## 更改、增强功能和解决的问题

- JupyterHub
  - Amazon S3 中添加了对笔记本持久性的支持。有关更多信息，请参阅[在 Amazon S3 中配置笔记本的持久性](#)。
- Spark
  - 添加了对 [S3 Select](#) 的支持。有关更多信息，请参阅 [将 S3 Select 与 Spark 结合使用以提高查询性能](#)。
- 解决了亚马逊EMR版本 5.14.0、5.15.0 或 5.16.0 中的 Cloudwatch 指标和自动扩展功能的问题。

## 已知问题

- 创建使用 Kerberos 的集群时，如果安装了 Livy，Livy 将失败，并显示未启用简单身份验证的错误。重新启动 Livy 服务器可解决此问题。解决方法是在集群创建过程中添加一个在主节点上运行 `sudo restart livy-server` 的步骤。
- 如果你使用AMI基于亚马逊 Linux 的自定义 Amazon LinuxAMI，创建日期为 2018-08-11，Oozie 服务器将无法启动。如果你使用 Oozie，请AMI根据创建日期不同的亚马逊 Linux AMI ID 创建自定义账号。您可以使用以下 AWS CLI 命令返回所有 HVM Amazon Linux AMIs 版本为 2018.03 的图片 IDs列表以及发布日期，以便您可以选择合适的 Amazon Linux AMI 作为基础。MyRegion 替换为您的地区标识符，例如 us-west-2。

```
aws ec2 --region MyRegion describe-images --owner amazon --query 'Images[?
Name!=`null`][[?starts_with(Name, `amzn-ami-hvm-2018.03`) == `true`].
[CreationDate,ImageId,Name]' --output text | sort -rk1
```

## 5.17.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.1.3	亚马逊 SageMaker Spark SDK
emr-ddb	4.6.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.5.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.10.0	针对 Amazon S3 优化的分布式复制应用程序。
emr-s3-select	1.0.0	EMRS3Select 连接器
emrfs	2.26.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.5.2	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。



组件	版本	描述
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.8.4-amzn-1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.8.4-amzn-1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.4-amzn-1	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.4-amzn-1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-https-server	2.8.4-amzn-1	HTTPHDFS操作端点。
hadoop-kms-server	2.8.4-amzn-1	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.8.4-amzn-1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.8.4-amzn-1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.8.4-amzn-1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.4-amzn-1	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.6	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.6	为一个或多个HBase地区提供服务的服务。

组件	版本	描述
hbase-client	1.4.6	HBase命令行客户端。
hbase-rest-server	1.4.6	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.6	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.3.3-amzn-1	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.3-amzn-1	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.3-amzn-1	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.3-amzn-1	Hive 命令行客户端。
hive-hbase	2.3.3-amzn-1	Hive-hbase 客户端。
hive-metastore-server	2.3.3-amzn-1	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.3-amzn-1	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	4.2.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	0.8.1	Jupyter notebook 的多用户服务器

组件	版本	描述
livy-server	0.5.0-incubating	REST用于与 Apache Spark 交互的接口
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.2.0	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
nvidia-cuda	9.2.88	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.0.0	Oozie 命令行客户端。
oozie-server	5.0.0	用于接受 Oozie 工作流请求的服务。
opencv	3.4.0	开源计算机视觉库。
phoenix-library	4.14.0--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.14.0--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.206	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.206	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。
r	3.4.1	用于统计计算的 R 项目
spark-client	2.3.1	Spark 命令行客户端。

组件	版本	描述
spark-history-server	2.3.1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.3.1	的内存中执行引擎。YARN
spark-yarn-slave	2.3.1	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tensorflow	1.9.0	TensorFlow 用于高性能数值计算的开源软件库。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.7.3	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.12	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.12	ZooKeeper 命令行客户端。

### 5.17.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

## emr-5.17.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。

分类	描述
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。

分类	描述
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 jupyterhub_config.py 文件 JupyterHubs 中的值。
jupyter-s3-conf	配置 Jupyter notebook S3 持久性。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。

分类	描述
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。



分类	描述
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。

分类	描述
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.16.1

### 5.16.1 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#) 和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

## 应用程序版本信息

	emr-5.16.1	emr-5.16.0	emr-5.15.1	emr-5.15.0
适用于 Java 的 AWS SDK	1.11.336	1.11.336	1.11.333	1.11.333
Python	2.7、3.4	2.7、3.4	2.7、3.4	2.7、3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.5.0	1.5.0	1.4.2	1.4.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.4	1.4.4	1.4.4	1.4.4
HCatalog	2.3.3	2.3.3	2.3.3	2.3.3
Hadoop	2.8.4	2.8.4	2.8.3	2.8.3
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0.5.0	0.5.0	0.4.0	0.4.0
MXNet	1.2.0	1.2.0	1.1.0	1.1.0

	emr-5.16.1	emr-5.16.0	emr-5.15.1	emr-5.15.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.203	0.203	0.194	0.194
Spark	2.3.1	2.3.1	2.3.0	2.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.12	3.4.12	3.4.12

### 5.16.1 发布说明

这是一个补丁版本，用于为向 Amazon S3 发出的请求添加 AWS 签名版本 4 身份验证。所有应用程序和组件都与之前的 Amazon EMR 版本相同。

#### Important

在此版本中，亚马逊仅EMR使用 AWS 签名版本 4 来验证向 Amazon S3 发出的请求。有关更多信息，请参阅[新功能](#)。

## 5.16.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>aws-sagemaker-spark-sdk</code>	1.1.0	亚马逊 SageMaker Spark SDK
<code>emr-ddb</code>	4.6.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.4.0	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
<code>emr-s3-dist-cp</code>	2.10.0	针对 Amazon S3 优化的分布式复制应用程序。
<code>emrfs</code>	2.25.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
<code>flink-client</code>	1.5.0	Apache Flink 命令行客户端脚本和应用程序。
<code>ganglia-monitor</code>	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。

组件	版本	描述
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.8.4-amzn-0	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.8.4-amzn-0	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.4-amzn-0	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.4-amzn-0	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.8.4-amzn-0	HTTPHDFS操作端点。
hadoop-kms-server	2.8.4-amzn-0	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.8.4-amzn-0	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.8.4-amzn-0	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.8.4-amzn-0	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.4-amzn-0	用于检索YARN应用程序的当前和历史信息的服务。

组件	版本	描述
hbase-hmaster	1.4.4	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.4	为一个或多个HBase地区提供服务的服务器。
hbase-client	1.4.4	HBase命令行客户端。
hbase-rest-server	1.4.4	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.4	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.3.3-amzn-1	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.3-amzn-1	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.3-amzn-1	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.3-amzn-1	Hive 命令行客户端。
hive-hbase	2.3.3-amzn-1	Hive-hbase 客户端。
hive-metastore-server	2.3.3-amzn-1	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.3-amzn-1	用于将 Hive 查询作为 Web 请求接受的服务。

组件	版本	描述
hue-server	4.2.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	0.8.1	Jupyter notebook 的多用户服务器
livy-server	0.5.0-incubating	REST用于与 Apache Spark 交互的接口
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.2.0	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
nvidia-cuda	9.2.88	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.0.0	Oozie 命令行客户端。
oozie-server	5.0.0	用于接受 Oozie  workflow 请求的服务。
opencv	3.4.0	开源计算机视觉库。
phoenix-library	4.14.0--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.14.0--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.203	用于在 presto-worker 之中接受查询并管理查询的服务。



组件	版本	描述
presto-worker	0.203	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。
r	3.4.1	用于统计计算的 R 项目
spark-client	2.3.1	Spark 命令行客户端。
spark-history-server	2.3.1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.3.1	的内存中执行引擎。YARN
spark-yarn-slave	2.3.1	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.7.3	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.12	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.12	ZooKeeper 命令行客户端。

### 5.16.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

## emr-5.16.1 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。

分类	描述
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。

分类	描述
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 jupyterhub_config.py 文件 JupyterHubs 中的值。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。

分类	描述
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。

分类	描述
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。

分类	描述
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.16.0

### 5.16.0 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#) 和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

## 应用程序版本信息

	emr-5.16.0	emr-5.15.1	emr-5.15.0	emr-5.14.2
适用于 Java 的 AWS SDK	1.11.336	1.11.333	1.11.333	1.11.297
Python	2.7、3.4	2.7、3.4	2.7、3.4	2.7、3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.5.0	1.4.2	1.4.2	1.4.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.4	1.4.4	1.4.4	1.4.2
HCatalog	2.3.3	2.3.3	2.3.3	2.3.2
Hadoop	2.8.4	2.8.3	2.8.3	2.8.3
Hive	2.3.3	2.3.3	2.3.3	2.3.2
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0.5.0	0.4.0	0.4.0	0.4.0
MXNet	1.2.0	1.1.0	1.1.0	1.1.0



	emr-5.16.0	emr-5.15.1	emr-5.15.0	emr-5.14.2
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	4.3.0
Phoenix	4.14.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.203	0.194	0.194	0.194
Spark	2.3.1	2.3.0	2.3.0	2.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.12	3.4.12	3.4.10

## 5.16.0 发布说明

以下发行说明包含有关亚马逊 5.16. EMR 0 版本的信息。更改与 5.15.0 有关。

首次发布日期：2018 年 7 月 19 日

### 升级

- Hadoop 2.8.4
- Flink 1.5.0
- Livy 0.5.0
- MXNet1.2.0
- Phoenix 4.14.0

- Presto 0.203
- Spark 2.3.1
- AWS SDK for Java 1.11.336
- CUDA9.2
- Redshift JDBC Driver 1.2.15.1025

### 更改、增强功能和解决的问题

- HBase
  - [向后移植 HBASE -20723](#)
- Presto
  - 配置更改为支持LDAP身份验证。有关更多信息，请参阅 [在亚马逊上使用 Presto 的LDAP身份验证 EMR](#)。
- Spark
  - [Apache Spark 版本 2.3.1，从亚马逊EMR版本 5.16.0 开始推出，地址CVE为 -2018-8024 和 -2018-1334。CVE建议您将 Spark 的早期版本迁移到 Spark 2.3.1 版本或更高版本。](#)

### 已知问题

- 此发行版不支持 c1.medium 或 m1.small 实例类型。使用这些实例类型的集群将无法启动。解决方法：指定其它实例类型或使用其它发行版。
- 创建使用 Kerberos 的集群时，如果安装了 Livy，Livy 将失败，并显示未启用简单身份验证的错误。重新启动 Livy 服务器可解决此问题。解决方法是在集群创建过程中添加一个在主节点上运行 `sudo restart livy-server` 的步骤。
- 在主节点重启或实例控制器重新启动后，将不会收集 CloudWatch 指标，并且自动扩展功能将在 Amazon EMR 版本 5.14.0、5.15.0 或 5.16.0 中不可用。此问题已在亚马逊 EMR 5.17.0 中修复。

## 5.16.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.1.0	亚马逊 SageMaker Spark SDK
emr-ddb	4.6.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.4.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.10.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.25.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.5.0	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。

组件	版本	描述
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.8.4-amzn-0	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.8.4-amzn-0	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.4-amzn-0	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.4-amzn-0	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.8.4-amzn-0	HTTPHDFS操作端点。
hadoop-kms-server	2.8.4-amzn-0	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.8.4-amzn-0	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.8.4-amzn-0	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.8.4-amzn-0	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.4-amzn-0	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.4	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.4	为一个或多个HBase地区提供服务的服务。

组件	版本	描述
hbase-client	1.4.4	HBase命令行客户端。
hbase-rest-server	1.4.4	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.4	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	2.3.3-amzn-1	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.3-amzn-1	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.3-amzn-1	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.3-amzn-1	Hive 命令行客户端。
hive-hbase	2.3.3-amzn-1	Hive-hbase 客户端。
hive-metastore-server	2.3.3-amzn-1	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.3-amzn-1	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	4.2.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	0.8.1	Jupyter notebook 的多用户服务器

组件	版本	描述
livy-server	0.5.0-incubating	REST用于与 Apache Spark 交互的接口
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.2.0	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
nvidia-cuda	9.2.88	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.0.0	Oozie 命令行客户端。
oozie-server	5.0.0	用于接受 Oozie 工作流请求的服务。
opencv	3.4.0	开源计算机视觉库。
phoenix-library	4.14.0--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.14.0--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.203	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.203	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。
r	3.4.1	用于统计计算的 R 项目
spark-client	2.3.1	Spark 命令行客户端。

组件	版本	描述
spark-history-server	2.3.1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.3.1	的内存中执行引擎。YARN
spark-yarn-slave	2.3.1	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.7.3	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.12	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.12	ZooKeeper 命令行客户端。

### 5.16.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.16.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。

分类	描述
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。



分类	描述
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值

分类	描述
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 jupyterhub_config.py 文件 JupyterHubs 中的值。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。

分类	描述
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-password-authenticator	更改 Presto 的 password-authenticator.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。

分类	描述
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。

分类	描述
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.15.1

### 5.15.1 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#)和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-5.15.1	emr-5.15.0	emr-5.14.2	emr-5.14.1
适用于 Java 的 AWS SDK	1.11.333	1.11.333	1.11.297	1.11.297

	emr-5.15.1	emr-5.15.0	emr-5.14.2	emr-5.14.1
Python	2.7、3.4	2.7、3.4	2.7、3.4	2.7、3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.2	1.4.2	1.4.2	1.4.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.4	1.4.4	1.4.2	1.4.2
HCatalog	2.3.3	2.3.3	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.3	2.3.3	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.1.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	1.1.0	1.1.0	1.1.0	1.1.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	4.3.0	4.3.0

	emr-5.15.1	emr-5.15.0	emr-5.14.2	emr-5.14.1
Phoenix	4.13.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.194	0.194	0.194	0.194
Spark	2.3.0	2.3.0	2.3.0	2.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.12	3.4.10	3.4.10

### 5.15.1 发布说明

这是一个补丁版本，用于为向 Amazon S3 发出的请求添加 AWS 签名版本 4 身份验证。所有应用程序和组件都与之前的 Amazon EMR 版本相同。

#### Important

在此版本中，亚马逊仅EMR使用 AWS 签名版本 4 来验证向 Amazon S3 发出的请求。有关更多信息，请参阅[新功能](#)。

### 5.15.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.0.1	亚马逊 SageMaker Spark SDK
emr-ddb	4.5.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.4.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.10.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.24.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.4.2	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。



组件	版本	描述
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.8.3-amzn-1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.8.3-amzn-1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.3-amzn-1	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.3-amzn-1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.8.3-amzn-1	HTTPHDFS操作端点。
hadoop-kms-server	2.8.3-amzn-1	基于 Hadoop 的加密密钥管理服务器。 KeyProvider API
hadoop-mapred	2.8.3-amzn-1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.8.3-amzn-1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.8.3-amzn-1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.3-amzn-1	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.4	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.4	为一个或多个HBase地区提供服务的服务。

组件	版本	描述
hbase-client	1.4.4	HBase命令行客户端。
hbase-rest-server	1.4.4	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.4	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	2.3.3-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.3-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.3-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.3-amzn-0	Hive 命令行客户端。
hive-hbase	2.3.3-amzn-0	Hive-hbase 客户端。
hive-metastore-server	2.3.3-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.3-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	4.2.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	0.8.1	Jupyter notebook 的多用户服务器

组件	版本	描述
livy-server	0.4.0-incubating	REST用于与 Apache Spark 交互的接口
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.1.0	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
nvidia-cuda	9.1.85	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.0.0	Oozie 命令行客户端。
oozie-server	5.0.0	用于接受 Oozie 工作流请求的服务。
opencv	3.4.0	开源计算机视觉库。
phoenix-library	4.13.0--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.13.0--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.194	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.194	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。
r	3.4.1	用于统计计算的 R 项目
spark-client	2.3.0	Spark 命令行客户端。

组件	版本	描述
spark-history-server	2.3.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.3.0	的内存中执行引擎。YARN
spark-yarn-slave	2.3.0	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.7.3	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.12	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.12	ZooKeeper 命令行客户端。

### 5.15.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.15.1 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。

分类	描述
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。

分类	描述
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值

分类	描述
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 jupyterhub_config.py 文件 JupyterHubs 中的值。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。

分类	描述
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。



分类	描述
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。

分类	描述
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper 中的值。
zookeeper-log4j	更改 ZooKeeper 的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.15.0

### 5.15.0 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#) 和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-5.15.0	emr-5.14.2	emr-5.14.1	emr-5.14.0
适用于 Java 的 AWS SDK	1.11.333	1.11.297	1.11.297	1.11.297
Python	2.7、3.4	2.7、3.4	2.7、3.4	2.7、3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8

	emr-5.15.0	emr-5.14.2	emr-5.14.1	emr-5.14.0
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.2	1.4.2	1.4.2	1.4.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.4	1.4.2	1.4.2	1.4.2
HCatalog	2.3.3	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.3	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.2.0	4.1.0	4.1.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	1.1.0	1.1.0	1.1.0	1.1.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0

	emr-5.15.0	emr-5.14.2	emr-5.14.1	emr-5.14.0
Presto	0.194	0.194	0.194	0.194
Spark	2.3.0	2.3.0	2.3.0	2.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.10	3.4.10	3.4.10

## 5.15.0 发布说明

以下发行说明包含有关亚马逊 5.15. EMR 0 版本的信息。更改与 5.14.0 有关。

首次发布日期：2018 年 6 月 21 日

### 升级

- 已升级HBase到 1.4.4
- 已将 Hive 升级到 2.3.3
- 已将 Hue 升级到 4.2.0
- 已将 Oozie 升级到 5.0.0
- 已将 Zookeeper 升级到 3.4.12
- 已升级 AWS SDK到 1.11.333

### 更改、增强功能和解决的问题

- Hive
  - [向后移植 HIVE -18069](#)

- Hue
  - 更新了 Hue，启用 Kerberos 后可以使用 Livy 正确地进行身份验证。在亚马逊上使用 Kerberos 时，现在支持 Livy。EMR
- JupyterHub
  - 已更新，以 JupyterHub 便 Amazon 默认EMR安装LDAP客户端库。
  - 修复了生成自签名凭证的脚本中的错误。

## 已知问题

- 此发行版不支持 c1.medium 或 m1.small 实例类型。使用这些实例类型的集群将无法启动。解决方法：指定其它实例类型或使用其它发行版。
- 在主节点重启或实例控制器重新启动后，将不会收集 CloudWatch 指标，并且自动扩展功能将在 Amazon EMR 版本 5.14.0、5.15.0 或 5.16.0 中不可用。此问题已在亚马逊 EMR 5.17.0 中修复。

## 5.15.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 emr 或 aws 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.0.1	亚马逊 SageMaker Spark SDK
emr-ddb	4.5.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.4.0	适用于 Hadoop 生态系统的方便易用的库。

组件	版本	描述
emr-kinesis	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.10.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.24.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.4.2	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.8.3-amzn-1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.8.3-amzn-1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.3-amzn-1	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.3-amzn-1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.8.3-amzn-1	HTTPHDFS操作端点。

组件	版本	描述
hadoop-kms-server	2.8.3-amzn-1	基于 Hadoop 的加密密钥管理服务。KeyProvider API
hadoop-mapred	2.8.3-amzn-1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.8.3-amzn-1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.8.3-amzn-1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.3-amzn-1	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.4	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.4	为一个或多个HBase地区提供服务的服务。
hbase-client	1.4.4	HBase命令行客户端。
hbase-rest-server	1.4.4	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.4	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	2.3.3-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.3-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。

组件	版本	描述
hcatalog-webhcat-server	2.3.3-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.3-amzn-0	Hive 命令行客户端。
hive-hbase	2.3.3-amzn-0	Hive-hbase 客户端。
hive-metastore-server	2.3.3-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.3-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	4.2.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	0.8.1	Jupyter notebook 的多用户服务器
livy-server	0.4.0-incubating	REST用于与 Apache Spark 交互的接口
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.1.0	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
nvidia-cuda	9.1.85	Nvidia 驱动程序和 Cuda 工具包
oozie-client	5.0.0	Oozie 命令行客户端。



组件	版本	描述
oozie-server	5.0.0	用于接受 Oozie 工作流请求的服务。
opencv	3.4.0	开源计算机视觉库。
phoenix-library	4.13.0--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.13.0--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.194	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.194	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。
r	3.4.1	用于统计计算的 R 项目
spark-client	2.3.0	Spark 命令行客户端。
spark-history-server	2.3.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.3.0	的内存中执行引擎。YARN
spark-yarn-slave	2.3.0	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。

组件	版本	描述
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.7.3	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.12	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.12	ZooKeeper 命令行客户端。

### 5.15.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.15.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。

分类	描述
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。

分类	描述
hcatalog-webhcat-site	更改 HCatalog WebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。

分类	描述
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 jupyterhub_config.py 文件 JupyterHubs 中的值。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。

分类	描述
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。

分类	描述
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.14.2

### 5.14.2 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#)和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-5.14.2	emr-5.14.1	emr-5.14.0	emr-5.13.1
适用于 Java 的 AWS SDK	1.11.297	1.11.297	1.11.297	1.11.297
Python	2.7、3.4	2.7、3.4	2.7、3.4	2.7、3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.2	1.4.2	1.4.2	1.4.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.2	1.4.2	1.4.2	1.4.2



	emr-5.14.2	emr-5.14.1	emr-5.14.0	emr-5.13.1
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.1.0	4.1.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	-
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	1.1.0	1.1.0	1.1.0	1.0.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.194	0.194	0.194	0.194
Spark	2.3.0	2.3.0	2.3.0	2.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.14.2	emr-5.14.1	emr-5.14.0	emr-5.13.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

## 5.14.2 发布说明

这是一个补丁版本，用于为向 Amazon S3 发出的请求添加 AWS 签名版本 4 身份验证。所有应用程序和组件都与之前的 Amazon EMR 版本相同。

### Important

在此版本中，亚马逊仅EMR使用 AWS 签名版本 4 来验证向 Amazon S3 发出的请求。有关更多信息，请参阅[新功能](#)。

## 5.14.2 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.0.1	亚马逊 SageMaker Spark SDK

组件	版本	描述
emr-ddb	4.5.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.4.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.10.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.23.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.4.2	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.8.3-amzn-1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.8.3-amzn-1	HDFS用于存储区块的节点级服务。

组件	版本	描述
hadoop-hdfs-library	2.8.3-amzn-1	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.3-amzn-1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httfs-server	2.8.3-amzn-1	HTTPHDFS操作端点。
hadoop-kms-server	2.8.3-amzn-1	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.8.3-amzn-1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.8.3-amzn-1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.8.3-amzn-1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.3-amzn-1	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.2	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.2	为一个或多个HBase地区提供服务的服务。
hbase-client	1.4.2	HBase命令行客户端。
hbase-rest-server	1.4.2	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.2	为其提供 Thrift 端点的HBase服务。

组件	版本	描述
hcatalog-client	2.3.2-amzn-2	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.2-amzn-2	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.2-amzn-2	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.2-amzn-2	Hive 命令行客户端。
hive-hbase	2.3.2-amzn-2	Hive-hbase 客户端。
hive-metastore-server	2.3.2-amzn-2	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.2-amzn-2	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	4.1.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	0.8.1	Jupyter notebook 的多用户服务器
livy-server	0.4.0-incubating	REST用于与 Apache Spark 交互的接口
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.1.0	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的SQL数据库服务器。

组件	版本	描述
nvidia-cuda	9.1.85	Nvidia 驱动程序和 Cuda 工具包
oozie-client	4.3.0	Oozie 命令行客户端。
oozie-server	4.3.0	用于接受 Oozie 工作流请求的服务。
opencv	3.4.0	开源计算机视觉库。
phoenix-library	4.13.0--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.13.0--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.194	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.194	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。
r	3.4.1	用于统计计算的 R 项目
spark-client	2.3.0	Spark 命令行客户端。
spark-history-server	2.3.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.3.0	的内存中执行引擎。YARN
spark-yarn-slave	2.3.0	YARN从属服务器需要的 Apache Spark 库。

组件	版本	描述
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.7.3	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.10	ZooKeeper 命令行客户端。

## 5.14.2 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

### emr-5.14.2 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。

分类	描述
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。



分类	描述
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。

分类	描述
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 jupyterhub_config.py 文件 JupyterHubs 中的值。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。

分类	描述
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。

分类	描述
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-orahome-site	更改 Sqoop OraOop 的 orahome-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.14.1

### 5.14.1 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#)和[ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-5.14.1	emr-5.14.0	emr-5.13.1	emr-5.13.0
适用于 Java 的 AWS SDK	1.11.297	1.11.297	1.11.297	1.11.297
Python	2.7、3.4	2.7、3.4	2.7、3.4	2.7、3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.2	1.4.2	1.4.0	1.4.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.2	1.4.2	1.4.2	1.4.2

	emr-5.14.1	emr-5.14.0	emr-5.13.1	emr-5.13.0
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.1.0	4.1.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	-	-
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	1.1.0	1.1.0	1.0.0	1.0.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.194	0.194	0.194	0.194
Spark	2.3.0	2.3.0	2.3.0	2.3.0
Sqoop	1.4.7	1.4.7	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.14.1	emr-5.14.0	emr-5.13.1	emr-5.13.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

### 5.14.1 发布说明

以下发行说明包含有关亚马逊 5.14.1 EMR 版本的信息。更改与 5.14.0 有关。

首次发布日期：2018 年 10 月 17 日

更新了 Amazon AMI 的默认设置EMR以解决潜在的安全漏洞。

### 5.14.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>aws-sagemaker-spark-sdk</code>	1.0.1	亚马逊 SageMaker Spark SDK
<code>emr-ddb</code>	4.5.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.4.0	适用于 Hadoop 生态系统的方便易用的库。

组件	版本	描述
emr-kinesis	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.10.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.23.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.4.2	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.8.3-amzn-1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.8.3-amzn-1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.3-amzn-1	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.3-amzn-1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.8.3-amzn-1	HTTPHDFS操作端点。



组件	版本	描述
hadoop-kms-server	2.8.3-amzn-1	基于 Hadoop 的加密密钥管理服务。KeyProvider API
hadoop-mapred	2.8.3-amzn-1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.8.3-amzn-1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.8.3-amzn-1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.3-amzn-1	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.2	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.2	为一个或多个HBase地区提供服务的服务。
hbase-client	1.4.2	HBase命令行客户端。
hbase-rest-server	1.4.2	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.2	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	2.3.2-amzn-2	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.2-amzn-2	提供服务HCatalog，为分布式应用程序提供表和存储管理层。

组件	版本	描述
hcatalog-webhcat-server	2.3.2-amzn-2	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.2-amzn-2	Hive 命令行客户端。
hive-hbase	2.3.2-amzn-2	Hive-hbase 客户端。
hive-metastore-server	2.3.2-amzn-2	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.2-amzn-2	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	4.1.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	0.8.1	Jupyter notebook 的多用户服务器
livy-server	0.4.0-incubating	REST用于与 Apache Spark 交互的接口
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.1.0	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
nvidia-cuda	9.1.85	Nvidia 驱动程序和 Cuda 工具包
oozie-client	4.3.0	Oozie 命令行客户端。

组件	版本	描述
oozie-server	4.3.0	用于接受 Oozie 工作流请求的服务。
opencv	3.4.0	开源计算机视觉库。
phoenix-library	4.13.0--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.13.0--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.194	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.194	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。
r	3.4.1	用于统计计算的 R 项目
spark-client	2.3.0	Spark 命令行客户端。
spark-history-server	2.3.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.3.0	的内存中执行引擎。YARN
spark-yarn-slave	2.3.0	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。

组件	版本	描述
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.7.3	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.10	ZooKeeper 命令行客户端。

### 5.14.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.14.1 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。

分类	描述
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。

分类	描述
hcatalog-webhcat-site	更改 HCatalog WebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。

分类	描述
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 jupyterhub_config.py 文件 JupyterHubs 中的值。
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。

分类	描述
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。



分类	描述
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.14.0

### 5.14.0 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[JupyterHub](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#)和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-5.14.0	emr-5.13.1	emr-5.13.0	emr-5.12.3
适用于 Java 的 AWS SDK	1.11.297	1.11.297	1.11.297	1.11.267
Python	2.7、3.4	2.7、3.4	2.7、3.4	2.7、3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.2	1.4.0	1.4.0	1.4.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.2	1.4.2	1.4.2	1.4.0

	emr-5.14.0	emr-5.13.1	emr-5.13.0	emr-5.12.3
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.1.0	4.1.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	-	-	-
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	1.1.0	1.0.0	1.0.0	1.0.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.194	0.194	0.194	0.188
Spark	2.3.0	2.3.0	2.3.0	2.2.1
Sqoop	1.4.7	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.14.0	emr-5.13.1	emr-5.13.0	emr-5.12.3
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

## 5.14.0 发布说明

以下发行说明包含有关亚马逊 5.14. EMR 0 版本的信息。更改与 5.13.0 有关。

首次发布日期：2018 年 6 月 4 日

### 升级

- 已将 Apache Flink 升级到 1.4.2
- 将 Apache 升级MXnet到 1.1.0
- 已将 Apache Sqoop 升级到 1.4.7

### 新功能

- 增加了 JupyterHub 支持。有关更多信息，请参阅 [JupyterHub](#)。

### 更改、增强功能和解决的问题

- EMRFS
  - 对 Amazon S3 的请求中的 userAgent 字符串已更新为包含调用委托人的用户和组信息。这可以与 AWS CloudTrail 日志一起使用，以实现更全面的请求跟踪。
- HBase
  - 包括 [HBASE-20447](#)，它解决了可能导致缓存问题，尤其是在拆分区域的情况下。
- MXnet
  - 新增了 OpenCV 库。
- Spark

- 当 Spark 使用将 Parquet 文件写入 Amazon S3 位置时 EMRFS，FileOutputCommitter 算法已更新为使用版本 2 而不是版本 1。这将减少重命名的数量，从而提高应用程序性能。此更改不会影响：
  - Spark 以外的应用程序。
  - 写入其他文件系统的应用程序，例如 HDFS（仍使用版本 1 FileOutputCommitter）。
  - 使用其他输出格式（例如文本或 csv）且已经使用 EMRFS 直接写入的应用程序。

## 已知问题

### • JupyterHub

- 不支持在创建集群时使用配置分类来设置 JupyterHub 和单个 Jupyter 笔记本。手动编辑每个用户的 jupyterhub\_config.py 文件和 jupyter\_notebook\_config.py 文件。有关更多信息，请参阅 [正在配置 JupyterHub](#)。
- JupyterHub 无法在私有子网内的群集上启动，失败并显示消息 `Error: ENOENT: no such file or directory, open '/etc/jupyter/conf/server.crt'`。这由生成自签名凭证的脚本中的错误所致。使用以下解决方法生成自签名凭证。在连接到主节点时执行所有命令。
  1. 将凭证生成脚本从容器复制到主节点：

```
sudo docker cp jupyterhub:/tmp/gen_self_signed_cert.sh ./
```

2. 使用文本编辑器更改第 23 行，将公有主机名更改为本地主机名，如下所示：

```
local hostname=$(curl -s $EC2_METADATA_SERVICE_URI/local-hostname)
```

3. 运行脚本，生成自签名凭证：

```
sudo bash ./gen_self_signed_cert.sh
```

4. 将脚本生成的凭证文件移至 `/etc/jupyter/conf/` 目录：

```
sudo mv /tmp/server.crt /tmp/server.key /etc/jupyter/conf/
```

您可以通过 `tail` 该 `jupyter.log` 文件来验证是否已 JupyterHub 重新启动并且正在返回 200 响应码。例如：

```
tail -f /var/log/jupyter/jupyter.log
```

该命令应返回与以下示例类似的响应：

```
# [I 2018-06-14 18:56:51.356 JupyterHub app:1581] JupyterHub is now running at
https://:9443/
# 19:01:51.359 - info: [ConfigProxy] 200 GET /api/routes
```

- 在主节点重启或实例控制器重新启动后，将不会收集 CloudWatch 指标，并且自动扩展功能将在 Amazon EMR 版本 5.14.0、5.15.0 或 5.16.0 中不可用。此问题已在亚马逊 EMR 5.17.0 中修复。

## 5.14.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>aws-sagemaker-spark-sdk</code>	1.0.1	亚马逊 SageMaker Spark SDK
<code>emr-ddb</code>	4.5.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.4.0	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
<code>emr-s3-dist-cp</code>	2.10.0	针对 Amazon S3 优化的分布式复制应用程序。

组件	版本	描述
emrfs	2.23.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.4.2	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.8.3-amzn-1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.8.3-amzn-1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.3-amzn-1	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.3-amzn-1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.8.3-amzn-1	HTTPHDFS操作端点。
hadoop-kms-server	2.8.3-amzn-1	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.8.3-amzn-1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。

组件	版本	描述
hadoop-yarn-nodemanager	2.8.3-amzn-1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.8.3-amzn-1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.3-amzn-1	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.2	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.2	为一个或多个HBase地区提供服务的服务。
hbase-client	1.4.2	HBase命令行客户端。
hbase-rest-server	1.4.2	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.2	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.3.2-amzn-2	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.2-amzn-2	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.2-amzn-2	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.2-amzn-2	Hive 命令行客户端。
hive-hbase	2.3.2-amzn-2	Hive-hbase 客户端。



组件	版本	描述
hive-metastore-server	2.3.2-amzn-2	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的 SQL 元数据。
hive-server2	2.3.2-amzn-2	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	4.1.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
jupyterhub	0.8.1	Jupyter notebook 的多用户服务器
livy-server	0.4.0-incubating	REST 用于与 Apache Spark 交互的接口
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.1.0	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的 SQL 数据库服务器。
nvidia-cuda	9.1.85	Nvidia 驱动程序和 Cuda 工具包
oozie-client	4.3.0	Oozie 命令行客户端。
oozie-server	4.3.0	用于接受 Oozie 工作流请求的服务。
opencv	3.4.0	开源计算机视觉库。
phoenix-library	4.13.0--1.4 HBase	服务器和客户端的 phoenix 库

组件	版本	描述
phoenix-query-server	4.13.0--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限 API
presto-coordinator	0.194	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.194	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。
r	3.4.1	用于统计计算的 R 项目
spark-client	2.3.0	Spark 命令行客户端。
spark-history-server	2.3.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.3.0	的内存中执行引擎。YARN
spark-yarn-slave	2.3.0	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.7	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.7.3	支持交互式数据分析的基于 Web 的笔记本电脑。

组件	版本	描述
zookeeper-server	3.4.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.10	ZooKeeper 命令行客户端。

### 5.14.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.14.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
container-log4j	更改 Hadoop 的 container-log4j.prop YARN erties 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。

分类	描述
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。

分类	描述
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
jupyter-notebook-conf	更改 Jupyter notebook 的 jupyter_notebook_config.py 文件中的值。
jupyter-hub-conf	更改 jupyterhub_config.py 文件 JupyterHubs 中的值。

分类	描述
jupyter-sparkmagic-conf	更改 Sparkmagic 的 config.json 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。

分类	描述
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。

分类	描述
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.13.1

### 5.13.1 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：



- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

## 应用程序版本信息

	emr-5.13.1	emr-5.13.0	emr-5.12.3	emr-5.12.2
适用于 Java 的 AWS SDK	1.11.297	1.11.297	1.11.267	1.11.267
Python	2.7、3.4	2.7、3.4	2.7、3.4	2.7、3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.0	1.4.0	1.4.0	1.4.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.2	1.4.2	1.4.0	1.4.0
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.1.0	4.1.0	4.1.0
Iceberg	-	-	-	-

	emr-5.13.1	emr-5.13.0	emr-5.12.3	emr-5.12.2
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	1.0.0	1.0.0	1.0.0	1.0.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.194	0.194	0.188	0.188
Spark	2.3.0	2.3.0	2.2.1	2.2.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

### 5.13.1 发布说明

这是一个补丁版本，用于为向 Amazon S3 发出的请求添加 AWS 签名版本 4 身份验证。所有应用程序和组件都与之前的 Amazon EMR 版本相同。

**⚠ Important**

在此版本中，亚马逊仅EMR使用 AWS 签名版本 4 来验证向 Amazon S3 发出的请求。有关更多信息，请参阅[新功能](#)。

### 5.13.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>aws-sagemaker-spark-sdk</code>	1.0.1	亚马逊 SageMaker Spark SDK
<code>emr-ddb</code>	4.5.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.4.0	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
<code>emr-s3-dist-cp</code>	2.10.0	针对 Amazon S3 优化的分布式复制应用程序。
<code>emrfs</code>	2.22.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。

组件	版本	描述
flink-client	1.4.0	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.8.3-amzn-0	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.8.3-amzn-0	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.3-amzn-0	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.3-amzn-0	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.8.3-amzn-0	HTTPHDFS操作端点。
hadoop-kms-server	2.8.3-amzn-0	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.8.3-amzn-0	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.8.3-amzn-0	YARN用于管理单个节点上的容器的服务。

组件	版本	描述
hadoop-yarn-resourcemanager	2.8.3-amzn-0	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.3-amzn-0	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.2	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.2	为一个或多个HBase地区提供服务的服务。
hbase-client	1.4.2	HBase命令行客户端。
hbase-rest-server	1.4.2	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.2	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.3.2-amzn-2	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.2-amzn-2	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.2-amzn-2	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.2-amzn-2	Hive 命令行客户端。
hive-hbase	2.3.2-amzn-2	Hive-hbase 客户端。
hive-metastore-server	2.3.2-amzn-2	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。

组件	版本	描述
hive-server2	2.3.2-amzn-2	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	4.1.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
livy-server	0.4.0-incubating	REST用于与 Apache Spark 交互的接口
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.0.0	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
nvidia-cuda	9.1.85	Nvidia 驱动程序和 Cuda 工具包
oozie-client	4.3.0	Oozie 命令行客户端。
oozie-server	4.3.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.13.0--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.13.0--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.194	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.194	用于执行查询的各个部分的服务。

组件	版本	描述
pig-client	0.17.0	Pig 命令行客户端。
r	3.4.1	用于统计计算的 R 项目
spark-client	2.3.0	Spark 命令行客户端。
spark-history-server	2.3.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.3.0	的内存中执行引擎。YARN
spark-yarn-slave	2.3.0	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.7.3	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.10	ZooKeeper 命令行客户端。

### 5.13.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

## emr-5.13.1 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。



分类	描述
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值

分类	描述
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。

分类	描述
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。

分类	描述
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-orahome-site	更改 Sqoop OraOop 的 orahome-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.13.0

### 5.13.0 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-5.13.0	emr-5.12.3	emr-5.12.2	emr-5.12.1
适用于 Java 的 AWS SDK	1.11.297	1.11.267	1.11.267	1.11.267
Python	2.7、3.4	2.7、3.4	2.7、3.4	2.7、3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.0	1.4.0	1.4.0	1.4.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.2	1.4.0	1.4.0	1.4.0

	emr-5.13.0	emr-5.12.3	emr-5.12.2	emr-5.12.1
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.1.0	4.1.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	1.0.0	1.0.0	1.0.0	1.0.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.194	0.188	0.188	0.188
Spark	2.3.0	2.2.1	2.2.1	2.2.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.13.0	emr-5.12.3	emr-5.12.2	emr-5.12.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

## 5.13.0 发布说明

以下发行说明包含有关亚马逊 5.13.0 EMR 版本的信息。更改与 5.12.0 有关。

### 升级

- 已将 Spark 升级到 2.3.0
- 已升级HBase到 1.4.2
- 已将 Presto 升级到 0.194
- 已升级 AWS SDK for Java 到 1.11.297

### 更改、增强功能和解决的问题

- Hive
  - 向后移植 [HIVE-15436](#)。增强了 HiveAPIs，仅返回视图。

### 已知问题

- MXNet目前没有 OpenCV 库。

## 5.13.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 emr 或 aws 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.0.1	亚马逊 SageMaker Spark SDK
emr-ddb	4.5.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.4.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.10.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.22.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.4.0	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。



组件	版本	描述
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.8.3-amzn-0	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.8.3-amzn-0	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.3-amzn-0	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.3-amzn-0	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.8.3-amzn-0	HTTPHDFS操作端点。
hadoop-kms-server	2.8.3-amzn-0	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.8.3-amzn-0	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.8.3-amzn-0	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.8.3-amzn-0	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.3-amzn-0	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.2	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.2	为一个或多个HBase地区提供服务的服务。

组件	版本	描述
hbase-client	1.4.2	HBase命令行客户端。
hbase-rest-server	1.4.2	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.2	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	2.3.2-amzn-2	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.2-amzn-2	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.2-amzn-2	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.2-amzn-2	Hive 命令行客户端。
hive-hbase	2.3.2-amzn-2	Hive-hbase 客户端。
hive-metastore-server	2.3.2-amzn-2	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.2-amzn-2	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	4.1.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
livy-server	0.4.0-incubating	REST用于与 Apache Spark 交互的接口
mahout-client	0.13.0	用于机器学习的库。

组件	版本	描述
mxnet	1.0.0	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
nvidia-cuda	9.1.85	Nvidia 驱动程序和 Cuda 工具包
oozie-client	4.3.0	Oozie 命令行客户端。
oozie-server	4.3.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.13.0--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.13.0--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.194	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.194	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。
r	3.4.1	用于统计计算的 R 项目
spark-client	2.3.0	Spark 命令行客户端。
spark-history-server	2.3.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.3.0	的内存中执行引擎。YARN

组件	版本	描述
spark-yarn-slave	2.3.0	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.7.3	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.10	ZooKeeper 命令行客户端。

### 5.13.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.13.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。

分类	描述
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。

分类	描述
hcatalog-webhcat-log4j2	更改 HCatalog WebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog WebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改 HTTPFS 环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。

分类	描述
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。

分类	描述
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。



分类	描述
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.12.3

### 5.12.3 应用程序版本

此版本支持以下应用程

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本 ( 如果适用 )。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-5.12.3	emr-5.12.2	emr-5.12.1	emr-5.12.0
适用于 Java 的 AWS SDK	1.11.267	1.11.267	1.11.267	1.11.267
Python	2.7、3.4	2.7、3.4	2.7、3.4	2.7、3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.0	1.4.0	1.4.0	1.4.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.0	1.4.0	1.4.0	1.4.0
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-

	emr-5.12.3	emr-5.12.2	emr-5.12.1	emr-5.12.0
Hue	4.1.0	4.1.0	4.1.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	1.0.0	1.0.0	1.0.0	1.0.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.188	0.188	0.188	0.188
Spark	2.2.1	2.2.1	2.2.1	2.2.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

### 5.12.3 发布说明

这是一个补丁版本，用于为向 Amazon S3 发出的请求添加 AWS 签名版本 4 身份验证。所有应用程序和组件都与之前的 Amazon EMR 版本相同。

#### Important

在此版本中，亚马逊仅EMR使用 AWS 签名版本 4 来验证向 Amazon S3 发出的请求。有关更多信息，请参阅[新功能](#)。

### 5.12.3 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>aws-sagemaker-spark-sdk</code>	1.0.1	亚马逊 SageMaker Spark SDK
<code>emr-ddb</code>	4.5.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.4.0	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。

组件	版本	描述
emr-s3-dist-cp	2.9.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.21.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.4.0	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.8.3-amzn-0	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.8.3-amzn-0	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.3-amzn-0	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.3-amzn-0	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.8.3-amzn-0	HTTPHDFS操作端点。
hadoop-kms-server	2.8.3-amzn-0	基于 Hadoop 的加密密钥管理服务器。 KeyProvider API

组件	版本	描述
hadoop-mapred	2.8.3-amzn-0	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.8.3-amzn-0	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.8.3-amzn-0	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.3-amzn-0	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.0	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.0	为一个或多个HBase地区提供服务的服务。
hbase-client	1.4.0	HBase命令行客户端。
hbase-rest-server	1.4.0	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.0	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	2.3.2-amzn-1	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.2-amzn-1	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.2-amzn-1	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.2-amzn-1	Hive 命令行客户端。

组件	版本	描述
hive-hbase	2.3.2-amzn-1	Hive-hbase 客户端。
hive-metastore-server	2.3.2-amzn-1	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的 SQL 元数据。
hive-server2	2.3.2-amzn-1	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	4.1.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
livy-server	0.4.0-incubating	REST 用于与 Apache Spark 交互的接口
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.0.0	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的 SQL 数据库服务器。
nvidia-cuda	9.1.85	Nvidia 驱动程序和 Cuda 工具包
oozie-client	4.3.0	Oozie 命令行客户端。
oozie-server	4.3.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.13.0--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.13.0--1.4 HBase	一款轻量级服务器，提供 JDBC 对 Avatica 的访问权限以及协议缓冲区和 JSON 格式访问权限 API

组件	版本	描述
presto-coordinator	0.188	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.188	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。
spark-client	2.2.1	Spark 命令行客户端。
spark-history-server	2.2.1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.2.1	的内存中执行引擎。YARN
spark-yarn-slave	2.2.1	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.7.3	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.10	ZooKeeper 命令行客户端。



### 5.12.3 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.12.3 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。

分类	描述
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。

分类	描述
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。

分类	描述
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。

分类	描述
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。

分类	描述
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper 中的值。
zookeeper-log4j	更改 ZooKeeper 的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.12.2

### 5.12.2 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-5.12.2	emr-5.12.1	emr-5.12.0	emr-5.11.4
适用于 Java 的 AWS SDK	1.11.267	1.11.267	1.11.267	1.11.238
Python	2.7、3.4	2.7、3.4	2.7、3.4	2.7、3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8

	emr-5.12.2	emr-5.12.1	emr-5.12.0	emr-5.11.4
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.0	1.4.0	1.4.0	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.0	1.4.0	1.4.0	1.3.1
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.7.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.1.0	4.1.0	4.0.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	1.0.0	1.0.0	1.0.0	0.12.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0

	emr-5.12.2	emr-5.12.1	emr-5.12.0	emr-5.11.4
Presto	0.188	0.188	0.188	0.187
Spark	2.2.1	2.2.1	2.2.1	2.2.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

## 5.12.2 发布说明

以下发行说明包含有关亚马逊 5.12.2 EMR 版本的信息。更改与 5.12.1 有关。

首次发布日期：2018年 8 月 29 日

更改、增强功能和解决的问题

- 此版本解决了潜在的安全漏洞。

## 5.12.2 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。



组件	版本	描述
aws-sagemaker-spark-sdk	1.0.1	亚马逊 SageMaker Spark SDK
emr-ddb	4.5.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.4.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.9.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.21.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.4.0	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.8.3-amzn-0	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。

组件	版本	描述
hadoop-hdfs-datanode	2.8.3-amzn-0	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.3-amzn-0	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.3-amzn-0	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httfs-server	2.8.3-amzn-0	HTTPHDFS操作端点。
hadoop-kms-server	2.8.3-amzn-0	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.8.3-amzn-0	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.8.3-amzn-0	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.8.3-amzn-0	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.3-amzn-0	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.0	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.0	为一个或多个HBase地区提供服务的服务。
hbase-client	1.4.0	HBase命令行客户端。
hbase-rest-server	1.4.0	为其提供RESTfulHTTP终端节点的服务HBase。

组件	版本	描述
hbase-thrift-server	1.4.0	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	2.3.2-amzn-1	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.2-amzn-1	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.2-amzn-1	HTTP端点为其提供REST接口 HCatalog。
hive-client	2.3.2-amzn-1	Hive 命令行客户端。
hive-hbase	2.3.2-amzn-1	Hive-hbase 客户端。
hive-metastore-server	2.3.2-amzn-1	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.2-amzn-1	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	4.1.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
livy-server	0.4.0-incubating	REST用于与 Apache Spark 交互的接口
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.0.0	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的SQL数据库服务器。

组件	版本	描述
nvidia-cuda	9.1.85	Nvidia 驱动程序和 Cuda 工具包
oozie-client	4.3.0	Oozie 命令行客户端。
oozie-server	4.3.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.13.0--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.13.0--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.188	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.188	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。
spark-client	2.2.1	Spark 命令行客户端。
spark-history-server	2.2.1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.2.1	的内存中执行引擎。YARN
spark-yarn-slave	2.2.1	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。

组件	版本	描述
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.7.3	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.10	ZooKeeper 命令行客户端。

### 5.12.2 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.12.2 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。

分类	描述
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。

分类	描述
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。

分类	描述
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。



分类	描述
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。

分类	描述
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.12.1

### 5.12.1 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)

- [Amazon EMR 4.x 发行版中的应用程序版本](#)

## 应用程序版本信息

	emr-5.12.1	emr-5.12.0	emr-5.11.4	emr-5.11.3
适用于 Java 的 AWS SDK	1.11.267	1.11.267	1.11.238	1.11.238
Python	2.7、3.4	2.7、3.4	2.7、3.4	2.7、3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.0	1.4.0	1.3.2	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.0	1.4.0	1.3.1	1.3.1
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.7.3	2.7.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.1.0	4.0.1	4.0.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-

	emr-5.12.1	emr-5.12.0	emr-5.11.4	emr-5.11.3
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	1.0.0	1.0.0	0.12.0	0.12.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.188	0.188	0.187	0.187
Spark	2.2.1	2.2.1	2.2.1	2.2.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

## 5.12.1 发布说明

以下发行说明包含有关亚马逊 5.12.1 EMR 版本的信息。更改与 5.12.0 有关。

首次发布日期：2018 年 3 月 29 日

更改、增强功能和解决的问题

- 更新了亚马逊 Linux 版的亚马逊 defaultAmazon Linux AMI 内核EMR以修复潜在漏洞。

## 5.12.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>aws-sagemaker-spark-sdk</code>	1.0.1	亚马逊 SageMaker Spark SDK
<code>emr-ddb</code>	4.5.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.4.0	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
<code>emr-s3-dist-cp</code>	2.9.0	针对 Amazon S3 优化的分布式复制应用程序。
<code>emrfs</code>	2.21.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
<code>flink-client</code>	1.4.0	Apache Flink 命令行客户端脚本和应用程序。
<code>ganglia-monitor</code>	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。

组件	版本	描述
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.8.3-amzn-0	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.8.3-amzn-0	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.3-amzn-0	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.3-amzn-0	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.8.3-amzn-0	HTTPHDFS操作端点。
hadoop-kms-server	2.8.3-amzn-0	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.8.3-amzn-0	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.8.3-amzn-0	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.8.3-amzn-0	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.3-amzn-0	用于检索YARN应用程序的当前和历史信息的服务。

组件	版本	描述
hbase-hmaster	1.4.0	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.0	为一个或多个HBase地区提供服务的服务器。
hbase-client	1.4.0	HBase命令行客户端。
hbase-rest-server	1.4.0	为其提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.0	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.3.2-amzn-1	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.2-amzn-1	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.2-amzn-1	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.2-amzn-1	Hive 命令行客户端。
hive-hbase	2.3.2-amzn-1	Hive-hbase 客户端。
hive-metastore-server	2.3.2-amzn-1	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.2-amzn-1	用于将 Hive 查询作为 Web 请求接受的服务。

组件	版本	描述
hue-server	4.1.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
livy-server	0.4.0-incubating	REST用于与 Apache Spark 交互的接口
mahout-client	0.13.0	用于机器学习的库。
mxnet	1.0.0	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
nvidia-cuda	9.1.85	Nvidia 驱动程序和 Cuda 工具包
oozie-client	4.3.0	Oozie 命令行客户端。
oozie-server	4.3.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.13.0--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.13.0--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.188	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.188	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。
spark-client	2.2.1	Spark 命令行客户端。



组件	版本	描述
spark-history-server	2.2.1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.2.1	的内存中执行引擎。YARN
spark-yarn-slave	2.2.1	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.7.3	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.10	ZooKeeper 命令行客户端。

### 5.12.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.12.1 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。

分类	描述
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改的 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。

分类	描述
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。

分类	描述
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。

分类	描述
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。

分类	描述
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.12.0

### 5.12.0 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-5.12.0	emr-5.11.4	emr-5.11.3	emr-5.11.2
适用于 Java 的 AWS SDK	1.11.267	1.11.238	1.11.238	1.11.238
Python	2.7、3.4	2.7、3.4	2.7、3.4	2.7、3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.0	1.3.2	1.3.2	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.0	1.3.1	1.3.1	1.3.1

	emr-5.12.0	emr-5.11.4	emr-5.11.3	emr-5.11.2
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.7.3	2.7.3	2.7.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.0.1	4.0.1	4.0.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	1.0.0	0.12.0	0.12.0	0.12.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.188	0.187	0.187	0.187
Spark	2.2.1	2.2.1	2.2.1	2.2.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4



	emr-5.12.0	emr-5.11.4	emr-5.11.3	emr-5.11.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

## 5.12.0 发布说明

以下发行说明包含有关亚马逊 5.12.0 EMR 版本的信息。更改与 5.11.1 有关。

### 升级

- AWS SDK适用于 Java 1.11.238 ⇒ 1.11.267。有关更多信息，请参阅适用于 [Java AWS SDK 的 Change Log](#) on GitHub。
- Hadoop 2.7.3 升级到 2.8.3。有关更多信息，请参阅 [Apache Hadoop 发行版](#)。
- Fink 1.3.2 升级到 1.4.0。有关详细信息，请参阅 [Apache Flink 1.4.0 版本公告](#)。
- HBase1.3.1 ⇒ 1.4.0。有关更多信息，请参阅[HBase发布公告](#)。
- Hue 4.0.1 升级到 4.1.0。有关更多信息，请参阅[发布说明](#)。
- MxNet 0.12.0 ⇒ 1.0.0。有关更多信息，请参阅“[MXNet更改日志](#)” GitHub。
- Presto 0.187 升级到 0.188。有关更多信息，请参阅[发布说明](#)。

### 更改、增强功能和解决的问题

- Hadoop
  - `yarn.resourcemanager.decommissioning.timeout` 属性已更改为 `yarn.resourcemanager.nodemanager-graceful-decommission-timeout-secs`。您可以使用此属性自定义集群缩减。有关更多信息，请参阅 Amazon EMR 管理指南中的[集群扩容](#)。
  - Hadoop 在 `cp` (复制) 命令中CLI添加了该 `-d` 选项，该命令指定了直接复制。可以使用它来避免创建中间 `.COPYING` 文件，这加快了在 Amazon S3 之间复制数据的速度。有关更多信息，请参阅 [HADOOP-12384](#)。
- Pig

- 添加了 `pig-env` 配置分类，这简化了 Pig 环境属性的配置。有关更多信息，请参阅[配置应用程序](#)。
- Presto
  - 新增 `presto-connector-redshift` 配置分类，您可以将其用于配置 Presto `redshift.properties` 配置文件中的值。有关更多信息，请参阅 Presto 文档中[Redshift 连接器](#)以及[配置应用程序](#)。
  - EMRFS已添加对 Presto 的支持，这是默认配置。亚马逊早期EMR版本使用的是 `PrestoS3FileSystem`，这是唯一的选择。有关更多信息，请参阅[EMRFS和 Presto FileSystem S3 配置](#)。

#### Note

如果您使用亚马逊EMR版本 5.12.0 在 Amazon S3 中查询基础数据，则可能会出现 Presto 错误。这是因为 Presto 无法从 `emrfs-site.xml` 提取配置分类值。解决方法是在 `usr/lib/presto/plugin/hive-hadoop2/` 下创建一个 `emrfs` 子目录，并在 `usr/lib/presto/plugin/hive-hadoop2/emrfs` 中创建一个指向现有 `/usr/share/aws/emr/emrfs/conf/emrfs-site.xml` 文件的符号链接。然后重新启动 `presto-server` 进程（首先执行 `sudo presto-server stop`，然后执行 `sudo presto-server start`）。

- Spark
  - 向后移植的 [SPARK-22036 : BigDecimal 乘法有时会返回空值](#)。

#### 已知问题

- MXNet不包括 OpenCV 库。
- SparkR 不适用于使用自定义创建的集群，因为默认情况下 AMI R 未安装在群集节点上。

## 5.12.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.0.1	亚马逊 SageMaker Spark SDK
emr-ddb	4.5.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.4.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.9.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.21.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.4.0	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。

组件	版本	描述
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.8.3-amzn-0	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.8.3-amzn-0	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.8.3-amzn-0	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.8.3-amzn-0	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.8.3-amzn-0	HTTPHDFS操作端点。
hadoop-kms-server	2.8.3-amzn-0	基于 Hadoop 的加密密钥管理服务器。 KeyProvider API
hadoop-mapred	2.8.3-amzn-0	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.8.3-amzn-0	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.8.3-amzn-0	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.8.3-amzn-0	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.4.0	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.4.0	为一个或多个HBase地区提供服务的服务。

组件	版本	描述
hbase-client	1.4.0	HBase命令行客户端。
hbase-rest-server	1.4.0	为其提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.4.0	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	2.3.2-amzn-1	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.2-amzn-1	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.2-amzn-1	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.2-amzn-1	Hive 命令行客户端。
hive-hbase	2.3.2-amzn-1	Hive-hbase 客户端。
hive-metastore-server	2.3.2-amzn-1	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.2-amzn-1	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	4.1.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
livy-server	0.4.0-incubating	REST用于与 Apache Spark 交互的接口
mahout-client	0.13.0	用于机器学习的库。

组件	版本	描述
mxnet	1.0.0	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
nvidia-cuda	9.1.85	Nvidia 驱动程序和 Cuda 工具包
oozie-client	4.3.0	Oozie 命令行客户端。
oozie-server	4.3.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.13.0--1.4 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.13.0--1.4 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.188	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.188	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。
spark-client	2.2.1	Spark 命令行客户端。
spark-history-server	2.2.1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.2.1	的内存中执行引擎。YARN
spark-yarn-slave	2.2.1	YARN从属服务器需要的 Apache Spark 库。

组件	版本	描述
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.7.3	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.10	ZooKeeper 命令行客户端。

## 5.12.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

### emr-5.12.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。

分类	描述
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。



分类	描述
hcatalog-webhcat-site	更改 HCatalog WebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。

分类	描述
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-env	更改 Pig 环境中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。

分类	描述
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-redshift	更改 Presto 的 redshift.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值

分类	描述
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.11.4

### 5.11.4 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)

- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

## 应用程序版本信息

	emr-5.11.4	emr-5.11.3	emr-5.11.2	emr-5.11.1
适用于 Java 的 AWS SDK	1.11.238	1.11.238	1.11.238	1.11.238
Python	2.7、3.4	2.7、3.4	2.7、3.4	2.7、3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.2	1.3.2	1.3.2	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.0.1	4.0.1	4.0.1	4.0.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-

	emr-5.11.4	emr-5.11.3	emr-5.11.2	emr-5.11.1
JupyterHub	-	-	-	-
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	0.12.0	0.12.0	0.12.0	0.12.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.187	0.187	0.187	0.187
Spark	2.2.1	2.2.1	2.2.1	2.2.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

## 5.11.4 发布说明

这是一个补丁版本，用于为向 Amazon S3 发出的请求添加 AWS 签名版本 4 身份验证。所有应用程序和组件都与之前的 Amazon EMR 版本相同。

**⚠ Important**

在此版本中，亚马逊仅EMR使用 AWS 签名版本 4 来验证向 Amazon S3 发出的请求。有关更多信息，请参阅[新功能](#)。

### 5.11.4 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>aws-sagemaker-spark-sdk</code>	1.0	亚马逊 SageMaker Spark SDK
<code>emr-ddb</code>	4.5.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.4.0	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
<code>emr-s3-dist-cp</code>	2.8.0	针对 Amazon S3 优化的分布式复制应用程序。
<code>emrfs</code>	2.20.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。

组件	版本	描述
flink-client	1.3.2	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-6	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-6	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.3-amzn-6	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.3-amzn-6	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.3-amzn-6	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-6	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-6	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.3-amzn-6	YARN用于管理单个节点上的容器的服务。



组件	版本	描述
hadoop-yarn-resourcemanager	2.7.3-amzn-6	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.3-amzn-6	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.3.1	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.3.1	为一个或多个HBase地区提供服务的服务。
hbase-client	1.3.1	HBase命令行客户端。
hbase-rest-server	1.3.1	为其提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.3.1	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.3.2-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.2-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.2-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.2-amzn-0	Hive 命令行客户端。
hive-hbase	2.3.2-amzn-0	Hive-hbase 客户端。
hive-metastore-server	2.3.2-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。

组件	版本	描述
hive-server2	2.3.2-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	4.0.1	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
livy-server	0.4.0-incubating	REST用于与 Apache Spark 交互的接口
mahout-client	0.13.0	用于机器学习的库。
mxnet	0.12.0	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
nvidia-cuda	9.0.176	Nvidia 驱动程序和 Cuda 工具包
oozie-client	4.3.0	Oozie 命令行客户端。
oozie-server	4.3.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.11.0--1.3 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.11.0--1.3 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.187	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.187	用于执行查询的各个部分的服务。

组件	版本	描述
pig-client	0.17.0	Pig 命令行客户端。
spark-client	2.2.1	Spark 命令行客户端。
spark-history-server	2.2.1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.2.1	的内存中执行引擎。YARN
spark-yarn-slave	2.2.1	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.7.3	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.10	ZooKeeper 命令行客户端。

#### 5.11.4 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

## emr-5.11.4 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。

分类	描述
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值

分类	描述
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。

分类	描述
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。

分类	描述
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。



## 亚马逊EMR版本 5.11.3

### 5.11.3 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-5.11.3	emr-5.11.2	emr-5.11.1	emr-5.11.0
适用于 Java 的 AWS SDK	1.11.238	1.11.238	1.11.238	1.11.238
Python	2.7、3.4	2.7、3.4	2.7、3.4	2.7、3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.2	1.3.2	1.3.2	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1

	emr-5.11.3	emr-5.11.2	emr-5.11.1	emr-5.11.0
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.0.1	4.0.1	4.0.1	4.0.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	0.12.0	0.12.0	0.12.0	0.12.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.187	0.187	0.187	0.187
Spark	2.2.1	2.2.1	2.2.1	2.2.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.11.3	emr-5.11.2	emr-5.11.1	emr-5.11.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

### 5.11.3 发布说明

以下发行说明包含有关亚马逊 5.11.3 EMR 版本的信息。更改与 5.11.2 有关。

首次发布日期：2019 年 7 月 18 日

更改、增强功能和解决的问题

- 更新了亚马逊的默认 Amazon Linux AMI，EMR使其包含重要的 Linux 内核安全更新，包括 TCPSACK拒绝服务问题 ([AWS-2019-005](#))。

### 5.11.3 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 `CommunityVersion-amzn-EmrVersion` 的发行版标注。`EmrVersion` 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.0	亚马逊 SageMaker Spark SDK

组件	版本	描述
emr-ddb	4.5.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.4.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.8.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.20.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.3.2	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-6	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-6	HDFS用于存储区块的节点级服务。

组件	版本	描述
hadoop-hdfs-library	2.7.3-amzn-6	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.3-amzn-6	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httfs-server	2.7.3-amzn-6	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-6	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-6	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.3-amzn-6	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-6	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.3-amzn-6	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.3.1	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.3.1	为一个或多个HBase地区提供服务的服务。
hbase-client	1.3.1	HBase命令行客户端。
hbase-rest-server	1.3.1	为其提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.3.1	为其提供 Thrift 端点的HBase 服务。

组件	版本	描述
hcatalog-client	2.3.2-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.2-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.2-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.2-amzn-0	Hive 命令行客户端。
hive-hbase	2.3.2-amzn-0	Hive-hbase 客户端。
hive-metastore-server	2.3.2-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.2-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	4.0.1	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
livy-server	0.4.0-incubating	REST用于与 Apache Spark 交互的接口
mahout-client	0.13.0	用于机器学习的库。
mxnet	0.12.0	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
nvidia-cuda	9.0.176	Nvidia 驱动程序和 Cuda 工具包

组件	版本	描述
oozie-client	4.3.0	Oozie 命令行客户端。
oozie-server	4.3.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.11.0--1.3 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.11.0--1.3 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.187	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.187	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。
spark-client	2.2.1	Spark 命令行客户端。
spark-history-server	2.2.1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.2.1	的内存中执行引擎。YARN
spark-yarn-slave	2.2.1	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。

组件	版本	描述
zeppelin-server	0.7.3	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.10	ZooKeeper 命令行客户端。

### 5.11.3 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.11.3 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。



分类	描述
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。

分类	描述
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。

分类	描述
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。

分类	描述
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。

分类	描述
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.11.2

### 5.11.2 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

## 应用程序版本信息

	emr-5.11.2	emr-5.11.1	emr-5.11.0	emr-5.10.1
适用于 Java 的 AWS SDK	1.11.238	1.11.238	1.11.238	1.11.221
Python	2.7、3.4	2.7、3.4	2.7、3.4	2.7、3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.2	1.3.2	1.3.2	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.2	2.3.2	2.3.2	2.3.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.2	2.3.2	2.3.2	2.3.1
Hudi	-	-	-	-
Hue	4.0.1	4.0.1	4.0.1	4.0.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	0.12.0	0.12.0	0.12.0	0.12.0

	emr-5.11.2	emr-5.11.1	emr-5.11.0	emr-5.10.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.187	0.187	0.187	0.187
Spark	2.2.1	2.2.1	2.2.1	2.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

## 5.11.2 发布说明

以下发行说明包含有关亚马逊 5.11.2 EMR 版本的信息。更改与 5.11.1 有关。

首次发布日期：2018年 8 月 29 日

更改、增强功能和解决的问题

- 此版本解决了潜在的安全漏洞。

## 5.11.2 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 emr 或 aws 开头。Amazon

最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.0	亚马逊 SageMaker Spark SDK
emr-ddb	4.5.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.4.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.8.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.20.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.3.2	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。



组件	版本	描述
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-6	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-6	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.3-amzn-6	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.3-amzn-6	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.3-amzn-6	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-6	基于 Hadoop 的加密密钥管理服务器。 KeyProvider API
hadoop-mapred	2.7.3-amzn-6	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.3-amzn-6	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-6	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.3-amzn-6	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.3.1	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.3.1	为一个或多个HBase地区提供服务的服务。

组件	版本	描述
hbase-client	1.3.1	HBase命令行客户端。
hbase-rest-server	1.3.1	为其提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.3.1	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	2.3.2-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.2-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.2-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.2-amzn-0	Hive 命令行客户端。
hive-hbase	2.3.2-amzn-0	Hive-hbase 客户端。
hive-metastore-server	2.3.2-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.2-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	4.0.1	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
livy-server	0.4.0-incubating	REST用于与 Apache Spark 交互的接口
mahout-client	0.13.0	用于机器学习的库。

组件	版本	描述
mxnet	0.12.0	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
nvidia-cuda	9.0.176	Nvidia 驱动程序和 Cuda 工具包
oozie-client	4.3.0	Oozie 命令行客户端。
oozie-server	4.3.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.11.0--1.3 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.11.0--1.3 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.187	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.187	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。
spark-client	2.2.1	Spark 命令行客户端。
spark-history-server	2.2.1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.2.1	的内存中执行引擎。YARN
spark-yarn-slave	2.2.1	YARN从属服务器需要的 Apache Spark 库。

组件	版本	描述
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.7.3	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.10	ZooKeeper 命令行客户端。

### 5.11.2 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.11.2 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。

分类	描述
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。

分类	描述
hcatalog-webhcat-site	更改 HCatalog WebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。

分类	描述
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。

分类	描述
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。



分类	描述
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.11.1

### 5.11.1 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

## 应用程序版本信息

	emr-5.11.1	emr-5.11.0	emr-5.10.1	emr-5.10.0
适用于 Java 的 AWS SDK	1.11.238	1.11.238	1.11.221	1.11.221
Python	2.7、3.4	2.7、3.4	2.7、3.4	2.7、3.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.2	1.3.2	1.3.2	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.2	2.3.2	2.3.1	2.3.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.2	2.3.2	2.3.1	2.3.1
Hudi	-	-	-	-
Hue	4.0.1	4.0.1	4.0.1	4.0.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	0.12.0	0.12.0	0.12.0	0.12.0

	emr-5.11.1	emr-5.11.0	emr-5.10.1	emr-5.10.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.187	0.187	0.187	0.187
Spark	2.2.1	2.2.1	2.2.0	2.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

### 5.11.1 发布说明

以下发行说明包含 Amazon EMR 5.11.1 版本的信息。更改与 Amazon EMR 5.8.0 版本有关。

首次发布日期：2018 年 1 月 22 日

#### 更改、增强功能和解决的问题

- 更新了亚马逊版 Linux 的亚马逊 L defaultAmazon inux AMI 内核，EMR以解决与推测执行相关的漏洞 ( CVE-2017-5715、-2017-5753 和 -2017 CVE -5754 )。CVE有关更多信息，请参阅 <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>。

### 5.11.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.0	亚马逊 SageMaker Spark SDK
emr-ddb	4.5.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.4.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.8.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.20.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.3.2	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。

组件	版本	描述
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-6	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-6	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.3-amzn-6	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.3-amzn-6	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.3-amzn-6	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-6	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-6	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.3-amzn-6	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-6	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.3-amzn-6	用于检索YARN应用程序的当前和历史信息的服务。

组件	版本	描述
hbase-hmaster	1.3.1	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.3.1	为一个或多个HBase地区提供服务的服务器。
hbase-client	1.3.1	HBase命令行客户端。
hbase-rest-server	1.3.1	为其提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.3.1	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	2.3.2-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.2-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.2-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.2-amzn-0	Hive 命令行客户端。
hive-hbase	2.3.2-amzn-0	Hive-hbase 客户端。
hive-metastore-server	2.3.2-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.2-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。

组件	版本	描述
hue-server	4.0.1	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
livy-server	0.4.0-incubating	REST用于与 Apache Spark 交互的接口
mahout-client	0.13.0	用于机器学习的库。
mxnet	0.12.0	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
nvidia-cuda	9.0.176	Nvidia 驱动程序和 Cuda 工具包
oozie-client	4.3.0	Oozie 命令行客户端。
oozie-server	4.3.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.11.0--1.3 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.11.0--1.3 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.187	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.187	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。
spark-client	2.2.1	Spark 命令行客户端。

组件	版本	描述
spark-history-server	2.2.1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.2.1	的内存中执行引擎。YARN
spark-yarn-slave	2.2.1	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.7.3	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.10	ZooKeeper 命令行客户端。

### 5.11.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.11.1 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。



分类	描述
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。

分类	描述
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。

分类	描述
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。

分类	描述
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。

分类	描述
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.11.0

### 5.11.0 应用程序版本

此版本支持以下应用程

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本 ( 如果适用 )。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-5.11.0	emr-5.10.1	emr-5.10.0	emr-5.9.1
适用于 Java 的 AWS SDK	1.11.238	1.11.221	1.11.221	1.11.183
Python	2.7、3.4	2.7、3.4	2.7、3.4	未跟踪
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.2	1.3.2	1.3.2	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.2	2.3.1	2.3.1	2.3.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.2	2.3.1	2.3.1	2.3.0
Hudi	-	-	-	-

	emr-5.11.0	emr-5.10.1	emr-5.10.0	emr-5.9.1
Hue	4.0.1	4.0.1	4.0.1	4.0.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	0.12.0	0.12.0	0.12.0	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.187	0.187	0.187	0.184
Spark	2.2.1	2.2.0	2.2.0	2.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.2
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

## 5.11.0 发布说明

以下发行说明包含有关亚马逊 5.11.0 EMR 版本的信息。更改与 5.10.0 有关。

### 升级

- Hive 2.3.2
- Spark 2.2.1
- SDK适用于 Java 1.11.238

### 新功能

- Spark
  - 增加了 `spark.decommissioning.timeout.threshold` 设置，这将改进使用竞价型实例时的 Spark 停用行为。有关更多信息，请参阅 [配置节点停用行为](#)。
  - 在 Spark 中添加了 `aws-sagemaker-spark-sdk` 组件，它安装了 Amazon SageMaker Spark 以及与 Spark 与[亚马逊](#)集成的相关依赖项 SageMaker。您可以使用 Amazon SageMaker Spark 通过亚马逊 SageMaker 阶段构建 Spark 机器学习 (ML) 管道。有关更多信息，请参阅《亚马逊 SageMaker 开发者指南》中的 [SageMaker Spark 自述文件](#) GitHub 和 SageMaker在亚马逊[上使用 Apache Spark](#)。

### 已知问题

- MXNet不包括 OpenCV 库。
- 默认情况下，Hive 2.3.2 设置 `hive.compute.query.using.stats=true`。这会导致查询从现有统计数据而不是直接从数据中获取数据，这可能会造成混淆。例如，如果您有一个包含 `hive.compute.query.using.stats=true` 的表并且将新文件上载到表 LOCATION，则在该表上运行 `SELECT COUNT(*)` 查询会返回来自统计数据的计数，而不是选择已添加的行。

作为解决方法，请使用 `ANALYZE TABLE` 命令收集新的统计数据，或者设置 `hive.compute.query.using.stats=false`。有关更多信息，请参阅 Apache Hive 文档中的 [Hive 中的统计数据](#)。

## 5.11.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon



最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
aws-sagemaker-spark-sdk	1.0	亚马逊 SageMaker Spark SDK
emr-ddb	4.5.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.4.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.8.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.20.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.3.2	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。

组件	版本	描述
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-6	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-6	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.3-amzn-6	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.3-amzn-6	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.3-amzn-6	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-6	基于 Hadoop 的加密密钥管理服务器。 KeyProvider API
hadoop-mapred	2.7.3-amzn-6	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.3-amzn-6	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-6	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.3-amzn-6	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.3.1	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.3.1	为一个或多个HBase地区提供服务的服务。

组件	版本	描述
hbase-client	1.3.1	HBase命令行客户端。
hbase-rest-server	1.3.1	为其提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.3.1	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	2.3.2-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.2-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.2-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.2-amzn-0	Hive 命令行客户端。
hive-hbase	2.3.2-amzn-0	Hive-hbase 客户端。
hive-metastore-server	2.3.2-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.2-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	4.0.1	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
livy-server	0.4.0-incubating	REST用于与 Apache Spark 交互的接口
mahout-client	0.13.0	用于机器学习的库。

组件	版本	描述
mxnet	0.12.0	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
nvidia-cuda	9.0.176	Nvidia 驱动程序和 Cuda 工具包
oozie-client	4.3.0	Oozie 命令行客户端。
oozie-server	4.3.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.11.0--1.3 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.11.0--1.3 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.187	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.187	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。
spark-client	2.2.1	Spark 命令行客户端。
spark-history-server	2.2.1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.2.1	的内存中执行引擎。YARN
spark-yarn-slave	2.2.1	YARN从属服务器需要的 Apache Spark 库。

组件	版本	描述
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.7.3	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.10	ZooKeeper 命令行客户端。

### 5.11.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.11.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。

分类	描述
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。

分类	描述
hcatalog-webhcat-site	更改 HCatalog WebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。

分类	描述
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。



分类	描述
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。

分类	描述
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.10.1

### 5.10.1 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

## 应用程序版本信息

	emr-5.10.1	emr-5.10.0	emr-5.9.1	emr-5.9.0
适用于 Java 的 AWS SDK	1.11.221	1.11.221	1.11.183	1.11.183
Python	2.7、3.4	2.7、3.4	未跟踪	未跟踪
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.2	1.3.2	1.3.2	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.1	2.3.1	2.3.0	2.3.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.1	2.3.1	2.3.0	2.3.0
Hudi	-	-	-	-
Hue	4.0.1	4.0.1	4.0.1	4.0.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0.4.0	0.4.0	0.4.0
MXNet	0.12.0	0.12.0	-	-

	emr-5.10.1	emr-5.10.0	emr-5.9.1	emr-5.9.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.187	0.187	0.184	0.184
Spark	2.2.0	2.2.0	2.2.0	2.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.2	0.7.2
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

### 5.10.1 发布说明

这是一个补丁版本，用于为向 Amazon S3 发出的请求添加 AWS 签名版本 4 身份验证。所有应用程序和组件都与之前的 Amazon EMR 版本相同。

#### Important

在此版本中，亚马逊仅EMR使用 AWS 签名版本 4 来验证向 Amazon S3 发出的请求。有关更多信息，请参阅[新功能](#)。

## 5.10.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>emr-ddb</code>	4.5.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.4.0	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
<code>emr-s3-dist-cp</code>	2.7.0	针对 Amazon S3 优化的分布式复制应用程序。
<code>emrfs</code>	2.20.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
<code>flink-client</code>	1.3.2	Apache Flink 命令行客户端脚本和应用程序。
<code>ganglia-monitor</code>	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。

组件	版本	描述
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-5	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-5	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.3-amzn-5	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.3-amzn-5	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.3-amzn-5	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-5	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-5	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.3-amzn-5	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-5	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.3-amzn-5	用于检索YARN应用程序的当前和历史信息的服务。

组件	版本	描述
hbase-hmaster	1.3.1	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.3.1	为一个或多个HBase地区提供服务的服务器。
hbase-client	1.3.1	HBase命令行客户端。
hbase-rest-server	1.3.1	为其提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.3.1	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.3.1-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.1-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.1-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.1-amzn-0	Hive 命令行客户端。
hive-hbase	2.3.1-amzn-0	Hive-hbase 客户端。
hive-metastore-server	2.3.1-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.1-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。

组件	版本	描述
hue-server	4.0.1	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
livy-server	0.4.0-incubating	REST用于与 Apache Spark 交互的接口
mahout-client	0.13.0	用于机器学习的库。
mxnet	0.12.0	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
nvidia-cuda	9.0.176	Nvidia 驱动程序和 Cuda 工具包
oozie-client	4.3.0	Oozie 命令行客户端。
oozie-server	4.3.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.11.0--1.3 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.11.0--1.3 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.187	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.187	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。
spark-client	2.2.0	Spark 命令行客户端。



组件	版本	描述
spark-history-server	2.2.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.2.0	的内存中执行引擎。YARN
spark-yarn-slave	2.2.0	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.7.3	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.10	ZooKeeper 命令行客户端。

### 5.10.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.10.1 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。

分类	描述
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。

分类	描述
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。

分类	描述
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。

分类	描述
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。

分类	描述
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.10.0

### 5.10.0 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Livy](#)、[MXNet](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本 ( 如果适用 )。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-5.10.0	emr-5.9.1	emr-5.9.0	emr-5.8.3
适用于 Java 的 AWS SDK	1.11.221	1.11.183	1.11.183	1.11.160
Python	2.7、3.4	未跟踪	未跟踪	未跟踪
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.2	1.3.2	1.3.2	1.3.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.1	2.3.0	2.3.0	2.3.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.1	2.3.0	2.3.0	2.3.0
Hudi	-	-	-	-

	emr-5.10.0	emr-5.9.1	emr-5.9.0	emr-5.8.3
Hue	4.0.1	4.0.1	4.0.1	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0.4.0	0.4.0	-
MXNet	0.12.0	-	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.16.0
Presto	0.187	0.184	0.184	0.170
Spark	2.2.0	2.2.0	2.2.0	2.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.2	0.7.2	0.7.2
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10



## 5.10.0 发布说明

以下发行说明包含有关亚马逊 5.10.0 EMR 版本的信息。更改与 Amazon EMR 5.9.0 版本有关。

### 升级

- AWS SDK for Java 1.11.221
- Hive 2.3.1
- Presto 0.187

### 新功能

- 添加了对 Kerberos 身份验证的支持。有关更多信息，请参阅《Amazon EMR 管理指南》中的“[使用 Kerberos 身份验证](#)”
- 添加了对IAM角色的支持EMRFS。有关更多信息，请参阅《亚马逊EMR管理指南》中的为向 Amazon S3 发出的EMRFS[请求配置IAM角色](#)
- 增加了对GPU基于 P2 和 P3 的实例类型的支持。有关更多信息，请参阅[亚马逊 EC2 P2 实例](#)和[亚马逊 EC2 P3 实例](#)。NVIDIA默认情况下，驱动程序 384.81 和CUDA驱动程序 9.0.176 安装在这些实例类型上。
- 增加了对 [Apache MXNet](#) 的支持。

### 更改、增强功能和解决的问题

- Presto
  - 增加了对使用 Glue 数据 AWS 目录作为默认 Hive 元数据仓库的支持。有关更多信息，请参阅将[Presto 与 AWS Glue 数据目录配合使用](#)。
  - 增加了对[地理空间函数](#)的支持。
  - 为联接添加了[溢出到磁盘](#)支持。
  - 增加了对 [Redshift 连接器](#)的支持。
- Spark
  - 向后移植了 [SPARK-20640](#)，这使得使用和属性可以配置 rpc 超时和洗牌注册值的重试次数。spark.shuffle.registration.timeout  
spark.shuffle.registration.maxAttempts
  - 向后移植的 [SPARK-21549](#)，它更正了向非位置写入自定义 OutputFormat 内容时发生的错误。HDFS

- 已逆向移植 [Hadoop 13270](#)
- Numpy、Scipy 和 Matplotlib 库已从亚马逊基础库中删除。EMR AMI如果您的应用程序需要这些库，应用程序存储库中提供了它们，因此您可以通过引导操作使用 `yum install` 在所有节点上安装它们。
- Amazon EMR 基础AMI不再包含应用程序RPM包，因此这些RPM软件包不再存在于集群节点上。Custom AMIs 和 Amazon EMR 基础AMI现在引用 Amazon S3 中的RPM软件包存储库。
- 由于 Amazon 引入了按秒计费EC2，因此默认的缩减行为现在是任务完成时终止，而不是在实例时间终止。有关更多信息，请参阅[配置集群缩减](#)。

## 已知问题

- MXNet不包括 OpenCV 库。
- 默认情况下，Hive 2.3.1 设置 `hive.compute.query.using.stats=true`。这会导致查询从现有统计数据而不是直接从数据中获取数据，这可能会造成混淆。例如，如果您有一个包含 `hive.compute.query.using.stats=true` 的表并且将新文件上载到表 LOCATION，则在该表上运行 `SELECT COUNT(*)` 查询会返回来自统计数据的计数，而不是选择已添加的行。

作为解决方法，请使用 `ANALYZE TABLE` 命令收集新的统计数据，或者设置 `hive.compute.query.using.stats=false`。有关更多信息，请参阅 Apache Hive 文档中的 [Hive 中的统计数据](#)。

## 5.10.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
emr-ddb	4.5.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.4.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.7.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.20.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.3.2	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-5	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-5	HDFS用于存储区块的节点级服务。

组件	版本	描述
hadoop-hdfs-library	2.7.3-amzn-5	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.3-amzn-5	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httfs-server	2.7.3-amzn-5	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-5	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-5	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.3-amzn-5	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-5	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.3-amzn-5	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.3.1	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.3.1	为一个或多个HBase地区提供服务的服务。
hbase-client	1.3.1	HBase命令行客户端。
hbase-rest-server	1.3.1	为其提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.3.1	为其提供 Thrift 端点的HBase服务。

组件	版本	描述
hcatalog-client	2.3.1-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.1-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.1-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.1-amzn-0	Hive 命令行客户端。
hive-hbase	2.3.1-amzn-0	Hive-hbase 客户端。
hive-metastore-server	2.3.1-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.1-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	4.0.1	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
livy-server	0.4.0-incubating	REST用于与 Apache Spark 交互的接口
mahout-client	0.13.0	用于机器学习的库。
mxnet	0.12.0	用于深度学习的灵活的、可扩展且高效的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
nvidia-cuda	9.0.176	Nvidia 驱动程序和 Cuda 工具包

组件	版本	描述
oozie-client	4.3.0	Oozie 命令行客户端。
oozie-server	4.3.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.11.0--1.3 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.11.0--1.3 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.187	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.187	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。
spark-client	2.2.0	Spark 命令行客户端。
spark-history-server	2.2.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.2.0	的内存中执行引擎。YARN
spark-yarn-slave	2.2.0	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。

组件	版本	描述
zeppelin-server	0.7.3	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.10	ZooKeeper 命令行客户端。

### 5.10.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.10.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。

分类	描述
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。



分类	描述
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。

分类	描述
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。

分类	描述
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。

分类	描述
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.9.1

### 5.9.1 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Livy](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、[Pig](#)、[Presto](#)、[Spark](#)、[Sqoop](#)、[Tez](#)、[Zeppelin](#) 和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

## 应用程序版本信息

	emr-5.9.1	emr-5.9.0	emr-5.8.3	emr-5.8.2
适用于 Java 的 AWS SDK	1.11.183	1.11.183	1.11.160	1.11.160
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.2	1.3.2	1.3.1	1.3.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.0	2.3.0	2.3.0	2.3.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.0	2.3.0	2.3.0	2.3.0
Hudi	-	-	-	-
Hue	4.0.1	4.0.1	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	0.4.0	-	-
MXNet	-	-	-	-

	emr-5.9.1	emr-5.9.0	emr-5.8.3	emr-5.8.2
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.16.0	0.16.0
Presto	0.184	0.184	0.170	0.170
Spark	2.2.0	2.2.0	2.2.0	2.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.2	0.7.2	0.7.2	0.7.2
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

### 5.9.1 发布说明

这是一个补丁版本，用于为向 Amazon S3 发出的请求添加 AWS 签名版本 4 身份验证。所有应用程序和组件都与之前的 Amazon EMR 版本相同。

#### Important

在此版本中，亚马逊仅EMR使用 AWS 签名版本 4 来验证向 Amazon S3 发出的请求。有关更多信息，请参阅[新功能](#)。

## 5.9.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>emr-ddb</code>	4.4.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.4.0	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
<code>emr-s3-dist-cp</code>	2.7.0	针对 Amazon S3 优化的分布式复制应用程序。
<code>emrfs</code>	2.19.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
<code>flink-client</code>	1.3.2	Apache Flink 命令行客户端脚本和应用程序。
<code>ganglia-monitor</code>	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。

组件	版本	描述
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-4	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-4	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.3-amzn-4	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.3-amzn-4	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.3-amzn-4	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-4	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-4	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.3-amzn-4	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-4	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.3-amzn-4	用于检索YARN应用程序的当前和历史信息的服务。



组件	版本	描述
hbase-hmaster	1.3.1	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.3.1	为一个或多个HBase地区提供服务的服务。
hbase-client	1.3.1	HBase命令行客户端。
hbase-rest-server	1.3.1	为其提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.3.1	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	2.3.0-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.0-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.0-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.0-amzn-0	Hive 命令行客户端。
hive-hbase	2.3.0-amzn-0	Hive-hbase 客户端。
hive-metastore-server	2.3.0-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.0-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。

组件	版本	描述
hue-server	4.0.1	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
livy-server	0.4.0-incubating	REST用于与 Apache Spark 交互的接口
mahout-client	0.13.0	用于机器学习的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
oozie-client	4.3.0	Oozie 命令行客户端。
oozie-server	4.3.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.11.0--1.3 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.11.0--1.3 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.184	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.184	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。
spark-client	2.2.0	Spark 命令行客户端。
spark-history-server	2.2.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.2.0	的内存中执行引擎。YARN

组件	版本	描述
spark-yarn-slave	2.2.0	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.7.2	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.10	ZooKeeper 命令行客户端。

### 5.9.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.9.1 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。

分类	描述
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改 hbase-policy.xml 文件HBase中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改 jndi.prop HCatalog erties 中的值。
hcatalog-server-props-hive-site	更改 props-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。

分类	描述
hcatalog-webhcat-log4j2	更改 HCatalog WebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog WebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改 HTTPFS 环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。

分类	描述
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。

分类	描述
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。

分类	描述
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改 yarn-site.xml 文件YARN中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.9.0

### 5.9.0 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Livy](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、[Pig](#)、[Presto](#)、[Spark](#)、[Sqoop](#)、[Tez](#)、[Zeppelin](#) 和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：



- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-5.9.0	emr-5.8.3	emr-5.8.2	emr-5.8.1
适用于 Java 的 AWS SDK	1.11.183	1.11.160	1.11.160	1.11.160
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.2	1.3.1	1.3.1	1.3.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.0	2.3.0	2.3.0	2.3.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.0	2.3.0	2.3.0	2.3.0
Hudi	-	-	-	-
Hue	4.0.1	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-

	emr-5.9.0	emr-5.8.3	emr-5.8.2	emr-5.8.1
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0.4.0	-	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.16.0	0.16.0	0.16.0
Presto	0.184	0.170	0.170	0.170
Spark	2.2.0	2.2.0	2.2.0	2.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.2	0.7.2	0.7.2	0.7.2
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

## 5.9.0 发布说明

以下发行说明包含 Amazon 5.9.0 EMR 版本的信息。更改与 Amazon EMR 5.8.0 版本有关。

发布日期：2017 年 10 月 5 日

最近功能更新时间：2017 年 10 月 12 日

## 升级

- AWS SDK for Java 版本 1.11.183
- Flink 1.3.2
- Hue 4.0.1
- Pig 0.17.0
- Presto 0.184

## 新功能

- 添加了 Livy 支持 (0.4.0-incubating 版)。有关更多信息，请参阅[Apache Livy](#)。
- 添加了对 Hue Notebook for Spark 的支持。
- 增加了对 i3 系列亚马逊EC2实例的支持 (2017 年 10 月 12 日)。

## 更改、增强功能和解决的问题

- Spark
  - 添加了一组新功能，有助于确保 Spark 能够更为正常地处理因手动调整大小或自动扩展策略请求导致的节点终止。有关更多信息，请参阅 [配置节点停用行为](#)。
  - SSL用于区块传输服务的传输中加密，而不是 3DES，当使用带 AES-NI 的 Amazon EC2 实例类型时，这可以提高性能。
  - 向后移植 [SPARK-21494](#)。
- Zeppelin
  - 向后移植 [ZEPPELIN-2377](#)。
- HBase
  - 添加了 [HBASE-18533](#) 补丁，它允许使用HBase BucketCache 配置分类进行更多配置值。hbase-site
- Hue
  - 为 H AWS ue 中的 Hive 查询编辑器添加了 Glue 数据目录支持。
  - 默认情况下，Hue 中的超级用户可以访问允许 Amazon EMR IAM 角色访问的所有文件。新建用户不会自动拥有对 Amazon S3 filebrowser 的访问权限，并且必须为其组启用 filebrowser.s3\_access 权限。

- 解决了导致无法访问使用 AWS Glue 数据目录创建的底层JSON数据的问题。

## 已知问题

- 当所有应用程序都安装完毕并且默认的 Amazon EBS 根卷大小未更改时，集群启动将失败。解决方法是，使用中的 `aws emr create-cluster` 命令 AWS CLI 并指定更大的 `--ebs-root-volume-size` 参数。
- 默认情况下，Hive 2.3.0 设置 `hive.compute.query.using.stats=true`。这会导致查询从现有统计数据而不是直接从数据中获取数据，这可能会造成混淆。例如，如果您有一个包含 `hive.compute.query.using.stats=true` 的表并且将新文件上载到表 LOCATION，则在该表上运行 `SELECT COUNT(*)` 查询会返回来自统计数据的计数，而不是选择已添加的行。

作为解决方法，请使用 `ANALYZE TABLE` 命令收集新的统计数据，或者设置 `hive.compute.query.using.stats=false`。有关更多信息，请参阅 Apache Hive 文档中的 [Hive 中的统计数据](#)。

## 5.9.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>emr-ddb</code>	4.4.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.4.0	适用于 Hadoop 生态系统的方便易用的库。

组件	版本	描述
emr-kinesis	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.7.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.19.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.3.2	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-4	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-4	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.3-amzn-4	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.3-amzn-4	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.3-amzn-4	HTTPHDFS操作端点。

组件	版本	描述
hadoop-kms-server	2.7.3-amzn-4	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-4	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.3-amzn-4	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-4	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.3-amzn-4	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.3.1	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.3.1	为一个或多个HBase地区提供服务的服务。
hbase-client	1.3.1	HBase命令行客户端。
hbase-rest-server	1.3.1	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.3.1	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	2.3.0-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.0-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。

组件	版本	描述
hcatalog-webhcat-server	2.3.0-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.0-amzn-0	Hive 命令行客户端。
hive-hbase	2.3.0-amzn-0	Hive-hbase 客户端。
hive-metastore-server	2.3.0-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.0-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	4.0.1	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
livy-server	0.4.0-incubating	REST用于与 Apache Spark 交互的接口
mahout-client	0.13.0	用于机器学习的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
oozie-client	4.3.0	Oozie 命令行客户端。
oozie-server	4.3.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.11.0--1.3 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.11.0--1.3 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API

组件	版本	描述
presto-coordinator	0.184	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.184	用于执行查询的各个部分的服务。
pig-client	0.17.0	Pig 命令行客户端。
spark-client	2.2.0	Spark 命令行客户端。
spark-history-server	2.2.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.2.0	的内存中执行引擎。YARN
spark-yarn-slave	2.2.0	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.7.2	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.10	ZooKeeper 命令行客户端。



## 5.9.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

### emr-5.9.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。

分类	描述
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改 hbase-policy.xml 文件HBase中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改 jndi.prop HCatalog erties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。

分类	描述
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
livy-conf	更改 Livy 的 livy.conf 文件中的值。
livy-env	更改 Livy 环境中的值。
livy-log4j	更改 Livy log4j.properties 设置。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。

分类	描述
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。

分类	描述
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改 yarn-site.xml 文件YARN中的值。
zeppelin-env	更改 Zeppelin 环境中的值。

分类	描述
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper 中的值。
zookeeper-log4j	更改 ZooKeeper 的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.8.3

### 5.8.3 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、[Pig](#)、[Presto](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-5.8.3	emr-5.8.2	emr-5.8.1	emr-5.8.0
适用于 Java 的 AWS SDK	1.11.160	1.11.160	1.11.160	1.11.160
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-

	emr-5.8.3	emr-5.8.2	emr-5.8.1	emr-5.8.0
Delta	-	-	-	-
Flink	1.3.1	1.3.1	1.3.1	1.3.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.0	2.3.0	2.3.0	2.3.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.0	2.3.0	2.3.0	2.3.0
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.170	0.170	0.170	0.170
Spark	2.2.0	2.2.0	2.2.0	2.2.0

	emr-5.8.3	emr-5.8.2	emr-5.8.1	emr-5.8.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.2	0.7.2	0.7.2	0.7.2
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

### 5.8.3 发布说明

这是一个补丁版本，用于为向 Amazon S3 发出的请求添加 AWS 签名版本 4 身份验证。所有应用程序和组件都与之前的 Amazon EMR 版本相同。

#### Important

在此版本中，亚马逊仅EMR使用 AWS 签名版本 4 来验证向 Amazon S3 发出的请求。有关更多信息，请参阅[新功能](#)。

### 5.8.3 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 emr 或 aws 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。



组件	版本	描述
emr-ddb	4.4.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.4.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.6.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.18.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.3.1	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-3	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-3	HDFS用于存储区块的节点级服务。

组件	版本	描述
hadoop-hdfs-library	2.7.3-amzn-3	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.3-amzn-3	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.3-amzn-3	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-3	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-3	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.3-amzn-3	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-3	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.3-amzn-3	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.3.1	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.3.1	为一个或多个HBase地区提供服务的服务。
hbase-client	1.3.1	HBase命令行客户端。
hbase-rest-server	1.3.1	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.3.1	为其提供 Thrift 端点的HBase服务。

组件	版本	描述
hcatalog-client	2.3.0-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.0-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.0-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.0-amzn-0	Hive 命令行客户端。
hive-hbase	2.3.0-amzn-0	Hive-hbase 客户端。
hive-metastore-server	2.3.0-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.0-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.12.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.13.0	用于机器学习的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
oozie-client	4.3.0	Oozie 命令行客户端。
oozie-server	4.3.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.11.0--1.3 HBase	服务器和客户端的 phoenix 库

组件	版本	描述
phoenix-query-server	4.11.0--1.3 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.170	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.170	用于执行查询的各个部分的服务。
pig-client	0.16.0-amzn-1	Pig 命令行客户端。
spark-client	2.2.0	Spark 命令行客户端。
spark-history-server	2.2.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.2.0	的内存中执行引擎。YARN
spark-yarn-slave	2.2.0	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.7.2	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。

组件	版本	描述
zookeeper-client	3.4.10	ZooKeeper 命令行客户端。

### 5.8.3 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.8.3 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。

分类	描述
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改 hbase-policy.xml 文件HBase中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改 jndi.prop HCatalog erties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。

分类	描述
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。

分类	描述
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。



分类	描述
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改 yarn-site.xml 文件YARN中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。

分类	描述
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.8.2

### 5.8.2 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、[Pig](#)、[Presto](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-5.8.2	emr-5.8.1	emr-5.8.0	emr-5.7.1
适用于 Java 的 AWS SDK	1.11.160	1.11.160	1.11.160	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-

	emr-5.8.2	emr-5.8.1	emr-5.8.0	emr-5.7.1
Flink	1.3.1	1.3.1	1.3.1	1.3.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.0	2.3.0	2.3.0	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.0	2.3.0	2.3.0	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.170	0.170	0.170	0.170
Spark	2.2.0	2.2.0	2.2.0	2.1.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6

	emr-5.8.2	emr-5.8.1	emr-5.8.0	emr-5.7.1
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.2	0.7.2	0.7.2	0.7.2
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

## 5.8.2 发布说明

以下发行说明包含有关亚马逊 5.8.2 EMR 版本的信息。更改与 5.8.1 有关。

首次发布日期：2018 年 3 月 29 日

更改、增强功能和解决的问题

- 更新了亚马逊 Linux 版的亚马逊 defaultAmazon Linux AMI 内核EMR以修复潜在漏洞。

## 5.8.2 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
emr-ddb	4.4.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.4.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.6.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.18.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.3.1	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-3	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-3	HDFS用于存储区块的节点级服务。

组件	版本	描述
hadoop-hdfs-library	2.7.3-amzn-3	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.3-amzn-3	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.3-amzn-3	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-3	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-3	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.3-amzn-3	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-3	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.3-amzn-3	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.3.1	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.3.1	为一个或多个HBase地区提供服务的服务。
hbase-client	1.3.1	HBase命令行客户端。
hbase-rest-server	1.3.1	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.3.1	为其提供 Thrift 端点的HBase服务。

组件	版本	描述
hcatalog-client	2.3.0-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.0-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.0-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.0-amzn-0	Hive 命令行客户端。
hive-hbase	2.3.0-amzn-0	Hive-hbase 客户端。
hive-metastore-server	2.3.0-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.0-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.12.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.13.0	用于机器学习的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
oozie-client	4.3.0	Oozie 命令行客户端。
oozie-server	4.3.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.11.0--1.3 HBase	服务器和客户端的 phoenix 库

组件	版本	描述
phoenix-query-server	4.11.0--1.3 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.170	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.170	用于执行查询的各个部分的服务。
pig-client	0.16.0-amzn-1	Pig 命令行客户端。
spark-client	2.2.0	Spark 命令行客户端。
spark-history-server	2.2.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.2.0	的内存中执行引擎。YARN
spark-yarn-slave	2.2.0	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.7.2	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。



组件	版本	描述
zookeeper-client	3.4.10	ZooKeeper 命令行客户端。

## 5.8.2 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

### emr-5.8.2 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。

分类	描述
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改 hbase-policy.xml 文件HBase中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改 jndi.prop HCatalog erties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。

分类	描述
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。

分类	描述
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。

分类	描述
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改 yarn-site.xml 文件YARN中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。

分类	描述
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.8.1

### 5.8.1 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、[Pig](#)、[Presto](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-5.8.1	emr-5.8.0	emr-5.7.1	emr-5.7.0
适用于 Java 的 AWS SDK	1.11.160	1.11.160	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-

	emr-5.8.1	emr-5.8.0	emr-5.7.1	emr-5.7.0
Flink	1.3.1	1.3.1	1.3.0	1.3.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.0	2.3.0	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.0	2.3.0	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.170	0.170	0.170	0.170
Spark	2.2.0	2.2.0	2.1.1	2.1.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6

	emr-5.8.1	emr-5.8.0	emr-5.7.1	emr-5.7.0
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.2	0.7.2	0.7.2	0.7.2
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

## 5.8.1 发布说明

以下发行说明包含 Amazon EMR 5.8.1 版本的信息。更改与 Amazon EMR 5.8.0 版本有关。

首次发布日期：2018 年 1 月 22 日

更改、增强功能和解决的问题

- 更新了亚马逊版 Linux 的亚马逊 L defaultAmazon inux AMI 内核，EMR以解决与推测执行相关的漏洞 ( CVE-2017-5715、-2017-5753 和 -2017 CVE -5754 )。CVE有关更多信息，请参阅 <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>。

## 5.8.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 emr 或 aws 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。



组件	版本	描述
emr-ddb	4.4.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.4.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.6.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.18.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.3.1	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-3	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-3	HDFS用于存储区块的节点级服务。

组件	版本	描述
hadoop-hdfs-library	2.7.3-amzn-3	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.3-amzn-3	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.3-amzn-3	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-3	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-3	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.3-amzn-3	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-3	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.3-amzn-3	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.3.1	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.3.1	为一个或多个HBase地区提供服务的服务。
hbase-client	1.3.1	HBase命令行客户端。
hbase-rest-server	1.3.1	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.3.1	为其提供 Thrift 端点的HBase服务。

组件	版本	描述
hcatalog-client	2.3.0-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.0-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.0-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.0-amzn-0	Hive 命令行客户端。
hive-hbase	2.3.0-amzn-0	Hive-hbase 客户端。
hive-metastore-server	2.3.0-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.0-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.12.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.13.0	用于机器学习的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
oozie-client	4.3.0	Oozie 命令行客户端。
oozie-server	4.3.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.11.0--1.3 HBase	服务器和客户端的 phoenix 库

组件	版本	描述
phoenix-query-server	4.11.0--1.3 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.170	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.170	用于执行查询的各个部分的服务。
pig-client	0.16.0-amzn-1	Pig 命令行客户端。
spark-client	2.2.0	Spark 命令行客户端。
spark-history-server	2.2.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.2.0	的内存中执行引擎。YARN
spark-yarn-slave	2.2.0	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.7.2	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。

组件	版本	描述
zookeeper-client	3.4.10	ZooKeeper 命令行客户端。

### 5.8.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.8.1 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。

分类	描述
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改 hbase-policy.xml 文件HBase中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改 jndi.prop HCatalog erties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。

分类	描述
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。

分类	描述
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。



分类	描述
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改 yarn-site.xml 文件YARN中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。

分类	描述
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.8.0

### 5.8.0 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、[Pig](#)、[Presto](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-5.8.0	emr-5.7.1	emr-5.7.0	emr-5.6.1
适用于 Java 的 AWS SDK	1.11.160	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-

	emr-5.8.0	emr-5.7.1	emr-5.7.0	emr-5.6.1
Flink	1.3.1	1.3.0	1.3.0	1.2.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.0
HCatalog	2.3.0	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.0	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.9.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.170	0.170	0.170	0.170
Spark	2.2.0	2.1.1	2.1.1	2.1.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6

	emr-5.8.0	emr-5.7.1	emr-5.7.0	emr-5.6.1
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.2	0.7.2	0.7.2	0.7.1
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

## 5.8.0 发布说明

以下发行说明包含 Amazon 5.8.0 EMR 版本的信息。更改与 Amazon EMR 5.7.0 版本有关。

首次发布日期：2017 年 8 月 10 日

最近功能更新时间：2017 年 9 月 25 日

### 升级

- AWS SDK1.11.160
- Flink 1.3.1
- Hive 2.3.0。有关更多信息，请参阅 Apache Hive 站点上的[发布说明](#)。
- Spark 2.2.0。有关更多信息，请参阅 Apache Spark 站点上的[发布说明](#)。

### 新功能

- 添加了对查看应用程序历史记录的支持 (2017 年 9 月 25 日)。有关更多信息，请参阅《Amazon EMR 管理指南》中的[查看应用程序历史记录](#)。

### 更改、增强功能和解决的问题

- 与 AWS Glue 数据目录集成

- 添加了 Hive 和 Spark SQL 使用 AWS Glue 数据目录作为 Hive 元数据存储的功能。有关更多信息，请参阅 [使用 AWS Glue 数据目录作为 Hive 的元数据库](#) 和 [使用 AWS Glue 数据目录作为 Spark 的元数据库 SQL](#)。
- 在集群详细信息中添加了应用程序历史记录，允许您查看 YARN 应用程序的历史数据和 Spark 应用程序的其他详细信息。有关更多信息，请参阅《Amazon EMR 管理指南》中的“[查看应用程序历史记录](#)”。
- Oozie
  - 向后移植 [OOZIE-2748](#)。
- Hue
  - [向后移植 HUE -5859](#)
- HBase
  - 添加了用于通过 Java 管理扩展 (JMX) 公开 HBase 主服务器启动时间的补丁 `TgetMasterInitializedTime`。
  - 添加了改进集群启动时间的补丁。

## 已知问题

- 当所有应用程序都安装完毕并且默认的 Amazon EBS 根卷大小未更改时，集群启动将失败。解决方法是，使用中的 `aws emr create-cluster` 命令 AWS CLI 并指定更大的 `--ebs-root-volume-size` 参数。
- 默认情况下，Hive 2.3.0 设置 `hive.compute.query.using.stats=true`。这会导致查询从现有统计数据而不是直接从数据中获取数据，这可能会造成混淆。例如，如果您有一个包含 `hive.compute.query.using.stats=true` 的表并且将新文件上载到表 LOCATION，则在该表上运行 `SELECT COUNT(*)` 查询会返回来自统计数据的计数，而不是选择已添加的行。

作为解决方法，请使用 `ANALYZE TABLE` 命令收集新的统计数据，或者设置 `hive.compute.query.using.stats=false`。有关更多信息，请参阅 Apache Hive 文档中的 [Hive 中的统计数据](#)。

- Spark – 在使用 Spark 时，`appusher` 进程守护程序存在文件处理程序泄漏问题，长时间运行的 Spark 任务在几个小时或几天后可能会出现此情况。要解决此问题，请连接到主节点并键入 `sudo /etc/init.d/appusher stop`。这将停止 `appusher` 守护程序，亚马逊 EMR 将自动重启该守护程序。
- 应用程序历史记录
  - 死 Spark 执行程序的历史数据不可用。

- 应用程序历史记录对使用安全配置来启用传输中加密的集群不可用。

## 5.8.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>emr-ddb</code>	4.4.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.4.0	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.4.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
<code>emr-s3-dist-cp</code>	2.6.0	针对 Amazon S3 优化的分布式复制应用程序。
<code>emrfs</code>	2.18.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
<code>flink-client</code>	1.3.1	Apache Flink 命令行客户端脚本和应用程序。

组件	版本	描述
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-3	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-3	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.3-amzn-3	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.3-amzn-3	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.3-amzn-3	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-3	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-3	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.3-amzn-3	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-3	YARN用于分配和管理群集资源和分布式应用程序的服务。

组件	版本	描述
hadoop-yarn-timeline-server	2.7.3-amzn-3	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.3.1	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.3.1	为一个或多个HBase地区提供服务的服务。
hbase-client	1.3.1	HBase命令行客户端。
hbase-rest-server	1.3.1	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.3.1	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.3.0-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.3.0-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.3.0-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.3.0-amzn-0	Hive 命令行客户端。
hive-hbase	2.3.0-amzn-0	Hive-hbase 客户端。
hive-metastore-server	2.3.0-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.3.0-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。



组件	版本	描述
hue-server	3.12.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.13.0	用于机器学习的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
oozie-client	4.3.0	Oozie 命令行客户端。
oozie-server	4.3.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.11.0--1.3 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.11.0--1.3 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.170	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.170	用于执行查询的各个部分的服务。
pig-client	0.16.0-amzn-1	Pig 命令行客户端。
spark-client	2.2.0	Spark 命令行客户端。
spark-history-server	2.2.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.2.0	的内存中执行引擎。YARN
spark-yarn-slave	2.2.0	YARN从属服务器需要的 Apache Spark 库。

组件	版本	描述
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.7.2	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.10	ZooKeeper 命令行客户端。

## 5.8.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

### emr-5.8.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。

分类	描述
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改 hbase-policy.xml 文件HBase中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改 jndi.prop HCatalog erties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。

分类	描述
hcatalog-webhcat-site	更改 HCatalog WebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。

分类	描述
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。

分类	描述
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。

分类	描述
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改 yarn-site.xml 文件YARN中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.7.1

### 5.7.1 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、[Pig](#)、[Presto](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

## 应用程序版本信息

	emr-5.7.1	emr-5.7.0	emr-5.6.1	emr-5.6.0
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.0	1.3.0	1.2.1	1.2.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-



	emr-5.7.1	emr-5.7.0	emr-5.6.1	emr-5.6.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.9.0	4.9.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.170	0.170	0.170	0.170
Spark	2.1.1	2.1.1	2.1.1	2.1.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.2	0.7.2	0.7.1	0.7.1
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

### 5.7.1 发布说明

这是一个补丁版本，用于为向 Amazon S3 发出的请求添加 AWS 签名版本 4 身份验证。所有应用程序和组件都与之前的 Amazon EMR 版本相同。

#### Important

在此版本中，亚马逊仅EMR使用 AWS 签名版本 4 来验证向 Amazon S3 发出的请求。有关更多信息，请参阅[新功能](#)。

## 5.7.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>emr-ddb</code>	4.3.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.3.0	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.3.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
<code>emr-s3-dist-cp</code>	2.5.0	针对 Amazon S3 优化的分布式复制应用程序。
<code>emrfs</code>	2.18.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
<code>flink-client</code>	1.3.0	Apache Flink 命令行客户端脚本和应用程序。
<code>ganglia-monitor</code>	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。

组件	版本	描述
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-2	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.3-amzn-2	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.3-amzn-2	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-2	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.3-amzn-2	用于检索YARN应用程序的当前和历史信息的服务。

组件	版本	描述
hbase-hmaster	1.3.1	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.3.1	为一个或多个HBase地区提供服务的服务。
hbase-client	1.3.1	HBase命令行客户端。
hbase-rest-server	1.3.1	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.3.1	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.1.1-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.1.1-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.1.1-amzn-0	Hive 命令行客户端。
hive-hbase	2.1.1-amzn-0	Hive-hbase 客户端。
hive-metastore-server	2.1.1-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.1.1-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。

组件	版本	描述
hue-server	3.12.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.13.0	用于机器学习的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
oozie-client	4.3.0	Oozie 命令行客户端。
oozie-server	4.3.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.11.0--1.3 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.11.0--1.3 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限 API
presto-coordinator	0.170	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.170	用于执行查询的各个部分的服务。
pig-client	0.16.0-amzn-0	Pig 命令行客户端。
spark-client	2.1.1	Spark 命令行客户端。
spark-history-server	2.1.1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.1.1	的内存中执行引擎。YARN
spark-yarn-slave	2.1.1	YARN从属服务器需要的 Apache Spark 库。

组件	版本	描述
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.7.2	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.10	ZooKeeper 命令行客户端。

### 5.7.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.7.1 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。

分类	描述
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改 hbase-policy.xml 文件HBase中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改 jndi.prop HCatalog erties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。

分类	描述
hcatalog-webhcat-site	更改 HCatalog WebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。



分类	描述
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。

分类	描述
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。

分类	描述
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改 yarn-site.xml 文件YARN中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.7.0

### 5.7.0 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、[Pig](#)、[Presto](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

## 应用程序版本信息

	emr-5.7.0	emr-5.6.1	emr-5.6.0	emr-5.5.4
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.0	1.2.1	1.2.1	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.0	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-

	emr-5.7.0	emr-5.6.1	emr-5.6.0	emr-5.5.4
Mahout	0.13.0	0.13.0	0.13.0	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.9.0	4.9.0	4.9.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.170	0.170	0.170	0.170
Spark	2.1.1	2.1.1	2.1.1	2.1.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.2	0.7.1	0.7.1	0.7.1
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

## 5.7.0 发布说明

以下发行说明包含 Amazon EMR 5.7.0 版本的信息。更改与 Amazon EMR 5.6.0 版本有关。

发布日期：2017 年 7 月 13 日

### 升级

- Flink 1.3.0
- Phoenix 4.11.0
- Zeppelin 0.7.2

## 新功能

- 添加了在创建集群AMI时指定 customAmazon Linux 的功能。有关更多信息，请参阅[使用自定义AMI](#)。

## 更改、增强功能和解决的问题

- HBase
  - 增加了配置HBase只读副本集群的功能。请参阅[使用只读副本集群](#)。
  - 多个错误修复和增强功能
- Presto – 增加了配置 `node.properties` 的功能。
- YARN-增加了配置功能 `container-log4j.properties`
- Sqoop-向后移植的 [SQOOP-2880](#)，它引入了一个允许你设置 Sqoop 临时目录的参数。

## 5.7.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>emr-ddb</code>	4.3.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.3.0	适用于 Hadoop 生态系统的方便易用的库。

组件	版本	描述
emr-kinesis	3.3.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.5.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.18.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.3.0	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-2	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.3-amzn-2	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.3-amzn-2	HTTPHDFS操作端点。

组件	版本	描述
hadoop-kms-server	2.7.3-amzn-2	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.3-amzn-2	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.3.1	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.3.1	为一个或多个HBase地区提供服务的服务。
hbase-client	1.3.1	HBase命令行客户端。
hbase-rest-server	1.3.1	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.3.1	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	2.1.1-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.1.1-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。



组件	版本	描述
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.1.1-amzn-0	Hive 命令行客户端。
hive-hbase	2.1.1-amzn-0	Hive-hbase 客户端。
hive-metastore-server	2.1.1-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.1.1-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.12.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.13.0	用于机器学习的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
oozie-client	4.3.0	Oozie 命令行客户端。
oozie-server	4.3.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.11.0--1.3 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.11.0--1.3 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.170	用于在 presto-worker 之中接受查询并管理查询的服务。

组件	版本	描述
presto-worker	0.170	用于执行查询的各个部分的服务。
pig-client	0.16.0-amzn-0	Pig 命令行客户端。
spark-client	2.1.1	Spark 命令行客户端。
spark-history-server	2.1.1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.1.1	的内存中执行引擎。YARN
spark-yarn-slave	2.1.1	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.7.2	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.10	ZooKeeper 命令行客户端。

### 5.7.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

## emr-5.7.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改 hbase-policy.xml 文件HBase中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。

分类	描述
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改 jndi.prop HCatalog erties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值

分类	描述
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。

分类	描述
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。

分类	描述
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改 yarn-site.xml 文件YARN中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.6.1

### 5.6.1 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、[Pig](#)、[Presto](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本 ( 如果适用 )。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-5.6.1	emr-5.6.0	emr-5.5.4	emr-5.5.3
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.2.1	1.2.1	1.2.0	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-



	emr-5.6.1	emr-5.6.0	emr-5.5.4	emr-5.5.3
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.13.0	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.9.0	4.9.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.170	0.170	0.170	0.170
Spark	2.1.1	2.1.1	2.1.0	2.1.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.1	0.7.1	0.7.1	0.7.1
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

## 5.6.1 发布说明

这是一个补丁版本，用于为向 Amazon S3 发出的请求添加 AWS 签名版本 4 身份验证。所有应用程序和组件都与之前的 Amazon EMR 版本相同。

### Important

在此版本中，亚马逊仅EMR使用 AWS 签名版本 4 来验证向 Amazon S3 发出的请求。有关更多信息，请参阅[新功能](#)。

## 5.6.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>emr-ddb</code>	4.3.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.3.0	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.3.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
<code>emr-s3-dist-cp</code>	2.5.0	针对 Amazon S3 优化的分布式复制应用程序。

组件	版本	描述
emrfs	2.17.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.2.1	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-2	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.3-amzn-2	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.3-amzn-2	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-2	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。

组件	版本	描述
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.3-amzn-2	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.3.1	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.3.1	为一个或多个HBase地区提供服务的服务。
hbase-client	1.3.1	HBase命令行客户端。
hbase-rest-server	1.3.1	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.3.1	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.1.1-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.1.1-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.1.1-amzn-0	Hive 命令行客户端。
hive-hbase	2.1.1-amzn-0	Hive-hbase 客户端。

组件	版本	描述
hive-metastore-server	2.1.1-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的 SQL 元数据。
hive-server2	2.1.1-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.12.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.13.0	用于机器学习的库。
mysql-server	5.5.54+	我的 SQL 数据库服务器。
oozie-client	4.3.0	Oozie 命令行客户端。
oozie-server	4.3.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.9.0--1.2 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.9.0--1.2 HBase	一款轻量级服务器，提供 JDBC 对 Avatica 的访问权限以及协议缓冲区和 JSON 格式访问权限 API
presto-coordinator	0.170	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.170	用于执行查询的各个部分的服务。
pig-client	0.16.0-amzn-0	Pig 命令行客户端。
spark-client	2.1.1	Spark 命令行客户端。

组件	版本	描述
spark-history-server	2.1.1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.1.1	的内存中执行引擎。YARN
spark-yarn-slave	2.1.1	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.7.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.10	ZooKeeper 命令行客户端。

### 5.6.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.6.1 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。

分类	描述
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改 hbase-policy.xml 文件HBase中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。

分类	描述
hcatalog-server-jndi	更改 jndi.prop HCatalog erties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。



分类	描述
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。

分类	描述
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值

分类	描述
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改 yarn-site.xml 文件YARN中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.6.0

### 5.6.0 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、[Pig](#)、[Presto](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)

- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

## 应用程序版本信息

	emr-5.6.0	emr-5.5.4	emr-5.5.3	emr-5.5.2
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.2.1	1.2.0	1.2.0	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-

	emr-5.6.0	emr-5.5.4	emr-5.5.3	emr-5.5.2
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.9.0	4.9.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.170	0.170	0.170	0.170
Spark	2.1.1	2.1.0	2.1.0	2.1.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.1	0.7.1	0.7.1	0.7.1
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

## 5.6.0 发布说明

以下发行说明包含 Amazon EMR 5.6.0 版本的信息。更改与亚马逊 EMR 5.5.0 版本有关。

发布日期：2017 年 6 月 5 日

### 升级

- Flink 1.2.1

- HBase1.3.1
- Mahout 0.13.0。这是第一个在亚马逊 5.0 及更高版本中支持 Spark 2.x 的 Mahout EMR 版本。
- Spark 2.1.1

## 更改、增强功能和解决的问题

- Presto
  - 通过使用TLS安全配置启用传输中加密，增加了在 Presto 节点之间启用 SSL /安全通信的功能。有关更多信息，请参阅[传输中的数据加密](#)。
  - 已逆向移植 [Presto 7661](#)，它向 VERBOSE 语句添加了 EXPLAIN ANALYZE 选项，以报告有关查询计划的更详细、更低级别的统计数据。

## 5.6.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 emr 或 aws 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
emr-ddb	4.3.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.3.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.3.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。

组件	版本	描述
emr-s3-dist-cp	2.5.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.17.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.2.1	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-2	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.3-amzn-2	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.3-amzn-2	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-2	基于 Hadoop 的加密密钥管理服务。 KeyProvider API

组件	版本	描述
hadoop-mapred	2.7.3-amzn-2	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.3-amzn-2	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.3.1	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.3.1	为一个或多个HBase地区提供服务的服务。
hbase-client	1.3.1	HBase命令行客户端。
hbase-rest-server	1.3.1	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.3.1	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	2.1.1-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.1.1-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.1.1-amzn-0	Hive 命令行客户端。



组件	版本	描述
hive-hbase	2.1.1-amzn-0	Hive-hbase 客户端。
hive-metastore-server	2.1.1-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的 SQL 元数据。
hive-server2	2.1.1-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.12.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.13.0	用于机器学习的库。
mysql-server	5.5.54+	我的 SQL 数据库服务器。
oozie-client	4.3.0	Oozie 命令行客户端。
oozie-server	4.3.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.9.0--1.2 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.9.0--1.2 HBase	一款轻量级服务器，提供 JDBC 对 Avatica 的访问权限以及协议缓冲区和 JSON 格式访问权限 API
presto-coordinator	0.170	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.170	用于执行查询的各个部分的服务。
pig-client	0.16.0-amzn-0	Pig 命令行客户端。

组件	版本	描述
spark-client	2.1.1	Spark 命令行客户端。
spark-history-server	2.1.1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.1.1	的内存中执行引擎。YARN
spark-yarn-slave	2.1.1	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.7.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.10	ZooKeeper 命令行客户端。

### 5.6.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.6.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。

分类	描述
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-h HBase base.properties 文件中的值。
hbase-policy	更改 hbase-policy.xml 文件HBase中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。

分类	描述
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改 jndi.prop HCatalog erties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。

分类	描述
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。

分类	描述
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-env	更改 Presto 的 presto-env.sh 文件中的值。
presto-node	更改 Presto 的 node.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。

分类	描述
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改 yarn-site.xml 文件YARN中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.5.4

### 5.5.4 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、[Pig](#)、[Presto](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-5.5.4	emr-5.5.3	emr-5.5.2	emr-5.5.1
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.2.0	1.2.0	1.2.0	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-



	emr-5.5.4	emr-5.5.3	emr-5.5.2	emr-5.5.1
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.9.0	4.9.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.170	0.170	0.170	0.170
Spark	2.1.0	2.1.0	2.1.0	2.1.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.1	0.7.1	0.7.1	0.7.1
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

## 5.5.4 发布说明

这是一个补丁版本，用于为向 Amazon S3 发出的请求添加 AWS 签名版本 4 身份验证。所有应用程序和组件都与之前的 Amazon EMR 版本相同。

**⚠ Important**

在此版本中，亚马逊仅EMR使用 AWS 签名版本 4 来验证向 Amazon S3 发出的请求。有关更多信息，请参阅[新功能](#)。

### 5.5.4 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>emr-ddb</code>	4.3.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.3.0	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.3.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
<code>emr-s3-dist-cp</code>	2.5.0	针对 Amazon S3 优化的分布式复制应用程序。
<code>emrfs</code>	2.16.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
<code>flink-client</code>	1.2.0	Apache Flink 命令行客户端脚本和应用程序。

组件	版本	描述
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-2	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.3-amzn-2	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.3-amzn-2	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-2	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN用于分配和管理群集资源和分布式应用程序的服务。

组件	版本	描述
hadoop-yarn-timeline-server	2.7.3-amzn-2	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.3.0	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.3.0	为一个或多个HBase地区提供服务的服务。
hbase-client	1.3.0	HBase命令行客户端。
hbase-rest-server	1.3.0	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.3.0	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.1.1-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.1.1-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.1.1-amzn-0	Hive 命令行客户端。
hive-hbase	2.1.1-amzn-0	Hive-hbase 客户端。
hive-metastore-server	2.1.1-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.1.1-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。

组件	版本	描述
hue-server	3.12.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.12.2	用于机器学习的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
oozie-client	4.3.0	Oozie 命令行客户端。
oozie-server	4.3.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.9.0--1.2 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.9.0--1.2 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.170	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.170	用于执行查询的各个部分的服务。
pig-client	0.16.0-amzn-0	Pig 命令行客户端。
spark-client	2.1.0	Spark 命令行客户端。
spark-history-server	2.1.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.1.0	的内存中执行引擎。YARN
spark-yarn-slave	2.1.0	YARN从属服务器需要的 Apache Spark 库。

组件	版本	描述
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.7.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.10	ZooKeeper 命令行客户端。

#### 5.5.4 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

##### emr-5.5.4 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。

分类	描述
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。
hbase-policy	更改 hbase-policy.xml 文件HBase中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改 jndi.prop HCatalog erties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。

分类	描述
hcatalog-webhcat-site	更改 HCatalog WebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。



分类	描述
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。

分类	描述
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。

分类	描述
yarn-site	更改 yarn-site.xml 文件YARN中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.5.3

### 5.5.3 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、[Pig](#)、[Presto](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-5.5.3	emr-5.5.2	emr-5.5.1	emr-5.5.0
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪

	emr-5.5.3	emr-5.5.2	emr-5.5.1	emr-5.5.0
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.2.0	1.2.0	1.2.0	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.9.0	4.9.0

	emr-5.5.3	emr-5.5.2	emr-5.5.1	emr-5.5.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.170	0.170	0.170	0.170
Spark	2.1.0	2.1.0	2.1.0	2.1.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.1	0.7.1	0.7.1	0.7.1
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

### 5.5.3 发布说明

以下发行说明包含有关亚马逊 5.5.3 EMR 版本的信息。更改与 5.5.2 有关。

首次发布日期：2018年 8 月 29 日

更改、增强功能和解决的问题

- 此版本解决了潜在的安全漏洞。

### 5.5.3 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 `CommunityVersion-amzn-EmrVersion` 的发行版标注。`EmrVersion` 从 0 开始。例如，如果 `myapp-component` 以 2.2

版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
emr-ddb	4.3.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.3.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.3.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.5.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.16.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.2.0	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-2	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。

组件	版本	描述
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.3-amzn-2	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.3-amzn-2	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-2	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.3-amzn-2	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.3.0	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.3.0	为一个或多个HBase地区提供服务的服务。
hbase-client	1.3.0	HBase命令行客户端。
hbase-rest-server	1.3.0	为提供RESTfulHTTP终端节点的服务HBase。

组件	版本	描述
hbase-thrift-server	1.3.0	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	2.1.1-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.1.1-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP端点为其提供REST接口 HCatalog。
hive-client	2.1.1-amzn-0	Hive 命令行客户端。
hive-hbase	2.1.1-amzn-0	Hive-hbase 客户端。
hive-metastore-server	2.1.1-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.1.1-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.12.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.12.2	用于机器学习的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
oozie-client	4.3.0	Oozie 命令行客户端。
oozie-server	4.3.0	用于接受 Oozie 工作流请求的服务。



组件	版本	描述
phoenix-library	4.9.0--1.2 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.9.0--1.2 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.170	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.170	用于执行查询的各个部分的服务。
pig-client	0.16.0-amzn-0	Pig 命令行客户端。
spark-client	2.1.0	Spark 命令行客户端。
spark-history-server	2.1.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.1.0	的内存中执行引擎。YARN
spark-yarn-slave	2.1.0	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.7.1	支持交互式数据分析的基于 Web 的笔记本电脑。

组件	版本	描述
zookeeper-server	3.4.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.10	ZooKeeper 命令行客户端。

### 5.5.3 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.5.3 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置

分类	描述
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。
hbase-policy	更改 hbase-policy.xml 文件HBase中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改 jndi.prop HCatalog erties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。

分类	描述
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。

分类	描述
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。

分类	描述
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改 yarn-site.xml 文件YARN中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。

分类	描述
zookeeper-log4j	更改 ZooKeeper 的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.5.2

### 5.5.2 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、[Pig](#)、[Presto](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-5.5.2	emr-5.5.1	emr-5.5.0	emr-5.4.1
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-

	emr-5.5.2	emr-5.5.1	emr-5.5.0	emr-5.4.1
Flink	1.2.0	1.2.0	1.2.0	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.11.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.9.0	4.9.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.170	0.170	0.170	0.166
Spark	2.1.0	2.1.0	2.1.0	2.1.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6



	emr-5.5.2	emr-5.5.1	emr-5.5.0	emr-5.4.1
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.1	0.7.1	0.7.1	0.7.0
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.9

## 5.5.2 发布说明

以下发行说明包含 Amazon EMR 版本 5.5.2 的信息。更改与 5.5.1 有关。

首次发布日期：2018 年 3 月 29 日

更改、增强功能和解决的问题

- 更新了亚马逊 Linux 版的亚马逊 defaultAmazon Linux AMI 内核EMR以修复潜在漏洞。

## 5.5.2 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
emr-ddb	4.3.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.3.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.3.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.5.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.16.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.2.0	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-2	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS用于存储区块的节点级服务。

组件	版本	描述
hadoop-hdfs-library	2.7.3-amzn-2	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.3-amzn-2	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-2	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.3-amzn-2	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.3.0	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.3.0	为一个或多个HBase地区提供服务的服务。
hbase-client	1.3.0	HBase命令行客户端。
hbase-rest-server	1.3.0	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.3.0	为其提供 Thrift 端点的HBase服务。

组件	版本	描述
hcatalog-client	2.1.1-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.1.1-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.1.1-amzn-0	Hive 命令行客户端。
hive-hbase	2.1.1-amzn-0	Hive-hbase 客户端。
hive-metastore-server	2.1.1-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.1.1-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.12.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.12.2	用于机器学习的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
oozie-client	4.3.0	Oozie 命令行客户端。
oozie-server	4.3.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.9.0--1.2 HBase	服务器和客户端的 phoenix 库

组件	版本	描述
phoenix-query-server	4.9.0--1.2 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.170	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.170	用于执行查询的各个部分的服务。
pig-client	0.16.0-amzn-0	Pig 命令行客户端。
spark-client	2.1.0	Spark 命令行客户端。
spark-history-server	2.1.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.1.0	的内存中执行引擎。YARN
spark-yarn-slave	2.1.0	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.7.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。

组件	版本	描述
zookeeper-client	3.4.10	ZooKeeper 命令行客户端。

## 5.5.2 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

### emr-5.5.2 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。

分类	描述
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。
hbase-policy	更改 hbase-policy.xml 文件HBase中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改 jndi.prop HCatalog erties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。

分类	描述
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。



分类	描述
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。

分类	描述
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.5.1

### 5.5.1 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、[Pig](#)、[Presto](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-5.5.1	emr-5.5.0	emr-5.4.1	emr-5.4.0
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.2.0	1.2.0	1.2.0	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.3.0	1.3.0

	emr-5.5.1	emr-5.5.0	emr-5.4.1	emr-5.4.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.11.0	3.11.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.9.0	4.9.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.170	0.170	0.166	0.166
Spark	2.1.0	2.1.0	2.1.0	2.1.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.5.1	emr-5.5.0	emr-5.4.1	emr-5.4.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.1	0.7.1	0.7.0	0.7.0
ZooKeeper	3.4.10	3.4.10	3.4.9	3.4.9

## 5.5.1 发布说明

以下发行说明包含有关 Amazon EMR 5.5.1 版本的信息。更改与亚马逊 EMR 5.5.0 版本有关。

首次发布日期：2018 年 1 月 22 日

更改、增强功能和解决的问题

- 更新了亚马逊版 Linux 的亚马逊 L defaultAmazon inux AMI 内核，EMR以解决与推测执行相关的漏洞 ( CVE-2017-5715、-2017-5753 和 -2017 CVE -5754 )。CVE有关更多信息，请参阅 <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>。

## 5.5.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 emr 或 aws 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
emr-ddb	4.3.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。

组件	版本	描述
emr-goodies	2.3.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.3.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.5.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.16.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.2.0	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-2	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.3-amzn-2	HDFS命令行客户端和库

组件	版本	描述
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.3-amzn-2	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-2	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce 用于运行 MapReduce 应用程序的执行引擎。
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.3-amzn-2	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.3.0	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.3.0	为一个或多个HBase地区提供服务的服务。
hbase-client	1.3.0	HBase命令行客户端。
hbase-rest-server	1.3.0	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.3.0	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.1.1-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。

组件	版本	描述
hcatalog-server	2.1.1-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.1.1-amzn-0	Hive 命令行客户端。
hive-hbase	2.1.1-amzn-0	Hive-hbase 客户端。
hive-metastore-server	2.1.1-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.1.1-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.12.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.12.2	用于机器学习的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
oozie-client	4.3.0	Oozie 命令行客户端。
oozie-server	4.3.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.9.0--1.2 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.9.0--1.2 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API



组件	版本	描述
presto-coordinator	0.170	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.170	用于执行查询的各个部分的服务。
pig-client	0.16.0-amzn-0	Pig 命令行客户端。
spark-client	2.1.0	Spark 命令行客户端。
spark-history-server	2.1.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.1.0	的内存中执行引擎。YARN
spark-yarn-slave	2.1.0	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.7.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.10	ZooKeeper 命令行客户端。

## 5.5.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

### emr-5.5.1 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。

分类	描述
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。
hbase-policy	更改 hbase-policy.xml 文件HBase中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改 jndi.prop HCatalog erties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。

分类	描述
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。

分类	描述
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。

分类	描述
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.5.0

### 5.5.0 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、[Pig](#)、[Presto](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-5.5.0	emr-5.4.1	emr-5.4.0	emr-5.3.2
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.2.0	1.2.0	1.2.0	flink-client
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.3.0	1.2.3

	emr-5.5.0	emr-5.4.1	emr-5.4.0	emr-5.3.2
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.11.0	3.11.0	3.11.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.9.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.170	0.166	0.166	0.157.1
Spark	2.1.0	2.1.0	2.1.0	2.1.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4



	emr-5.5.0	emr-5.4.1	emr-5.4.0	emr-5.3.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.1	0.7.0	0.7.0	0.6.2
ZooKeeper	3.4.10	3.4.9	3.4.9	3.4.9

## 5.5.0 发布说明

以下发行说明包含有关 Amazon EMR 5.5.0 版本的信息。更改与亚马逊 EMR 5.4.0 版本有关。

发布日期：2017 年 4 月 26 日

### 升级

- Hue 3.12
- Presto 0.170
- Zeppelin 0.7.1
- ZooKeeper 3.4.10

### 更改、增强功能和解决的问题

- Spark
  - 向后移植的 Spark Patch ([SPARK-20115](#)) 修复 DAGScheduler 了在本版本中包含的 [Spark 2.1.0 版本中无法使用外部洗牌服务时，重新计算所有丢失的随机播放块](#)的问题。
- Flink
  - Flink 现在使用 Scala 2.11 进行构建。如果您使用 Scala API 和库，我们建议您在项目中使用 Scala 2.11。
  - 解决了 HADOOP\_CONF\_DIR 和 YARN\_CONF\_DIR 默认值未正确设置，因此 start-scala-shell.sh 无法工作的问题。此外，还添加了使用 env.hadoop.conf.dir 或 env.yarn.conf.dir 配置类别中的 /etc/flink/conf/flink-conf.yaml 和 flink-conf 设置这些值的功能。

- 引入了一个新的EMR特定于命令 `flink-scala-shell` 的封装器。 `start-scala-shell.sh` 我们建议使用此命令而不是 `start-scala-shell`。新命令可简化执行。例如，`flink-scala-shell -n 2` 将使用任务并行度 2 启动 Flink Scala Shell。
- 引入了一个新的EMR特定于命令 `flink-yarn-session` 的封装器。 `yarn-session.sh` 我们建议使用此命令而不是 `yarn-session`。新命令可简化执行。例如，`flink-yarn-session -d -n 2` 将使用两个任务管理器在分离状态下启动长时间运行的 Flink 会话。
- 已解决 [\(FLINK-6125\) Commons httpclient 在 Flink 1.2 中不再显示阴影。](#)
- Presto
  - 增加了对LDAP身份验证的支持。在亚马逊上LDAP与 Presto 一起使用EMR需要启用 Presto 协调器的HTTPS访问权限 ( `http-server.https.enabled=true` 中 )。 `config.properties` 有关配置的详细信息，请参阅 Presto 文档中的 [LDAP身份验证](#)。
  - 增加了对 SHOW GRANTS 的支持。
- 亚马逊 EMR Base Linux AMI
  - 亚马逊EMR版本现在基于亚马逊 Linux 2017.03。有关更多信息，请参阅 [亚马逊 Linux AMI 2017.03 发行说明](#)。
  - 从亚马逊EMR基本的 Linux 镜像中移除了 Python 2.6。默认安装 Python 2.7 和 3.4。如果需要，您可以手动安装 Python 2.6。

### 5.5.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>emr-ddb</code>	4.3.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。

组件	版本	描述
emr-goodies	2.3.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.3.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.5.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.16.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.2.0	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-2	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.3-amzn-2	HDFS命令行客户端和库

组件	版本	描述
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httfs-server	2.7.3-amzn-2	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-2	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce 用于运行 MapReduce 应用程序的执行引擎。
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.3-amzn-2	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.3.0	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.3.0	为一个或多个HBase地区提供服务的服务。
hbase-client	1.3.0	HBase命令行客户端。
hbase-rest-server	1.3.0	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.3.0	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.1.1-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。

组件	版本	描述
hcatalog-server	2.1.1-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.1.1-amzn-0	Hive 命令行客户端。
hive-hbase	2.1.1-amzn-0	Hive-hbase 客户端。
hive-metastore-server	2.1.1-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.1.1-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.12.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.12.2	用于机器学习的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
oozie-client	4.3.0	Oozie 命令行客户端。
oozie-server	4.3.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.9.0--1.2 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.9.0--1.2 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API

组件	版本	描述
presto-coordinator	0.170	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.170	用于执行查询的各个部分的服务。
pig-client	0.16.0-amzn-0	Pig 命令行客户端。
spark-client	2.1.0	Spark 命令行客户端。
spark-history-server	2.1.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.1.0	的内存中执行引擎。YARN
spark-yarn-slave	2.1.0	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.7.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.10	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.10	ZooKeeper 命令行客户端。

## 5.5.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

### emr-5.5.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。

分类	描述
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。
hbase-policy	更改 hbase-policy.xml 文件HBase中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改 jndi.prop HCatalog erties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。



分类	描述
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。

分类	描述
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。

分类	描述
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.4.1

### 5.4.1 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、[Pig](#)、[Presto](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-5.4.1	emr-5.4.0	emr-5.3.2	emr-5.3.1
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.2.0	1.2.0	flink-client	flink-client
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.2.3	1.2.3

	emr-5.4.1	emr-5.4.0	emr-5.3.2	emr-5.3.1
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.11.0	3.11.0	3.11.0	3.11.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.7.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.166	0.166	0.157.1	0.157.1
Spark	2.1.0	2.1.0	2.1.0	2.1.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.4.1	emr-5.4.0	emr-5.3.2	emr-5.3.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.0	0.7.0	0.6.2	0.6.2
ZooKeeper	3.4.9	3.4.9	3.4.9	3.4.9

### 5.4.1 发布说明

这是一个补丁版本，用于为向 Amazon S3 发出的请求添加 AWS 签名版本 4 身份验证。所有应用程序和组件都与之前的 Amazon EMR 版本相同。

#### Important

在此版本中，亚马逊仅EMR使用 AWS 签名版本 4 来验证向 Amazon S3 发出的请求。有关更多信息，请参阅[新功能](#)。

### 5.4.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 emr 或 aws 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
emr-ddb	4.2.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。

组件	版本	描述
emr-goodies	2.3.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.2.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.4.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.15.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.2.0	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.3-amzn-1	HDFS命令行客户端和库

组件	版本	描述
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.3-amzn-1	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-1	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-1	MapReduce 用于运行 MapReduce 应用程序的执行引擎。
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.3-amzn-1	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.3.0	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.3.0	为一个或多个HBase地区提供服务的服务。
hbase-client	1.3.0	HBase命令行客户端。
hbase-rest-server	1.3.0	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.3.0	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.1.1-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。



组件	版本	描述
hcatalog-server	2.1.1-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.1.1-amzn-0	Hive 命令行客户端。
hive-hbase	2.1.1-amzn-0	Hive-hbase 客户端。
hive-metastore-server	2.1.1-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.1.1-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.11.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.12.2	用于机器学习的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
oozie-client	4.3.0	Oozie 命令行客户端。
oozie-server	4.3.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.9.0--1.2 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.9.0--1.2 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API

组件	版本	描述
presto-coordinator	0.166	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.166	用于执行查询的各个部分的服务。
pig-client	0.16.0-amzn-0	Pig 命令行客户端。
spark-client	2.1.0	Spark 命令行客户端。
spark-history-server	2.1.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.1.0	的内存中执行引擎。YARN
spark-yarn-slave	2.1.0	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.7.0	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.9	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.9	ZooKeeper 命令行客户端。

## 5.4.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

### emr-5.4.1 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。

分类	描述
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。
hbase-policy	更改 hbase-policy.xml 文件HBase中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改 jndi.prop HCatalog erties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。

分类	描述
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。

分类	描述
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。

分类	描述
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.4.0

### 5.4.0 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、[Pig](#)、[Presto](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-5.4.0	emr-5.3.2	emr-5.3.1	emr-5.3.0
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.2.0	flink-client	flink-client	flink-client
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.2.3	1.2.3	1.2.3



	emr-5.4.0	emr-5.3.2	emr-5.3.1	emr-5.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.11.0	3.11.0	3.11.0	3.11.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.166	0.157.1	0.157.1	0.157.1
Spark	2.1.0	2.1.0	2.1.0	2.1.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.4.0	emr-5.3.2	emr-5.3.1	emr-5.3.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.0	0.6.2	0.6.2	0.6.2
ZooKeeper	3.4.9	3.4.9	3.4.9	3.4.9

## 5.4.0 发布说明

以下发行说明包含 Amazon EMR 5.4.0 版本的信息。更改与亚马逊 EMR 5.3.0 版本有关。

发布日期：2017 年 3 月 8 日

### 升级

- 已升级到 Flink 1.2.0
- 已升级到 HBase 1.3.0
- 已升级到 Phoenix 4.9.0

#### Note

如果您从亚马逊的早期版本升级EMR到亚马逊版本 5.4.0 或更高EMR版本并使用二级索引，请按照 A [pache Phoenix](#) 文档中的说明升级本地索引。Amazon 从hbase-site分类中EMR删除了所需的配置，但需要重新填充索引。支持在线和离线升级索引。在线升级为默认值，这意味着，在从版本 4.8.0 或更高版本的 Phoenix 客户端初始化时重新填充索引。要指定离线升级，请在phoenix-site分类中将`phoenix.client.localIndexUpgrade`配置设置为 `false`，然后SSH将配置设置为要运行的主节点`psql [zookeeper] -1`。

- 已升级到 Presto 0.166
- 已升级到 Zeppelin 0.7.0

### 更改和增强功能

- 增加了对 r4 实例的支持。参见 [Amazon EC2 实例类型](#)。

## 5.4.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>emr-ddb</code>	4.2.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.3.0	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.2.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
<code>emr-s3-dist-cp</code>	2.4.0	针对 Amazon S3 优化的分布式复制应用程序。
<code>emrfs</code>	2.15.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
<code>flink-client</code>	1.2.0	Apache Flink 命令行客户端脚本和应用程序。
<code>ganglia-monitor</code>	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。

组件	版本	描述
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.3-amzn-1	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.3-amzn-1	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-1	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.3-amzn-1	用于检索YARN应用程序的当前和历史信息的服务。

组件	版本	描述
hbase-hmaster	1.3.0	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.3.0	为一个或多个HBase地区提供服务的服务。
hbase-client	1.3.0	HBase命令行客户端。
hbase-rest-server	1.3.0	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.3.0	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.1.1-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.1.1-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.1.1-amzn-0	Hive 命令行客户端。
hive-hbase	2.1.1-amzn-0	Hive-hbase 客户端。
hive-metastore-server	2.1.1-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server2	2.1.1-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。

组件	版本	描述
hue-server	3.11.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.12.2	用于机器学习的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
oozie-client	4.3.0	Oozie 命令行客户端。
oozie-server	4.3.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.9.0--1.2 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.9.0--1.2 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限 API
presto-coordinator	0.166	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.166	用于执行查询的各个部分的服务。
pig-client	0.16.0-amzn-0	Pig 命令行客户端。
spark-client	2.1.0	Spark 命令行客户端。
spark-history-server	2.1.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.1.0	的内存中执行引擎。YARN
spark-yarn-slave	2.1.0	YARN从属服务器需要的 Apache Spark 库。

组件	版本	描述
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.7.0	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.9	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.9	ZooKeeper 命令行客户端。

## 5.4.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

### emr-5.4.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。

分类	描述
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。
hbase-policy	更改 hbase-policy.xml 文件HBase中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改 jndi.prop HCatalog erties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。



分类	描述
hcatalog-webhcat-site	更改 HCatalog WebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。

分类	描述
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。

分类	描述
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。

分类	描述
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.3.2

### 5.3.2 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、[Pig](#)、[Presto](#)、[Tez](#) 和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-5.3.2	emr-5.3.1	emr-5.3.0	emr-5.2.3
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪

	emr-5.3.2	emr-5.3.1	emr-5.3.0	emr-5.2.3
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	flink-client	flink-client	flink-client	1.1.3
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.3	1.2.3
HCatalog	2.1.1	2.1.1	2.1.1	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.0
Hudi	-	-	-	-
Hue	3.11.0	3.11.0	3.11.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0

	emr-5.3.2	emr-5.3.1	emr-5.3.0	emr-5.2.3
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.157.1	0.157.1	0.157.1	0.157.1
Spark	2.1.0	2.1.0	2.1.0	2.0.2
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.2	0.6.2	0.6.2
ZooKeeper	3.4.9	3.4.9	3.4.9	3.4.9

### 5.3.2 发布说明

这是一个补丁版本，用于为向 Amazon S3 发出的请求添加 AWS 签名版本 4 身份验证。所有应用程序和组件都与之前的 Amazon EMR 版本相同。

#### Important

在此版本中，亚马逊仅EMR使用 AWS 签名版本 4 来验证向 Amazon S3 发出的请求。有关更多信息，请参阅[新功能](#)。

### 5.3.2 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 emr 或 aws 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
emr-ddb	4.2.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.2.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.2.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.4.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	emrfs	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	flink-client	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。

组件	版本	描述
hadoop-client	2.7.3-amzn-1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.3-amzn-1	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-https-server	2.7.3-amzn-1	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-1	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.3-amzn-1	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.2.3	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.2.3	为一个或多个HBase地区提供服务的服务。
hbase-client	1.2.3	HBase命令行客户端。



组件	版本	描述
hbase-rest-server	1.2.3	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.2.3	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	2.1.1-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.1.1-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.1.1-amzn-0	Hive 命令行客户端。
hive-metastore-server	2.1.1-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server	2.1.1-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.11.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.12.2	用于机器学习的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
oozie-client	4.3.0	Oozie 命令行客户端。
oozie-server	4.3.0	用于接受 Oozie 工作流请求的服务。

组件	版本	描述
phoenix-library	4.7.0--1.2 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.7.0--1.2 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.157.1	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.157.1	用于执行查询的各个部分的服务。
pig-client	0.16.0-amzn-0	Pig 命令行客户端。
spark-client	2.1.0	Spark 命令行客户端。
spark-history-server	2.1.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.1.0	的内存中执行引擎。YARN
spark-yarn-slave	2.1.0	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.6.2	支持交互式数据分析的基于 Web 的笔记本电脑。

组件	版本	描述
zookeeper-server	3.4.9	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.9	ZooKeeper 命令行客户端。

### 5.3.2 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.3.2 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置

分类	描述
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。
hbase-policy	更改 hbase-policy.xml 文件HBase中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改 jndi.prop HCatalog erties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。

分类	描述
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。

分类	描述
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。

分类	描述
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。

分类	描述
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.3.1

### 5.3.1 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、[Pig](#)、[Presto](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-5.3.1	emr-5.3.0	emr-5.2.3	emr-5.2.2
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-



	emr-5.3.1	emr-5.3.0	emr-5.2.3	emr-5.2.2
Flink	flink-client	flink-client	1.1.3	1.1.3
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.3	1.2.3
HCatalog	2.1.1	2.1.1	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.11.0	3.11.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.157.1	0.157.1	0.157.1	0.157.1
Spark	2.1.0	2.1.0	2.0.2	2.0.2
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6

	emr-5.3.1	emr-5.3.0	emr-5.2.3	emr-5.2.2
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.2	0.6.2	0.6.2
ZooKeeper	3.4.9	3.4.9	3.4.9	3.4.9

### 5.3.1 发布说明

以下发行说明包含 Amazon EMR 5.3.1 版本的信息。更改与亚马逊 EMR 5.3.0 版本有关。

发布日期：2017 年 2 月 7 日

对向后移植齐柏林飞艇补丁进行了细微的更改，并更新了亚马逊的默认补丁。AMI EMR

### 5.3.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
emr-ddb	4.2.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。

组件	版本	描述
emr-goodies	2.2.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.2.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.4.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	emrfs	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	flink-client	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.3-amzn-1	HDFS命令行客户端和库

组件	版本	描述
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httfs-server	2.7.3-amzn-1	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-1	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-1	MapReduce 用于运行 MapReduce 应用程序的执行引擎。
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.3-amzn-1	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.2.3	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.2.3	为一个或多个HBase地区提供服务的服务。
hbase-client	1.2.3	HBase命令行客户端。
hbase-rest-server	1.2.3	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.2.3	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.1.1-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。

组件	版本	描述
hcatalog-server	2.1.1-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.1.1-amzn-0	Hive 命令行客户端。
hive-metastore-server	2.1.1-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server	2.1.1-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.11.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.12.2	用于机器学习的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
oozie-client	4.3.0	Oozie 命令行客户端。
oozie-server	4.3.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.7.0--1.2 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.7.0--1.2 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API

组件	版本	描述
presto-coordinator	0.157.1	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.157.1	用于执行查询的各个部分的服务。
pig-client	0.16.0-amzn-0	Pig 命令行客户端。
spark-client	2.1.0	Spark 命令行客户端。
spark-history-server	2.1.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.1.0	的内存中执行引擎。YARN
spark-yarn-slave	2.1.0	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.6.2	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.9	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.9	ZooKeeper 命令行客户端。

### 5.3.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.3.1 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。

分类	描述
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。
hbase-policy	更改 hbase-policy.xml 文件HBase中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改 jndi.prop HCatalog erties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。



分类	描述
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。

分类	描述
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。

分类	描述
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.3.0

### 5.3.0 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、[Pig](#)、[Presto](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-5.3.0	emr-5.2.3	emr-5.2.2	emr-5.2.1
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	flink-client	1.1.3	1.1.3	1.1.3
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.3	1.2.3

	emr-5.3.0	emr-5.2.3	emr-5.2.2	emr-5.2.1
HCatalog	2.1.1	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.11.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.157.1	0.157.1	0.157.1	0.157.1
Spark	2.1.0	2.0.2	2.0.2	2.0.2
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.3.0	emr-5.2.3	emr-5.2.2	emr-5.2.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.2	0.6.2	0.6.2
ZooKeeper	3.4.9	3.4.9	3.4.9	3.4.9

## 5.3.0 发布说明

以下发行说明包含 Amazon EMR 5.3.0 版本的信息。更改与 Amazon EMR 5.2.1 版本有关。

发布日期：2017 年 1 月 26 日

### 升级

- 已升级到 Hive 2.1.1
- 已升级到 Hue 3.11.0
- 已升级到 Spark 2.1.0
- 已升级到 Oozie 4.3.0
- 已升级到 Flink 1.1.4

### 更改和增强功能

- Hue 新增补丁可使您使用 `interpreters_shown_on_wheel` 设置配置解释器在笔记本选择轮盘上最先显示的内容，而不受 `hue.ini` 文件中排序的限制。
- 新增 `hive-parquet-logging` 配置分类，您可以将其用于配置 Hive `parquet-logging.properties` 文件中的值。

## 5.3.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
emr-ddb	4.2.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.2.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.2.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.4.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	emrfs	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	flink-client	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。

组件	版本	描述
hadoop-client	2.7.3-amzn-1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.3-amzn-1	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-https-server	2.7.3-amzn-1	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-1	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.3-amzn-1	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.2.3	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.2.3	为一个或多个HBase地区提供服务的服务。
hbase-client	1.2.3	HBase命令行客户端。



组件	版本	描述
hbase-rest-server	1.2.3	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.2.3	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	2.1.1-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.1.1-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.1.1-amzn-0	Hive 命令行客户端。
hive-metastore-server	2.1.1-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server	2.1.1-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.11.0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.12.2	用于机器学习的库。
mysql-server	5.5.52	我的SQL数据库服务器。
oozie-client	4.3.0	Oozie 命令行客户端。
oozie-server	4.3.0	用于接受 Oozie 工作流请求的服务。

组件	版本	描述
phoenix-library	4.7.0--1.2 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.7.0--1.2 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限 API
presto-coordinator	0.157.1	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.157.1	用于执行查询的各个部分的服务。
pig-client	0.16.0-amzn-0	Pig 命令行客户端。
spark-client	2.1.0	Spark 命令行客户端。
spark-history-server	2.1.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.1.0	的内存中执行引擎。YARN
spark-yarn-slave	2.1.0	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.6.2	支持交互式数据分析的基于 Web 的笔记本电脑。

组件	版本	描述
zookeeper-server	3.4.9	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.9	ZooKeeper 命令行客户端。

### 5.3.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.3.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置

分类	描述
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。
hbase-policy	更改 hbase-policy.xml 文件HBase中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改 jndi.prop HCatalog erties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-parquet-logging	更改 Hive parquet-logging.properties 文件中的值。

分类	描述
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。

分类	描述
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。

分类	描述
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。

分类	描述
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.2.3

### 5.2.3 用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、[Pig](#)、[Presto](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-5.2.3	emr-5.2.2	emr-5.2.1	emr-5.2.0
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-



	emr-5.2.3	emr-5.2.2	emr-5.2.1	emr-5.2.0
Flink	1.1.3	1.1.3	1.1.3	1.1.3
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.3	1.2.3
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.157.1	0.157.1	0.157.1	0.152.3
Spark	2.0.2	2.0.2	2.0.2	2.0.2
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6

	emr-5.2.3	emr-5.2.2	emr-5.2.1	emr-5.2.0
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.2	0.6.2	0.6.2
ZooKeeper	3.4.9	3.4.9	3.4.9	3.4.8

### 5.2.3 发布说明

这是一个补丁版本，用于为向 Amazon S3 发出的请求添加 AWS 签名版本 4 身份验证。所有应用程序和组件都与之前的 Amazon EMR 版本相同。

#### Important

在此版本中，亚马逊仅EMR使用 AWS 签名版本 4 来验证向 Amazon S3 发出的请求。有关更多信息，请参阅[新功能](#)。

### 5.2.3 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
emr-ddb	4.2.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.2.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.2.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.4.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.13.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.1.3	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS用于存储区块的节点级服务。

组件	版本	描述
hadoop-hdfs-library	2.7.3-amzn-1	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.3-amzn-1	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-1	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.3-amzn-1	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.2.3	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.2.3	为一个或多个HBase地区提供服务的服务。
hbase-client	1.2.3	HBase命令行客户端。
hbase-rest-server	1.2.3	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.2.3	为其提供 Thrift 端点的HBase服务。

组件	版本	描述
hcatalog-client	2.1.0-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.1.0-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.1.0-amzn-0	Hive 命令行客户端。
hive-metastore-server	2.1.0-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server	2.1.0-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.10.0-amzn-0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.12.2	用于机器学习的库。
mysql-server	5.5.52	我的SQL数据库服务器。
oozie-client	4.2.0	Oozie 命令行客户端。
oozie-server	4.2.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.7.0--1.2 HBase	服务器和客户端的 phoenix 库

组件	版本	描述
phoenix-query-server	4.7.0--1.2 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.157.1	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.157.1	用于执行查询的各个部分的服务。
pig-client	0.16.0-amzn-0	Pig 命令行客户端。
spark-client	2.0.2	Spark 命令行客户端。
spark-history-server	2.0.2	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.0.2	的内存中执行引擎。YARN
spark-yarn-slave	2.0.2	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.23	Apache HTTP 服务器。
zeppelin-server	0.6.2	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.9	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。

组件	版本	描述
zookeeper-client	3.4.9	ZooKeeper 命令行客户端。

### 5.2.3 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.2.3 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。

分类	描述
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。
hbase-policy	更改 hbase-policy.xml 文件HBase中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改 jndi.prop HCatalog erties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。



分类	描述
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。

分类	描述
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。

分类	描述
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.2.2

### 5.2.2 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、[Pig](#)、[Presto](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-5.2.2	emr-5.2.1	emr-5.2.0	emr-5.1.1
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.1.3	1.1.3	1.1.3	1.1.3
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.3	1.2.3

	emr-5.2.2	emr-5.2.1	emr-5.2.0	emr-5.1.1
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.157.1	0.157.1	0.152.3	0.152.3
Spark	2.0.2	2.0.2	2.0.2	2.0.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.2.2	emr-5.2.1	emr-5.2.0	emr-5.1.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.2	0.6.2	0.6.2
ZooKeeper	3.4.9	3.4.9	3.4.8	3.4.8

## 5.2.2 发布说明

以下发行说明包含 Amazon EMR 5.2.2 版本的信息。更改与 Amazon EMR 5.2.1 版本有关。

发布日期：2017 年 5 月 2 日

早期版本中已解决的已知问题

- 向后移植 [SPARK-194459](#)，它解决了从包含 char/varchar ORC 列的表中读取可能失败的问题。

## 5.2.2 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 emr 或 aws 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
emr-ddb	4.2.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。

组件	版本	描述
emr-goodies	2.2.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.2.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.4.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.13.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.1.3	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.3-amzn-1	HDFS命令行客户端和库

组件	版本	描述
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.3-amzn-1	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-1	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-1	MapReduce 用于运行 MapReduce 应用程序的执行引擎。
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.3-amzn-1	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.2.3	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.2.3	为一个或多个HBase地区提供服务的服务。
hbase-client	1.2.3	HBase命令行客户端。
hbase-rest-server	1.2.3	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.2.3	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.1.0-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。



组件	版本	描述
hcatalog-server	2.1.0-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.1.0-amzn-0	Hive 命令行客户端。
hive-metastore-server	2.1.0-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server	2.1.0-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.10.0-amzn-0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.12.2	用于机器学习的库。
mysql-server	5.5.52	我的SQL数据库服务器。
oozie-client	4.2.0	Oozie 命令行客户端。
oozie-server	4.2.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.7.0--1.2 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.7.0--1.2 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API

组件	版本	描述
presto-coordinator	0.157.1	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.157.1	用于执行查询的各个部分的服务。
pig-client	0.16.0-amzn-0	Pig 命令行客户端。
spark-client	2.0.2	Spark 命令行客户端。
spark-history-server	2.0.2	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.0.2	的内存中执行引擎。YARN
spark-yarn-slave	2.0.2	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.23	Apache HTTP 服务器。
zeppelin-server	0.6.2	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.9	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.9	ZooKeeper 命令行客户端。

## 5.2.2 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

### emr-5.2.2 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。

分类	描述
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改 jndi.properties HCatalog 中的值。
hcatalog-server-protocol-hive-site	更改 protocol-hive-site.xml HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。

分类	描述
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。

分类	描述
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。

分类	描述
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.2.1

### 5.2.1 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、[Pig](#)、[Presto](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-5.2.1	emr-5.2.0	emr-5.1.1	emr-5.1.0
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.1.3	1.1.3	1.1.3	1.1.3
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.3	1.2.3



	emr-5.2.1	emr-5.2.0	emr-5.1.1	emr-5.1.0
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.157.1	0.152.3	0.152.3	0.152.3
Spark	2.0.2	2.0.2	2.0.1	2.0.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.2.1	emr-5.2.0	emr-5.1.1	emr-5.1.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.2	0.6.2	0.6.2
ZooKeeper	3.4.9	3.4.8	3.4.8	3.4.8

## 5.2.1 发布说明

以下发行说明包含 Amazon EMR 5.2.1 版本的信息。更改与亚马逊 EMR 5.2.0 版本有关。

发布日期：2016 年 12 月 29 日

### 升级

- 已升级到 Presto 0.157.1。有关更多信息，请参阅 Presto 文档中的 [Presto 发布说明](#)。
- 已升级到 Zookeeper 3.4.9。有关更多信息，请参阅 Apache ZooKeeper 文档中的 [ZooKeeper 发行说明](#)。

### 更改和增强功能

- 在亚马逊 4.8.3 及更高EMR版本中增加了对亚马逊 EC2 m4.16xlarge 实例类型的支持，不包括 5.0.0、5.0.3 和 5.2.0。
- 亚马逊EMR发布的版本现在基于亚马逊 Linux 2016.09。有关更多信息，请参阅 <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>。
- 现在，Flink 和YARN配置路径的位置是默认设置的/etc/default/flink，因为您无需设置环境变量FLINK\_CONF\_DIR，运行flink或yarn-session.sh驱动脚本即可启动 Flink 作业。HADOOP\_CONF\_DIR
- 增加了对 FlinkKinesisConsumer 类的支持。

### 早期版本中已解决的已知问题

- 修复了 Hadoop 中的一个问题，即由于在大型集群中复制和删除同一文件之间存在竞争，ReplicationMonitor 线程可能会长时间卡住。
- 修复了任务状态未成功更新时 ControlledJob # toString 失败并出现空指针异常 (NPE) 的问题。

## 5.2.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>emr-ddb</code>	4.2.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.2.0	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.2.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
<code>emr-s3-dist-cp</code>	2.4.0	针对 Amazon S3 优化的分布式复制应用程序。
<code>emrfs</code>	2.13.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
<code>flink-client</code>	1.1.3	Apache Flink 命令行客户端脚本和应用程序。
<code>ganglia-monitor</code>	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。

组件	版本	描述
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.3-amzn-1	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.3-amzn-1	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-1	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.3-amzn-1	用于检索YARN应用程序的当前和历史信息的服务。

组件	版本	描述
hbase-hmaster	1.2.3	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.2.3	为一个或多个HBase地区提供服务的服务。
hbase-client	1.2.3	HBase命令行客户端。
hbase-rest-server	1.2.3	为其提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.2.3	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.1.0-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.1.0-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.1.0-amzn-0	Hive 命令行客户端。
hive-metastore-server	2.1.0-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server	2.1.0-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.10.0-amzn-0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序

组件	版本	描述
mahout-client	0.12.2	用于机器学习的库。
mysql-server	5.5.52	我的SQL数据库服务器。
oozie-client	4.2.0	Oozie 命令行客户端。
oozie-server	4.2.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.7.0--1.2 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.7.0--1.2 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.157.1	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.157.1	用于执行查询的各个部分的服务。
pig-client	0.16.0-amzn-0	Pig 命令行客户端。
spark-client	2.0.2	Spark 命令行客户端。
spark-history-server	2.0.2	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.0.2	的内存中执行引擎。YARN
spark-yarn-slave	2.0.2	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。

组件	版本	描述
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.23	Apache HTTP 服务器。
zeppelin-server	0.6.2	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.9	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.9	ZooKeeper 命令行客户端。

## 5.2.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

### emr-5.2.1 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。

分类	描述
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改 jndi.prop HCatalog erties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。



分类	描述
hcatalog-webhcat-site	更改 HCatalog WebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。

分类	描述
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。

分类	描述
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。

分类	描述
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper 中的值。
zookeeper-log4j	更改 ZooKeeper 的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.2.0

### 5.2.0 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、[Pig](#)、[Presto](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-5.2.0	emr-5.1.1	emr-5.1.0	emr-5.0.3
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	2.11.8	2.11.8	2.11.8	2.11.8

	emr-5.2.0	emr-5.1.1	emr-5.1.0	emr-5.0.3
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.1.3	1.1.3	1.1.3	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.3	1.2.2
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0

	emr-5.2.0	emr-5.1.1	emr-5.1.0	emr-5.0.3
Presto	0.152.3	0.152.3	0.152.3	0.152.3
Spark	2.0.2	2.0.1	2.0.1	2.0.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.2	0.6.2	0.6.1
ZooKeeper	3.4.8	3.4.8	3.4.8	3.4.8

## 5.2.0 发布说明

以下发行说明包含 Amazon EMR 5.2.0 版本的信息。更改与亚马逊 EMR 5.1.0 版本有关。

发布日期：2016 年 11 月 21 日

### 更改和增强功能

- 为添加了 Amazon S3 存储模式HBase。
- 允许您为 r HBase oodir 指定 Amazon S3 的位置。有关更多信息，请参阅 [Amaz HBase on S3](#)。

### 升级

- 已升级到 Spark 2.0.2

### 早期版本中已解决的已知问题

- 修复了EBS仅限实例类型的 /mnt 限制为 2 TB 的问题。

- 修复了输出到相应的 .out 文件而不是常规 log4j 配置的 .log 文件 (每小时转动一次) 的 instance-controller 和 logpusher 日志的问题。 .out 文件不会转动，因此这最终将填满 /emr 分区。此问题仅影响硬件虚拟机 (HVM) 实例类型。

## 5.2.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 emr 或 aws 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
emr-ddb	4.1.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.1.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.2.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.4.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.12.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.1.3	Apache Flink 命令行客户端脚本和应用程序。

组件	版本	描述
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-0	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-0	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.3-amzn-0	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.3-amzn-0	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.3-amzn-0	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-0	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-0	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.3-amzn-0	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-0	YARN用于分配和管理群集资源和分布式应用程序的服务。



组件	版本	描述
hadoop-yarn-timeline-server	2.7.3-amzn-0	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.2.3	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.2.3	为一个或多个HBase地区提供服务的服务。
hbase-client	1.2.3	HBase命令行客户端。
hbase-rest-server	1.2.3	为其提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.2.3	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.1.0-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.1.0-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.1.0-amzn-0	Hive 命令行客户端。
hive-metastore-server	2.1.0-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server	2.1.0-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。

组件	版本	描述
hue-server	3.10.0-amzn-0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.12.2	用于机器学习的库。
mysql-server	5.5.52	我的SQL数据库服务器。
oozie-client	4.2.0	Oozie 命令行客户端。
oozie-server	4.2.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.7.0--1.2 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.7.0--1.2 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.152.3	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.152.3	用于执行查询的各个部分的服务。
pig-client	0.16.0-amzn-0	Pig 命令行客户端。
spark-client	2.0.2	Spark 命令行客户端。
spark-history-server	2.0.2	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.0.2	的内存中执行引擎。YARN
spark-yarn-slave	2.0.2	YARN从属服务器需要的 Apache Spark 库。

组件	版本	描述
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.23	Apache HTTP 服务器。
zeppelin-server	0.6.2	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.8	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.8	ZooKeeper 命令行客户端。

## 5.2.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

### emr-5.2.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。

分类	描述
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase	亚马逊EMR精心策划的 Apache 设置。HBase
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改 jndi.prop HCatalog erties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。

分类	描述
hcatalog-webhcat-site	更改 HCatalog WebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。

分类	描述
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。

分类	描述
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。

分类	描述
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper 中的值。
zookeeper-log4j	更改 ZooKeeper 的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.1.1

### 5.1.1 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、[Pig](#)、[Presto](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-5.1.1	emr-5.1.0	emr-5.0.3	emr-5.0.2
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	2.11.8	2.11.8	2.11.8	2.11.8



	emr-5.1.1	emr-5.1.0	emr-5.0.3	emr-5.0.2
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.1.3	1.1.3	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.2	1.2.2
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.2
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0

	emr-5.1.1	emr-5.1.0	emr-5.0.3	emr-5.0.2
Presto	0.152.3	0.152.3	0.152.3	0.150
Spark	2.0.1	2.0.1	2.0.1	2.0.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.2	0.6.1	0.6.1
ZooKeeper	3.4.8	3.4.8	3.4.8	3.4.8

### 5.1.1 发布说明

这是一个补丁版本，用于为向 Amazon S3 发出的请求添加 AWS 签名版本 4 身份验证。所有应用程序和组件都与之前的 Amazon EMR 版本相同。

#### Important

在此版本中，亚马逊仅EMR使用 AWS 签名版本 4 来验证向 Amazon S3 发出的请求。有关更多信息，请参阅[新功能](#)。

### 5.1.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 emr 或 aws 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2

版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
emr-ddb	4.1.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.1.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.2.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.4.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.11.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.1.3	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-0	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。

组件	版本	描述
hadoop-hdfs-datanode	2.7.3-amzn-0	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.3-amzn-0	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.3-amzn-0	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httfs-server	2.7.3-amzn-0	HTTPhDFS操作端点。
hadoop-kms-server	2.7.3-amzn-0	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-0	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.3-amzn-0	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-0	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.3-amzn-0	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.2.3	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.2.3	为一个或多个HBase地区提供服务的服务。
hbase-client	1.2.3	HBase命令行客户端。
hbase-rest-server	1.2.3	为其提供RESTfulHTTP终端节点的服务HBase。

组件	版本	描述
hbase-thrift-server	1.2.3	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	2.1.0-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.1.0-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP端点为其提供REST接口 HCatalog。
hive-client	2.1.0-amzn-0	Hive 命令行客户端。
hive-metastore-server	2.1.0-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server	2.1.0-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.10.0-amzn-0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.12.2	用于机器学习的库。
mysql-server	5.5.52	我的SQL数据库服务器。
oozie-client	4.2.0	Oozie 命令行客户端。
oozie-server	4.2.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.7.0--1.2 HBase	服务器和客户端的 phoenix 库

组件	版本	描述
phoenix-query-server	4.7.0--1.2 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.152.3	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.152.3	用于执行查询的各个部分的服务。
pig-client	0.16.0-amzn-0	Pig 命令行客户端。
spark-client	2.0.1	Spark 命令行客户端。
spark-history-server	2.0.1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.0.1	的内存中执行引擎。YARN
spark-yarn-slave	2.0.1	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.23	Apache HTTP 服务器。
zeppelin-server	0.6.2	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.8	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。

组件	版本	描述
zookeeper-client	3.4.8	ZooKeeper 命令行客户端。

### 5.1.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.1.1 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase-env	更改环境HBase中的值。

分类	描述
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改 jndi.prop HCatalog erties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。



分类	描述
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。

分类	描述
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。

分类	描述
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.1.0

### 5.1.0 应用程序版本

此版本支持以下应用程序

序：[Flink](#)、[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、[Pig](#)、[Presto](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-5.1.0	emr-5.0.3	emr-5.0.2	emr-5.0.1
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.1.3	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.2	1.2.2	1.2.2

	emr-5.1.0	emr-5.0.3	emr-5.0.2	emr-5.0.1
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.2	2.7.2
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.152.3	0.152.3	0.150	0.150
Spark	2.0.1	2.0.1	2.0.0	2.0.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.1.0	emr-5.0.3	emr-5.0.2	emr-5.0.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.1	0.6.1	0.6.1
ZooKeeper	3.4.8	3.4.8	3.4.8	3.4.8

## 5.1.0 发布说明

以下发行说明包含 Amazon EMR 5.1.0 版本的信息。更改与 Amazon EMR 5.0.3 版本有关。

发布日期：2016 年 11 月 3 日

### 更改和增强功能

- 增加了对 Flink 1.1.3 的支持。
- Presto 已作为 Hue 的记事本部分中的选项添加。

### 升级

- 已升级到 HBase 1.2.3
- 已升级到 Zeppelin 0.6.2

### 早期版本中已解决的已知问题

- 修复了 Amazon S3 上的 Tez 查询问题，其中 ORC 文件表现不如早期的 Amazon EMR 4.x 版本。

## 5.1.0 组件版本

下面列出了 Amazon 随此版本 EMR 安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的 EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新 EMR 版本中的大数据应用程序包通常是社区中的最新版本。我们会 EMR 尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本 EMR 不同。这些组件具有以下形式的 `CommunityVersion-amzn-EmrVersion` 的发行版标注。`EmrVersion` 从 0 开始。例如，如果 `myapp-component` 以 2.2

版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
emr-ddb	4.1.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.1.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.2.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.4.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.11.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
flink-client	1.1.3	Apache Flink 命令行客户端脚本和应用程序。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-0	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。

组件	版本	描述
hadoop-hdfs-datanode	2.7.3-amzn-0	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.3-amzn-0	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.3-amzn-0	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httfs-server	2.7.3-amzn-0	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-0	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-0	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.3-amzn-0	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-0	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.3-amzn-0	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.2.3	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.2.3	为一个或多个HBase地区提供服务的服务。
hbase-client	1.2.3	HBase命令行客户端。
hbase-rest-server	1.2.3	为其提供RESTfulHTTP终端节点的服务HBase。



组件	版本	描述
hbase-thrift-server	1.2.3	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	2.1.0-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.1.0-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP端点为其提供REST接口 HCatalog。
hive-client	2.1.0-amzn-0	Hive 命令行客户端。
hive-metastore-server	2.1.0-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server	2.1.0-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.10.0-amzn-0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.12.2	用于机器学习的库。
mysql-server	5.5.52	我的SQL数据库服务器。
oozie-client	4.2.0	Oozie 命令行客户端。
oozie-server	4.2.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.7.0--1.2 HBase	服务器和客户端的 phoenix 库

组件	版本	描述
phoenix-query-server	4.7.0--1.2 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.152.3	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.152.3	用于执行查询的各个部分的服务。
pig-client	0.16.0-amzn-0	Pig 命令行客户端。
spark-client	2.0.1	Spark 命令行客户端。
spark-history-server	2.0.1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.0.1	的内存中执行引擎。YARN
spark-yarn-slave	2.0.1	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.23	Apache HTTP 服务器。
zeppelin-server	0.6.2	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.8	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。

组件	版本	描述
zookeeper-client	3.4.8	ZooKeeper 命令行客户端。

### 5.1.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.1.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
flink-conf	更改 flink-conf.yaml 设置。
flink-log4j	更改 Flink log4j.properties 设置。
flink-log4j-yarn-session	更改 Flink log4 j-yarn-session .properties 设置。
flink-log4j-cli	更改 Flink log4j-cli.properties 设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase-env	更改环境HBase中的值。

分类	描述
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改 jndi.prop HCatalog erties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。

分类	描述
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。

分类	描述
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。

分类	描述
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.0.3

### 5.0.3 应用程序版本

此版本支持以下应用程序：

[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、[Pig](#)、[Presto](#)、[Spark](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2



	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.2	2.7.2	2.7.2
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.152.3	0.150	0.150	0.150
Spark	2.0.1	2.0.0	2.0.0	2.0.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	3.4.8	3.4.8	3.4.8	3.4.8

### 5.0.3 发布说明

以下发行说明包含 Amazon EMR 5.0.3 版本的信息。更改与 Amazon EMR 5.0.0 版本有关。

发布日期：2016 年 10 月 24 日

#### 升级

- 已升级到 Hadoop 2.7.3
- 已升级到 Presto 0.152.3，它包括对 Presto Web 界面的支持。可使用端口 8889 访问 Presto 协调器上的 Presto Web 界面。有关 Presto Web 界面的更多信息，请参阅 Presto 文档中的 [Web 界面](#)。
- 已升级到 Spark 2.0.1
- 亚马逊EMR发布的版本现在基于亚马逊 Linux 2016.09。有关更多信息，请参阅 <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>。

### 5.0.3 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 emr 或 aws 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 myapp-component 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
emr-ddb	4.1.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.1.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.2.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.4.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.10.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-0	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-0	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.3-amzn-0	HDFS命令行客户端和库

组件	版本	描述
hadoop-hdfs-namenode	2.7.3-amzn-0	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.3-amzn-0	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-0	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-0	MapReduce 用于运行 MapReduce 应用程序的执行引擎。
hadoop-yarn-nodemanager	2.7.3-amzn-0	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-0	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.3-amzn-0	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.2.2	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.2.2	为一个或多个HBase地区提供服务的服务。
hbase-client	1.2.2	HBase命令行客户端。
hbase-rest-server	1.2.2	为其提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.2.2	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.1.0-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。

组件	版本	描述
hcatalog-server	2.1.0-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.1.0-amzn-0	Hive 命令行客户端。
hive-metastore-server	2.1.0-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server	2.1.0-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.10.0-amzn-0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.12.2	用于机器学习的库。
mysql-server	5.5.52	我的SQL数据库服务器。
oozie-client	4.2.0	Oozie 命令行客户端。
oozie-server	4.2.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.7.0--1.2 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.7.0--1.2 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API

组件	版本	描述
presto-coordinator	0.152.3	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.152.3	用于执行查询的各个部分的服务。
pig-client	0.16.0-amzn-0	Pig 命令行客户端。
spark-client	2.0.1	Spark 命令行客户端。
spark-history-server	2.0.1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.0.1	的内存中执行引擎。YARN
spark-yarn-slave	2.0.1	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.23	Apache HTTP 服务器。
zeppelin-server	0.6.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.8	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.8	ZooKeeper 命令行客户端。

### 5.0.3 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-5.0.3 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。

分类	描述
hcatalog-server-jndi	更改 jndi.prop HCatalog erties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。



分类	描述
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。

分类	描述
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。

分类	描述
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.0.2

### 5.0.2 应用程序版本

此版本支持以下应用程序：

[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、[Pig](#)、[Presto](#)、[Spark](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

## 应用程序版本信息

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.2	2.7.2	2.7.2
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.152.3	0.150	0.150	0.150
Spark	2.0.1	2.0.0	2.0.0	2.0.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	3.4.8	3.4.8	3.4.8	3.4.8

## 5.0.2 发布说明

### 5.0.2 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
emr-ddb	4.0.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.1.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.2.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.4.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.9.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.2-amzn-3	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.2-amzn-3	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.2-amzn-3	HDFS命令行客户端和库

组件	版本	描述
hadoop-hdfs-namenode	2.7.2-amzn-3	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.2-amzn-3	HTTPHDFS操作端点。
hadoop-kms-server	2.7.2-amzn-3	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.2-amzn-3	MapReduce 用于运行 MapReduce 应用程序的执行引擎。
hadoop-yarn-nodemanager	2.7.2-amzn-3	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.2-amzn-3	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.2-amzn-3	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.2.2	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.2.2	为一个或多个HBase地区提供服务的服务。
hbase-client	1.2.2	HBase命令行客户端。
hbase-rest-server	1.2.2	为其提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.2.2	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.1.0-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。

组件	版本	描述
hcatalog-server	2.1.0-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.1.0-amzn-0	Hive 命令行客户端。
hive-metastore-server	2.1.0-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server	2.1.0-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.10.0-amzn-0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.12.2	用于机器学习的库。
mysql-server	5.5.46	我的SQL数据库服务器。
oozie-client	4.2.0	Oozie 命令行客户端。
oozie-server	4.2.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.7.0--1.2 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.7.0--1.2 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API



组件	版本	描述
presto-coordinator	0.150	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.150	用于执行查询的各个部分的服务。
pig-client	0.16.0-amzn-0	Pig 命令行客户端。
spark-client	2.0.0	Spark 命令行客户端。
spark-history-server	2.0.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.0.0	的内存中执行引擎。YARN
spark-yarn-slave	2.0.0	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.23	Apache HTTP 服务器。
zeppelin-server	0.6.1-SNAPSHOT	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.8	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.8	ZooKeeper 命令行客户端。

## 5.0.2 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

### emr-5.0.2 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。

分类	描述
hcatalog-server-jndi	更改 jndi.prop HCatalog erties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。

分类	描述
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。

分类	描述
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。

分类	描述
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.0.1

### 5.0.1 应用程序版本

此版本支持以下应用程序：

[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、[Pig](#)、[Presto](#)、[Spark](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

## 应用程序版本信息

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.2	2.7.2	2.7.2
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.152.3	0.150	0.150	0.150
Spark	2.0.1	2.0.0	2.0.0	2.0.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	3.4.8	3.4.8	3.4.8	3.4.8

## 5.0.1 发布说明

### 5.0.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。



组件	版本	描述
emr-ddb	4.0.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.1.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.2.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.4.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.9.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.2-amzn-3	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.2-amzn-3	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.2-amzn-3	HDFS命令行客户端和库

组件	版本	描述
hadoop-hdfs-namenode	2.7.2-amzn-3	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.2-amzn-3	HTTPHDFS操作端点。
hadoop-kms-server	2.7.2-amzn-3	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.2-amzn-3	MapReduce 用于运行 MapReduce 应用程序的执行引擎。
hadoop-yarn-nodemanager	2.7.2-amzn-3	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.2-amzn-3	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.2-amzn-3	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.2.2	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.2.2	为一个或多个HBase地区提供服务的服务。
hbase-client	1.2.2	HBase命令行客户端。
hbase-rest-server	1.2.2	为其提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.2.2	为其提供 Thrift 端点的HBase服务。
hcatalog-client	2.1.0-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。

组件	版本	描述
hcatalog-server	2.1.0-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.1.0-amzn-0	Hive 命令行客户端。
hive-metastore-server	2.1.0-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server	2.1.0-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.10.0-amzn-0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.12.2	用于机器学习的库。
mysql-server	5.5.46	我的SQL数据库服务器。
oozie-client	4.2.0	Oozie 命令行客户端。
oozie-server	4.2.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.7.0--1.2 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.7.0--1.2 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API

组件	版本	描述
presto-coordinator	0.150	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.150	用于执行查询的各个部分的服务。
pig-client	0.16.0-amzn-0	Pig 命令行客户端。
spark-client	2.0.0	Spark 命令行客户端。
spark-history-server	2.0.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.0.0	的内存中执行引擎。YARN
spark-yarn-slave	2.0.0	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.23	Apache HTTP 服务器。
zeppelin-server	0.6.1-SNAPSHOT	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.8	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.8	ZooKeeper 命令行客户端。

## 5.0.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

### emr-5.0.1 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。

分类	描述
hcatalog-server-jndi	更改 jndi.prop HCatalog erties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。

分类	描述
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。

分类	描述
presto-connector-cassandra	更改 Presto 的 <code>cassandra.properties</code> 文件中的值。
presto-connector-hive	更改 Presto 的 <code>hive.properties</code> 文件中的值。
presto-connector-jmx	更改 Presto 的 <code>jmx.properties</code> 文件中的值。
presto-connector-kafka	更改 Presto 的 <code>kafka.properties</code> 文件中的值。
presto-connector-localfile	更改 Presto 的 <code>localfile.properties</code> 文件中的值。
presto-connector-mongodb	更改 Presto 的 <code>mongodb.properties</code> 文件中的值。
presto-connector-mysql	更改 Presto 的 <code>mysql.properties</code> 文件中的值。
presto-connector-postgresql	更改 Presto 的 <code>postgresql.properties</code> 文件中的值。
presto-connector-raptor	更改 Presto 的 <code>raptor.properties</code> 文件中的值。
presto-connector-redis	更改 Presto 的 <code>redis.properties</code> 文件中的值。
presto-connector-tpch	更改 Presto 的 <code>tpch.properties</code> 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 <code>spark-defaults.conf</code> 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 <code>hive-site.xml</code> 文件中的值
spark-log4j	更改 Spark 的 <code>log4j.properties</code> 文件中的值。
spark-metrics	更改 Spark 的 <code>metrics.properties</code> 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。



分类	描述
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 5.0.0

### 5.0.0 应用程序版本

此版本支持以下应用程序：

[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie](#)、[Phoenix](#)、[Pig](#)、[Presto](#)、[Spark](#)、和 [ZooKeeper](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

## 应用程序版本信息

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.2	2.7.2	2.7.2
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.16.0	0.16.0	0.16.0	0.16.0
Presto	0.152.3	0.150	0.150	0.150
Spark	2.0.1	2.0.0	2.0.0	2.0.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	3.4.8	3.4.8	3.4.8	3.4.8

## 5.0.0 发布说明

发布日期：2016 年 7 月 27 日

### 升级

- 已升级到 Hive 2.1
- 已升级到 Presto 0.150
- 已升级到 Spark 2.0
- 已升级到 Hue 3.10.0
- 已升级到 Pig 0.16.0
- 已升级到 Tez 0.8.4

- 已升级到 Zeppelin 0.6.1

## 更改和增强功能

- 亚马逊EMR支持 Hive ( 版本 2.1 ) 和 Pig ( 版本 0.16.0 ) 的最新开源版本。如果您过去曾在EMR在亚马逊上使用过 Hive 或 Pig，这可能会影响某些用例。有关更多信息，请参阅 [Hive](#) 和 [Pig](#)。
- Hive 和 Pig 的默认执行引擎现在是 Tez。要更改该设置，您可以在 hive-site 和 pig-properties 配置分类中分别编辑相应的值。
- 添加了增强型步骤调试功能，可让您查看步骤失败的根本原因 (如果服务可以确定原因)。有关更多信息，请参阅 Amazon EMR 管理指南中的 [增强型步骤调试](#)。
- 先前以“-Sandbox”结尾的应用程序不再拥有该后缀。这可能会中断您的自动化，例如，如果您使用脚本来启动具有这些应用程序的集群。下表显示了亚马逊 EMR 4.7.2 与亚马逊 EMR 5.0.0 中的应用程序名称。

### 应用程序名称更改

亚马逊 EMR 4.7.2	亚马逊 EMR 5.0.0
Oozie-Sandbox	Oozie
Presto-Sandbox	Presto
Sqoop-Sandbox	Sqoop
Zeppelin-Sandbox	Zeppelin
ZooKeeper-沙箱	ZooKeeper

- Spark 现在针对 Scala 2.11 进行编译。
- Java 8 现在是默认设置JVM。所有应用程序均使用 Java 8 runtime 运行。对任何应用程序的字节代码目标都没有进行更改。大多数应用程序继续运行 Java 7。
- Zeppelin 现在包括身份验证功能。有关更多信息，请参阅 [Zeppelin](#)。
- 添加了对安全配置的支持，这使您可以更轻松地创建和应用加密选项。有关更多信息，请参阅[数据加密](#)。

## 5.0.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>emr-ddb</code>	4.0.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.1.0	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.2.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
<code>emr-s3-dist-cp</code>	2.4.0	针对 Amazon S3 优化的分布式复制应用程序。
<code>emrfs</code>	2.9.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
<code>ganglia-monitor</code>	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
<code>ganglia-metadata-collector</code>	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。

组件	版本	描述
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.2-amzn-3	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.2-amzn-3	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.2-amzn-3	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.2-amzn-3	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.2-amzn-3	HTTPHDFS操作端点。
hadoop-kms-server	2.7.2-amzn-3	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.2-amzn-3	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.2-amzn-3	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.2-amzn-3	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.2-amzn-3	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.2.2	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.2.2	为一个或多个HBase地区提供服务的服务。

组件	版本	描述
hbase-client	1.2.2	HBase命令行客户端。
hbase-rest-server	1.2.2	为其提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.2.2	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	2.1.0-amzn-0	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	2.1.0-amzn-0	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP端点为其提供REST接口HCatalog。
hive-client	2.1.0-amzn-0	Hive 命令行客户端。
hive-metastore-server	2.1.0-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server	2.1.0-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.10.0-amzn-0	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.12.2	用于机器学习的库。
mysql-server	5.5.46	我的SQL数据库服务器。
oozie-client	4.2.0	Oozie 命令行客户端。

组件	版本	描述
oozie-server	4.2.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.7.0--1.2 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.7.0--1.2 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.150	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.150	用于执行查询的各个部分的服务。
pig-client	0.16.0-amzn-0	Pig 命令行客户端。
spark-client	2.0.0	Spark 命令行客户端。
spark-history-server	2.0.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	2.0.0	的内存中执行引擎。YARN
spark-yarn-slave	2.0.0	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.23	Apache HTTP 服务器。
zeppelin-server	0.6.1-SNAPSHOT	支持交互式数据分析的基于 Web 的笔记本电脑。



组件	版本	描述
zookeeper-server	3.4.8	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.8	ZooKeeper 命令行客户端。

## 5.0.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

### emr-5.0.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。

分类	描述
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改 jndi.prop HCatalog erties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j2	更改 HCatalog W ebHCat 的 log4j2.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-beeline-log4j2	更改 Hive 的 beeline-log4j2.properties 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j2	更改 Hive 的 hive-exec-log 4j2.properties 文件中的值。
hive-llap-daemon-log4j2	更改 Hive 的 llap-daemon-log 4j2.properties 文件中的值。
hive-log4j2	更改 Hive 的 hive-log4j2.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值

分类	描述
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。

分类	描述
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。

分类	描述
spark-env	更改 Spark 环境中的值。
spark-hive-site	更改 Spark 的 hive-site.xml 文件中的值
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊 EMR 4.x 发布版本

本部分包含每个 Amazon EMR 4.x 发行版本中可用的应用程序版本、发行说明、组件版本和配置分类。

启动集群时，您可以从 Amazon 的多个版本中进行选择EMR。这允许您测试和使用满足您解决方案兼容性需求的应用程序版本。您可以使用发行版标注指定版本号。版本标签的格式是 `emr-x.x.x`。例如，`emr-7.2.0`。

Amazon 的新EMR版本将在几天内在不同地区上市，从最初发布日期的第一个地区开始。在此期间，您所在区域可能无法提供最新发行版。

有关每个 Amazon EMR 4.x 版本中应用程序版本的完整表，请参阅[Amazon EMR 4.x 发行版中的应用程序版本](#)。

## 主题

- [Amazon EMR 4.x 发行版中的应用程序版本](#)
- [各 Amazon EMR 4.x 发行版之间的差异](#)
- [亚马逊EMR版本 4.9.6](#)
- [亚马逊EMR版本 4.9.5](#)
- [亚马逊EMR版本 4.9.4](#)
- [亚马逊EMR版本 4.9.3](#)
- [亚马逊EMR版本 4.9.2](#)
- [亚马逊EMR版本 4.9.1](#)
- [亚马逊EMR版本 4.8.5](#)
- [亚马逊EMR版本 4.8.4](#)
- [亚马逊EMR版本 4.8.3](#)
- [亚马逊EMR版本 4.8.2](#)
- [亚马逊EMR版本 4.8.1](#)
- [亚马逊EMR版本 4.8.0](#)
- [亚马逊EMR版本 4.7.4](#)
- [亚马逊EMR版本 4.7.3](#)
- [亚马逊EMR版本 4.7.2](#)
- [亚马逊EMR版本 4.7.1](#)
- [亚马逊EMR版本 4.7.0](#)
- [亚马逊EMR版本 4.6.1](#)
- [亚马逊EMR版本 4.6.0](#)
- [亚马逊EMR版本 4.5.0](#)
- [亚马逊EMR版本 4.4.0](#)
- [亚马逊EMR版本 4.3.0](#)

- [亚马逊EMR版本 4.2.0](#)
- [亚马逊EMR版本 4.1.0](#)
- [亚马逊EMR版本 4.0.0](#)

## Amazon EMR 4.x 发行版中的应用程序版本

有关列出每个 Amazon EMR 4.x 发行版中可用的应用程序版本的综合表格，请在浏览器打开 [Amazon EMR 4.x 发行版中的应用程序版本](#)。

## 各 Amazon EMR 4.x 发行版之间的差异

《Amazon EMR 管理指南》中有关 Amazon EMR 功能的文档指定开始提供某种功能的 Amazon EMR 发行版以及追溯至 4.0.0 的 Amazon EMR 功能之间的适用差异。

从 Amazon EMR 发行版 5.0.0 开始，一些应用程序已进行重大版本升级，安装或运行详细信息改变，另一些应用程序则从沙盒应用程序提升为本机应用程序。此节中的各个主题提供使用 Amazon EMR 4.x 发行版时特定于应用程序的明显差异。

### 主题

- [沙盒应用程序](#)
- [使用 Amazon EMR 4.x 上的 Hive 的注意事项](#)
- [使用 Amazon EMR 4.x 上的 Pig 的注意事项](#)

## 沙盒应用程序

使用 Amazon EMR 4.x 发行版时，一些应用程序会被视为沙盒应用程序。沙盒应用程序是我们在初始 Amazon EMR 发行版时因需求提供的应用程序的早期版本。您可以使用控制台、AWS CLI 或 API 让 Amazon EMR 安装沙盒应用程序（安装方式与本机应用程序的相同），但沙盒应用程序的支持和文档有限。沙盒应用程序在 Amazon EMR 发行版 5.0.0 及更高版本中变为完全受支持的本机应用程序。以下是 Amazon EMR 4.x 发行版中的沙盒应用程序。

- Oozie
- Presto
- Sqoop
- Zeppelin
- ZooKeeper

当您安装沙盒应用程序时，使用后缀 `-sandbox` 指示应用程序名称。例如，要安装沙盒版本的 *Presto*，请使用 `Presto-sandbox`。与完全受支持的应用程序相比，安装时间可能更长。此节中列出的每个应用程序的版本号对应于应用程序的社区版本。

## Oozie (沙盒版本)

Oozie 从 Amazon EMR 发行版 4.1.0 开始作为沙盒应用程序提供。

默认情况下，不会使用沙盒版本安装 Oozie 示例。要安装以上示例，可使用 SSH 连接到主节点并运行 `install-oozie-examples`。

## Oozie 沙盒版本信息

Amazon EMR 发行版标签	Oozie 沙盒版本	随 Oozie 沙盒安装的组件
emr-4.9.6	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.9.5	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server



Amazon EMR 发行版标签	Oozie 沙盒版本	随 Oozie 沙盒安装的组件
emr-4.9.4	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.9.3	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.9.2	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server

Amazon EMR 发行版标签	Oozie 沙盒版本	随 Oozie 沙盒安装的组件
emr-4.9.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.8.5	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.8.4	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server

Amazon EMR 发行版标签	Oozie 沙盒版本	随 Oozie 沙盒安装的组件
emr-4.8.3	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.8.2	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.8.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server

Amazon EMR 发行版标签	Oozie 沙盒版本	随 Oozie 沙盒安装的组件
emr-4.8.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.7.4	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.7.3	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server

Amazon EMR 发行版标签	Oozie 沙盒版本	随 Oozie 沙盒安装的组件
emr-4.7.2	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.7.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.7.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server

Amazon EMR 发行版标签	Oozie 沙盒版本	随 Oozie 沙盒安装的组件
emr-4.6.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.6.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.5.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server

Amazon EMR 发行版标签	Oozie 沙盒版本	随 Oozie 沙盒安装的组件
emr-4.4.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.3.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.2.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server

Amazon EMR 发行版标签	Oozie 沙盒版本	随 Oozie 沙盒安装的组件
emr-4.1.0	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server

### Presto ( 沙盒版本 )

Presto 从 Amazon EMR 发行版 4.1.0 开始作为沙盒应用程序提供。

### Presto 沙盒版本信息

Amazon EMR 发行版标签	Presto 沙盒版本	随 Presto 沙盒安装的组件
emr-4.9.6	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.9.5	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-



Amazon EMR 发行版标签	Presto 沙盒版本	随 Presto 沙盒安装的组件
		resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.9.4	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.9.3	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR 发行版标签	Presto 沙盒版本	随 Presto 沙盒安装的组件
emr-4.9.2	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.9.1	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.8.5	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR 发行版标签	Presto 沙盒版本	随 Presto 沙盒安装的组件
emr-4.8.4	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.8.3	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.8.2	0.152.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR 发行版标签	Presto 沙盒版本	随 Presto 沙盒安装的组件
emr-4.8.1	0.151	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.8.0	0.151	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.7.4	0.148	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR 发行版标签	Presto 沙盒版本	随 Presto 沙盒安装的组件
emr-4.7.3	0.148	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.7.2	0.148	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.7.1	0.147	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR 发行版标签	Presto 沙盒版本	随 Presto 沙盒安装的组件
emr-4.7.0	0.147	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.6.1	0.143	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, mysql-server, presto-coordinator, presto-worker
emr-4.6.0	0.143	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR 发行版标签	Presto 沙盒版本	随 Presto 沙盒安装的组件
emr-4.5.0	0.140	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, mysql-server, presto-coordinator, presto-worker
emr-4.4.0	0.136	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, mysql-server, presto-coordinator, presto-worker
emr-4.3.0	0.130	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, mysql-server, presto-coordinator, presto-worker

Amazon EMR 发行版标签	Presto 沙盒版本	随 Presto 沙盒安装的组件
emr-4.2.0	0.125	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, mysql-server, presto-coordinator, presto-worker
emr-4.1.0	0.119	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, mysql-server, presto-coordinator, presto-worker

### Sqoop ( 沙盒版本 )

Sqoop 从 Amazon EMR 发行版 4.4.0 开始作为沙盒应用程序提供。

### Sqoop 沙盒版本信息

Amazon EMR 发行版标签	Sqoop 沙盒版本	随 Sqoop 沙盒安装的组件
emr-4.9.6	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resour



Amazon EMR 发行版标签	Sqoop 沙盒版本	随 Sqoop 沙盒安装的组件
		cemanager, mysql-server, sqoop-client
emr-4.9.5	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, mysql-server, sqoop-client
emr-4.9.4	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, mysql-server, sqoop-client
emr-4.9.3	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, mysql-server, sqoop-client

Amazon EMR 发行版标签	Sqoop 沙盒版本	随 Sqoop 沙盒安装的组件
emr-4.9.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, mysql-server, sqoop-client
emr-4.9.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, mysql-server, sqoop-client
emr-4.8.5	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, mysql-server, sqoop-client

Amazon EMR 发行版标签	Sqoop 沙盒版本	随 Sqoop 沙盒安装的组件
emr-4.8.4	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.8.3	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.8.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Amazon EMR 发行版标签	Sqoop 沙盒版本	随 Sqoop 沙盒安装的组件
emr-4.8.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, mysql-server, sqoop-client
emr-4.8.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, mysql-server, sqoop-client
emr-4.7.4	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, mysql-server, sqoop-client

Amazon EMR 发行版标签	Sqoop 沙盒版本	随 Sqoop 沙盒安装的组件
emr-4.7.3	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.7.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.7.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Amazon EMR 发行版标签	Sqoop 沙盒版本	随 Sqoop 沙盒安装的组件
emr-4.7.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.6.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.6.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Amazon EMR 发行版标签	Sqoop 沙盒版本	随 Sqoop 沙盒安装的组件
emr-4.5.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, sqoop-client
emr-4.4.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, sqoop-client

## Zeppelin ( 沙盒版本 )

Zeppelin 从 Amazon EMR 发行版 4.1.0 开始作为沙盒应用程序提供。

### Zeppelin 沙盒版本信息

Amazon EMR 发行版标签	Zeppelin 沙盒版本	随 Zeppelin 沙盒安装的组件
emr-4.9.6	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client,

Amazon EMR 发行版标签	Zeppelin 沙盒版本	随 Zeppelin 沙盒安装的组件
		spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.9.5	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.9.4	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server



Amazon EMR 发行版标签	Zeppelin 沙盒版本	随 Zeppelin 沙盒安装的组件
emr-4.9.3	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.9.2	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.9.1	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发行版标签	Zeppelin 沙盒版本	随 Zeppelin 沙盒安装的组件
emr-4.8.5	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.8.4	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.8.3	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发行版标签	Zeppelin 沙盒版本	随 Zeppelin 沙盒安装的组件
emr-4.8.2	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.8.1	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.8.0	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发行版标签	Zeppelin 沙盒版本	随 Zeppelin 沙盒安装的组件
emr-4.7.4	0.5.6	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.7.3	0.5.6	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.7.2	0.5.6	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发行版标签	Zeppelin 沙盒版本	随 Zeppelin 沙盒安装的组件
emr-4.7.1	0.5.6	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.7.0	0.5.6	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.6.1	0.5.6	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发行版标签	Zeppelin 沙盒版本	随 Zeppelin 沙盒安装的组件
emr-4.6.0	0.5.6	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.5.0	0.5.6	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.4.0	0.5.6	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发行版标签	Zeppelin 沙盒版本	随 Zeppelin 沙盒安装的组件
emr-4.3.0	0.5.5	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resource manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.2.0	0.5.5	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resource manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.1.0	0.6.0-SNAPSHOT	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resource manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

## Zookeeper ( 沙盒版本 )

Zookeeper 从 Amazon EMR 发行版 4.6.0 开始作为沙盒应用程序提供。

## ZooKeeper 沙盒版本信息

Amazon EMR 发行版标签	Zookeeper 沙盒版本	随 ZooKeeper 沙盒安装的组件
emr-4.9.6	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server
emr-4.9.5	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server
emr-4.9.4	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server



Amazon EMR 发行版标签	Zookeeper 沙盒版本	随 ZooKeeper 沙盒安装的组件
emr-4.9.3	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-4.9.2	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-4.9.1	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server

Amazon EMR 发行版标签	Zookeeper 沙盒版本	随 ZooKeeper 沙盒安装的组件
emr-4.8.5	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server
emr-4.8.4	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server
emr-4.8.3	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server

Amazon EMR 发行版标签	Zookeeper 沙盒版本	随 ZooKeeper 沙盒安装的组件
emr-4.8.2	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-4.8.1	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-4.8.0	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server

Amazon EMR 发行版标签	Zookeeper 沙盒版本	随 ZooKeeper 沙盒安装的组件
emr-4.7.4	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-4.7.3	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-4.7.2	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server

Amazon EMR 发行版标签	Zookeeper 沙盒版本	随 ZooKeeper 沙盒安装的组件
emr-4.7.1	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-4.7.0	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-4.6.1	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server

Amazon EMR 发行版标签	Zookeeper 沙盒版本	随 ZooKeeper 沙盒安装的组件
emr-4.6.0	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server

## 使用 Amazon EMR 4.x 上的 Hive 的注意事项

本节介绍在使用 Amazon EMR 4.x 发行版上的 Hive 版本 1.0.0 时要注意的差异（与 Amazon EMR 5.x 发行版上的 Hive 2.x 相比）。

### 不支持 ACID 事务

使用 Amazon EMR 4.x 发行版时，4.x 发行版上的 Hive 不支持 Hive 数据存储在 Amazon S3 中的 ACID 事务。如果您尝试在 Amazon S3 中创建事务表，将出现异常。

### 对 Amazon S3 中的表的读写操作

Amazon EMR 4.x 发行版上的 Hive 可直接写入 Amazon S3，无需使用临时文件。这可以改善性能，但导致您无法在同一个 Hive 语句内读写 Amazon S3 中的同一个表。一种解决办法是在 HDFS 中创建并使用临时表。

以下示例显示了如何使用多 Hive 语句更新 Amazon S3 中的表。这些语句基于 Amazon S3 中一个名为 my\_s3\_table 的表在 HDFS 中创建一个名为 tmp 的临时表。之后将使用此临时表的内容更新 Amazon S3 中的表。

```
CREATE TEMPORARY TABLE tmp LIKE my_s3_table;
INSERT OVERWRITE TABLE tmp SELECT ....;
INSERT OVERWRITE TABLE my_s3_table SELECT * FROM tmp;
```

## Log4j 与 Log4j 2 的对比

Amazon EMR 4.x 发行版上的 Hive 使用 Log4j。从版本 5.0.0 开始，默认情况下使用 Log4j 2。这些版本可能需要不同的日志记录配置。有关更多信息，请参阅 [Apache Log4j 2](#)。

## MapReduce 是默认执行引擎

Amazon EMR 4.x 发行版上的 Hive 使用 MapReduce 作为默认执行引擎。从 Amazon EMR 版本 5.0.0 开始，Tez 为默认引擎，这将改善大多数工作流的性能。

## Hive 授权

Amazon EMR 4.x 发行版上的 Hive 对于 HDFS 支持 [Hive 授权](#)，但对于 EMRFS 和 Amazon S3 不支持此授权。默认情况下，Amazon EMR 集群在禁用授权的状态下运行。

## Amazon S3 中的 Hive 文件合并操作

如果 `hive.merge.mapfiles` 为 `true`，Amazon EMR 4.x 发行版上的 Hive 将在仅映射任务结束时合并小型文件。仅当任务的平均输出大小低于 `hive.merge.smallfiles.avgsize` 设置时，才会触发合并。如果最终输出路径位于 HDFS 中，那么 Amazon EMR Hive 的行为将完全相同。但是，如果输出路径位于 Amazon S3 中，将忽略 `hive.merge.smallfiles.avgsize` 参数。在那种情况下，如果 `hive.merge.mapfiles` 设置为 `true`，会始终触发合并任务。

## 使用 Amazon EMR 4.x 上的 Pig 的注意事项

Pig 版本 0.14.0 安装在使用 Amazon EMR 4.x 发行版创建的集群上。Pig 在 Amazon EMR 5.0.0 中已升级到版本 0.16.0。下面介绍了明显差异。

### 不同的默认执行引擎

Amazon EMR 4.x 发行版上的 Pig 版本 0.14.0 使用 MapReduce 作为默认执行引擎。Pig 0.16.0 及更高版本均使用 Apache Tez。您可在 `exectype=mapreduce` 配置分类中显式设置 `pig-properties` 以使用 MapReduce。

### 已删除的 Pig 用户定义函数 (UDF)

从 Pig 0.16.0 开始，删除了在 Amazon EMR 4.x 发行版上的 Pig 中可用的自定义 UDF。大多数 UDF 具有您可替换的等效函数。下表列出了已删除的 UDF 和等效函数。有关更多信息，请参阅 Apache Pig 网站上的 [内置函数](#)。

已删除的 UDF	等效函数
FORMAT_DT(dtformat, date)	GetHour(date)、GetMinute(date)、GetMonth(date)、GetSecond(date)、GetWeek(date)、GetYear(date)、GetDay(date)
EXTRACT(string, pattern)	REGEX_EXTRACT_ALL(string, pattern)
REPLACE(string, pattern, replacement)	REPLACE(string, pattern, replacement)
DATE_TIME()	ToDate()
DURATION(dt, dt2)	WeeksBetween(dt, dt2)、YearsBetween(dt, dt2)、SecondsBetween(dt, dt2)、MonthsBetween(dt, dt2)、MinutesBetween(dt, dt2)、HoursBetween(dt, dt2)
EXTRACT_DT(format, date)	GetHour(date)、GetMinute(date)、GetMonth(date)、GetSecond(date)、GetWeek(date)、GetYear(date)、GetDay(date)
OFFSET_DT(date, duration)	AddDuration(date, duration)、SubtractDuration(date, duration)
PERIOD(dt, dt2)	WeeksBetween(dt, dt2)、YearsBetween(dt, dt2)、SecondsBetween(dt, dt2)、MonthsBetween(dt, dt2)、MinutesBetween(dt, dt2)、HoursBetween(dt, dt2)
CAPITALIZE(string)	UCFIRST(string)
CONCAT_WITH()	CONCAT()
INDEX_OF()	INDEXOF()
LAST_INDEX_OF()	LAST_INDEXOF()
SPLIT_ON_REGEX()	STRSPLT()
UNCAPITALIZE()	LCFIRST()



以下 UDF 已被删除，没有等效函数：

FORMAT()、LOCAL\_DATE()、LOCAL\_TIME()、CENTER()、LEFT\_PAD()、REPEAT()、REPLACE\_ONCE

已停止使用 Grunt 命令

某些 Grunt 命令已从 Pig 0.16.0 开始停用。下表列出了 Pig 0.14.0 中的 Grunt 命令以及当前版本中的等效命令（如果适用）。

Pig 0.14.0 和等效的当前 Grunt 命令

Pig 0.14.0 Grunt 命令	0.16.0 及更高版本中的 Pig Grunt 命令
cat <non-hdfs-path>)	fs -cat <non-hdfs-path>;
cd <non-hdfs-path>;	无等效函数
ls <non-hdfs-path>;	fs -ls <non-hdfs-path>;
move <non-hdfs-path> <non-hdfs-path>;	fs -mv <non-hdfs-path> <non-hdfs-path>;
copy <non-hdfs-path> <non-hdfs-path>;	fs -cp <non-hdfs-path> <non-hdfs-path>;
copyToLocal <non-hdfs-path> <local-path>;	fs -copyToLocal <non-hdfs-path> <local-path>;
copyFromLocal <local-path> <non-hdfs-path>;	fs -copyFromLocal <local-path> <non-hdfs-path>;
mkdir <non-hdfs-path>;	fs -mkdir <non-hdfs-path>;
rm <non-hdfs-path>;	fs -rm -r -skipTrash <non-hdfs-path>;
rmf <non-hdfs-path>;	fs -rm -r -skipTrash <non-hdfs-path>;

针对非 HDFS 主目录删除的功能

Amazon EMR 4.x 发行版上的 Pig 0.14.0 具有两种机制，以允许无主目录的 hadoop 用户之外的用户运行 Pig 脚本。第一种机制是自动后备，将初始工作目录设置为根目录（如果主目录不存在）。第二种机制是 pig.initial.fs.name 属性，它允许您更改初始工作目录。

这两种机制从 Amazon EMR 版本 5.0.0 开始不可用，因此用户必须在 HDFS 上有一个主目录。这不适用于 hadoop 用户，因为在启动时会配置一个主目录。使用 Hadoop jar 步骤运行的脚本默认为由 Hadoop 用户运行，除非使用 command-runner.jar 显式指定了其它用户。

## 亚马逊EMR版本 4.9.6

### 4.9.6 应用程序版本

此版本支持以下应用程序：[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie-Sandbox](#)、[Phoenix](#)、[Pig](#)、[Presto-Sandbox](#)、[Spark](#)、[Sqoop-Sandbox](#)、[Tez](#)、[Zeppelin-Sandbox](#) 和 [ZooKeeper-Sandbox](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-4.9.6	emr-4.9.5	emr-4.9.4	emr-4.9.3
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	未跟踪	未跟踪	未跟踪	未跟踪
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2

	emr-4.9.6	emr-4.9.5	emr-4.9.4	emr-4.9.3
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.157.1	0.157.1	0.157.1	0.157.1
Spark	1.6.3	1.6.3	1.6.3	1.6.3
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6

	emr-4.9.6	emr-4.9.5	emr-4.9.4	emr-4.9.3
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.9	3.4.9	3.4.9

#### 4.9.6 发行说明

这是一个补丁版本，用于为向 Amazon S3 发出的请求添加 AWS 签名版本 4 身份验证。所有应用程序和组件都与之前的 Amazon EMR 版本相同。

##### Important

在此版本中，亚马逊仅EMR使用 AWS 签名版本 4 来验证向 Amazon S3 发出的请求。有关更多信息，请参阅[新功能](#)。

#### 4.9.6 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 `CommunityVersion-amzn-EmrVersion` 的发行版标注。`EmrVersion` 从 0 开始。例如，如果 `myapp-component` 以 2.2

版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
emr-ddb	4.3.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.2.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.3.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.4.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.17.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-2	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS用于存储区块的节点级服务。

组件	版本	描述
hadoop-hdfs-library	2.7.3-amzn-2	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httfs-server	2.7.3-amzn-2	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-2	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.3-amzn-2	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.2.2	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.2.2	为一个或多个HBase地区提供服务的服务。
hbase-client	1.2.2	HBase命令行客户端。
hbase-rest-server	1.2.2	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.2.2	为其提供 Thrift 端点的HBase服务。

组件	版本	描述
hcatalog-client	1.0.0-amzn-9	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	1.0.0-amzn-9	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	1.0.0-amzn-9	HTTP端点为其提供REST接口HCatalog。
hive-client	1.0.0-amzn-9	Hive 命令行客户端。
hive-metastore-server	1.0.0-amzn-9	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server	1.0.0-amzn-9	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.7.1-amzn-7	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.12.2	用于机器学习的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
oozie-client	4.2.0	Oozie 命令行客户端。
oozie-server	4.2.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.7.0--1.2 HBase	服务器和客户端的 phoenix 库

组件	版本	描述
phoenix-query-server	4.7.0--1.2 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.157.1	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.157.1	用于执行查询的各个部分的服务。
pig-client	0.14.0-amzn-0	Pig 命令行客户端。
spark-client	1.6.3	Spark 命令行客户端。
spark-history-server	1.6.3	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	1.6.3	的内存中执行引擎。YARN
spark-yarn-slave	1.6.3	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.6.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.9	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。



组件	版本	描述
zookeeper-client	3.4.9	ZooKeeper 命令行客户端。

#### 4.9.6 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

##### emr-4.9.6 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。

分类	描述
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j	更改 HCatalog W ebHCat 的 log4j.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j	更改 Hive 的 hive-exec-log 4j.properties 文件中的值。
hive-log4j	更改 Hive 的 hive-log4j.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。

分类	描述
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。

分类	描述
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。

分类	描述
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 4.9.5

### 4.9.5 应用程序版本

此版本支持以下应用程序：[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie-Sandbox](#)、[Phoenix](#)、[Pig](#)、[Presto-Sandbox](#)、[Spark](#)、[Sqoop-Sandbox](#)、[Tez](#)、[Zeppelin-Sandbox](#) 和 [ZooKeeper-Sandbox](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本 ( 如果适用 )。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

## 应用程序版本信息

	emr-4.9.5	emr-4.9.4	emr-4.9.3	emr-4.9.2
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	未跟踪	未跟踪	未跟踪	未跟踪
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-

	emr-4.9.5	emr-4.9.4	emr-4.9.3	emr-4.9.2
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.157.1	0.157.1	0.157.1	0.157.1
Spark	1.6.3	1.6.3	1.6.3	1.6.3
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.9	3.4.9	3.4.9

## 4.9.5 发行说明

以下发行说明包含有关亚马逊 4.9.5 EMR 版本的信息。更改与 4.9.4 有关。

首次发布日期：2018年 8 月 29 日

更改、增强功能和解决的问题

- HBase
  - 此版本解决了潜在的安全漏洞。

## 4.9.5 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
emr-ddb	4.3.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.2.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.3.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.4.0	针对 Amazon S3 优化的分布式复制应用程序。



组件	版本	描述
emrfs	2.17.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-2	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.3-amzn-2	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.3-amzn-2	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-2	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN用于管理单个节点上的容器的服务。

组件	版本	描述
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.3-amzn-2	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.2.2	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.2.2	为一个或多个HBase地区提供服务的服务。
hbase-client	1.2.2	HBase命令行客户端。
hbase-rest-server	1.2.2	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.2.2	为其提供 Thrift 端点的HBase服务。
hcatalog-client	1.0.0-amzn-9	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	1.0.0-amzn-9	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	1.0.0-amzn-9	HTTP端点为其提供REST接口HCatalog。
hive-client	1.0.0-amzn-9	Hive 命令行客户端。
hive-metastore-server	1.0.0-amzn-9	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。

组件	版本	描述
hive-server	1.0.0-amzn-9	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.7.1-amzn-7	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.12.2	用于机器学习的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
oozie-client	4.2.0	Oozie 命令行客户端。
oozie-server	4.2.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.7.0--1.2 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.7.0--1.2 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.157.1	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.157.1	用于执行查询的各个部分的服务。
pig-client	0.14.0-amzn-0	Pig 命令行客户端。
spark-client	1.6.3	Spark 命令行客户端。
spark-history-server	1.6.3	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	1.6.3	的内存中执行引擎。YARN

组件	版本	描述
spark-yarn-slave	1.6.3	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.6.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.9	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.9	ZooKeeper 命令行客户端。

#### 4.9.5 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

##### emr-4.9.5 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。

分类	描述
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j	更改 HCatalog W ebHCat 的 log4j.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-env	更改 Hive 环境中的值。

分类	描述
hive-exec-log4j	更改 Hive 的 hive-exec-log 4j.properties 文件中的值。
hive-log4j	更改 Hive 的 hive-log4j.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。

分类	描述
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。

分类	描述
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。



## 亚马逊EMR版本 4.9.4

### 4.9.4 应用程序版本

此版本支持以下应用程序：[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie-Sandbox](#)、[Phoenix](#)、[Pig](#)、[Presto-Sandbox](#)、[Spark](#)、[Sqoop-Sandbox](#)、[Tez](#)、[Zeppelin-Sandbox](#) 和 [ZooKeeper-Sandbox](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-4.9.4	emr-4.9.3	emr-4.9.2	emr-4.9.1
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	未跟踪	未跟踪	未跟踪	未跟踪
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2

	emr-4.9.4	emr-4.9.3	emr-4.9.2	emr-4.9.1
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.157.1	0.157.1	0.157.1	0.157.1
Spark	1.6.3	1.6.3	1.6.3	1.6.3
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6

	emr-4.9.4	emr-4.9.3	emr-4.9.2	emr-4.9.1
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.9	3.4.9	3.4.9

#### 4.9.4 发行说明

以下发行说明包含有关亚马逊 4.9.4 EMR 版本的信息。更改与 4.9.3 有关。

首次发布日期：2018 年 3 月 29 日

更改、增强功能和解决的问题

- 更新了亚马逊 Linux 版的亚马逊 defaultAmazon Linux AMI 内核EMR以修复潜在漏洞。

#### 4.9.4 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 emr 或 aws 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
emr-ddb	4.3.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.2.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.3.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.4.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.17.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-2	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.3-amzn-2	HDFS命令行客户端和库

组件	版本	描述
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.3-amzn-2	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-2	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce 用于运行 MapReduce 应用程序的执行引擎。
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.3-amzn-2	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.2.2	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.2.2	为一个或多个HBase地区提供服务的服务。
hbase-client	1.2.2	HBase命令行客户端。
hbase-rest-server	1.2.2	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.2.2	为其提供 Thrift 端点的HBase服务。
hcatalog-client	1.0.0-amzn-9	用于操作 hcatalog-server 的“hcat”命令行客户端。

组件	版本	描述
hcatalog-server	1.0.0-amzn-9	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	1.0.0-amzn-9	HTTP端点为其提供REST接口HCatalog。
hive-client	1.0.0-amzn-9	Hive 命令行客户端。
hive-metastore-server	1.0.0-amzn-9	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server	1.0.0-amzn-9	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.7.1-amzn-7	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.12.2	用于机器学习的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
oozie-client	4.2.0	Oozie 命令行客户端。
oozie-server	4.2.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.7.0--1.2 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.7.0--1.2 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API

组件	版本	描述
presto-coordinator	0.157.1	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.157.1	用于执行查询的各个部分的服务。
pig-client	0.14.0-amzn-0	Pig 命令行客户端。
spark-client	1.6.3	Spark 命令行客户端。
spark-history-server	1.6.3	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	1.6.3	的内存中执行引擎。YARN
spark-yarn-slave	1.6.3	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.6.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.9	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.9	ZooKeeper 命令行客户端。

## 4.9.4 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

### emr-4.9.4 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。



分类	描述
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j	更改 HCatalog W ebHCat 的 log4j.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j	更改 Hive 的 hive-exec-log 4j.properties 文件中的值。
hive-log4j	更改 Hive 的 hive-log4j.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。

分类	描述
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。

分类	描述
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。

分类	描述
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 4.9.3

### 4.9.3 应用程序版本

此版本支持以下应用程序：[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie-Sandbox](#)、[Phoenix](#)、[Pig](#)、[Presto-Sandbox](#)、[Spark](#)、[Sqoop-Sandbox](#)、[Tez](#)、[Zeppelin-Sandbox](#) 和 [ZooKeeper-Sandbox](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-4.9.3	emr-4.9.2	emr-4.9.1	emr-4.8.5
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	未跟踪	未跟踪	未跟踪	未跟踪

	emr-4.9.3	emr-4.9.2	emr-4.9.1	emr-4.8.5
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0

	emr-4.9.3	emr-4.9.2	emr-4.9.1	emr-4.8.5
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.157.1	0.157.1	0.157.1	0.157.1
Spark	1.6.3	1.6.3	1.6.3	1.6.3
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.9	3.4.9	3.4.9

### 4.9.3 发行说明

以下发行说明包含 Amazon EMR 4.9.3 版本的信息。更改与 Amazon EMR 4.9.2 版本有关。

首次发布日期：2018 年 1 月 22 日

## 更改、增强功能和解决的问题

- 更新了亚马逊版 Linux 的亚马逊 L defaultAmazon inux AMI 内核，EMR以解决与推测执行相关的漏洞 ( CVE-2017-5715、-2017-5753 和 -2017 CVE -5754 )。CVE有关更多信息，请参阅 <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>。

### 4.9.3 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
emr-ddb	4.3.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.2.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.3.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.4.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.17.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。

组件	版本	描述
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-2	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.3-amzn-2	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.3-amzn-2	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-2	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN用于分配和管理群集资源和分布式应用程序的服务。



组件	版本	描述
hadoop-yarn-timeline-server	2.7.3-amzn-2	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.2.2	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.2.2	为一个或多个HBase地区提供服务的服务。
hbase-client	1.2.2	HBase命令行客户端。
hbase-rest-server	1.2.2	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.2.2	为其提供 Thrift 端点的HBase服务。
hcatalog-client	1.0.0-amzn-9	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	1.0.0-amzn-9	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	1.0.0-amzn-9	HTTP端点为其提供REST接口HCatalog。
hive-client	1.0.0-amzn-9	Hive 命令行客户端。
hive-metastore-server	1.0.0-amzn-9	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server	1.0.0-amzn-9	用于将 Hive 查询作为 Web 请求接受的服务。

组件	版本	描述
hue-server	3.7.1-amzn-7	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.12.2	用于机器学习的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
oozie-client	4.2.0	Oozie 命令行客户端。
oozie-server	4.2.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.7.0--1.2 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.7.0--1.2 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.157.1	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.157.1	用于执行查询的各个部分的服务。
pig-client	0.14.0-amzn-0	Pig 命令行客户端。
spark-client	1.6.3	Spark 命令行客户端。
spark-history-server	1.6.3	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	1.6.3	的内存中执行引擎。YARN
spark-yarn-slave	1.6.3	YARN从属服务器需要的 Apache Spark 库。

组件	版本	描述
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.6.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.9	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.9	ZooKeeper 命令行客户端。

### 4.9.3 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-4.9.3 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置

分类	描述
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j	更改 HCatalog W ebHCat 的 log4j.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j	更改 Hive 的 hive-exec-log 4j.properties 文件中的值。
hive-log4j	更改 Hive 的 hive-log4j.properties 文件中的值。

分类	描述
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。

分类	描述
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。

分类	描述
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 4.9.2

### 4.9.2 应用程序版本

此版本支持以下应用程序：[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie-Sandbox](#)、[Phoenix](#)、[Pig](#)、[Presto-Sandbox](#)、[Spark](#)、[Sqoop-Sandbox](#)、[Tez](#)、[Zeppelin-Sandbox](#) 和 [ZooKeeper-Sandbox](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本 ( 如果适用 )。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-4.9.2	emr-4.9.1	emr-4.8.5	emr-4.8.4
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	未跟踪	未跟踪	未跟踪	未跟踪
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-



	emr-4.9.2	emr-4.9.1	emr-4.8.5	emr-4.8.4
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.157.1	0.157.1	0.157.1	0.157.1
Spark	1.6.3	1.6.3	1.6.3	1.6.3
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-

	emr-4.9.2	emr-4.9.1	emr-4.8.5	emr-4.8.4
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.9	3.4.9	3.4.9

## 4.9.2 发行说明

以下发行说明包含 Amazon EMR 4.9.2 版本的信息。更改与 Amazon EMR 4.9.1 版本有关。

发布日期：2017 年 7 月 13 日

此版本略微进行了一些改动、错误修复和增强。

## 4.9.2 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
emr-ddb	4.3.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。

组件	版本	描述
emr-goodies	2.2.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.3.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.4.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.17.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-2	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.3-amzn-2	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.3-amzn-2	HTTPHDFS操作端点。

组件	版本	描述
hadoop-kms-server	2.7.3-amzn-2	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.3-amzn-2	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.2.2	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.2.2	为一个或多个HBase地区提供服务的服务。
hbase-client	1.2.2	HBase命令行客户端。
hbase-rest-server	1.2.2	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.2.2	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	1.0.0-amzn-9	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	1.0.0-amzn-9	提供服务HCatalog，为分布式应用程序提供表和存储管理层。

组件	版本	描述
hcatalog-webhcat-server	1.0.0-amzn-9	HTTP端点为其提供REST接口HCatalog。
hive-client	1.0.0-amzn-9	Hive 命令行客户端。
hive-metastore-server	1.0.0-amzn-9	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server	1.0.0-amzn-9	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.7.1-amzn-7	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.12.2	用于机器学习的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
oozie-client	4.2.0	Oozie 命令行客户端。
oozie-server	4.2.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.7.0--1.2 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.7.0--1.2 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.157.1	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.157.1	用于执行查询的各个部分的服务。

组件	版本	描述
pig-client	0.14.0-amzn-0	Pig 命令行客户端。
spark-client	1.6.3	Spark 命令行客户端。
spark-history-server	1.6.3	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	1.6.3	的内存中执行引擎。YARN
spark-yarn-slave	1.6.3	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.6.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.9	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.9	ZooKeeper 命令行客户端。

## 4.9.2 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

## emr-4.9.2 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。

分类	描述
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j	更改 HCatalog W ebHCat 的 log4j.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j	更改 Hive 的 hive-exec-log 4j.properties 文件中的值。
hive-log4j	更改 Hive 的 hive-log4j.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。



分类	描述
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。

分类	描述
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。

分类	描述
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 4.9.1

### 4.9.1 应用程序版本

此版本支持以下应用程序：[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie-Sandbox](#)、[Phoenix](#)、[Pig](#)、[Presto-Sandbox](#)、[Spark](#)、[Sqoop-Sandbox](#)、[Tez](#)、[Zeppelin-Sandbox](#) 和 [ZooKeeper-Sandbox](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-4.9.1	emr-4.8.5	emr-4.8.4	emr-4.8.3
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	未跟踪	未跟踪	未跟踪	未跟踪
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-

	emr-4.9.1	emr-4.8.5	emr-4.8.4	emr-4.8.3
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.157.1	0.157.1	0.157.1	0.157.1

	emr-4.9.1	emr-4.8.5	emr-4.8.4	emr-4.8.3
Spark	1.6.3	1.6.3	1.6.3	1.6.3
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.9	3.4.9	3.4.9

## 4.9.1 发行说明

以下发行说明包含 Amazon EMR 4.9.1 版本的信息。更改与 Amazon EMR 4.8.4 版本有关。

发布日期：2017 年 4 月 10 日

早期版本中已解决的已知问题

- [HIVE-9976 和- 10106 的向后移植 HIVE](#)
- 修复了大量节点（大于 2,000）和容器（大于 5,000）会导致 out-of-memory 错误的问题，例如："Exception in thread main java.lang.OutOfMemoryError"。YARN

## 更改和增强功能

- 亚马逊EMR版本现在基于亚马逊 Linux 2017.03。有关更多信息，请参阅 <https://aws.amazon.com/amazon-linux-ami/2017.03-release-notes/>。
- 从亚马逊EMR基本的 Linux 镜像中移除了 Python 2.6。如果需要，您可以手动安装 Python 2.6。

### 4.9.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>emr-ddb</code>	4.2.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.2.0	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.3.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
<code>emr-s3-dist-cp</code>	2.4.0	针对 Amazon S3 优化的分布式复制应用程序。
<code>emrfs</code>	2.15.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。

组件	版本	描述
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-2	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.3-amzn-2	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.3-amzn-2	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-2	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN用于分配和管理群集资源和分布式应用程序的服务。

组件	版本	描述
hadoop-yarn-timeline-server	2.7.3-amzn-2	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.2.2	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.2.2	为一个或多个HBase地区提供服务的服务。
hbase-client	1.2.2	HBase命令行客户端。
hbase-rest-server	1.2.2	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.2.2	为其提供 Thrift 端点的HBase服务。
hcatalog-client	1.0.0-amzn-9	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	1.0.0-amzn-9	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	1.0.0-amzn-9	HTTP端点为其提供REST接口HCatalog。
hive-client	1.0.0-amzn-9	Hive 命令行客户端。
hive-metastore-server	1.0.0-amzn-9	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server	1.0.0-amzn-9	用于将 Hive 查询作为 Web 请求接受的服务。



组件	版本	描述
hue-server	3.7.1-amzn-7	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.12.2	用于机器学习的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
oozie-client	4.2.0	Oozie 命令行客户端。
oozie-server	4.2.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.7.0--1.2 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.7.0--1.2 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限 API
presto-coordinator	0.157.1	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.157.1	用于执行查询的各个部分的服务。
pig-client	0.14.0-amzn-0	Pig 命令行客户端。
spark-client	1.6.3	Spark 命令行客户端。
spark-history-server	1.6.3	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	1.6.3	的内存中执行引擎。YARN
spark-yarn-slave	1.6.3	YARN从属服务器需要的 Apache Spark 库。

组件	版本	描述
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.6.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.9	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.9	ZooKeeper 命令行客户端。

#### 4.9.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

##### emr-4.9.1 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置

分类	描述
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j	更改 HCatalog W ebHCat 的 log4j.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j	更改 Hive 的 hive-exec-log 4j.properties 文件中的值。
hive-log4j	更改 Hive 的 hive-log4j.properties 文件中的值。

分类	描述
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。

分类	描述
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。

分类	描述
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 4.8.5

### 4.8.5 应用程序版本

此版本支持以下应用程序：[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie-Sandbox](#)、[Phoenix](#)、[Pig](#)、[Presto-Sandbox](#)、[Spark](#)、[Sqoop-Sandbox](#)、[Tez](#)、[Zeppelin-Sandbox](#) 和 [ZooKeeper-Sandbox](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本 ( 如果适用 )。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-4.8.5	emr-4.8.4	emr-4.8.3	emr-4.8.2
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	未跟踪	未跟踪	未跟踪	未跟踪
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-

	emr-4.8.5	emr-4.8.4	emr-4.8.3	emr-4.8.2
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.157.1	0.157.1	0.157.1	0.152.3
Spark	1.6.3	1.6.3	1.6.3	1.6.2
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-



	emr-4.8.5	emr-4.8.4	emr-4.8.3	emr-4.8.2
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.9	3.4.9	3.4.8

### 4.8.5 发行说明

这是一个补丁版本，用于为向 Amazon S3 发出的请求添加 AWS 签名版本 4 身份验证。所有应用程序和组件都与之前的 Amazon EMR 版本相同。

#### Important

在此版本中，亚马逊仅EMR使用 AWS 签名版本 4 来验证向 Amazon S3 发出的请求。有关更多信息，请参阅[新功能](#)。

### 4.8.5 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
emr-ddb	4.2.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.2.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.2.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.4.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	emrfs	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.3-amzn-1	HDFS命令行客户端和库

组件	版本	描述
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.3-amzn-1	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-1	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-1	MapReduce 用于运行 MapReduce 应用程序的执行引擎。
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.3-amzn-1	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.2.2	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.2.2	为一个或多个HBase地区提供服务的服务。
hbase-client	1.2.2	HBase命令行客户端。
hbase-rest-server	1.2.2	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.2.2	为其提供 Thrift 端点的HBase服务。
hcatalog-client	1.0.0-amzn-8	用于操作 hcatalog-server 的“hcat”命令行客户端。

组件	版本	描述
hcatalog-server	1.0.0-amzn-8	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	1.0.0-amzn-8	HTTP端点为其提供REST接口HCatalog。
hive-client	1.0.0-amzn-8	Hive 命令行客户端。
hive-metastore-server	1.0.0-amzn-8	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server	1.0.0-amzn-8	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.7.1-amzn-7	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.12.2	用于机器学习的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
oozie-client	4.2.0	Oozie 命令行客户端。
oozie-server	4.2.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.7.0--1.2 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.7.0--1.2 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API

组件	版本	描述
presto-coordinator	0.157.1	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.157.1	用于执行查询的各个部分的服务。
pig-client	0.14.0-amzn-0	Pig 命令行客户端。
spark-client	1.6.3	Spark 命令行客户端。
spark-history-server	1.6.3	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	1.6.3	的内存中执行引擎。YARN
spark-yarn-slave	1.6.3	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.6.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.9	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.9	ZooKeeper 命令行客户端。

## 4.8.5 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

### emr-4.8.5 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。

分类	描述
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j	更改 HCatalog W ebHCat 的 log4j.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j	更改 Hive 的 hive-exec-log 4j.properties 文件中的值。
hive-log4j	更改 Hive 的 hive-log4j.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。

分类	描述
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。



分类	描述
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。

分类	描述
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 4.8.4

### 4.8.4 应用程序版本

此版本支持以下应用程序：[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie-Sandbox](#)、[Phoenix](#)、[Pig](#)、[Presto-Sandbox](#)、[Spark](#)、[Sqoop-Sandbox](#)、[Tez](#)、[Zeppelin-Sandbox](#) 和 [ZooKeeper-Sandbox](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-4.8.4	emr-4.8.3	emr-4.8.2	emr-4.8.1
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	未跟踪	未跟踪	未跟踪	未跟踪

	emr-4.8.4	emr-4.8.3	emr-4.8.2	emr-4.8.1
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0

	emr-4.8.4	emr-4.8.3	emr-4.8.2	emr-4.8.1
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.157.1	0.157.1	0.152.3	0.151
Spark	1.6.3	1.6.3	1.6.2	1.6.2
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.9	3.4.8	3.4.8

#### 4.8.4 发行说明

以下发行说明包含 Amazon EMR 4.8.4 版本的信息。更改与亚马逊 EMR 4.8.3 版本有关。

发布日期：2017 年 2 月 7 日

此版本略微进行了一些改动、错误修复和增强。

#### 4.8.4 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>emr-ddb</code>	4.2.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.2.0	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.2.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
<code>emr-s3-dist-cp</code>	2.4.0	针对 Amazon S3 优化的分布式复制应用程序。
<code>emrfs</code>	<code>emrfs</code>	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
<code>ganglia-monitor</code>	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
<code>ganglia-metadata-collector</code>	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。

组件	版本	描述
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.3-amzn-1	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.3-amzn-1	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-1	基于 Hadoop 的加密密钥管理服务器。 KeyProvider API
hadoop-mapred	2.7.3-amzn-1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.3-amzn-1	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.2.2	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.2.2	为一个或多个HBase地区提供服务的服务。

组件	版本	描述
hbase-client	1.2.2	HBase命令行客户端。
hbase-rest-server	1.2.2	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.2.2	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	1.0.0-amzn-8	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	1.0.0-amzn-8	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	1.0.0-amzn-8	HTTP端点为其提供REST接口HCatalog。
hive-client	1.0.0-amzn-8	Hive 命令行客户端。
hive-metastore-server	1.0.0-amzn-8	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server	1.0.0-amzn-8	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.7.1-amzn-7	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.12.2	用于机器学习的库。
mysql-server	5.5.54+	我的SQL数据库服务器。
oozie-client	4.2.0	Oozie 命令行客户端。

组件	版本	描述
oozie-server	4.2.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.7.0--1.2 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.7.0--1.2 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.157.1	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.157.1	用于执行查询的各个部分的服务。
pig-client	0.14.0-amzn-0	Pig 命令行客户端。
spark-client	1.6.3	Spark 命令行客户端。
spark-history-server	1.6.3	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	1.6.3	的内存中执行引擎。YARN
spark-yarn-slave	1.6.3	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.25+	Apache HTTP 服务器。
zeppelin-server	0.6.1	支持交互式数据分析的基于 Web 的笔记本电脑。



组件	版本	描述
zookeeper-server	3.4.9	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.9	ZooKeeper 命令行客户端。

#### 4.8.4 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

##### emr-4.8.4 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。

分类	描述
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j	更改 HCatalog W ebHCat 的 log4j.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j	更改 Hive 的 hive-exec-log 4j.properties 文件中的值。
hive-log4j	更改 Hive 的 hive-log4j.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。

分类	描述
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。

分类	描述
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。

分类	描述
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 4.8.3

### 4.8.3 应用程序版本

此版本支持以下应用程序：[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie-Sandbox](#)、[Phoenix](#)、[Pig](#)、[Presto-Sandbox](#)、[Spark](#)、[Sqoop-Sandbox](#)、[Tez](#)、[Zeppelin-Sandbox](#) 和 [ZooKeeper-Sandbox](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

## 应用程序版本信息

	emr-4.8.3	emr-4.8.2	emr-4.8.1	emr-4.8.0
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	未跟踪	未跟踪	未跟踪	未跟踪
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.2	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-

	emr-4.8.3	emr-4.8.2	emr-4.8.1	emr-4.8.0
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.157.1	0.152.3	0.151	0.151
Spark	1.6.3	1.6.2	1.6.2	1.6.2
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.8	3.4.8	3.4.8

### 4.8.3 发行说明

以下发行说明包含 Amazon EMR 4.8.3 版本的信息。更改与亚马逊 EMR 4.8.2 版本有关。

发布日期：2016 年 12 月 29 日

#### 升级

- 已升级到 Presto 0.157.1。有关更多信息，请参阅 Presto 文档中的 [Presto 发布说明](#)。
- 已升级到 Spark 1.6.3。有关更多信息，请参阅 Apache Spark 文档中的 [Spark 发布说明](#)。
- 已升级到 ZooKeeper 3.4.9。有关更多信息，请参阅 Apache ZooKeeper 文档中的 [ZooKeeper 发行说明](#)。

#### 更改和增强功能

- 在亚马逊 4.8.3 及更高EMR版本中增加了对亚马逊 EC2 m4.16xlarge 实例类型的支持，不包括 5.0.0、5.0.3 和 5.2.0。
- 亚马逊EMR版本现在基于亚马逊 Linux 2016.09。有关更多信息，请参阅 <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>。

#### 早期版本中已解决的已知问题

- 修复了 Hadoop 中的一个问题，即由于在大型集群中复制和删除同一文件之间存在竞争，ReplicationMonitor 线程可能会长时间卡住。
- 修复了任务状态未成功更新时 ControlledJob # toString 失败并出现空指针异常 (NPE) 的问题。

### 4.8.3 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 emr 或 aws 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 myapp-component 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。



组件	版本	描述
emr-ddb	4.2.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.2.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.2.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.4.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.13.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.3-amzn-1	HDFS命令行客户端和库

组件	版本	描述
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.3-amzn-1	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-1	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-1	MapReduce 用于运行 MapReduce 应用程序的执行引擎。
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.3-amzn-1	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.2.2	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.2.2	为一个或多个HBase地区提供服务的服务。
hbase-client	1.2.2	HBase命令行客户端。
hbase-rest-server	1.2.2	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.2.2	为其提供 Thrift 端点的HBase服务。
hcatalog-client	1.0.0-amzn-8	用于操作 hcatalog-server 的“hcat”命令行客户端。

组件	版本	描述
hcatalog-server	1.0.0-amzn-8	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	1.0.0-amzn-8	HTTP端点为其提供REST接口HCatalog。
hive-client	1.0.0-amzn-8	Hive 命令行客户端。
hive-metastore-server	1.0.0-amzn-8	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server	1.0.0-amzn-8	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.7.1-amzn-7	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.12.2	用于机器学习的库。
mysql-server	5.5.52	我的SQL数据库服务器。
oozie-client	4.2.0	Oozie 命令行客户端。
oozie-server	4.2.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.7.0--1.2 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.7.0--1.2 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API

组件	版本	描述
presto-coordinator	0.157.1	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.157.1	用于执行查询的各个部分的服务。
pig-client	0.14.0-amzn-0	Pig 命令行客户端。
spark-client	1.6.3	Spark 命令行客户端。
spark-history-server	1.6.3	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	1.6.3	的内存中执行引擎。YARN
spark-yarn-slave	1.6.3	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.23	Apache HTTP 服务器。
zeppelin-server	0.6.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.9	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.9	ZooKeeper 命令行客户端。

### 4.8.3 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-4.8.3 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。

分类	描述
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j	更改 HCatalog W ebHCat 的 log4j.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j	更改 Hive 的 hive-exec-log 4j.properties 文件中的值。
hive-log4j	更改 Hive 的 hive-log4j.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。

分类	描述
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。

分类	描述
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。



分类	描述
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 4.8.2

### 4.8.2 应用程序版本

此版本支持以下应用程序：[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie-Sandbox](#)、[Phoenix](#)、[Pig](#)、[Presto-Sandbox](#)、[Spark](#)、[Sqoop-Sandbox](#)、[Tez](#)、[Zeppelin-Sandbox](#) 和 [ZooKeeper-Sandbox](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-4.8.2	emr-4.8.1	emr-4.8.0	emr-4.7.4
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	未跟踪	未跟踪	未跟踪	未跟踪

	emr-4.8.2	emr-4.8.1	emr-4.8.0	emr-4.7.4
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.1
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.2	2.7.2	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0

	emr-4.8.2	emr-4.8.1	emr-4.8.0	emr-4.7.4
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.152.3	0.151	0.151	0.148
Spark	1.6.2	1.6.2	1.6.2	1.6.2
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.3
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.5.6
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	3.4.8	3.4.8	3.4.8

## 4.8.2 发行说明

以下发行说明包含 Amazon EMR 4.8.2 版本的信息。更改与 Amazon EMR 4.8.0 版本有关。

发布日期：2016 年 10 月 24 日

### 升级

- 已升级到 Hadoop 2.7.3

- 已升级到 Presto 0.152.3，它包括对 Presto Web 界面的支持。可使用端口 8889 访问 Presto 协调器上的 Presto Web 界面。有关 Presto Web 界面的更多信息，请参阅 Presto 文档中的 [Web 界面](#)。
- 亚马逊EMR版本现在基于亚马逊 Linux 2016.09。有关更多信息，请参阅 <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>。

## 4.8.2 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>emr-ddb</code>	4.1.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.1.0	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.2.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
<code>emr-s3-dist-cp</code>	2.4.0	针对 Amazon S3 优化的分布式复制应用程序。
<code>emrfs</code>	2.10.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。

组件	版本	描述
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.3-amzn-0	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.3-amzn-0	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.3-amzn-0	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.3-amzn-0	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.3-amzn-0	HTTPHDFS操作端点。
hadoop-kms-server	2.7.3-amzn-0	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.3-amzn-0	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.3-amzn-0	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.3-amzn-0	YARN用于分配和管理群集资源和分布式应用程序的服务。

组件	版本	描述
hadoop-yarn-timeline-server	2.7.3-amzn-0	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.2.2	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.2.2	为一个或多个HBase地区提供服务的服务。
hbase-client	1.2.2	HBase命令行客户端。
hbase-rest-server	1.2.2	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.2.2	为其提供 Thrift 端点的HBase服务。
hcatalog-client	1.0.0-amzn-7	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	1.0.0-amzn-7	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	1.0.0-amzn-7	HTTP端点为其提供REST接口HCatalog。
hive-client	1.0.0-amzn-7	Hive 命令行客户端。
hive-metastore-server	1.0.0-amzn-7	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server	1.0.0-amzn-7	用于将 Hive 查询作为 Web 请求接受的服务。

组件	版本	描述
hue-server	3.7.1-amzn-7	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.12.2	用于机器学习的库。
mysql-server	5.5.52	我的SQL数据库服务器。
oozie-client	4.2.0	Oozie 命令行客户端。
oozie-server	4.2.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.7.0--1.2 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.7.0--1.2 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.152.3	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.152.3	用于执行查询的各个部分的服务。
pig-client	0.14.0-amzn-0	Pig 命令行客户端。
spark-client	1.6.2	Spark 命令行客户端。
spark-history-server	1.6.2	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	1.6.2	的内存中执行引擎。YARN
spark-yarn-slave	1.6.2	YARN从属服务器需要的 Apache Spark 库。

组件	版本	描述
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.23	Apache HTTP 服务器。
zeppelin-server	0.6.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.8	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.8	ZooKeeper 命令行客户端。

## 4.8.2 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

### emr-4.8.2 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置



分类	描述
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j	更改 HCatalog W ebHCat 的 log4j.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j	更改 Hive 的 hive-exec-log 4j.properties 文件中的值。
hive-log4j	更改 Hive 的 hive-log4j.properties 文件中的值。

分类	描述
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。

分类	描述
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。

分类	描述
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 4.8.1

### 4.8.1 应用程序版本

此版本支持以下应用程序：[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie-Sandbox](#)、[Phoenix](#)、[Pig](#)、[Presto-Sandbox](#)、[Spark](#)、[Sqoop-Sandbox](#)、[Tez](#)、[Zeppelin-Sandbox](#) 和 [ZooKeeper-Sandbox](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本 ( 如果适用 )。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-4.8.1	emr-4.8.0	emr-4.7.4	emr-4.7.3
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	未跟踪	未跟踪	未跟踪	未跟踪
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.1	1.2.1
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.2	2.7.2	2.7.2	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-

	emr-4.8.1	emr-4.8.0	emr-4.7.4	emr-4.7.3
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.151	0.151	0.148	0.148
Spark	1.6.2	1.6.2	1.6.2	1.6.2
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.3	0.8.3
Trino (PrestoSQL)	-	-	-	-

	emr-4.8.1	emr-4.8.0	emr-4.7.4	emr-4.7.3
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.5.6	0.5.6
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	3.4.8	3.4.8	3.4.8

## 4.8.1 发行说明

### 4.8.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>emr-ddb</code>	3.2.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.1.0	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.2.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。

组件	版本	描述
emr-s3-dist-cp	2.4.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.9.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.2-amzn-4	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.2-amzn-4	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.2-amzn-4	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.2-amzn-4	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.2-amzn-4	HTTPHDFS操作端点。
hadoop-kms-server	2.7.2-amzn-4	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.2-amzn-4	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。



组件	版本	描述
hadoop-yarn-nodemanager	2.7.2-amzn-4	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.2-amzn-4	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.2-amzn-4	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.2.2	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.2.2	为一个或多个HBase地区提供服务的服务。
hbase-client	1.2.2	HBase命令行客户端。
hbase-rest-server	1.2.2	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.2.2	为其提供 Thrift 端点的HBase服务。
hcatalog-client	1.0.0-amzn-7	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	1.0.0-amzn-7	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	1.0.0-amzn-7	HTTP端点为其提供REST接口HCatalog。
hive-client	1.0.0-amzn-7	Hive 命令行客户端。

组件	版本	描述
hive-metastore-server	1.0.0-amzn-7	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的 SQL 元数据。
hive-server	1.0.0-amzn-7	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.7.1-amzn-7	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.12.2	用于机器学习的库。
mysql-server	5.5.51	我的 SQL 数据库服务器。
oozie-client	4.2.0	Oozie 命令行客户端。
oozie-server	4.2.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.7.0--1.2 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.7.0--1.2 HBase	一款轻量级服务器，提供 JDBC 对 Avatica 的访问权限以及协议缓冲区和 JSON 格式访问权限 API
presto-coordinator	0.151	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.151	用于执行查询的各个部分的服务。
pig-client	0.14.0-amzn-0	Pig 命令行客户端。
spark-client	1.6.2	Spark 命令行客户端。

组件	版本	描述
spark-history-server	1.6.2	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	1.6.2	的内存中执行引擎。YARN
spark-yarn-slave	1.6.2	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.23	Apache HTTP 服务器。
zeppelin-server	0.6.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.8	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.8	ZooKeeper 命令行客户端。

#### 4.8.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

##### emr-4.8.1 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。

分类	描述
emrfs-site	更改EMRFS设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j	更改 HCatalog W ebHCat 的 log4j.properties 中的值。

分类	描述
hcatalog-webhcat-site	更改 HCatalog WebHCat 的 webhcat-site.xml 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j	更改 Hive 的 hive-exec-log4j.properties 文件中的值。
hive-log4j	更改 Hive 的 hive-log4j.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。

分类	描述
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。

分类	描述
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 4.8.0

### 4.8.0 应用程序版本

此版本支持以下应用程序：[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie-Sandbox](#)、[Phoenix](#)、[Pig](#)、[Presto-Sandbox](#)、[Spark](#)、[Sqoop-Sandbox](#)、[Tez](#)、[Zeppelin-Sandbox](#) 和 [ZooKeeper-Sandbox](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-4.8.0	emr-4.7.4	emr-4.7.3	emr-4.7.2
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	未跟踪	未跟踪	未跟踪	未跟踪
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.1	1.2.1	1.2.1



	emr-4.8.0	emr-4.7.4	emr-4.7.3	emr-4.7.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.2	2.7.2	2.7.2	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.151	0.148	0.148	0.148
Spark	1.6.2	1.6.2	1.6.2	1.6.2
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6

	emr-4.8.0	emr-4.7.4	emr-4.7.3	emr-4.7.2
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.3	0.8.3	0.8.3
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.5.6	0.5.6	0.5.6
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	3.4.8	3.4.8	3.4.8

## 4.8.0 版本说明

以下发行说明包含 Amazon EMR 4.8.0 版本的信息。更改与亚马逊 EMR 4.7.2 版本有关。

发布日期：2016 年 9 月 7 日

### 升级

- 已升级到 HBase 1.2.2
- 已升级到 Presto-Sandbox 0.151
- 已升级到 Tez 0.8.4
- 已升级到 Zeppelin-Sandbox 0.6.1

### 更改和增强功能

- YARN修复了 ApplicationMaster 尝试清理因实例已终止而不再存在的容器的问题。
- 在 Oozie 示例中更正了 Hive-server2 URL 的 Hive2 动作。
- 添加了对其它 Presto 目录的支持。
- [向后移植的补丁：HIVE-8948、-12679、-13405、HIVE-3116、-12689 HIVE PHOENIX HADOOP](#)

- 添加了对安全配置的支持，这使您可以更轻松地创建和应用加密选项。有关更多信息，请参阅[数据加密](#)。

#### 4.8.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>emr-ddb</code>	3.2.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.1.0	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.2.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
<code>emr-s3-dist-cp</code>	2.4.0	针对 Amazon S3 优化的分布式复制应用程序。
<code>emrfs</code>	2.9.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
<code>ganglia-monitor</code>	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。

组件	版本	描述
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.2-amzn-4	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.2-amzn-4	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.2-amzn-4	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.2-amzn-4	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.2-amzn-4	HTTPHDFS操作端点。
hadoop-kms-server	2.7.2-amzn-4	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.2-amzn-4	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.2-amzn-4	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.2-amzn-4	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.2-amzn-4	用于检索YARN应用程序的当前和历史信息的服务。

组件	版本	描述
hbase-hmaster	1.2.2	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.2.2	为一个或多个HBase地区提供服务的服务器。
hbase-client	1.2.2	HBase命令行客户端。
hbase-rest-server	1.2.2	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.2.2	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	1.0.0-amzn-7	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	1.0.0-amzn-7	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	1.0.0-amzn-7	HTTP端点为其提供REST接口HCatalog。
hive-client	1.0.0-amzn-7	Hive 命令行客户端。
hive-metastore-server	1.0.0-amzn-7	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server	1.0.0-amzn-7	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.7.1-amzn-7	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序

组件	版本	描述
mahout-client	0.12.2	用于机器学习的库。
mysql-server	5.5.51	我的SQL数据库服务器。
oozie-client	4.2.0	Oozie 命令行客户端。
oozie-server	4.2.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.7.0--1.2 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.7.0--1.2 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.151	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.151	用于执行查询的各个部分的服务。
pig-client	0.14.0-amzn-0	Pig 命令行客户端。
spark-client	1.6.2	Spark 命令行客户端。
spark-history-server	1.6.2	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	1.6.2	的内存中执行引擎。YARN
spark-yarn-slave	1.6.2	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。

组件	版本	描述
tez-on-yarn	0.8.4	tez YARN 应用程序和库。
webserver	2.4.23	Apache HTTP 服务器。
zeppelin-server	0.6.1	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.8	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.8	ZooKeeper 命令行客户端。

#### 4.8.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

##### emr-4.8.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置

分类	描述
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j	更改 HCatalog W ebHCat 的 log4j.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j	更改 Hive 的 hive-exec-log 4j.properties 文件中的值。
hive-log4j	更改 Hive 的 hive-log4j.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值



分类	描述
hiveserver2-site	更改 Hive Server2 的 hiveserver2-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。

分类	描述
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-blackhole	更改 Presto 的 blackhole.properties 文件中的值。
presto-connector-cassandra	更改 Presto 的 cassandra.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
presto-connector-jmx	更改 Presto 的 jmx.properties 文件中的值。
presto-connector-kafka	更改 Presto 的 kafka.properties 文件中的值。
presto-connector-localfile	更改 Presto 的 localfile.properties 文件中的值。
presto-connector-mongodb	更改 Presto 的 mongodb.properties 文件中的值。
presto-connector-mysql	更改 Presto 的 mysql.properties 文件中的值。
presto-connector-postgresql	更改 Presto 的 postgresql.properties 文件中的值。
presto-connector-raptor	更改 Presto 的 raptor.properties 文件中的值。
presto-connector-redis	更改 Presto 的 redis.properties 文件中的值。
presto-connector-tpch	更改 Presto 的 tpch.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。

分类	描述
spark-env	更改 Spark 环境中的值。
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 4.7.4

### 4.7.4 应用程序版本

此版本支持以下应用程序：[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie-Sandbox](#)、[Phoenix](#)、[Pig](#)、[Presto-Sandbox](#)、[Spark](#)、[Sqoop-Sandbox](#)、[Tez](#)、[Zeppelin-Sandbox](#) 和 [ZooKeeper-Sandbox](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-4.7.4	emr-4.7.3	emr-4.7.2	emr-4.7.1
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	未跟踪	未跟踪	未跟踪	未跟踪
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.1	1.2.1	1.2.1	1.2.1
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.2	2.7.2	2.7.2	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-

	emr-4.7.4	emr-4.7.3	emr-4.7.2	emr-4.7.1
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.148	0.148	0.148	0.147
Spark	1.6.2	1.6.2	1.6.2	1.6.1
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.3	0.8.3	0.8.3	0.8.3
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.6	0.5.6	0.5.6	0.5.6

	emr-4.7.4	emr-4.7.3	emr-4.7.2	emr-4.7.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	3.4.8	3.4.8	3.4.8

#### 4.7.4 发行说明

这是一个补丁版本，用于为向 Amazon S3 发出的请求添加 AWS 签名版本 4 身份验证。所有应用程序和组件都与之前的 Amazon EMR 版本相同。

##### Important

在此版本中，亚马逊仅EMR使用 AWS 签名版本 4 来验证向 Amazon S3 发出的请求。有关更多信息，请参阅[新功能](#)。

#### 4.7.4 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 emr 或 aws 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
emr-ddb	3.2.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.1.0	适用于 Hadoop 生态系统的方便易用的库。

组件	版本	描述
emr-kinesis	3.2.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.4.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.8.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.2-amzn-3	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.2-amzn-3	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.2-amzn-3	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.2-amzn-3	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.2-amzn-3	HTTPHDFS操作端点。
hadoop-kms-server	2.7.2-amzn-3	基于 Hadoop 的加密密钥管理服务。 KeyProvider API

组件	版本	描述
hadoop-mapred	2.7.2-amzn-3	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.2-amzn-3	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.2-amzn-3	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.2-amzn-3	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.2.1	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.2.1	为一个或多个HBase地区提供服务的服务。
hbase-client	1.2.1	HBase命令行客户端。
hbase-rest-server	1.2.1	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.2.1	为其提供 Thrift 端点的HBase服务。
hcatalog-client	1.0.0-amzn-6	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	1.0.0-amzn-6	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	1.0.0-amzn-6	HTTP端点为其提供REST接口HCatalog。
hive-client	1.0.0-amzn-6	Hive 命令行客户端。



组件	版本	描述
hive-metastore-server	1.0.0-amzn-6	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的 SQL 元数据。
hive-server	1.0.0-amzn-6	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.7.1-amzn-7	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.12.2	用于机器学习的库。
mysql-server	5.5.46	我的 SQL 数据库服务器。
oozie-client	4.2.0	Oozie 命令行客户端。
oozie-server	4.2.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.7.0--1.2 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.7.0--1.2 HBase	一款轻量级服务器，提供 JDBC 对 Avatica 的访问权限以及协议缓冲区和 JSON 格式访问权限 API
presto-coordinator	0.148	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.148	用于执行查询的各个部分的服务。
pig-client	0.14.0-amzn-0	Pig 命令行客户端。
spark-client	1.6.2	Spark 命令行客户端。

组件	版本	描述
spark-history-server	1.6.2	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	1.6.2	的内存中执行引擎。YARN
spark-yarn-slave	1.6.2	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.3	tez YARN 应用程序和库。
webserver	2.4.23	Apache HTTP 服务器。
zeppelin-server	0.5.6-incubating	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.8	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.8	ZooKeeper 命令行客户端。

#### 4.7.4 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

##### emr-4.7.4 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。

分类	描述
emrfs-site	更改EMRFS设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j	更改 HCatalog W ebHCat 的 log4j.properties 中的值。

分类	描述
hcatalog-webhcat-site	更改 HCatalog WebHCat 的 webhcat-site.xml 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j	更改 Hive 的 hive-exec-log4j.properties 文件中的值。
hive-log4j	更改 Hive 的 hive-log4j.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。

分类	描述
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。

分类	描述
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 4.7.3

### 4.7.3 应用程序版本

此版本支持以下应用程序：[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie-Sandbox](#)、[Phoenix](#)、[Pig](#)、[Presto-Sandbox](#)、[Spark](#)、[Sqoop-Sandbox](#)、[Tez](#)、[Zeppelin-Sandbox](#) 和 [ZooKeeper-Sandbox](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-4.7.3	emr-4.7.2	emr-4.7.1	emr-4.7.0
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.75
Python	未跟踪	未跟踪	未跟踪	未跟踪

	emr-4.7.3	emr-4.7.2	emr-4.7.1	emr-4.7.0
Scala	未跟踪	未跟踪	未跟踪	未跟踪
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.1	1.2.1	1.2.1	1.2.1
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.2	2.7.2	2.7.2	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.0	0.12.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0

	emr-4.7.3	emr-4.7.2	emr-4.7.1	emr-4.7.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.148	0.148	0.147	0.147
Spark	1.6.2	1.6.2	1.6.1	1.6.1
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.3	0.8.3	0.8.3	0.8.3
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.6	0.5.6	0.5.6	0.5.6
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	3.4.8	3.4.8	3.4.8

## 4.7.3 发行说明

### 4.7.3 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 emr 或 aws 开头。Amazon



最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
emr-ddb	3.2.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.1.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.2.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.4.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.8.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。

组件	版本	描述
hadoop-client	2.7.2-amzn-3	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.2-amzn-3	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.2-amzn-3	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.2-amzn-3	HDFS用于跟踪文件名和区块位置的服务。
hadoop-https-server	2.7.2-amzn-3	HTTPHDFS操作端点。
hadoop-kms-server	2.7.2-amzn-3	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.2-amzn-3	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.2-amzn-3	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.2-amzn-3	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.2-amzn-3	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.2.1	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.2.1	为一个或多个HBase地区提供服务的服务。
hbase-client	1.2.1	HBase命令行客户端。

组件	版本	描述
hbase-rest-server	1.2.1	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.2.1	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	1.0.0-amzn-6	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	1.0.0-amzn-6	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	1.0.0-amzn-6	HTTP端点为其提供REST接口HCatalog。
hive-client	1.0.0-amzn-6	Hive 命令行客户端。
hive-metastore-server	1.0.0-amzn-6	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server	1.0.0-amzn-6	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.7.1-amzn-7	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.12.2	用于机器学习的库。
mysql-server	5.5.46	我的SQL数据库服务器。
oozie-client	4.2.0	Oozie 命令行客户端。
oozie-server	4.2.0	用于接受 Oozie 工作流请求的服务。

组件	版本	描述
phoenix-library	4.7.0--1.2 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.7.0--1.2 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.148	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.148	用于执行查询的各个部分的服务。
pig-client	0.14.0-amzn-0	Pig 命令行客户端。
spark-client	1.6.2	Spark 命令行客户端。
spark-history-server	1.6.2	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	1.6.2	的内存中执行引擎。YARN
spark-yarn-slave	1.6.2	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.3	tez YARN 应用程序和库。
webserver	2.4.23	Apache HTTP 服务器。
zeppelin-server	0.5.6-incubating	支持交互式数据分析的基于 Web 的笔记本电脑。

组件	版本	描述
zookeeper-server	3.4.8	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.8	ZooKeeper 命令行客户端。

### 4.7.3 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-4.7.3 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。

分类	描述
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j	更改 HCatalog W ebHCat 的 log4j.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j	更改 Hive 的 hive-exec-log 4j.properties 文件中的值。
hive-log4j	更改 Hive 的 hive-log4j.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。

分类	描述
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。

分类	描述
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 ZooKeeper的 zoo.cfg 文件中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 4.7.2

### 4.7.2 应用程序版本

此版本支持以下应用程序：[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie-Sandbox](#)、[Phoenix](#)、[Pig](#)、[Presto-Sandbox](#)、[Spark](#)、[Sqoop-Sandbox](#)、[Tez](#)、[Zeppelin-Sandbox](#) 和 [ZooKeeper-Sandbox](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。



要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-4.7.2	emr-4.7.1	emr-4.7.0	emr-4.6.1
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.75	1.10.27
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	未跟踪	未跟踪	未跟踪	未跟踪
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.1	1.2.1	1.2.1	1.2.0
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.2	2.7.2	2.7.2	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-

	emr-4.7.2	emr-4.7.1	emr-4.7.0	emr-4.6.1
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.0	0.12.0	0.11.1
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	-
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.148	0.147	0.147	0.143
Spark	1.6.2	1.6.1	1.6.1	1.6.1
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.3	0.8.3	0.8.3	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.6	0.5.6	0.5.6	0.5.6

	emr-4.7.2	emr-4.7.1	emr-4.7.0	emr-4.6.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	3.4.8	3.4.8	3.4.8

## 4.7.2 发行说明

以下发行说明包含亚马逊 EMR 4.7.2 的信息。

发布日期：2016 年 7 月 15 日

### 功能

- 已升级到 Mahout 0.12.2
- 已升级到 Presto 0.148
- 已升级到 Spark 1.6.2
- 现在，您可以使用 a URI 作为参数来创建，以 `AWSCredentialsProvider` 供 EMRFS 使用。有关更多信息，请参阅 [AWSCredentialsProvider 为创建 EMRFS](#)。
- EMRFS 现在允许用户使用中的属性为其一致视图元数据配置自定义 DynamoDB 终端节点。`fs.s3.consistent.dynamodb.endpoint emrfs-site.xml`
- 在 `/usr/bin` 中添加了一个名为 `spark-example` 的脚本，它将包装 `/usr/lib/spark/spark/bin/run-example`，因此您可以直接运行示例。例如，要运行 Spark 发行版附带的 SparkPi 示例，可以 `spark-example SparkPi 100` 从命令行运行，也可以 `command-runner.jar` 将其用作其中的一个步骤 API。

### 早期版本中已解决的已知问题

- 修复了 Oozie 在安装 Spark 后拥有的 `spark-assembly.jar` 未位于正确位置 (这导致使用 Oozie 启动 Spark 应用程序失败) 的问题。
- 修复了容器中基于 Spark Log4J 的日志记录的问题。YARN

## 4.7.2 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>emr-ddb</code>	3.2.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.1.0	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.2.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
<code>emr-s3-dist-cp</code>	2.4.0	针对 Amazon S3 优化的分布式复制应用程序。
<code>emrfs</code>	2.8.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
<code>ganglia-monitor</code>	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
<code>ganglia-metadata-collector</code>	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。

组件	版本	描述
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.2-amzn-3	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.2-amzn-3	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.2-amzn-3	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.2-amzn-3	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.2-amzn-3	HTTPHDFS操作端点。
hadoop-kms-server	2.7.2-amzn-3	基于 Hadoop 的加密密钥管理服务器。 KeyProvider API
hadoop-mapred	2.7.2-amzn-3	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.2-amzn-3	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.2-amzn-3	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.2-amzn-3	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.2.1	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.2.1	为一个或多个HBase地区提供服务的服务。

组件	版本	描述
hbase-client	1.2.1	HBase命令行客户端。
hbase-rest-server	1.2.1	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.2.1	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	1.0.0-amzn-6	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	1.0.0-amzn-6	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	1.0.0-amzn-6	HTTP端点为其提供REST接口HCatalog。
hive-client	1.0.0-amzn-6	Hive 命令行客户端。
hive-metastore-server	1.0.0-amzn-6	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server	1.0.0-amzn-6	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.7.1-amzn-7	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.12.2	用于机器学习的库。
mysql-server	5.5.46	我的SQL数据库服务器。
oozie-client	4.2.0	Oozie 命令行客户端。

组件	版本	描述
oozie-server	4.2.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.7.0--1.2 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.7.0--1.2 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.148	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.148	用于执行查询的各个部分的服务。
pig-client	0.14.0-amzn-0	Pig 命令行客户端。
spark-client	1.6.2	Spark 命令行客户端。
spark-history-server	1.6.2	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	1.6.2	的内存中执行引擎。YARN
spark-yarn-slave	1.6.2	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.3	tez YARN 应用程序和库。
webserver	2.4.23	Apache HTTP 服务器。
zeppelin-server	0.5.6-incubating	支持交互式数据分析的基于 Web 的笔记本电脑。

组件	版本	描述
zookeeper-server	3.4.8	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.8	ZooKeeper 命令行客户端。

## 4.7.2 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

### emr-4.7.2 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hadoop-ssl-server	更改 hadoop ssl 服务器配置
hadoop-ssl-client	更改 hadoop ssl 客户端配置
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。



分类	描述
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j	更改 HCatalog W ebHCat 的 log4j.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j	更改 Hive 的 hive-exec-log 4j.properties 文件中的值。
hive-log4j	更改 Hive 的 hive-log4j.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。

分类	描述
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。

分类	描述
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 4.7.1

### 4.7.1 应用程序版本

此版本支持以下应用程序：[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie-Sandbox](#)、[Phoenix](#)、[Pig](#)、[Presto-Sandbox](#)、[Spark](#)、[Sqoop-Sandbox](#)、[Tez](#)、[Zeppelin-Sandbox](#) 和 [ZooKeeper-Sandbox](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-4.7.1	emr-4.7.0	emr-4.6.1	emr-4.6.0
适用于 Java 的 AWS SDK	1.10.75	1.10.75	1.10.27	1.10.27
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	未跟踪	未跟踪	未跟踪	未跟踪
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.1	1.2.1	1.2.0	1.2.0
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.2	2.7.2	2.7.2	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-

	emr-4.7.1	emr-4.7.0	emr-4.6.1	emr-4.6.0
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.0	0.12.0	0.11.1	0.11.1
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	-	-
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.147	0.147	0.143	0.143
Spark	1.6.1	1.6.1	1.6.1	1.6.1
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.3	0.8.3	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.6	0.5.6	0.5.6	0.5.6

	emr-4.7.1	emr-4.7.0	emr-4.6.1	emr-4.6.0
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	3.4.8	3.4.8	3.4.8

### 4.7.1 发行说明

以下发行说明包含亚马逊 EMR 4.7.1 的信息。

发布日期：2016 年 6 月 10 日

早期版本中已解决的已知问题

- 修复了延长VPC使用私有子网启动的集群的启动时间的问题。该错误仅影响在 Amazon EMR 4.7.0 版本中启动的集群。
- 修复了在 Amazon EMR 4.7.0 版本中启动EMR的集群在亚马逊中未正确处理文件列表的问题。

### 4.7.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
emr-ddb	3.1.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。

组件	版本	描述
emr-goodies	2.0.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.2.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.4.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.7.1	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.2-amzn-2	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.2-amzn-2	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.2-amzn-2	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.2-amzn-2	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.2-amzn-2	HTTPHDFS操作端点。

组件	版本	描述
hadoop-kms-server	2.7.2-amzn-2	基于 Hadoop 的加密密钥管理服务。KeyProvider API
hadoop-mapred	2.7.2-amzn-2	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.2-amzn-2	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.2-amzn-2	YARN用于分配和管理群集资源和分布式应用程序的服务。
hadoop-yarn-timeline-server	2.7.2-amzn-2	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.2.1	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.2.1	为一个或多个HBase地区提供服务的服务。
hbase-client	1.2.1	HBase命令行客户端。
hbase-rest-server	1.2.1	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.2.1	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	1.0.0-amzn-5	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	1.0.0-amzn-5	提供服务HCatalog，为分布式应用程序提供表和存储管理层。



组件	版本	描述
hcatalog-webhcat-server	1.0.0-amzn-5	HTTP端点为其提供REST接口HCatalog。
hive-client	1.0.0-amzn-5	Hive 命令行客户端。
hive-metastore-server	1.0.0-amzn-5	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server	1.0.0-amzn-5	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.7.1-amzn-7	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.12.0	用于机器学习的库。
mysql-server	5.5.46	我的SQL数据库服务器。
oozie-client	4.2.0	Oozie 命令行客户端。
oozie-server	4.2.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.7.0--1.2 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.7.0--1.2 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.147	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.147	用于执行查询的各个部分的服务。

组件	版本	描述
pig-client	0.14.0-amzn-0	Pig 命令行客户端。
spark-client	1.6.1	Spark 命令行客户端。
spark-history-server	1.6.1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	1.6.1	的内存中执行引擎。YARN
spark-yarn-slave	1.6.1	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.3	tez YARN 应用程序和库。
webserver	2.4.18	Apache HTTP 服务器。
zeppelin-server	0.5.6-incubating	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.8	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.8	ZooKeeper 命令行客户端。

#### 4.7.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

## emr-4.7.1 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j	更改 HCatalog W ebHCat 的 log4j.properties 中的值。

分类	描述
hcatalog-webhcat-site	更改 HCatalog WebHCat 的 webhcat-site.xml 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j	更改 Hive 的 hive-exec-log4j.properties 文件中的值。
hive-log4j	更改 Hive 的 hive-log4j.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。

分类	描述
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。

分类	描述
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 4.7.0

### 4.7.0 应用程序版本

此版本支持以下应用程序：[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie-Sandbox](#)、[Phoenix](#)、[Pig](#)、[Presto-Sandbox](#)、[Spark](#)、[Sqoop-Sandbox](#)、[Tez](#)、[Zeppelin-Sandbox](#) 和 [ZooKeeper-Sandbox](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-4.7.0	emr-4.6.1	emr-4.6.0	emr-4.5.0
适用于 Java 的 AWS SDK	1.10.75	1.10.27	1.10.27	1.10.27
Python	未跟踪	未跟踪	未跟踪	未跟踪

	emr-4.7.0	emr-4.6.1	emr-4.6.0	emr-4.5.0
Scala	未跟踪	未跟踪	未跟踪	未跟踪
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.1	1.2.0	1.2.0	-
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.2	2.7.2	2.7.2	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.0	0.11.1	0.11.1	0.11.1
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0

	emr-4.7.0	emr-4.6.1	emr-4.6.0	emr-4.5.0
Phoenix	4.7.0	-	-	-
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.147	0.143	0.143	0.140
Spark	1.6.1	1.6.1	1.6.1	1.6.1
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.3	-	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.6	0.5.6	0.5.6	0.5.6
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	3.4.8	3.4.8	-

## 4.7.0 版本说明

### Important

亚马逊 EMR 4.7.0 已被弃用。改用亚马逊 EMR 4.7.1 或更高版本。



发布日期：2016年6月2日

## 功能

- 已添加 Apache Phoenix 4.7.0
- 已添加 Apache Tez 0.8.3
- 已升级到 HBase 1.2.1
- 已升级到 Mahout 0.12.0
- 已升级到 Presto 0.147
- 已将升级 AWS SDK for Java 到 1.10.75
- 已从 `mapreduce.cluster.local.dir` 中的 `mapred-site.xml` 属性中删除最终标志以允许用户以本地模式运行 Pig。
- 集群上提供亚马逊 Redshift JDBC 驱动程序

亚马逊 Redshift JDBC 驱动程序现已包含在 `/usr/share/aws/redshift/jdbc` 中。`/usr/share/aws/redshift/jdbc/RedshiftJDBC41.jar` 是兼容 JDBC 4.1 的亚马逊 Redshift 驱动程序，也是兼容 4. JDBC 0 的亚马逊 Redshift 驱动程序 `/usr/share/aws/redshift/jdbc/RedshiftJDBC4.jar`。有关更多信息，请参阅 Amazon Redshift 管理 [指南中的配置 JDBC 连接](#)。

- Java 8

除了 Presto 之外，Open JDK 1.7 是所有应用程序的 JDK 默认设置。但是，Open JDK 1.7 和 1.8 都已安装。有关如何为应用程序设置 `JAVA_HOME` 的信息，请参阅 [配置应用程序以使用 Java 8](#)。

## 早期版本中已解决的已知问题

- 修复了 emr-4.6.0 中显著影响亚马逊 EMR 吞吐量优化 HDD (st1) EBS 卷性能的内核问题。
- 修复了如果在未选择 Hadoop 作为应用程序的情况下指定了任何 HDFS 加密区域，则集群会失败的问题。
- 将默认 HDFS 写入策略从更改 RoundRobin 为 AvailableSpaceVolumeChoosingPolicy。某些卷在 RoundRobin 配置中未得到正确利用，这导致核心节点出现故障且不可靠 HDFS。
- 修复了在 EMRFSCLI 为一致视图创建默认 DynamoDB 元数据表时会导致异常的问题。
- 修复了在 EMRFS 多部分重命名和复制操作期间可能发生的死锁问题。
- 修复了 EMRFS 导致默认 CopyPart 大小为 5 MB 的问题。默认值现已相应地设置为 128 MB。
- 修复了与 Zeppelin upstart 配置有关的问题，此问题可能会阻止您停止服务。

- 修复了 Spark 和 Zeppelin 的问题，该问题使你无法使用该s3a://URI方案，/usr/lib/hadoop/hadoop-aws.jar因为它们各自的职业路径中没有正确加载。
- 向后移植 [HUE-2484](#)。
- 向后移植了 Hue 3.9.0 中的[提交](#)（不JIRA存在），以修复浏览器示例的问题。HBase
- 向后移植 [HIVE-9073](#)。

## 4.7.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
emr-ddb	3.1.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.0.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.2.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.4.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.7.1	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。

组件	版本	描述
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.2-amzn-2	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.2-amzn-2	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.2-amzn-2	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.2-amzn-2	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.2-amzn-2	HTTPHDFS操作端点。
hadoop-kms-server	2.7.2-amzn-2	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.2-amzn-2	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.2-amzn-2	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.2-amzn-2	YARN用于分配和管理群集资源和分布式应用程序的服务。

组件	版本	描述
hadoop-yarn-timeline-server	2.7.2-amzn-2	用于检索YARN应用程序的当前和历史信息的服务。
hbase-hmaster	1.2.1	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.2.1	为一个或多个HBase地区提供服务的服务。
hbase-client	1.2.1	HBase命令行客户端。
hbase-rest-server	1.2.1	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.2.1	为其提供 Thrift 端点的HBase服务。
hcatalog-client	1.0.0-amzn-5	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	1.0.0-amzn-5	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	1.0.0-amzn-5	HTTP端点为其提供REST接口HCatalog。
hive-client	1.0.0-amzn-5	Hive 命令行客户端。
hive-metastore-server	1.0.0-amzn-5	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server	1.0.0-amzn-5	用于将 Hive 查询作为 Web 请求接受的服务。

组件	版本	描述
hue-server	3.7.1-amzn-7	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.12.0	用于机器学习的库。
mysql-server	5.5.46	我的SQL数据库服务器。
oozie-client	4.2.0	Oozie 命令行客户端。
oozie-server	4.2.0	用于接受 Oozie 工作流请求的服务。
phoenix-library	4.7.0--1.2 HBase	服务器和客户端的 phoenix 库
phoenix-query-server	4.7.0--1.2 HBase	一款轻量级服务器，提供JDBC对 Avatica 的访问权限以及协议缓冲区和JSON格式访问权限API
presto-coordinator	0.147	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.147	用于执行查询的各个部分的服务。
pig-client	0.14.0-amzn-0	Pig 命令行客户端。
spark-client	1.6.1	Spark 命令行客户端。
spark-history-server	1.6.1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	1.6.1	的内存中执行引擎。YARN
spark-yarn-slave	1.6.1	YARN从属服务器需要的 Apache Spark 库。

组件	版本	描述
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
tez-on-yarn	0.8.3	tez YARN 应用程序和库。
webserver	2.4.18	Apache HTTP 服务器。
zeppelin-server	0.5.6-incubating	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.8	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.8	ZooKeeper 命令行客户端。

#### 4.7.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

##### emr-4.7.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hbase-env	更改环境HBase中的值。

分类	描述
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j	更改 HCatalog W ebHCat 的 log4j.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j	更改 Hive 的 hive-exec-log 4j.properties 文件中的值。
hive-log4j	更改 Hive 的 hive-log4j.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值

分类	描述
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
phoenix-hbase-metrics	更改 Phoenix 的 hadoop-metrics2-hbase.properties 文件中的值。
phoenix-hbase-site	更改 Phoenix 的 hbase-site.xml 文件中的值。
phoenix-log4j	更改 Phoenix 的 log4j.properties 文件中的值。
phoenix-metrics	更改 Phoenix 的 hadoop-metrics2-phoenix.properties 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。



分类	描述
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
tez-site	更改 Tez 的 tez-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 4.6.1

### 4.6.1 应用程序版本

此发行版支持以下应用程序：[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie-Sandbox](#)、[Pig](#)、[Presto-Sandbox](#)、[Spark](#)、[Sqoop-Sandbox](#)、[Zeppelin-Sandbox](#) 和 [ZooKeeper-Sandbox](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-4.6.1	emr-4.6.0	emr-4.5.0	emr-4.4.0
适用于 Java 的 AWS SDK	1.10.27	1.10.27	1.10.27	1.10.27
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	未跟踪	未跟踪	未跟踪	未跟踪
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.0	1.2.0	-	-

	emr-4.6.1	emr-4.6.0	emr-4.5.0	emr-4.4.0
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.2	2.7.2	2.7.2	2.7.1
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.11.1	0.11.1	0.11.1	0.11.1
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	-	-	-	-
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.143	0.143	0.140	0.136
Spark	1.6.1	1.6.1	1.6.1	1.6.0
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6

	emr-4.6.1	emr-4.6.0	emr-4.5.0	emr-4.4.0
TensorFlow	-	-	-	-
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.6	0.5.6	0.5.6	0.5.6
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	3.4.8	-	-

## 4.6.1 发行说明

### 4.6.1 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
emr-ddb	3.0.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。

组件	版本	描述
emr-goodies	2.0.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.1.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.3.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.6.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.2-amzn-1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.2-amzn-1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.2-amzn-1	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.2-amzn-1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.2-amzn-1	HTTPHDFS操作端点。

组件	版本	描述
hadoop-kms-server	2.7.2-amzn-1	基于 Hadoop 的加密密钥管理服务器。 KeyProvider API
hadoop-mapred	2.7.2-amzn-1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.2-amzn-1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.2-amzn-1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hbase-hmaster	1.2.0	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.2.0	为一个或多个HBase地区提供服务的服务。
hbase-client	1.2.0	HBase命令行客户端。
hbase-rest-server	1.2.0	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.2.0	为其提供 Thrift 端点的HBase 服务。
hcatalog-client	1.0.0-amzn-4	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	1.0.0-amzn-4	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	1.0.0-amzn-4	HTTP端点为其提供REST接口HCatalog。
hive-client	1.0.0-amzn-4	Hive 命令行客户端。

组件	版本	描述
hive-metastore-server	1.0.0-amzn-4	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的 SQL 元数据。
hive-server	1.0.0-amzn-4	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.7.1-amzn-6	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.11.1	用于机器学习的库。
mysql-server	5.5	我的 SQL 数据库服务器。
oozie-client	4.2.0	Oozie 命令行客户端。
oozie-server	4.2.0	用于接受 Oozie 工作流请求的服务。
presto-coordinator	0.143	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.143	用于执行查询的各个部分的服务。
pig-client	0.14.0-amzn-0	Pig 命令行客户端。
spark-client	1.6.1	Spark 命令行客户端。
spark-history-server	1.6.1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	1.6.1	的内存中执行引擎。YARN

组件	版本	描述
spark-yarn-slave	1.6.1	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
webserver	2.4	Apache HTTP 服务器。
zeppelin-server	0.5.6-incubating	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.8	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.8	ZooKeeper 命令行客户端。

### 4.6.1 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-4.6.1 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。



分类	描述
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j	更改 HCatalog W ebHCat 的 log4j.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j	更改 Hive 的 hive-exec-log 4j.properties 文件中的值。
hive-log4j	更改 Hive 的 hive-log4j.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值

分类	描述
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。

分类	描述
spark-env	更改 Spark 环境中的值。
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 4.6.0

### 4.6.0 应用程序版本

此发行版支持以下应用程序：[Ganglia](#)、[HBase](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie-Sandbox](#)、[Pig](#)、[Presto-Sandbox](#)、[Spark](#)、[Sqoop-Sandbox](#)、[Zeppelin-Sandbox](#) 和 [ZooKeeper-Sandbox](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)

- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

## 应用程序版本信息

	emr-4.6.0	emr-4.5.0	emr-4.4.0	emr-4.3.0
适用于 Java 的 AWS SDK	1.10.27	1.10.27	1.10.27	1.10.27
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	未跟踪	未跟踪	未跟踪	未跟踪
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.0	-	-	-
HCatalog	1.0.0	1.0.0	1.0.0	-
Hadoop	2.7.2	2.7.2	2.7.1	2.7.1
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-

	emr-4.6.0	emr-4.5.0	emr-4.4.0	emr-4.3.0
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.11.1	0.11.1	0.11.1	0.11.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	-	-	-	-
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.143	0.140	0.136	0.130
Spark	1.6.1	1.6.1	1.6.0	1.6.0
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	-
TensorFlow	-	-	-	-
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.6	0.5.6	0.5.6	0.5.5
ZooKeeper	-	-	-	-

	emr-4.6.0	emr-4.5.0	emr-4.4.0	emr-4.3.0
ZooKeeper-Sandbox	3.4.8	-	-	-

## 4.6.0 版本说明

以下发行说明包含 Amazon EMR 4.6.0 版本的信息。

- 已添加 HBase 1.2.0
- 已添加 Zookeeper-Sandbox 3.4.8
- 已升级到 Presto-Sandbox 0.143
- 亚马逊EMR版本现在基于亚马逊 Linux 2016.03.0。有关更多信息，请参阅 <https://aws.amazon.com/amazon-linux-ami/2016.03-release-notes/>。
- 影响吞吐量优化 HDD (st1) EBS 卷类型的问题

Linux 内核版本 4.2 及更高版本中的问题会严重影响吞吐量优化 HDD (st1) EBS 卷的EMR性能。此版本 (emr-4.6.0) 使用内核版本 4.4.5，因此会受到影响。因此，如果您想使用 st1 卷，我们建议不要使用 emr-4.6.0。EBS您可以将 emr-4.5.0 或之前的亚马逊EMR版本与 st1 一起使用，而不会产生任何影响。此外，我们将随将来版本一起提供修复程序。

- Python 默认值

现在，默认情况下已安装 Python 3.4，但 Python 2.7 将保留系统默认值。您可以使用引导操作将 Python 3.4 配置为系统默认值；您可以使用配置API在spark-env分类/usr/bin/python3.4中将 PYSARK \_ e PYTHON xport 设置为以影响所 PySpark使用的 Python 版本。

- Java 8

除了 Presto 之外，Open JDK 1.7 是所有应用程序的JDK默认设置。但是，Open JDK 1.7 和 1.8 都已安装。有关如何为应用程序设置 JAVA\_HOME 的信息，请参阅[配置应用程序以使用 Java 8](#)。

### 早期版本中已解决的已知问题

- 修复了应用程序预置有时会因生成的密码导致随机失败的问题。
- 之前，mysqld 已安装在所有节点上。现在，它仅安装在主实例上，而且仅在所选应用程序将 mysql-server 作为组件包含时安装。目前，以下应用程序包括该mysql-server组件：、Hive HCatalog、Hue、Presto-Sandbox 和 Sqoop-Sandbox。

- 从默认值 32 更改 `yarn.scheduler.maximum-allocation-vcores` 为 80，这修复了 emr-4.0 中引入的一个问题，该问题主要发生在 Spark 中，该问题主要发生在 Spark 中，核心实例类型是 YARN vcore 设置为 32 的少数大型实例类型之一；即 `c4.8xlarge`、`cc2.8xlarge`、`hs1.8xlarge`、`i2.8xlarge`、`i2.8xlarge`、`i2.8xlarge`、`m2.4xlarge`、`d2.8xlarge`，或 `m4.10xlarge` 受到了此问题的影响。`maximizeResourceAllocation`
- `s3-dist-cp` 现在用于所有 EMRFS Amazon S3 提名，不再暂存到临时目录。HDFS
- 修复了与针对客户端加密分段上载的异常处理有关的问题。
- 添加了允许用户更改 Amazon S3 存储类的选项。默认情况下，此设置为 STANDARD。`emrfs-site` 配置分类设置为 `fs.s3.storageClass`，可能的值为 STANDARD、STANDARD\_IA 和 REDUCED\_REDUNDANCY。有关存储类的更多信息，请参阅《Amazon Simple Storage Service 用户指南》中的[存储类](#)。

#### 4.6.0 组件版本

下面列出了 Amazon 随此版本 EMR 安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的 EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新 EMR 版本中的大数据应用程序包通常是社区中的最新版本。我们会 EMR 尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本 EMR 不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>emr-ddb</code>	3.0.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.0.0	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.1.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。

组件	版本	描述
emr-s3-dist-cp	2.3.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.6.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.2-amzn-1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.2-amzn-1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.2-amzn-1	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.2-amzn-1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.2-amzn-1	HTTPHDFS操作端点。
hadoop-kms-server	2.7.2-amzn-1	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.2-amzn-1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。



组件	版本	描述
hadoop-yarn-nodemanager	2.7.2-amzn-1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.2-amzn-1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hbase-hmaster	1.2.0	为负责协调区域和执行管理命令的HBase集群提供服务。
hbase-region-server	1.2.0	为一个或多个HBase地区提供服务的服务。
hbase-client	1.2.0	HBase命令行客户端。
hbase-rest-server	1.2.0	为提供RESTfulHTTP终端节点的服务HBase。
hbase-thrift-server	1.2.0	为其提供 Thrift 端点的HBase服务。
hcatalog-client	1.0.0-amzn-4	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	1.0.0-amzn-4	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	1.0.0-amzn-4	HTTP端点为其提供REST接口HCatalog。
hive-client	1.0.0-amzn-4	Hive 命令行客户端。
hive-metastore-server	1.0.0-amzn-4	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。

组件	版本	描述
hive-server	1.0.0-amzn-4	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.7.1-amzn-6	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.11.1	用于机器学习的库。
mysql-server	5.5	我的SQL数据库服务器。
oozie-client	4.2.0	Oozie 命令行客户端。
oozie-server	4.2.0	用于接受 Oozie 工作流请求的服务。
presto-coordinator	0.143	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.143	用于执行查询的各个部分的服务。
pig-client	0.14.0-amzn-0	Pig 命令行客户端。
spark-client	1.6.1	Spark 命令行客户端。
spark-history-server	1.6.1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	1.6.1	的内存中执行引擎。YARN
spark-yarn-slave	1.6.1	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。

组件	版本	描述
webserver	2.4	Apache HTTP 服务器。
zeppelin-server	0.5.6-incubating	支持交互式数据分析的基于 Web 的笔记本电脑。
zookeeper-server	3.4.8	用于维护配置信息、命名、提供分布式同步以及提供组服务的集中式服务。
zookeeper-client	3.4.8	ZooKeeper 命令行客户端。

#### 4.6.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

##### emr-4.6.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hbase-env	更改环境HBase中的值。
hbase-log4j	更改 hbase-HBase log4j.properties 文件中的值。

分类	描述
hbase-metrics	更改 hadoop-metrics2-hb HBase aase.properties 文件中的值。
hbase-policy	更改HBase的 hbase-policy.xml 文件中的值。
hbase-site	更改HBase的 hbase-site.xml 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j	更改 HCatalog W ebHCat 的 log4j.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j	更改 Hive 的 hive-exec-log 4j.properties 文件中的值。
hive-log4j	更改 Hive 的 hive-log4j.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。

分类	描述
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。

分类	描述
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。
zookeeper-config	更改 zoo.cfg 文件 ZooKeeper中的值。
zookeeper-log4j	更改 ZooKeeper的 log4j.properties 文件中的值。

## 亚马逊EMR版本 4.5.0

### 4.5.0 应用程序版本

此发行版支持以下应用程序：[Ganglia](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie-Sandbox](#)、[Pig](#)、[Presto-Sandbox](#)、[Spark](#)、[Sqoop-Sandbox](#) 和 [Zeppelin-Sandbox](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

## 应用程序版本信息

	emr-4.5.0	emr-4.4.0	emr-4.3.0	emr-4.2.0
适用于 Java 的 AWS SDK	1.10.27	1.10.27	1.10.27	1.10.27
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	未跟踪	未跟踪	未跟踪	未跟踪
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.6.0
HBase	-	-	-	-
HCatalog	1.0.0	1.0.0	-	-
Hadoop	2.7.2	2.7.1	2.7.1	2.6.0
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-

	emr-4.5.0	emr-4.4.0	emr-4.3.0	emr-4.2.0
Mahout	0.11.1	0.11.1	0.11.0	0.11.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	-	-	-	-
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.140	0.136	0.130	0.125
Spark	1.6.1	1.6.0	1.6.0	1.5.2
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	-	-
TensorFlow	-	-	-	-
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.6	0.5.6	0.5.5	0.5.5
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	-	-	-	-



## 4.5.0 发行说明

以下发行说明包含 Amazon EMR 4.5.0 版本的信息。

发布日期：2016 年 4 月 4 日

### 功能

- 已升级到 Spark 1.6.1
- 已升级到 Hadoop 2.7.2
- 已升级到 Presto 0.140
- 增加了对 Amazon S3 服务器端加密的 AWS KMS 支持。

### 早期版本中已解决的已知问题

- 修复了节点重启后 My SQL 和 Apache 服务器无法启动的问题。
- 修复IMPORT了存储在 Amazon S3 中的非分区表无法正常运行问题
- 修复了与 Presto 有关的问题，此问题导致在写入 Hive 表时要求暂存目录为 /mnt/tmp 而不是 /tmp。

## 4.5.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 emr 或 aws 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
emr-ddb	3.0.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。

组件	版本	描述
emr-goodies	2.0.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.1.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.2.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.5.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.2-amzn-0	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.2-amzn-0	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.2-amzn-0	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.2-amzn-0	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.2-amzn-0	HTTPHDFS操作端点。

组件	版本	描述
hadoop-kms-server	2.7.2-amzn-0	基于 Hadoop 的加密密钥管理服务。KeyProvider API
hadoop-mapred	2.7.2-amzn-0	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.2-amzn-0	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.2-amzn-0	YARN用于分配和管理群集资源和分布式应用程序的服务。
hcatalog-client	1.0.0-amzn-4	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	1.0.0-amzn-4	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	1.0.0-amzn-4	HTTP端点为其提供REST接口HCatalog。
hive-client	1.0.0-amzn-4	Hive 命令行客户端。
hive-metastore-server	1.0.0-amzn-4	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server	1.0.0-amzn-4	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.7.1-amzn-5	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序

组件	版本	描述
mahout-client	0.11.1	用于机器学习的库。
mysql-server	5.5	我的SQL数据库服务器。
oozie-client	4.2.0	Oozie 命令行客户端。
oozie-server	4.2.0	用于接受 Oozie 工作流请求的服务。
presto-coordinator	0.140	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.140	用于执行查询的各个部分的服务。
pig-client	0.14.0-amzn-0	Pig 命令行客户端。
spark-client	1.6.1	Spark 命令行客户端。
spark-history-server	1.6.1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	1.6.1	的内存中执行引擎。YARN
spark-yarn-slave	1.6.1	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
webserver	2.4	Apache HTTP 服务器。
zeppelin-server	0.5.6-incubating	支持交互式数据分析的基于 Web 的笔记本电脑。

## 4.5.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

### emr-4.5.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j	更改 HCatalog W ebHCat 的 log4j.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-env	更改 Hive 环境中的值。

分类	描述
hive-exec-log4j	更改 Hive 的 hive-exec-log 4j.properties 文件中的值。
hive-log4j	更改 Hive 的 hive-log4j.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。

分类	描述
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。

## 亚马逊EMR版本 4.4.0

### 4.4.0 应用程序版本

此发行版支持以下应用程序：[Ganglia](#)、[HCatalog](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie-Sandbox](#)、[Pig](#)、[Presto-Sandbox](#)、[Spark](#)、[Sqoop-Sandbox](#) 和 [Zeppelin-Sandbox](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-4.4.0	emr-4.3.0	emr-4.2.0	emr-4.1.0
适用于 Java 的 AWS SDK	1.10.27	1.10.27	1.10.27	未跟踪
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	未跟踪	未跟踪	未跟踪	未跟踪
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.6.0	-
HBase	-	-	-	-
HCatalog	1.0.0	-	-	-
Hadoop	2.7.1	2.7.1	2.6.0	2.6.0
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-



	emr-4.4.0	emr-4.3.0	emr-4.2.0	emr-4.1.0
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.11.1	0.11.0	0.11.0	0.11.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.0.1
Phoenix	-	-	-	-
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.136	0.130	0.125	0.119
Spark	1.6.0	1.6.0	1.5.2	1.5.0
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	-	-	-
TensorFlow	-	-	-	-
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.6	0.5.5	0.5.5	0.6.0-SNA PSHOT

	emr-4.4.0	emr-4.3.0	emr-4.2.0	emr-4.1.0
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	-	-	-	-

## 4.4.0 版本说明

以下发行说明包含 Amazon EMR 4.4.0 版本的信息。

发布日期：2016 年 3 月 14 日

### 功能

- 已添加 HCatalog 1.0.0
- 已添加 Sqoop-Sandbox 1.4.6
- 已升级到 Presto 0.136
- 已升级到 Zeppelin 0.5.6
- 已升级到 Mahout 0.11.1
- 默认情况下已启用 dynamicResourceAllocation。
- 已添加针对此版本的所有配置分类的表。有关更多信息，请参阅[配置应用程序](#)中的“配置分类”表。

### 早期版本中已解决的已知问题

- 修复了该maximizeResourceAllocation设置无法为YARN ApplicationMaster守护程序保留足够内存的问题。
- 修复了自定义时遇到的问题DNS。如果 resolve.conf 中的任何条目位于提供的自定义条目之前，则自定义条目不可解析。此行为受到集群的影响，VPC其中默认VPC名称服务器作为顶部条目插入resolve.conf。
- 修复了默认 Python 已移至版本 2.7 且未为该版本安装 boto 的问题。
- 修复了YARN容器和 Spark 应用程序生成唯一的 Ganglia 循环数据库 (rrd) 文件的问题，该文件会导致连接到实例的第一个磁盘已满。由于此修复，YARN容器级别的指标已被禁用，Spark 应用程序级别的指标已被禁用。
- 修复了导致日志推送程序中删除所有空日志文件夹的问题。结果是 Hive 无法登录CLI，因为日志推送器正在删除其下的空user文件夹。/var/log/hive

- 修复了影响 Hive 导入的问题，此问题影响分区并导致导入期间出现错误。
- 修复了EMRFS和 s3-dist-cp 无法正确处理包含句点的存储桶名称的问题。
- 更改了中的行为，EMRFS以便在启用版本控制的存储桶中不会连续创建\_`$folder$`标记文件，这可能有助于提高启用版本控制的存储桶的性能。
- 更改了中的EMRFS行为，使其不使用指令文件，除非启用了客户端加密。如果您要在使用客户端加密时删除说明文件，可将 `emrfs-site.xml` 属性 `fs.s3.cse.cryptoStorageMode.deleteInstructionFiles.enabled` 设置为 `true`。
- 更改了YARN日志聚合，将聚合目标处的日志保留两天。默认目标是您的集群HDFS存储。如果您要更改此持续时间，请在创建集群时使用 `yarn.log-aggregation.retain-seconds` 配置分类来更改 `yarn-site` 的值。与往常一样，您可以在创建集群时使用 `log-uri` 参数将应用程序日志保存到 Amazon S3。

#### 已应用的修补程序

- [HIVE-9655](#)
- [HIVE-9183](#)
- [HADOOP-12810](#)

#### 4.4.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 `CommunityVersion-amzn-EmrVersion` 的发行版标注。`EmrVersion` 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>emr-ddb</code>	3.0.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。

组件	版本	描述
emr-goodies	2.0.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.1.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.2.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.4.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.1-amzn-1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.1-amzn-1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.1-amzn-1	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.1-amzn-1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.1-amzn-1	HTTPHDFS操作端点。

组件	版本	描述
hadoop-kms-server	2.7.1-amzn-1	基于 Hadoop 的加密密钥管理服务。KeyProvider API
hadoop-mapred	2.7.1-amzn-1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.7.1-amzn-1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.1-amzn-1	YARN用于分配和管理群集资源和分布式应用程序的服务。
hcatalog-client	1.0.0-amzn-3	用于操作 hcatalog-server 的“hcat”命令行客户端。
hcatalog-server	1.0.0-amzn-3	提供服务HCatalog，为分布式应用程序提供表和存储管理层。
hcatalog-webhcat-server	1.0.0-amzn-3	HTTP端点为其提供REST接口HCatalog。
hive-client	1.0.0-amzn-3	Hive 命令行客户端。
hive-metastore-server	1.0.0-amzn-3	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server	1.0.0-amzn-3	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.7.1-amzn-5	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序

组件	版本	描述
mahout-client	0.11.1	用于机器学习的库。
mysql-server	5.5	我的SQL数据库服务器。
oozie-client	4.2.0	Oozie 命令行客户端。
oozie-server	4.2.0	用于接受 Oozie 工作流请求的服务。
presto-coordinator	0.136	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.136	用于执行查询的各个部分的服务。
pig-client	0.14.0-amzn-0	Pig 命令行客户端。
spark-client	1.6.0	Spark 命令行客户端。
spark-history-server	1.6.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	1.6.0	的内存中执行引擎。YARN
spark-yarn-slave	1.6.0	YARN从属服务器需要的 Apache Spark 库。
sqoop-client	1.4.6	Apache Sqoop 命令行客户端。
webserver	2.4	Apache HTTP 服务器。
zeppelin-server	0.5.6-incubating	支持交互式数据分析的基于 Web 的笔记本电脑。

## 4.4.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

### emr-4.4.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hcatalog-env	更改环境HCatalog中的值。
hcatalog-server-jndi	更改HCatalog的 jndi.properties 中的值。
hcatalog-server-proto-hive-site	更改 proto-hive-site .x HCatalog ml 中的值。
hcatalog-webhcat-env	更改 HCatalog W 环境ebHCat中的值。
hcatalog-webhcat-log4j	更改 HCatalog W ebHCat 的 log4j.properties 中的值。
hcatalog-webhcat-site	更改 HCatalog W ebHCat 的 webhcat-site.xml 文件中的值。
hive-env	更改 Hive 环境中的值。

分类	描述
hive-exec-log4j	更改 Hive 的 hive-exec-log 4j.properties 文件中的值。
hive-log4j	更改 Hive 的 hive-log4j.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。



分类	描述
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
sqoop-env	更改 Sqoop 的环境中的值。
sqoop-oraoop-site	更改 Sqoop OraOop 的 oraoop-site.xml 文件中的值。
sqoop-site	更改 Sqoop 的 sqoop-site.xml 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。

## 亚马逊EMR版本 4.3.0

### 4.3.0 应用程序版本

此版本支持以下应用程序：[Ganglia](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie-Sandbox](#)、[Pig](#)、[Presto-Sandbox](#)、[Spark](#) 和 [Zeppelin-Sandbox](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
适用于 Java 的 AWS SDK	1.10.27	1.10.27	未跟踪	未跟踪
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	未跟踪	未跟踪	未跟踪	未跟踪
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.6.0	-	-
HBase	-	-	-	-
HCatalog	-	-	-	-
Hadoop	2.7.1	2.6.0	2.6.0	2.6.0
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	-
Iceberg	-	-	-	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.11.0	0.11.0	0.11.0	0.10.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.0.1	-
Phoenix	-	-	-	-
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.130	0.125	0.119	-
Spark	1.6.0	1.5.2	1.5.0	1.4.1
Sqoop	-	-	-	-
Sqoop-Sandbox	-	-	-	-
TensorFlow	-	-	-	-
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.5	0.5.5	0.6.0-SNA PSHOT	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	-	-	-	-

### 4.3.0 发行说明

以下发行说明包含 Amazon EMR 4.3.0 版本的信息。

发布日期：2016 年 1 月 19 日

#### 功能

- 已升级到 Hadoop 2.7.1
- 已升级到 Spark 1.6.0
- 已将 Ganglia 升级到 3.7.2
- 已将 Presto 升级到 0.130
- Amazon 对 `spark.dynamicAllocation.enabled` 何时设置为 true EMR 进行了一些更改；默认情况下设置为 false。如果设置为 true，则会影响到 `maximizeResourceAllocation` 设置设置的默认值：
  - 若 `spark.dynamicAllocation.enabled` 设为 true，则 `spark.executor.instances` 将不被 `maximizeResourceAllocation` 设置。
  - 目前，`spark.driver.memory` 设置根据集群中的实例类型进行配置，与 `spark.executors.memory` 设置的方式类似。但是，由于 Spark 驱动程序应用程序可以在主实例或其中一个核心实例上运行（例如，分别在 YARN 客户端和集群模式下），因此该 `spark.driver.memory` 设置是根据这两个实例组之间较小实例类型的实例类型进行设置的。
  - 现在，该 `spark.default.parallelism` 设置设置为 YARN 容器可用 CPU 内核数量的两倍。在上一版本中，这是该值的一半。
  - 对为 Spark YARN 进程预留的内存开销的计算进行了调整，使其更加准确，从而使 Spark 可用的内存总量（即 `spark.executor.memory`）略有增加。

#### 早期版本中已解决的已知问题

- YARN 默认情况下，日志聚合现在处于启用状态。

- 修复了启用日志聚合后，无法将日志推送到集群的 Amazon S3 YARN 日志存储桶的问题。
- YARN现在，在所有节点类型中，容器大小的最小值均为 32。
- 修复了与 Ganglia 有关的问题，此问题已导致大型集群中主节点上的磁盘 I/O 过多。
- 修复了在关闭集群时阻止应用程序日志推送至 Amazon S3 的问题。
- 修复了导致EMRFSCLI某些命令失败的问题。
- 修复了齐柏林飞艇无法在底层加载依赖项的问题。 SparkContext
- 修复了因发出尝试添加实例的调整大小命令导致的问题。
- 修复了 Hive 中出现的 AS 对 Amazon S3 SELECT 进行过多列表调用的问题。 CREATE TABLE
- 修复了在安装 Hue、Oozie 和 Ganglia 时无法正常预置大型集群的问题。
- 修复了 s3-dist-cp 中的问题，此问题导致即使在失败并出现错误的情况下仍将返回零退出代码。

### 4.3.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
<code>emr-ddb</code>	3.0.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
<code>emr-goodies</code>	2.0.0	适用于 Hadoop 生态系统的方便易用的库。
<code>emr-kinesis</code>	3.1.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。

组件	版本	描述
emr-s3-dist-cp	2.1.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.3.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
ganglia-monitor	3.7.2	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.7.2	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.7.1	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.7.1-amzn-0	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.7.1-amzn-0	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.7.1-amzn-0	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.7.1-amzn-0	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.7.1-amzn-0	HTTPHDFS操作端点。
hadoop-kms-server	2.7.1-amzn-0	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.7.1-amzn-0	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。

组件	版本	描述
hadoop-yarn-nodemanager	2.7.1-amzn-0	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.7.1-amzn-0	YARN用于分配和管理群集资源和分布式应用程序的服务。
hive-client	1.0.0-amzn-2	Hive 命令行客户端。
hive-metastore-server	1.0.0-amzn-2	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server	1.0.0-amzn-2	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.7.1-amzn-5	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.11.0	用于机器学习的库。
mysql-server	5.5	我的SQL数据库服务器。
oozie-client	4.2.0	Oozie 命令行客户端。
oozie-server	4.2.0	用于接受 Oozie 工作流请求的服务。
presto-coordinator	0.130	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.130	用于执行查询的各个部分的服务。
pig-client	0.14.0-amzn-0	Pig 命令行客户端。
spark-client	1.6.0	Spark 命令行客户端。

组件	版本	描述
spark-history-server	1.6.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	1.6.0	的内存中执行引擎。YARN
spark-yarn-slave	1.6.0	YARN从属服务器需要的 Apache Spark 库。
webserver	2.4	Apache HTTP 服务器。
zeppelin-server	0.5.5-incubating-amzn-1	支持交互式数据分析的基于 Web 的笔记本电脑。

### 4.3.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

#### emr-4.3.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。



分类	描述
hive-env	更改 Hive 环境中的值。
hive-exec-log4j	更改 Hive 的 hive-exec-log 4j.properties 文件中的值。
hive-log4j	更改 Hive 的 hive-log4j.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。

分类	描述
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。

## 亚马逊EMR版本 4.2.0

### 4.2.0 应用程序版本

此版本支持以下应用程序：[Ganglia](#)、[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie-Sandbox](#)、[Pig](#)、[Presto-Sandbox](#)、[Spark](#) 和 [Zeppelin-Sandbox](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

## 应用程序版本信息

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
适用于 Java 的 AWS SDK	1.10.27	1.10.27	未跟踪	未跟踪
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	未跟踪	未跟踪	未跟踪	未跟踪
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.6.0	-	-
HBase	-	-	-	-
HCatalog	-	-	-	-
Hadoop	2.7.1	2.6.0	2.6.0	2.6.0
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	-
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
Mahout	0.11.0	0.11.0	0.11.0	0.10.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.0.1	-
Phoenix	-	-	-	-
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.130	0.125	0.119	-
Spark	1.6.0	1.5.2	1.5.0	1.4.1
Sqoop	-	-	-	-
Sqoop-Sandbox	-	-	-	-
TensorFlow	-	-	-	-
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.5	0.5.5	0.6.0-SNA PSHOT	-
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	-	-	-	-

## 4.2.0 版本说明

以下发行说明包含 Amazon EMR 4.2.0 版本的信息。

发布日期：2015 年 11 月 18 日

### 功能

- 已添加 Ganglia 支持
- 已升级到 Spark 1.5.2
- 已升级到 Presto 0.125
- 已将 Oozie 升级到 4.2.0
- 已将 Zeppelin 升级到 0.5.5
- 已将升级 AWS SDK for Java 到 1.10.27

### 早期版本中已解决的已知问题

- 修复了未使用默认元数据表名称的问题。EMRFS CLI
- 修复了在 Amazon S3 中使用 ORC 由支持的表时遇到的问题。
- 修复了遇到的 Python 版本在 Spark 配置中不匹配的问题。
- 修复了由于中的群集问题而无法报告 YARN 节点状态 DNS 的问题 VPC。
- 修复了节点 YARN 停用 时遇到的问题，该问题导致应用程序挂起或无法调度新应用程序。
- 修复了集群以 TIMED OUT \_ 状态终止时遇到的问题 STARTING。
- 修复了在其他版本中包含 EMRFS Scala 依赖项时遇到的问题。Scala 依赖项已被删除。

## 4.2.0 组件版本

下面列出了 Amazon 随此版本 EMR 安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的 EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新 EMR 版本中的大数据应用程序包通常是社区中的最新版本。我们会 EMR 尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本 EMR 不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
emr-ddb	3.0.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.0.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.1.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.0.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.2.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
ganglia-monitor	3.6.0	适用于 Hadoop 生态系统应用程序的嵌入式 Ganglia 代理以及 Ganglia 监控代理。
ganglia-metadata-collector	3.6.0	用于从 Ganglia 监控代理中聚合指标的 Ganglia 元数据收集器。
ganglia-web	3.5.10	用于查看由 Ganglia 元数据收集器收集的指标的 Web 应用程序。
hadoop-client	2.6.0-amzn-2	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.6.0-amzn-2	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.6.0-amzn-2	HDFS命令行客户端和库

组件	版本	描述
hadoop-hdfs-namenode	2.6.0-amzn-2	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.6.0-amzn-2	HTTPHDFS操作端点。
hadoop-kms-server	2.6.0-amzn-2	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.6.0-amzn-2	MapReduce 用于运行 MapReduce 应用程序的执行引擎。
hadoop-yarn-nodemanager	2.6.0-amzn-2	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.6.0-amzn-2	YARN用于分配和管理群集资源和分布式应用程序的服务。
hive-client	1.0.0-amzn-1	Hive 命令行客户端。
hive-metastore-server	1.0.0-amzn-1	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server	1.0.0-amzn-1	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.7.1-amzn-5	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.11.0	用于机器学习的库。
mysql-server	5.5	我的SQL数据库服务器。
oozie-client	4.2.0	Oozie 命令行客户端。

组件	版本	描述
oozie-server	4.2.0	用于接受 Oozie 工作流请求的服务。
presto-coordinator	0.125	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.125	用于执行查询的各个部分的服务。
pig-client	0.14.0-amzn-0	Pig 命令行客户端。
spark-client	1.5.2	Spark 命令行客户端。
spark-history-server	1.5.2	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	1.5.2	的内存中执行引擎。YARN
spark-yarn-slave	1.5.2	YARN从属服务器需要的 Apache Spark 库。
webserver	2.4	Apache HTTP 服务器。
zeppelin-server	0.5.5-incubating-amzn-0	支持交互式数据分析的基于 Web 的笔记本电脑。

## 4.2.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

### emr-4.2.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。



分类	描述
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j	更改 Hive 的 hive-exec-log 4j.properties 文件中的值。
hive-log4j	更改 Hive 的 hive-log4j.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。

分类	描述
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
presto-connector-hive	更改 Presto 的 hive.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
spark-metrics	更改 Spark 的 metrics.properties 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。

## 亚马逊EMR版本 4.1.0

### 4.1.0 应用程序版本

此版本支持以下应用程序：[Hadoop](#)、[Hive](#)、[Hue](#)、[Mahout](#)、[Oozie-Sandbox](#)、[Pig](#)、[Presto-Sandbox](#)、[Spark](#) 和 [Zeppelin-Sandbox](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

#### 应用程序版本信息

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
适用于 Java 的 AWS SDK	1.10.27	1.10.27	未跟踪	未跟踪
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	未跟踪	未跟踪	未跟踪	未跟踪
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.6.0	-	-
HBase	-	-	-	-
HCatalog	-	-	-	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
Hadoop	2.7.1	2.6.0	2.6.0	2.6.0
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	-
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.11.0	0.11.0	0.11.0	0.10.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.0.1	-
Phoenix	-	-	-	-
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.130	0.125	0.119	-
Spark	1.6.0	1.5.2	1.5.0	1.4.1
Sqoop	-	-	-	-
Sqoop-Sandbox	-	-	-	-
TensorFlow	-	-	-	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.5	0.5.5	0.6.0-SNA PSHOT	-
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	-	-	-	-

## 4.1.0 发行说明

### 4.1.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果 `myapp-component` 以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 `2.2-amzn-2`。

组件	版本	描述
emr-ddb	3.0.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。

组件	版本	描述
emr-goodies	2.0.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.1.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.0.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.1.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
hadoop-client	2.6.0-amzn-1	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.6.0-amzn-1	HDFS用于存储区块的节点级服务。
hadoop-hdfs-library	2.6.0-amzn-1	HDFS命令行客户端和库
hadoop-hdfs-namenode	2.6.0-amzn-1	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.6.0-amzn-1	HTTPHDFS操作端点。
hadoop-kms-server	2.6.0-amzn-1	基于 Hadoop 的加密密钥管理服务。 KeyProvider API
hadoop-mapred	2.6.0-amzn-1	MapReduce 用于运行 MapReduce 应用程序的执行引擎库。
hadoop-yarn-nodemanager	2.6.0-amzn-1	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.6.0-amzn-1	YARN用于分配和管理群集资源和分布式应用程序的服务。

组件	版本	描述
hive-client	1.0.0-amzn-1	Hive 命令行客户端。
hive-metastore-server	1.0.0-amzn-1	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的 SQL 元数据。
hive-server	1.0.0-amzn-1	用于将 Hive 查询作为 Web 请求接受的服务。
hue-server	3.7.1-amzn-4	用于使用 Hadoop 生态系统应用程序分析数据的 Web 应用程序
mahout-client	0.11.0	用于机器学习的库。
mysql-server	5.5	我的 SQL 数据库服务器。
oozie-client	4.0.1	Oozie 命令行客户端。
oozie-server	4.0.1	用于接受 Oozie 工作流请求的服务。
presto-coordinator	0.119	用于在 presto-worker 之中接受查询并管理查询的服务。
presto-worker	0.119	用于执行查询的各个部分的服务。
pig-client	0.14.0-amzn-0	Pig 命令行客户端。
spark-client	1.5.0	Spark 命令行客户端。
spark-history-server	1.5.0	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	1.5.0	的内存中执行引擎。YARN

组件	版本	描述
spark-yarn-slave	1.5.0	YARN从属服务器需要的 Apache Spark 库。
zeppelin-server	0.6.0-正在孵化-SNAPSHOT	支持交互式数据分析的基于 Web 的笔记本电脑。

#### 4.1.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

##### emr-4.1.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hdfs-encryption-zones	配置HDFS加密区域。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j	更改 Hive 的 hive-exec-log 4j.properties 文件中的值。
hive-log4j	更改 Hive 的 hive-log4j.properties 文件中的值。



分类	描述
hive-site	更改 Hive 的 hive-site.xml 文件中的值
hue-ini	更改 Hue 的 ini 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
hadoop-kms-acls	更改 Hadoop 的 kms-acls.xml 文件中的值。
hadoop-kms-env	在 Hadoop KMS 环境中更改值。
hadoop-kms-log4j	更改 Hadoop 的 kms-log4j.properties 文件中的值。
hadoop-kms-site	更改 Hadoop 的 kms-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
oozie-env	更改 Oozie 的环境中的值。
oozie-log4j	更改 Oozie 的 oozie-log4j.properties 文件中的值。
oozie-site	更改 Oozie 的 oozie-site.xml 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
presto-log	更改 Presto 的 log.properties 文件中的值。
presto-config	更改 Presto 的 config.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。

分类	描述
spark-env	更改 Spark 环境中的值。
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。
zeppelin-env	更改 Zeppelin 环境中的值。

## 亚马逊EMR版本 4.0.0

### 4.0.0 应用程序版本

此版本支持以下应用程序：[Hadoop](#)、[Hive](#)、[Mahout](#)、[Pig](#) 和 [Spark](#)。

下表列出了此版本的 Amazon 中可用的应用程序版本EMR和前三个亚马逊版本中的应用程序EMR版本（如果适用）。

要全面了解每个 Amazon 版本的应用程序版本历史记录EMR，请参阅以下主题：

- [亚马逊 EMR 7.x 版本中的应用程序版本](#)
- [Amazon EMR 6.x 发行版中的应用程序版本](#)
- [Amazon EMR 5.x 发行版中的应用程序版本](#)
- [Amazon EMR 4.x 发行版中的应用程序版本](#)

### 应用程序版本信息

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
适用于 Java 的 AWS SDK	1.10.27	1.10.27	未跟踪	未跟踪
Python	未跟踪	未跟踪	未跟踪	未跟踪
Scala	未跟踪	未跟踪	未跟踪	未跟踪

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.6.0	-	-
HBase	-	-	-	-
HCatalog	-	-	-	-
Hadoop	2.7.1	2.6.0	2.6.0	2.6.0
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	-
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.11.0	0.11.0	0.11.0	0.10.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.0.1	-
Phoenix	-	-	-	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
Pig	0.14.0	0.14.0	0.14.0	0.14.0
Presto	-	-	-	-
Presto-Sandbox	0.130	0.125	0.119	-
Spark	1.6.0	1.5.2	1.5.0	1.4.1
Sqoop	-	-	-	-
Sqoop-Sandbox	-	-	-	-
TensorFlow	-	-	-	-
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.5	0.5.5	0.6.0-SNAPSHOT	-
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	-	-	-	-

## 4.0.0 版本说明

### 4.0.0 组件版本

下面列出了 Amazon 随此版本EMR安装的组件。一些组件作为大数据应用程序包的一部分安装。其他则是 Amazon 独有的EMR，是为系统流程和功能而安装的。它们通常以 `emr` 或 `aws` 开头。Amazon 最新EMR版本中的大数据应用程序包通常是社区中的最新版本。我们会EMR尽快在 Amazon 上发布社区版本。

Amazon 中的某些组件与社区版本EMR不同。这些组件具有以下形式的 *CommunityVersion-amzn-EmrVersion* 的发行版标注。*EmrVersion* 从 0 开始。例如，如果myapp-component以 2.2 版本命名的开源社区组件经过三次修改以包含在不同的 Amazon EMR 版本中，则其发布版本将列为 2.2-amzn-2。

组件	版本	描述
emr-ddb	3.0.0	适用于 Hadoop 生态系统应用程序的 Amazon DynamoDB 连接器。
emr-goodies	2.0.0	适用于 Hadoop 生态系统的方便易用的库。
emr-kinesis	3.0.0	适用于 Hadoop 生态系统应用程序的 Amazon Kinesis 连接器。
emr-s3-dist-cp	2.0.0	针对 Amazon S3 优化的分布式复制应用程序。
emrfs	2.0.0	适用于 Hadoop 生态系统应用程序的 Amazon S3 连接器。
hadoop-client	2.6.0-amzn-0	Hadoop 命令行客户端，如“hdfs”、“hadoop”或“yarn”。
hadoop-hdfs-datanode	2.6.0-amzn-0	HDFS用于存储区块的节点级服务。
hadoop-hdfs-namenode	2.6.0-amzn-0	HDFS用于跟踪文件名和区块位置的服务。
hadoop-httpfs-server	2.6.0-amzn-0	HTTPHDFS操作端点。
hadoop-mapred	2.6.0-amzn-0	MapReduce 用于运行 MapReduce 应用程序的执行引擎。

组件	版本	描述
hadoop-yarn-nodemanager	2.6.0-amzn-0	YARN用于管理单个节点上的容器的服务。
hadoop-yarn-resourcemanager	2.6.0-amzn-0	YARN用于分配和管理群集资源和分布式应用程序的服务。
hive-client	1.0.0-amzn-0	Hive 命令行客户端。
hive-metastore-server	1.0.0-amzn-0	用于访问 Hive 元数据仓库的服务，Hive 元存储库是一个语义存储库，用于存储 Hadoop 操作的SQL元数据。
hive-server	1.0.0-amzn-0	用于将 Hive 查询作为 Web 请求接受的服务。
mahout-client	0.10.0	用于机器学习的库。
mysql-server	5.5	我的SQL数据库服务器。
pig-client	0.14.0-amzn-0	Pig 命令行客户端。
spark-client	1.4.1	Spark 命令行客户端。
spark-history-server	1.4.1	用于查看完整的 Spark 应用程序的生命周期的已记录事件的 Web UI。
spark-on-yarn	1.4.1	的内存中执行引擎。YARN
spark-yarn-slave	1.4.1	YARN从属服务器需要的 Apache Spark 库。

#### 4.0.0 配置分类

配置分类允许您自定义应用程序。它们通常对应于应用程序的配置XML文件，例如hive-site.xml。有关更多信息，请参阅 [配置应用程序](#)。

## emr-4.0.0 分类

分类	描述
capacity-scheduler	更改 Hadoop 的 capacity-scheduler.xml 文件中的值。
core-site	更改 Hadoop 的 core-site.xml 文件中的值。
emrfs-site	更改EMRFS设置。
hadoop-env	更改适用于所有 Hadoop 组件的 Hadoop 环境中的值。
hadoop-log4j	更改 Hadoop 的 log4j.properties 文件中的值。
hdfs-site	更改 hdfs-site.xml HDFS 中的值。
hive-env	更改 Hive 环境中的值。
hive-exec-log4j	更改 Hive 的 hive-exec-log 4j.properties 文件中的值。
hive-log4j	更改 Hive 的 hive-log4j.properties 文件中的值。
hive-site	更改 Hive 的 hive-site.xml 文件中的值
httpfs-env	更改HTTPFS环境中的值。
httpfs-site	更改 Hadoop 的 httpfs-site.xml 文件中的值。
mapred-env	更改 MapReduce 应用程序环境中的值。
mapred-site	更改 MapReduce 应用程序的 mapred-site.xml 文件中的值。
pig-properties	更改 Pig 的 pig.properties 文件中的值。
pig-log4j	更改 Pig 的 log4j.properties 文件中的值。
spark	亚马逊EMR精心策划的 Apache Spark 设置。

分类	描述
spark-defaults	更改 Spark 的 spark-defaults.conf 文件中的值。
spark-env	更改 Spark 环境中的值。
spark-log4j	更改 Spark 的 log4j.properties 文件中的值。
yarn-env	更改YARN环境中的值。
yarn-site	更改YARN的 yarn-site.xml 文件中的值。

## 亚马逊 EMR 2.x 和 3. AMI x 版本

### Note

AWS 正在将所有 AWS API端点的TLS配置更新到 TLS 1.2 的最低版本。亚马逊EMR版本 3.10 及更低版本仅支持 TLS 1.0/1.1 连接。2023 年 12 月 4 日之后，您将无法使用亚马逊 EMR 3.10 及更低版本创建集群。

如果您使用的是 Amazon EMR 3.10 或更低版本，我们建议您立即测试您的工作负载并将其迁移到最新的 Amazon EMR 版本。有关更多信息，请参阅 [AWS 安全博客](#)。

Amazon EMR 2.x 和 3.x 版本 AMI ( 称为版本 ) 适用于出于兼容性原因需要这些版本的现有解决方案。我们建议不要使用这些发行版创建新集群或新解决方案。它们缺少更新的发行版功能并且包含过时的应用程序包。

我们建议您使用最新的 Amazon EMR 发行版来构建解决方案。

2.x 和 3.x 系列的发行版本与最新的 Amazon EMR 发行版本之间的差异范围很大。从您创建和配置集群的方式到集群上应用程序的端口和目录结构，均存在差异。

本节试图介绍 Amazon 最显著的差异EMR，以及具体的应用程序配置和管理差异。介绍并不全面。如果您在 2.x 或 3.x 系列中创建并使用集群，可能遇到此节中未涉及的差异。

### 主题

- [使用早期AMI版本的 Amazon 创建集群 EMR](#)
- [使用早期AMI版本的 Amazon 安装应用程序 EMR](#)



- [使用早期AMI版本的 Amazon 自定义集群和应用程序配置 EMR](#)
- [早期AMI版本的 Amazon 的 Hive 应用程序细节 EMR](#)
- [HBase早期AMI版本的 Amazon 的应用程序细节 EMR](#)
- [早期AMI版本的 Amazon 的 Pig 应用程序细节 EMR](#)
- [早期AMI版本的 Amazon 的 Spark 应用程序细节 EMR](#)
- [S3 DistCp 实用程序与早期AMI版本的 Amazon 的区别 EMR](#)

## 使用早期AMI版本的 Amazon 创建集群 EMR

按AMI版本引用亚马逊 EMR 2.x 和 3.x 版本。在 Amazon EMR 4.0.0 及更高版本中，版本使用发布标签按发布版本进行引用，例如。emr-5.11.0当您以编程方式使用 AWS CLI 或创建集群时，这种变化最为明显。

当您使用使用AMI发布版本创建集群时，请使用--ami-version选项，例如--ami-version 3.11.0。AWS CLI 当您指定时，Amazon EMR 4.0.0 及更高版本中引入的许多选项、功能和应用程序都不可用。--ami-version有关更多信息，请参阅《AWS CLI 命令参考》中的 [create-cluster](#)。

以下示例 AWS CLI 命令使用AMI版本启动集群。

### Note

为了便于读取，包含 Linux 行继续符 (\)。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (^)。

```
aws emr create-cluster --name "Test cluster" --ami-version 3.11.0 \  
--applications Name=Hue Name=Hive Name=Pig \  
--use-default-roles --ec2-attributes KeyName=myKey \  
--instance-groups InstanceGroupType=MASTER,InstanceCount=1,\  
InstanceType=m3.xlarge InstanceGroupType=CORE,InstanceCount=2,\  
InstanceType=m3.xlarge --bootstrap-actions Path=s3://elasticmapreduce/bootstrap-  
actions/configure-hadoop,\  
Name="Configuring infinite JVM reuse",Args=["-m","mapred.job.reuse.jvm.num.tasks=-1"]
```

在编程方面，所有 Amazon EMR 发行版本都使用中的RunJobFlowRequest操作EMRAPI来创建集群。以下示例 Java 代码使用AMI发行版 3.11.0 创建集群。

```
RunJobFlowRequest request = new RunJobFlowRequest()
    .withName("AmiVersion Cluster")
    .withAmiVersion("3.11.0")
    .withInstances(new JobFlowInstancesConfig()
        .withEc2KeyName("myKeyPair")
        .withInstanceCount(1)
        .withKeepJobFlowAliveWhenNoSteps(true)
        .withMasterInstanceType("m3.xlarge")
        .withSlaveInstanceType("m3.xlarge"));
```

以下 RunJobFlowRequest 调用改用版本标签：

```
RunJobFlowRequest request = new RunJobFlowRequest()
    .withName("ReleaseLabel Cluster")
    .withReleaseLabel("emr-7.2.0")
    .withInstances(new JobFlowInstancesConfig()
        .withEc2KeyName("myKeyPair")
        .withInstanceCount(1)
        .withKeepJobFlowAliveWhenNoSteps(true)
        .withMasterInstanceType("m3.xlarge")
        .withSlaveInstanceType("m3.xlarge"));
```

## 配置集群大小

当您的集群运行时，Hadoop 会决定处理数据所需的映射器和 Reducer 任务的数量。更大的集群应当有更多任务，以更好地利用资源并缩短处理时间。通常，在整个EMR集群中，集群的大小保持不变；您可以在创建集群时设置任务数。当您调整正在运行的集群大小时，您可以改变集群执行过程中的处理。因此，您可以在集群的生命周期内改变任务数量，而不是使用固定数量的任务。有两个配置选项可以帮助设置理想的任务数量：

- `mapred.map.tasksperslot`
- `mapred.reduce.tasksperslot`

您可以在 `mapred-conf.xml` 文件中设置这两个选项。当您向集群提交任务时，任务客户端会检查当前整个集群范围中可用的 map-reduce 槽位总数。然后，作业客户端会使用以下公式设置任务数量：

- `mapred.map.tasks = mapred.map.tasksperslot * 集群中的映射槽位数`
- `mapred.reduce.tasks = mapred.reduce.tasksperslot * 集群中的 reduce 槽位`

如果未配置任务数量，则作业客户端只会读取 `tasksperslot` 参数。通过引导操作 (对于全部集群) 或添加一个更改配置的步骤 (对每个作业单独执行)。您可以随时覆盖任务数量。

即使任务节点不可用，Amazon 也能EMR承受任务节点故障并继续执行集群。Amazon EMR 会自动配置额外的任务节点来替换失败的任务节点。

对于每个集群步骤，您可以有不同数量的任务节点。您还可以向正在运行的集群添加步骤，以修改任务节点的数量。由于所有步骤在默认情况下都保证按顺序运行，您可以为任何步骤指定运行任务节点的数量。

## 使用早期AMI版本的 Amazon 安装应用程序 EMR

使用AMI版本时，可以通过多种方式安装应用程序，包括使用操作的 `NewSupportedProducts` 参数、使用引导操作和使用 [Step](#) 操作。 [RunJobFlow](#)

## 使用早期AMI版本的 Amazon 自定义集群和应用程序配置 EMR

Amazon EMR 发行版 4.0.0 引入了一种使用配置分类配置应用程序的简化方法。有关更多信息，请参阅 [配置应用程序](#)。使用AMI版本时，您可以使用引导操作和传递的参数来配置应用程序。例如，`configure-hadoop`和`configure-daemons`引导操作设置了 Hadoop 和YARN特定环境属性，例如。`--namenode-heap-size`在更新的版本中，这些属性是使用 `hadoop-env` 和 `yarn-env` 配置分类配置的。有关执行常见配置的引导操作，请参阅 [Github 上的emr-bootstrap-actions 存储库](#)。

下表将引导操作映射到最新的 Amazon EMR 发行版本中的配置分类。

### Hadoop

受影响的应用程序文件名	AMI版本引导操作	配置分类
<code>core-site.xml</code>	<code>configure-hadoop -c</code>	<code>core-site</code>
<code>log4j.properties</code>	<code>configure-hadoop -l</code>	<code>hadoop-log4j</code>
<code>hdfs-site.xml</code>	<code>configure-hadoop -s</code>	<code>hdfs-site</code>
不适用	不适用	<code>hdfs-encryption-zones</code>
<code>mapred-site.xml</code>	<code>configure-hadoop -m</code>	<code>mapred-site</code>
<code>yarn-site.xml</code>	<code>configure-hadoop -y</code>	<code>yarn-site</code>

受影响的应用程序文件名	AMI版本引导操作	配置分类
httpfs-site.xml	configure-hadoop -t	httpfs-site
capacity-scheduler.xml	configure-hadoop -z	capacity-scheduler
yarn-env.sh	configure-daemons --resource-manager-opts	yarn-env

## Hive

受影响的应用程序文件名	AMI版本引导操作	配置分类
hive-env.sh	不适用	hive-env
hive-site.xml	hive-script --install -hive-site \${MY_HIVE_SITE_FILE}	hive-site
hive-exec-log4j.properties	不适用	hive-exec-log4j
hive-log4j.properties	不适用	hive-log4j

## EMRFS

受影响的应用程序文件名	AMI版本引导操作	配置分类
emrfs-site.xml	configure-hadoop -e	emrfs-site
不适用	s3get -s s3://customer-provider.jar -d /usr/share/aws/emr/auxlib/	emrfs-site (带新设置 fs.s3.cse.encryptedMaterialsProvider.uri )

有关所有分类的列表，请参阅[配置应用程序](#)。

## 应用程序环境变量

使用AMI版本时，将使用`hadoop-user-env.sh`脚本和`configure-daemons`引导操作来配置Hadoop环境。该脚本包括以下操作：

```
#!/bin/bash
export HADOOP_USER_CLASSPATH_FIRST=true;
echo "HADOOP_CLASSPATH=/path/to/my.jar" >> /home/hadoop/conf/hadoop-user-env.sh
```

在 Amazon 4.x EMR 版本中，您可以使用`hadoop-env`配置分类进行同样的操作，如以下示例所示：

```
[
  {
    "Classification": "hadoop-env",
    "Properties": {
    },
  },
  "Configurations": [
    {
      "Classification": "export",
      "Properties": {
        "HADOOP_USER_CLASSPATH_FIRST": "true",
        "HADOOP_CLASSPATH": "/path/to/my.jar"
      }
    }
  ]
}
```

再举一例，使用 `configure-daemons` 并传递 `--namenode-heap-size=2048` 和 `--namenode-opts=-XX:GCTimeRatio=19` 等效于以下配置分类。

```
[
  {
    "Classification": "hadoop-env",
    "Properties": {
    },
  },
  "Configurations": [
    {
```

```

        "Classification": "export",
        "Properties": {
            "HADOOP_DATANODE_HEAPSIZE": "2048",
            "HADOOP_NAMENODE_OPTS": "-XX:GCTimeRatio=19"
        }
    ]
}
]
]

```

其他应用程序环境变量不再在 `/home/hadoop/.bashrc` 中定义。相反，它们主要在 `/etc/default` 文件中基于每个组件或应用程序设置，例如 `/etc/default/hadoop`。应用程序 `/usr/bin/` 安装的 Wrapper 脚本也RPMs可以在涉及实际的 `bin` 脚本之前设置其他环境变量。

## 服务端口

使用AMI版本时，某些服务使用自定义端口。

### 端口设置中的更改

设置	AMI版本 3.x	开源默认值
<code>fs.default.name</code>	<code>hdfs://IP: 9000 emrDetermined</code>	默认 ( <code>hdfs://<i>emrDeterminedIP</i> :8020</code> )
<code>dfs.datanode.address</code>	<code>0.0.0.0:9200</code>	默认值 ( <code>0.0.0.0:50010</code> )
<code>dfs.datanode.http.address</code>	<code>0.0.0.0:9102</code>	默认值 ( <code>0.0.0.0:50075</code> )
<code>dfs.datanode.https.address</code>	<code>0.0.0.0:9402</code>	默认值 ( <code>0.0.0.0:50475</code> )
<code>dfs.datanode.ipc.address</code>	<code>0.0.0.0:9201</code>	默认值 ( <code>0.0.0.0:50020</code> )
<code>dfs.http.address</code>	<code>0.0.0.0:9101</code>	默认值 ( <code>0.0.0.0:50070</code> )
<code>dfs.https.address</code>	<code>0.0.0.0:9202</code>	默认值 ( <code>0.0.0.0:50470</code> )
<code>dfs.secondary.http.address</code>	<code>0.0.0.0:9104</code>	默认值 ( <code>0.0.0.0:50090</code> )
<code>yarn.nodemanager.address</code>	<code>0.0.0.0:9103</code>	默认值 ( <code>\${yarn.nodemanager.hostname}:0</code> )

设置	AMI版本 3.x	开源默认值
yarn.nodemanager.localizer.address	0.0.0.0:9033	默认值 (\${yarn.nodemanager.hostname}:8040)
yarn.nodemanager.webapp.address	0.0.0.0:9035	默认值 (\${yarn.nodemanager.hostname}:8042)
yarn.resourcemanager.address	<i>emrDeterminedIP</i> :9022	默认值 (\${yarn.resourcemanager.hostname}:8032)
yarn.resourcemanager.admin.address	<i>emrDeterminedIP</i> :9025	默认值 (\${yarn.resourcemanager.hostname}:8033)
yarn.resourcemanager.resource-tracker.address	<i>emrDeterminedIP</i> :9023	默认值 (\${yarn.resourcemanager.hostname}:8031)
yarn.resourcemanager.scheduler.address	<i>emrDeterminedIP</i> :9024	默认值 (\${yarn.resourcemanager.hostname}:8030)
yarn.resourcemanager.webapp.address	0.0.0.0:9026	默认值 (\${yarn.resourcemanager.hostname}:8088)
yarn.web-proxy.address	<i>emrDeterminedIP</i> :9046	默认值 (无值)
yarn.resourcemanager.hostname	0.0.0.0 (默认值)	<i>emrDeterminedIP</i>

### Note

这些区域有：*emrDeterminedIP* 是由亚马逊生成的 IP 地址EMR。

## 用户

使用AMI版本时，用户hadoop运行所有进程并拥有所有文件。在 Amazon EMR 发行版 4.0.0 及更高版本中，用户存在于应用程序和组件级别。

## 安装顺序、安装的构件和日志文件位置

使用AMI版本时，应用程序工件及其配置目录将安装在 `/home/hadoop/application` 目录中。例如，如果已安装 Hive，目录将是 `/home/hadoop/hive`。在 Amazon 4.0.0 及更高EMR版本中，应用程序项目安装在目录中。`/usr/lib/application` 使用某个AMI版本时，可以在不同的地方找到日志文件。下表列出了位置。

### Amazon S3 上的日志位置的更改

守护进程或应用程序	目录位置
instance-state	node/ <i>instance-id</i> /instance-state/
hadoop-hdfs-namenode	守护进程/ <i>instance-id</i> /hadoop-hadoop-namenode.log
hadoop-hdfs-datanode	守护进程/ <i>instance-id</i> /hadoop-hadoop-datanode.log
hadoop-yarn () ResourceManager	守护进程/ <i>instance-id</i> /yarn-hadoop-resourcemanager
hadoop-yarn (代理服务器)	守护进程/ <i>instance-id</i> /yarn-hadoop-proxyserver
mapred-historyserver	守护进程/ <i>instance-id</i> /
httpfs	守护进程/ <i>instance-id</i> /httpfs.log
hive-server	node/ <i>instance-id</i> /hive-server/hive-server.log
hive-metastore	node/ <i>instance-id</i> /apps/hive.log
Hive CLI	node/ <i>instance-id</i> /apps/hive.log
YARN应用程序用户日志和容器日志	task-attempts/
Mahout	不适用
Pig	不适用



守护进程或应用程序	目录位置
spark-historyserver	不适用
mapreduce 任务历史记录文件	jobs/

## Command Runner

使用某个AMI版本时，许多脚本或程序（例如/home/hadoop/contrib/streaming/hadoop-streaming.jar）并未放置在 shell 登录路径环境中，因此在使用 jar 文件（例如 command-runner.jar 或 script-runner.jar）执行脚本时，需要指定完整路径。command-runner.jar 位于上面，AMI 因此无需像以前那 URI 样知道完整的内容 script-runner.jar。

## 重复因子

重复因子允许您配置何时启动 Hadoop JVM。您可以 JVM 为每项任务启动一个新的 Hadoop，这样可以提供更好的任务隔离，也可以在任务 JVMs 之间共享，从而降低框架开销。如果您要处理许多小文件，那么多次重复使用这些文件来分摊启动成本是 JVM 有意义的。但是，如果每项任务花费很长时间或处理大量数据，则可以选择不重复使用，JVM 以确保为后续任务腾出所有内存。使用 AMI 版本时，您可以使用 configure-hadoop 引导操作来设置 mapred.job.reuse.jvm.num.tasks 属性来自定义复制因子。

以下示例演示了为无限 JVM JVM 重复使用设置重用系数。

### Note

为了便于读取，包含 Linux 行继续符 (\)。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (^)。


```
aws emr create-cluster --name "Test cluster" --ami-version 3.11.0 \
--applications Name=Hue Name=Hive Name=Pig \
--use-default-roles --ec2-attributes KeyName=myKey \
--instance-groups InstanceGroupType=MASTER,InstanceCount=1,InstanceType=m3.xlarge \
InstanceGroupType=CORE,InstanceCount=2,InstanceType=m3.xlarge \
--bootstrap-actions Path=s3://elasticmapreduce/bootstrap-actions/configure-hadoop,\
Name="Configuring infinite JVM reuse",Args=["-m", "mapred.job.reuse.jvm.num.tasks=-1"]
```

## 早期AMI版本的 Amazon 的 Hive 应用程序细节 EMR

### 日志文件

使用亚马逊 2.x 和 3.x EMR AMI 版本，Hive 日志将保存到 `/mnt/var/log/apps/`。为了支持 Hive 的并行版本，您运行的这个 Hive 版本还会确定日志文件名称，如下表所示。

Hive 版本	日志文件名称
0.13.1	hive.log <div data-bbox="592 651 1507 919" style="border: 1px solid #add8e6; border-radius: 10px; padding: 10px; margin-top: 10px;"> <p> <b>Note</b></p> <p>从该版本开始，Amazon EMR 使用未受版本控制的文件名。hive.log次要版本将与主要版本共享同一个日志位置。</p> </div>
0.11.0	hive_0110.log <div data-bbox="592 1039 1507 1255" style="border: 1px solid #add8e6; border-radius: 10px; padding: 10px; margin-top: 10px;"> <p> <b>Note</b></p> <p>0.11.0 的次要版本，如 Hive 0.11.0.1，和 Hive 0.11.0 一起共享相同的日志文件位置。</p> </div>
0.8.1	hive_081.log <div data-bbox="592 1375 1507 1591" style="border: 1px solid #add8e6; border-radius: 10px; padding: 10px; margin-top: 10px;"> <p> <b>Note</b></p> <p>Hive 0.8.1 的次要版本，如 Hive 0.8.1.1，和 Hive 0.8.1 一起共享相同的日志文件位置。</p> </div>
0.7.1	hive_07_1.log

Hive 版本	日志文件名称
	<div style="border: 1px solid #ccc; border-radius: 10px; padding: 10px; background-color: #f0f8ff;"> <p> <b>Note</b></p> <p>Hive 0.7.1 的次要版本，如 Hive 0.7.1.3 和 Hive 0.7.1.4，和 Hive 0.7.1 一起共享相同的日志文件位置。</p> </div>
0.7	hive_07.log
0.5	hive_05.log
0.4	hive.log

## 拆分输入功能

要使用低于 0.13.1 的 Hive 版本（亚马逊版本 3.11.0 之前的 EMRAMI 版本）实现拆分输入功能，请使用以下命令：

```
hive> set hive.input.format=org.apache.hadoop.hive.q1.io.HiveCombineSplitsInputFormat;
hive> set mapred.min.split.size=100000000;
```

Hive 0.13.1 已弃用此功能。要在 Amazon EMR AMI 版本 3.11.0 中获得相同的拆分输入格式功能，请使用以下命令：

```
set hive.hadoop.supports.plittable.combineinputformat=true;
```

## Thrift 服务端口

Thrift 是一个 RPC 框架，它定义了一种紧凑的二进制序列化格式，用于保留数据结构以供日后分析。通常，Hive 会配置服务器在以下端口上运行。

Hive 版本	端口号
Hive 0.13.1	10000
Hive 0.11.0	10004

Hive 版本	端口号
Hive 0.8.1	10003
Hive 0.7.1	10002
Hive 0.7	10001
Hive 0.5	10000

有关 Thrift 服务的更多信息，请参阅 <http://wiki.apache.org/thrift/>。

## 使用 Hive 恢复分区

Amazon EMR 包含一条采用 Hive 查询语言的语句，该语句用于从 Amazon S3 中的表数据中恢复表的分区。以下示例对此进行了介绍。

```
CREATE EXTERNAL TABLE (json string) raw_impression
PARTITIONED BY (dt string)
LOCATION 's3://elastic-mapreduce/samples/hive-ads/tables/impressions';
ALTER TABLE logs RECOVER PARTITIONS;
```

分区目录和数据必须处于表定义所指定的位置，而且必须根据 Hive 惯例命名：如 dt=2009-01-01。

### Note

Hive 0.13.1 版之后使用 `msck repair table` 在本机支持此功能，因此不支持 `recover partitions`。欲了解更多信息，请参阅 <https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DDL>。

## 将 Hive 变量传递到脚本

要使用将变量传递到 Hive 步骤 AWS CLI，请键入以下命令，替换 *myKey* 用你的 EC2 key pair 的名字，然后替换 *mybucket* 使用您的存储桶名称。在此示例中，SAMPLE 是 -d 开关后面的变量值。此变量在 Hive 脚本中的定义如下：`${SAMPLE}`。

**Note**

为了便于读取，包含 Linux 行继续符 ( \ )。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 ( ^ )。

```
aws emr create-cluster --name "Test cluster" --ami-version 3.9 \
--applications Name=Hue Name=Hive Name=Pig \
--use-default-roles --ec2-attributes KeyName=myKey \
--instance-type m3.xlarge --instance-count 3 \
--steps Type=Hive,Name="Hive Program",ActionOnFailure=CONTINUE,\
Args=[-f,s3://elasticmapreduce/samples/hive-ads/libs/response-time-stats.q,-d,\
INPUT=s3://elasticmapreduce/samples/hive-ads/tables,-d,OUTPUT=s3://mybucket/hive-ads/\
output/,\
-d,SAMPLE=s3://elasticmapreduce/samples/hive-ads/]
```

## 指定外部元存储位置

以下步骤介绍了如何覆盖 Hive 元数据仓库位置的默认配置值和使用重新配置的元数据仓库位置启动集群。

### 创建位于集群外部的元数据仓库 EMR

1. 使用亚马逊创建 My SQL 或 Aurora 数据库RDS。

有关如何创建亚马逊RDS数据库的信息，请参阅[亚马逊入门RDS](#)。

2. 修改您的安全组以允许您的数据库和 ElasticMapReduce-Master 安全组之间JDBC建立连接。

有关如何修改安全组以供访问的信息，请参阅[亚马逊RDS用户指南中的亚马逊RDS安全组](#)。

3. 在以下位置设置JDBC配置值hive-site.xml：

- a. 创建包含以下信息的 hive-site.xml 配置文件：

```
<configuration>
  <property>
    <name>javax.jdo.option.ConnectionURL</name>
    <value>jdbc:mariadb://hostname:3306/hive?createDatabaseIfNotExist=true</
value>
    <description>JDBC connect string for a JDBC metastore</description>
  </property>
  <property>
    <name>javax.jdo.option.ConnectionUserName</name>
```

```

    <value>hive</value>
    <description>Username to use against metastore database</description>
  </property>
</property>
  <name>javax.jdo.option.ConnectionPassword</name>
  <value>password</value>
  <description>Password to use against metastore database</description>
</property>
</configuration>

```

*hostname* 是运行数据库的 Amazon RDS 实例DNS的地址。*username* 以及 *password* 是您的数据库的凭证。有关连接到 My SQL 和 Aurora 数据库实例的更多信息，请参阅《亚马逊 RDS用户指南》中的[连接到运行“我的SQL数据库引擎”的数据库实例和连接 Aurora 数据库集群](#)。

JDBC驱动程序由 Amazon 安装EMR。

#### Note

值属性不应该包含任何空格或回车。所有内容应显示在一行中。

- b. 将 hive-site.xml 文件保存到 Amazon S3 中的位置上，如 `s3://mybucket/hive-site.xml`。
4. 创建一个集群，以指定自定义 hive-site.xml 文件的 Amazon S3 位置。

以下示例命令演示了 AWS CLI 执行此操作的命令。

#### Note

为了便于读取，包含 Linux 行继续符 (\)。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (^)。

```

aws emr create-cluster --name "Test cluster" --ami-version 3.10 \
--applications Name=Hue Name=Hive Name=Pig \
--use-default-roles --ec2-attributes KeyName=myKey \
--instance-type m3.xlarge --instance-count 3 \
--bootstrap-actions Name="Install Hive Site Configuration",\
Path="s3://region.elasticmapreduce/libs/hive/hive-script",\
Args=["--base-path","s3://elasticmapreduce/libs/hive","--install-hive-site",\

```

```
"--hive-site=s3://mybucket/hive-site.xml", "--hive-versions", "latest"]
```

## 使用 Connect 到 Hive JDBC

要通过连接到 Hive，JDBC需要下载JDBC驱动程序并安装SQL客户端。以下示例演示了使用 SQL Workbench/J 通过以下方式连接到 Hive。JDBC

### 要下载JDBC驱动程序

1. 下载并解压适用于您想访问的 Hive 版本的驱动程序。根据您在创建 Amazon EMR 集群时选择 AMI的版本，Hive 版本会有所不同。
  - [Hive 0.13.1 驱动程序：\\_1.0. JDBC 4.1004.zip](https://amazon-odbc-jdbc-drivers.s3.amazonaws.com/public/AmazonHiveJDBC) [https://amazon-odbc-jdbc-drivers.s3.amazonaws.com/public/ AmazonHive JDBC](https://amazon-odbc-jdbc-drivers.s3.amazonaws.com/public/AmazonHiveJDBC)
  - Hive 0.11.0 JDBC 驱动程序：<https://mvnrepository.com/artifact/org.apache.hive/hive-jdbc/0.11.0>
  - Hive 0.8.1 JDBC 驱动程序：<https://mvnrepository.com/artifact/org.apache.hive/hive-jdbc/0.8.1>
2. 安装 SQL Workbench/J。有关更多信息，请参阅 Workbench/J 手册用户手册中的[安装和启动 SQL Workbench/J](#)。SQL
3. 创建SSH通往集群主节点的隧道。连接端口因 Hive 版本而异。下表中提供了适用于 Linux ssh 用户的命令示例，为 Windows 用户提供了 Pu TTY 命令示例

### Linux SSH 命令

Hive 版本	命令
0.13.1	<code>ssh -o ServerAliveInterval=10 -i <i>path-to-key-file</i> -N -L 10000:localhost:10000 hadoop@ <i>master-public-dns-name</i></code>
0.11.0	<code>ssh -o ServerAliveInterval=10 -i <i>path-to-key-file</i> -N -L 10004:localhost:10004 hadoop@ <i>master-public-dns-name</i></code>
0.8.1	<code>ssh -o ServerAliveInterval=10 -i <i>path-to-key-file</i> -N -L 10003:localhost:10003 hadoop@ <i>master-public-dns-name</i></code>
0.7.1	<code>ssh -o ServerAliveInterval=10 -i <i>path-to-key-file</i> -N -L 10002:localhost:10002 hadoop@ <i>master-public-dns-name</i></code>

Hive 版本	命令
0.7	<code>ssh -o ServerAliveInterval=10 -i <i>path-to-key-file</i> -N -L 10001:localhost:10001 hadoop@ <i>master-public-dns-name</i></code>
0.5	<code>ssh -o ServerAliveInterval=10 -i <i>path-to-key-file</i> -N -L 10000:localhost:10000 hadoop@ <i>master-public-dns-name</i></code>

### Windows PuTTY 隧道设置

Hive 版本	隧道设置
0.13.1	源端口 : 10000 目的地 : <i>master-public-dns-name</i> :10000
0.11.0	源端口 : 10004 目的地 : <i>master-public-dns-name</i> :10004
0.8.1	源端口 : 10003 目的地 : <i>master-public-dns-name</i> :10003

4. 将JDBC驱动程序添加到 SQL Workbench。
  - a. 在 Select Connection Profile (选择连接配置文件) 对话框中，选择 Manage Drivers (管理驱动程序)。
  - b. 选择 Create a new entry (创建新条目) (空白页) 图标。
  - c. 在名称字段中，键入 **Hive JDBC**。
  - d. 在“库”中，单击“选择JAR文件”图标。
  - e. 选择下表所示的JAR文件。

Hive 驱动程序版本	JAR要添加的文件
0.13.1	<pre>hive_metastore.jar hive_service.jar HiveJDBC3.jar libfb303-0.9.0.jar libthrift-0.9.0.jar log4j-1.2.14.jar ql.jar slf4j-api-1.5.8.jar</pre>



Hive 驱动程序版本	JAR要添加的文件
	<pre>slf4j-log4j12-1.5.8.jar TCLIServiceClient.jar</pre>
0.11.0	<pre>hadoop-core-1.0.3.jar hive-exec-0.11.0.jar hive-jdbc-0.11.0.jar hive-metastore-0.11.0.jar hive-service-0.11.0.jar libfb303-0.9.0.jar commons-logging-1.0.4.jar slf4j-api-1.6.1.jar</pre>
0.8.1	<pre>hadoop-core-0.20.205.jar hive-exec-0.8.1.jar hive-jdbc-0.8.1.jar hive-metastore-0.8.1.jar hive-service-0.8.1.jar libfb303-0.7.0.jar libthrift-0.7.0.jar log4j-1.2.15.jar slf4j-api-1.6.1.jar slf4j-log4j12-1.6.1.jar</pre>
0.7.1	<pre>hadoop-0.20-core.jar hive-exec-0.7.1.jar hive-jdbc-0.7.1.jar hive-metastore-0.7.1.jar hive-service-0.7.1.jar libfb303.jar commons-logging-1.0.4.jar slf4j-api-1.6.1.jar slf4j-log4j12-1.6.1.jar</pre>

Hive 驱动程序版本	JAR要添加的文件
0.7	<pre> hadoop-0.20-core.jar hive-exec-0.7.0.jar hive-jdbc-0.7.0.jar hive-metastore-0.7.0.jar hive-service-0.7.0.jar libfb303.jar commons-logging-1.0.4.jar slf4j-api-1.5.6.jar slf4j-log4j12-1.5.6.jar </pre>
0.5	<pre> hadoop-0.20-core.jar hive-exec-0.5.0.jar hive-jdbc-0.5.0.jar hive-metastore-0.5.0.jar hive-service-0.5.0.jar libfb303.jar log4j-1.2.15.jar commons-logging-1.0.4.jar </pre>

- f. 在 Please select one driver (请选择一个驱动程序) 对话框中，根据下表选择一个驱动程序并单击 OK (确定)。

Hive 版本	驱动程序类名
0.13.1	<code>com.amazon.hive.jdbc3.HS2Driver</code>
0.11.0	<code>org.apache.hadoop.hive.jdbc.HiveDriver.jar</code>
0.8.1	<code>org.apache.hadoop.hive.jdbc.HiveDriver.jar</code>
0.7.1	<code>org.apache.hadoop.hive.jdbc.HiveDriver.jar</code>

Hive 版本	驱动程序类名
0.7	<code>org.apache.hadoop.hive.jdbc.HiveDriver.jar</code>
0.5	<code>org.apache.hadoop.hive.jdbc.HiveDriver.jar</code>

5. 返回到“选择连接配置文件”对话框时，请确认“驱动程序”字段已设置为 Hive，JDBC并根据下表在该URL字段中提供JDBC连接字符串。

Hive 版本	JDBC连接字符串
0.13.1	<code>jdbc:hive2://localhost:10000/default</code>
0.11.0	<code>jdbc:hive://localhost:10004/default</code>
0.8.1	<code>jdbc:hive://localhost:10003/default</code>

如果您的集群使用AMI版本 3.3.1 或更高版本，请在选择连接配置文件对话框**hadoop**中键入用户名字段。

## HBase早期AMI版本的 Amazon 的应用程序细节 EMR

### 支持的HBase版本

HBase版本	AMI版本	AWS CLI 配置参数	HBase版本详情
<a href="#">0.94.18</a>	3.1.0 和更高版本	<code>--ami-version 3.1</code>	<ul style="list-style-type: none"> <li>错误修复和增强功能。</li> </ul>
		<code>--ami-version 3.2</code>	
		<code>--ami-version 3.3</code>	

HBase版本	AMI版本	AWS CLI 配置参数	HBase版本详情
		--applications Name=HBase	
<a href="#">0.94.7</a>	3.0-3.0.4	--ami-version 3.0  --applications Name=HBase	
<a href="#">0.92</a>	2.2 和更高版本	--ami-version 2.2 or later  --applications Name=HBase	

## HBase集群先决条件

使用 Amazon 2.x 和 3.x EMR AMI 版本创建的集群应满足以下要求。HBase

- AWS CLI ( 可选 ) -要HBase使用命令行进行交互，请下载并安装最新版本的 AWS CLI。有关更多信息，请参阅AWS Command Line Interface 《用户指南》中的[安装 AWS Command Line Interface](#)。
- 至少两个实例 ( 可选 ) — 群集的主节点运行HBase主服务器和 Zookeeper，任务节点运行HBase区域服务器。为了获得最佳性能，HBase集群应至少在两个EC2实例上运行，但出于评估目的，您可以在单个节点HBase上运行。
- 长时间运行的集群-HBase 仅在长时间运行的集群上运行。默认情况下，CLI和 Amazon EMR 控制台会创建长时间运行的集群。
- Amazon EC2 密钥对集 ( 推荐 ) -要使用安全外壳 (SSH) 网络协议连接主节点并运行 HBase shell 命令，您必须在创建集群时使用 Amazon EC2 密钥对。
- 正确的版本AMI和 Hadoop 版本——目前只有 Hadoop 20.205 或更高版本支持HBase集群。
- Ganglia ( 可选 ) -要监控HBase性能指标，请在创建集群时安装 Ganglia。
- 用于存储日志的 Amazon S3 存储桶 ( 可选 ) -主节点上提供日志。HBase如果要将这些日志复制到 Amazon S3 中，请在创建集群时指定接收日志文件的 S3 存储桶。

## 使用 HBase 创建集群

下表列出了使用 Amazon EMR AMI 发行版使用控制台创建集群时可用的选项。HBase

字段	操作
Restore from backup (从备份中还原)	指定是否使用存储在 Amazon S3 中的数据预加载HBase集群。
Backup location (备份位置)	指定要从中恢复的备份在 Amazon S3 中的位置。URI
Backup version (备份版本)	在 Backup Location (备份位置) 处指定要使用的备份的版本名称 ( 可选 )。如果您将此字段留空，Amazon 将EMR使用备份位置的最新备份来填充新HBase集群。
Schedule Regular Backups (安排定期备份)	指定是否要计划自动增量备份。第一次备份是完整备份，为以后的增量备份创建基线。
Consistent backup (一致性备份)	请指定备份是否应该保持一致。一致性备份指的是，在初始备份阶段为保持节点同步而暂停写入操作的备份。将任何因此而暂停的写入操作放入队列中，然后在同步完成时恢复操作。
Backup frequency (备份频率)	计划备份间隔的天/小时/分钟数。
Backup location (备份位置)	存储备份URI的 Amazon S3。每个HBase群集的备份位置应不同，以确保差异备份保持正确。
Backup start time (备份启动时间)	指定进行首次备份的时间。您可以将其设置为now，这样在集群运行后立即启动第一次备份，或者以 <a href="#">ISO格式</a> 输入日期和时间。例如，2012-06-15T20:00Z 会将开始时间设置为 2012 年 6 月 15 日晚上 8 点。UTC

以下示例 AWS CLI 命令启动带有HBase和其他应用程序的集群：

**Note**

为了便于读取，包含 Linux 行继续符 (\)。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (^)。

```
aws emr create-cluster --name "Test cluster" --ami-version 3.3 \
  --applications Name=Hue Name=Hive Name=Pig Name=HBase \
  --use-default-roles --ec2-attributes KeyName=myKey \
  --instance-type c1.xlarge --instance-count 3 --termination-protected
```

在 Hive 和 HBase 集群之间建立连接后（如前面的过程所示），您可以通过在 Hive 中创建外部表来访问存储在 HBase 集群上的数据。

以下示例在 Hive 提示符下运行时，会创建一个外部表，该表引用存储在名为 inputTable 的 HBase 表中的数据。然后，您可以 inputTable 在 Hive 语句中引用以查询和修改存储在 HBase 集群中的数据。

**Note**

以下示例在 2. AMI 3.3 中使用 p rotobuf-java-2.4.0a.jar，但您应该修改示例以匹配您的版本。要检查 JAR 您拥有哪个版本的协议缓冲区，请在 Hive 命令提示符下运行命令：`! ls /home/hadoop/lib;`

```
add jar lib/emr-metrics-1.0.jar ;
add jar lib/protobuf-java-2.4.0a.jar ;

set hbase.zookeeper.quorum=ec2-107-21-163-157.compute-1.amazonaws.com ;

create external table inputTable (key string, value string)
  stored by 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
  with serdeproperties ("hbase.columns.mapping" = ":key,f1:col1")
  tblproperties ("hbase.table.name" = "t1");

select count(*) from inputTable ;
```

## 自定义配置 HBase

尽管默认设置应该适用于大多数应用程序，但您可以灵活地修改 HBase 配置设置。为此，请运行两种引导操作脚本之一：

- `configure-hbase-daemons`— 配置主守护程序、区域服务器守护程序和 zookeeper 守护程序的属性。这些属性包括堆大小和HBase守护程序启动时传递给 Java 虚拟机 (JVM) 的选项。将这些属性设置为引导操作中的参数。此引导操作修改了集群上的 `/home/hadoop/hbase-user-env/conf/.sh` 配置文件。HBase
- `configure HBase-hbase` —配置特定于站点的设置，例如HBase主服务器应绑定的端口以及客户端客户端应重试操作的最大次数。CLI您可以将它们 one-by-one设置为引导操作中的参数，也可以在 Amazon S3 中指定XML配置文件的位置。此引导操作会修改集群上的 `/home/hadoop/conf/hbase-site.xml` 配置文件。HBase

### Note

与其他引导操作一样，这些脚本只能在创建集群时运行；您不能使用它们来更改当前正在运行的HBase集群的配置。

运行 `configure re-hbase` 或 `configure-hbase-daemons` 引导操作时，您指定的值会覆盖默认值。任何未显式设置的值都会接受默认值。

使用这些引导操作HBase进行配置类似于在 Amazon 中使用引导操作EMR来配置 Hadoop 设置和 Hadoop 守护程序属性。不同之处在于HBase没有每进程内存选项。取而代之的是，使用 `--daemon-opts` 参数设置内存选项，其中 `daemon` 被要配置的守护程序的名称所取代。

## 配置HBase守护程序

Amazon EMR 提供了一个引导操作 `s3://region.elasticmapreduce/bootstrap-actions/configure-hbase-daemons`，您可以使用它来更改HBase守护程序的配置，其中 `region` 是您要在其中启动HBase集群的区域。

要使用配置HBase守护程序 AWS CLI，请在启动集群时添加 `configure-hbase-daemons` 引导操作以配置一个或多个HBase守护程序。您可以设置以下属性。

属性	描述
<code>hbase-master-opts</code>	控制主守护程序JVM运行方式的选项。如果已设置，则这些变量将覆盖默认的 <code>HBASE _ MASTER _ OPTS</code> 变量。

属性	描述
regionserver-opts	控制如何JVM运行区域服务器守护程序的选项。如果已设置，则这些变量将覆盖默认的 HBASE _ REGIONSERVER _ OPTS 变量。
zookeeper-opts	控制如何JVM运行 zookeeper 守护程序的选项。如果已设置，则这些变量将覆盖默认的 HBASE _ ZOOKEEPER _ OPTS 变量。

有关这些选项的更多信息，请参阅HBase文档中的 [hbase-env.sh](#)。

以下示例中显示的是为 zookeeper-opts 和 hbase-master-opts 配置值的引导操作。

#### Note

为了便于读取，包含 Linux 行继续符 ( \ )。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (^)。

```
aws emr create-cluster --name "Test cluster" --ami-version 3.3 \
--applications Name=Hue Name=Hive Name=Pig Name=HBase \
--use-default-roles --ec2-attributes KeyName=myKey \
--instance-type c1.xlarge --instance-count 3 --termination-protected \
--bootstrap-actions Path=s3://elasticmapreduce/bootstrap-actions/configure-hbase-
daemons,\
Args=["--hbase-zookeeper-opts=-Xmx1024m -XX:GCTimeRatio=19", "--hbase-master-opts=-
Xmx2048m", "--hbase-regionserver-opts=-Xmx4096m"]
```

## 配置HBase站点设置

Amazon EMR 提供了一个引导操作 `s3://elasticmapreduce/bootstrap-actions/configure-hbase`，您可以使用该操作来更改的配置。HBase 您可以将配置值 `one-by-one` 设置为引导操作中的参数，也可以在 Amazon S3 中指定 XML 配置文件的位置。如果您只需要设置几个配置设置，`one-by-one` 则设置配置值非常有用。如果您要进行许多更改，或者要保存配置设置以供重复使用，则使用 XML 文件进行设置会很有用。



**Note**

您可以在 Amazon S3 存储桶名称前加上区域前缀，例如 `s3://region.elasticmapreduce/bootstrap-actions/configure-hbase`，其中 *region* 是您要在其中启动HBase集群的区域。

此引导操作修改集群上的 `/home/hadoop/conf/hbase-site.xml` 配置文件。HBase引导操作只能在HBase集群启动时运行。

有关您可以配置的HBase站点设置的更多信息，请参阅HBase文档中的[默认配置](#)。

在 `configure-hbase` 启动HBase集群时设置引导操作，并在其中指定 `hbase-site.xml` 要更改的值。

要指定各个HBase站点的设置，请使用 AWS CLI

- 要更改 `hbase.hregion.max.filesize` 设置，请键入以下命令并替换 *myKey* 使用您的 Amazon EC2 密钥对名称。

**Note**


为了便于读取，包含 Linux 行继续符 (`\`)。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (`^`)。

```
aws emr create-cluster --name "Test cluster" --ami-version 3.3 \
--applications Name=Hue Name=Hive Name=Pig Name=HBase \
--use-default-roles --ec2-attributes KeyName=myKey \
--instance-type c1.xlarge --instance-count 3 --termination-protected \
--bootstrap-actions Path=s3://elasticmapreduce/bootstrap-actions/configure-
hbase,Args=["-s", "hbase.hregion.max.filesize=52428800"]
```

要使用XML文件指定HBase站点设置，请使用 AWS CLI

- 创建 `hbase-site.xml` 的自定义版本。您的自定义文件必须有效XML。为了减少引入错误的机会，请从位于 Amazon EMR HBase 主节点的 `hbase-site.xml` 默认副本开始 `/home/hadoop/conf/hbase-site.xml`，然后编辑该文件的副本，而不是从头开始创建文件。可以给您的新文件指定一个新名称，或保留 `hbase-site.xml`。

2. 将您的自定义 `hbase-site.xml` 文件上传到 Amazon S3 存储桶。它应该设置权限，以便启动集群的 AWS 账户可以访问该文件。如果启动集群的 AWS 账户还拥有 Amazon S3 存储桶，则该账户有权访问。
3. 在启动集HBase群时设置 `configure-hbase` 引导操作，并包括自定义文件的位置。`hbase-site.xml`以下示例将HBase站点配置值设置为文件中指定的值`s3://mybucket/my-hbase-site.xml`。键入以下命令，替换 *myKey* 用你的 EC2 key pair 的名字，然后替换 *mybucket* 使用您的 Amazon S3 存储桶的名称。

 Note

为了便于读取，包含 Linux 行继续符 (\)。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (^)。

```
aws emr create-cluster --name "Test cluster" --ami-version 3.3 \
  --applications Name=Hue Name=Hive Name=Pig Name=HBase \
  --use-default-roles --ec2-attributes KeyName=myKey \
  --instance-type c1.xlarge --instance-count 3 --termination-protected \
  --bootstrap-actions Path=s3://elasticmapreduce/bootstrap-actions/configure-
hbase,Args=["--site-config-file","s3://mybucket/config.xml"]
```

如果您指定多个选项来自定义HBase操作，则必须在每个键值对前面加上一个 `-s` 选项开关，如下示例所示：

```
--bootstrap-actions s3://elasticmapreduce/bootstrap-actions/configure-
hbase,Args=["-s","zookeeper.session.timeout=60000"]
```

设置代理并打开SSH连接后，您可以通过 `http://`打开浏览器窗口来查看HBase用户界面`master-public-dns-name:60010/master-status`，其中 `master-public-dns-name` 是HBase集群中主节点的公共DNS地址。

您可以使用连接到主节点并导航SSH到该`mnt/var/log/hbase`目录来查看当前HBase日志。除非您在集群启动时启用了针对 Amazon S3 的日志记录，否则这些日志将在集群终止后不再可用。

## 备份和恢复 HBase

Amazon EMR 允许您手动或按自动计划将您的HBase数据备份到 Amazon S3。您可以执行完整备份和增量备份。获得备份版本HBase的数据后，可以将该版本还原到集HBase群。您可以恢复到当前正在运行的HBase集群，也可以启动预先填充了备份数据的新集群。

在备份过程中，HBase继续执行写入命令。虽然这样可以确保集群在整个备份过程中都处于可用状态，但存在的风险是，正在备份的数据可能与并行执行的任何写入操作不一致。要了解可能出现的不一致性，您必须考虑在集群中的节点之间HBase分配写入操作。如果写入操作发生在轮询特定节点之后，则备份存档中不会包含该数据。您甚至可能会发现，早期对HBase集群的写入（发送到已经轮询过的节点）可能不在备份存档中，而后来的写入（在轮询之前发送到节点）则包括在内。

如果需要一致的备份，则必须在备份过程的HBase初始部分（跨节点同步）暂停写入。您可以通过在请求备份时指定 `--consistent` 参数完成此操作。指定此参数后，系统会将此期间的写入操作加入队列并在同步完成后立即执行这些操作。您还可以计划重复备份，一次备份过程中遗失的数据会在后续备份过程中得到备份，从而解决随时间推移而产生的任何不一致问题。

备份HBase数据时，应为每个群集指定不同的备份目录。完成此操作的一种简单方法是，将集群标识符用作备份目录指定路径的一部分。例如，`s3://mybucket/backups/j-3AEXXXXXX16F2`。这样可以确保将来的任何 future 增量备份都引用正确的HBase群集。

当您准备删除不再需要的旧备份文件时，我们建议您先对HBase数据进行完整备份。这样可以确保保留所有数据并为以后的增量备份提供基线。完成完整备份后，您可以导航到备份位置并手动删除旧备份文件。

HBase备份过程使用 S3 DistCp 进行复制操作，这在临时文件存储空间方面有一定的限制。

### HBase使用控制台进行备份和恢复

控制台允许启动新集群并使用先前HBase备份中的数据填充该集群。它还使您能够计划HBase数据的定期增量备份。其他备份和还原功能，例如将数据还原到已在运行的集群、进行手动备份和安排自动完整备份的功能，可使用使用CLI。

#### 使用控制台向新集群填充存档HBase数据

1. 导航到新的 Amazon EMR 控制台，然后从侧面导航栏中选择“切换到旧主机”。有关切换到旧控制台后预期情况的更多信息，请参阅 [Using the old console](#)。
2. 选择创建集群。
3. 在“软件配置”部分，对于“其他应用程序”，选择“HBase配置”，然后添加。

4. 在 Add Application (添加应用程序) 对话框中，选中 Restore From Backup (从备份中恢复)。
5. 在 Backup Location 中，指定要加载到新HBase集群中的备份的位置。这应该是这种形式的 Amazon S3 URL `s3://myawsbucket/backups/`。
6. 对于 Backup Version (备份版本)，您可以选择通过设置值来指定要加载的备份版本的名称。如果您没有为备份版本设置值，Amazon 会将最新备份EMR加载到指定位置。
7. 选择 Add (添加) 并继续使用所需的其他选项创建集群。

### 使用控制台安排HBase数据的自动备份

1. 在“软件配置”部分，对于“其他应用程序”，选择“HBase配置”，然后添加。
2. 选择 Schedule Regular Backups (安排定期备份)。
3. 请指定备份是否应该保持一致。一致性备份指的是，在初始备份阶段为保持节点同步而暂停写入操作的备份。将任何因此而暂停的写入操作放入队列中，然后在同步完成时恢复操作。
4. 通过在 Backup Frequency (备份频率) 中输入一个数字并选择 Days (天)、Hours (小时数) 或 Minutes (分钟数)来设置备份频率。运行的第一个自动备份是完整备份；之后，Amazon EMR 会根据您指定的计划保存增量备份。
5. 指定 Amazon S3 中应当存储备份的位置。应将每个HBase集群备份到 Amazon S3 中的一个单独位置，以确保正确计算增量备份。
6. 通过为 Backup Start Time (备份启动时间) 设置一个值，指定应当何时进行第一次备份。您可以将其设置为now，这样在集群运行后立即启动第一次备份，或者以[ISO格式](#)输入日期和时间。例如，2013-09-26T20:00Z 将开始时间设置为 2013 年 9 月 26 日晚上 8 点。UTC
7. 选择添加。
8. 根据需要使用其他选项创建集群。

### HBase使用监视器 CloudWatch

Amazon EMR 报告了三个指标 CloudWatch ，您可以使用这些指标来监控HBase备份。这些指标每隔五 CloudWatch 分钟推送一次，并且免费提供。

指标	描述
HBaseBackupFailed	最后一次备份是否失败。默认设置为 0，如果上一次备份尝试失败，则更新为 1。此指标仅针对HBase集群报告。  用例：监控HBase备份

指标	描述
	单位：计数
HBaseMostRecentBackupDuration	完成上一次备份所需的时长。无论最后完成的备份成功或失败，都会设置此指标。进行备份的同时，此指标返回备份开始之后的分钟数。此指标仅针对HBase集群报告。  用例：监控HBase备份  单位：分钟
HBaseTimeSinceLastSuccessfulBackup	在您的集群上次成功启动HBase备份后经过的分钟数。此指标仅针对HBase集群报告。  用例：监控HBase备份  单位：分钟

## 将 Ganglia 配置为 HBase

您可以将 Ganglia 配置为HBase使用configure-hbase-for-ganglia引导操作。此引导操作配置为向 Ganglia 发布指标。

启动集群时必须配置HBase和 Ganglia；无法将 Ganglia 报告添加到正在运行的集群中。

Ganglia 还可以将日志文件存储在 /mnt/var/log/ganglia/rrds 处的服务器上。如果您配置了集群以将日志文件保存到 Amazon S3 存储桶，Ganglia 日志文件也会保存在那里。

要启动带有 Ganglia 的集群HBase，请使用configure-hbase-for-ganglia引导操作，如下例所示。

### Note

为了便于读取，包含 Linux 行继续符 (\)。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (^)。

```
aws emr create-cluster --name "Test cluster" --ami-version 3.3 \
--applications Name=Hue Name=Hive Name=Pig Name=HBase Name=Ganglia \
--use-default-roles --ec2-attributes KeyName=myKey \
--instance-type c1.xlarge --instance-count 3 --termination-protected \
```

```
--bootstrap-actions Path=s3://elasticmapreduce/bootstrap-actions/configure-hbase-for-ganglia
```

在配置了 Ganglia 的情况下启动集群后，您就可以使用主节点上运行的图形界面来访问 Ganglia 图形和报告。

## 早期AMI版本的 Amazon 的 Pig 应用程序细节 EMR

### 支持的 Pig 版本

您可以添加到集群的 Pig 版本取决于您使用的 Amazon 版本EMRAMI和您使用的 Hadoop 版本。下表显示了哪些AMI版本和版本的 Hadoop 与不同版本的 Pig 兼容。我们建议使用最新版本的 Pig，以便利用各种性能增强和新的功能。

使用安装 Pig 时，API将使用默认版本，除非您在调用期间--pig-versions将 Pig 加载到集群的步骤中指定为参数[RunJobFlow](#)。

Pig 版本	AMI版本	配置参数	Pig 版本详细信息
0.12.0 <a href="#">发布说明</a> <a href="#">文档</a>	3.1.0 和更高版本	--ami-version 3.1  --ami-version 3.2  --ami-version 3.3	添加对以下各项的支持：  <ul style="list-style-type: none"> <li>UDFs无需JVM实现即可进行直播</li> <li>ASSERT和 IN 运算符</li> <li>CASE表情</li> <li>AvroStorage 作为 Pig 的内置函数。</li> <li>ParquetLoader 并 ParquetStorer 作为内置函数</li> <li>BigInteger 和 BigDecimal 类型</li> </ul>
0.11.1.1 <a href="#">发布说明</a>	2.2 和更高版本	--pig-versions 0.11.1.1	PigStorage 如果输入位于 Amazon S3 中，

Pig 版本	AMI版本	配置参数	Pig 版本详细信息
<a href="#">文档</a>		<code>--ami-version</code> 2.2	则可以提高LOAD命令的性能。
0.11.1 <a href="#">发布说明</a> <a href="#">文档</a>	2.2 和更高版本	<code>--pig-versions</code> 0.11.1  <code>--ami-version</code> 2.2	增加了对 JDK 7、Hadoop 2、Groovy 用户定义函数、SchemaTuple 优化、新运算符等的支持。有关更多信息，请参阅 <a href="#">Pig 0.11.1 更改日志</a> 。
0.9.2.2 <a href="#">发布说明</a> <a href="#">文档</a>	2.2 和更高版本	<code>--pig-versions</code> 0.9.2.2  <code>--ami-version</code> 2.2	添加了对于 Hadoop 1.0.3 的支持。
0.9.2.1 <a href="#">发布说明</a> <a href="#">文档</a>	2.2 和更高版本	<code>--pig-versions</code> 0.9.2.1  <code>--ami-version</code> 2.2	增加了对 MapR 的支持。
0.9.2 <a href="#">发布说明</a> <a href="#">文档</a>	2.2 和更高版本	<code>--pig-versions</code> 0.9.2  <code>--ami-version</code> 2.2	包括多项性能改进和错误修复。有关 Pig 0.9.2 的更改的完整信息，请转到 <a href="#">Pig 0.9.2 更改日志</a> 。
0.9.1 <a href="#">发布说明</a> <a href="#">文档</a>	2.0	<code>--pig-versions</code> 0.9.1  <code>--ami-version</code> 2.0	

Pig 版本	AMI版本	配置参数	Pig 版本详细信息
0.6 <a href="#">发布说明</a>	1.0	<code>--pig-versions 0.6</code> <code>--ami-version 1.0</code>	
0.3 <a href="#">发布说明</a>	1.0	<code>--pig-versions 0.3</code> <code>--ami-version 1.0</code>	

## Pig 版本详细信息

亚马逊EMR支持某些 Pig 版本，这些版本可能会应用其他亚马逊EMR补丁。您可以配置在 Amazon EMR 集群上运行哪个版本的 Pig。有关此操作的详细信息，请参阅 [Apache Pig](#)。以下各节介绍不同的 Pig 版本以及应用于 Amazon 上加载的版本的补丁EMR。

### Pig 修补

本节介绍应用于 Amazon 上提供的 Pig 版本的自定义补丁EMR。

#### Pig 0.11.1.1 修补

亚马逊EMR版本的 Pig 0.11.1.1 是一个维护版本，PigStorage 如果输入位于亚马逊 S3 中，则可以提高LOAD命令性能。

#### Pig 0.11.1 补丁

亚马逊EMR版本的 Pig 0.11.1 包含 Apache 软件基金会提供的所有更新以及 Pig 版本 0.9.2.2 中的累积亚马逊EMR补丁。但是，Pig 0.11.1 中没有新的亚马逊EMR专用补丁。

#### Pig 0.9.2 补丁

Apache Pig 0.9.2 是 Pig 的维护版。亚马逊EMR团队已对亚马逊EMR版本的 Pig 0.9.2 应用了以下补丁。



修补	描述
PIG-1429	<p>将布尔数据类型以第一个类数据类型的形式添加到 Pig。欲了解更多信息，请访问 <a href="https://issues.apache.org/jira/browse/PIG-1429">https://issues.apache.org/jira/browse/PIG-1429</a>。</p> <p>状态：Committed (已提交)</p> <p>已在以下 Apache Pig 版本中修复：0.10</p>
PIG-1824	<p>支持 Jython UDF 中的导入模块。欲了解更多信息，请访问 <a href="https://issues.apache.org/jira/browse/PIG-1824">https://issues.apache.org/jira/browse/PIG-1824</a>。</p> <p>状态：Committed (已提交)</p> <p>已在以下 Apache Pig 版本中修复：0.10</p>
PIG-2010	<p>在分布式缓存JARs上注册的捆绑包。欲了解更多信息，请访问 <a href="https://issues.apache.org/jira/browse/PIG-2010">https://issues.apache.org/jira/browse/PIG-2010</a>。</p> <p>状态：Committed (已提交)</p> <p>已在以下 Apache Pig 版本中修复：0.11</p>
PIG-2456	<p>添加 ~/.pigbootstrap 文件，用户可以在其中指定默认的 Pig 语句。欲了解更多信息，请访问 <a href="https://issues.apache.org/jira/browse/PIG-2456">https://issues.apache.org/jira/browse/PIG-2456</a>。</p> <p>状态：Committed (已提交)</p> <p>已在以下 Apache Pig 版本中修复：0.11</p>
PIG-2623	<p>Support 支持使用 Amazon S3 路径进行注册UDFs。欲了解更多信息，请访问 <a href="https://issues.apache.org/jira/browse/PIG-2623">https://issues.apache.org/jira/browse/PIG-2623</a>。</p> <p>状态：Committed (已提交)</p> <p>已在以下 Apache Pig 版本中修复：0.10、0.11</p>

## Pig 0.9.1 补丁

亚马逊EMR团队已对亚马逊EMR版本的 Pig 0.9.1 应用了以下补丁。

修补	描述
dfs 中的 Support JAR 文件和 Pig 脚本	<p>添加对运行脚本和注册JAR存储在 HDFS Amazon S3 或其他分布式文件系统中的文件的支持。欲了解更多信息，请访问 <a href="https://issues.apache.org/jira/browse/PIG-1505">https://issues.apache.org/jira/browse/ PIG -1505</a>。</p> <p>状态：Committed (已提交)</p> <p>已在以下 Apache Pig 版本中修复：0.8.0</p>
支持 Pig 中的多个文件系统	<p>添加对于 Pig 脚本的支持，以便从一个文件系统中读取数据，并写入另一个文件系统。欲了解更多信息，请访问 <a href="https://issues.apache.org/jira/browse/PIG-1564">https://issues.apache.org/jira/browse/ PIG -1564</a>。</p> <p>状态：Not Committed (未提交)</p> <p>已在以下 Apache Pig 版本中修复：不适用</p>
添加 Piggybank 日期时间和字符串 UDFs	<p>添加日期时间和字符串UDFs以支持自定义 Pig 脚本。欲了解更多信息，请访问 <a href="https://issues.apache.org/jira/browse/PIG-1565">https://issues.apache.org/jira/browse/ PIG -1565</a>。</p> <p>状态：Not Committed (未提交)</p> <p>已在以下 Apache Pig 版本中修复：不适用</p>

## 交互式的和批处理的 Pig 集群

Amazon EMR 允许您以两种模式运行 Pig 脚本：

- 交互式
- 批处理

当你使用控制台或启动长时间运行的集群时 AWS CLI，你可以以 Hadoop 用户的身份使用连接到主节点，ssh然后使用 Grunt shell 以交互方式开发和运行你的 Pig 脚本。以交互方式使用 Pig，您就可以比批处理方式更轻松地修改 Pig 脚本。以交互模式成功修改 Pig 脚本之后，可以将脚本上传到 Amazon

S3，并使用批处理模式在生产环境中运行该脚本。您还可以根据需要，在正在运行的集群中以交互方式提交 Pig 命令来分析和转换数据。

在批处理模式下，需要将 Pig 脚本上传到 Amazon S3，然后作为操作步骤将此工作提交到集群。Pig 步骤可提交到长时间运行的集群或临时集群。

## 早期AMI版本的 Amazon 的 Spark 应用程序细节 EMR

### 以交互方式或批处理模式使用 Spark

Amazon EMR 允许您在两种模式下运行 Spark 应用程序：

- 交互式
- 批处理

使用控制台或启动长时间运行的集群时 AWS CLI，您可以以 Hadoop 用户身份使用SSH连接到主节点，并使用 Spark 外壳以交互方式开发和运行 Spark 应用程序。与批处理环境相比，以交互方式使用 Spark 能够让您更轻松地对 Spark 应用程序进行原型设计或测试。在交互模式下成功修改 Spark 应用程序后，您可以将该应用程序JAR或 Python 程序放入 Amazon S3 上集群主节点本地的文件系统中。然后，您可以将应用程序作为批处理工作流程提交。

在批处理模式中，将 Spark 脚本上传到 Amazon S3 或本地主节点文件系统，然后将此工作作为步骤提交到集群。Spark 步骤可提交到长时间运行的集群或暂时性集群。

### 创建安装了 Spark 的集群

使用控制台启动安装了 Spark 的集群

1. 导航到新的 Amazon EMR 控制台，然后从侧面导航栏中选择“切换到旧主机”。有关切换到旧控制台后预期情况的更多信息，请参阅 [Using the old console](#)。
2. 选择创建集群。
3. 在“软件配置”中，选择所需的AMI发行版本。
4. 对于 Applications to be installed (要安装的应用程序)，从列表中选择 Spark，然后选择 Configure and add (配置并添加)。
5. 添加参数以按需更改 Spark 配置。有关更多信息，请参阅 [配置 Spark](#)。选择添加。
6. 根据需要选择其它选项，然后选择 Create cluster (创建集群)。

以下示例演示如何使用 Java 创建带 Spark 的集群：

```
AmazonElasticMapReduceClient emr = new AmazonElasticMapReduceClient(credentials);
SupportedProductConfig sparkConfig = new SupportedProductConfig()
    .withName("Spark");

RunJobFlowRequest request = new RunJobFlowRequest()
    .withName("Spark Cluster")
    .withAmiVersion("3.11.0")
    .withNewSupportedProducts(sparkConfig)
    .withInstances(new JobFlowInstancesConfig()
        .withEc2KeyName("myKeyName")
        .withInstanceCount(1)
        .withKeepJobFlowAliveWhenNoSteps(true)
        .withMasterInstanceType("m3.xlarge")
        .withSlaveInstanceType("m3.xlarge")
    );
RunJobFlowResult result = emr.runJobFlow(request);
```

## 配置 Spark

在创建集群时，您可以通过运行位于 Github 上的 [awslabs emr-bootstrap-actions //s park](#) 存储库中的引导操作来配置 Spark。有关引导操作接受的参数，请参阅该存储库 [README](#) 中的。引导操作配置 \$SPARK\_CONF\_DIR/spark-defaults.conf 文件中的属性。有关设置的更多信息，请参阅 Spark 文档中的 Spark 配置主题。例如，您可以将以下内容 URL 中的“最新”替换为要安装的 Spark 版本号 2.2.0 <http://spark.apache.org/docs/latest/configuration.html>。

您也可以在每次提交应用程序时动态配置 Spark。使用 spark 配置文件提供了便于执行程序自动充分利用资源分配的设置。有关更多信息，请参阅 [覆盖 Spark 默认配置设置](#)。

### 更改 Spark 默认设置

以下示例演示如何使用 AWS CLI 创建 spark.executor.memory 设置为 2G 的集群。

#### Note

为了便于读取，包含 Linux 行继续符 (\)。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (^)。

```
aws emr create-cluster --name "Spark cluster" --ami-version 3.11.0 \
--applications Name=Spark, Args=[-d,spark.executor.memory=2G] --ec2-attributes
KeyName=myKey \
```

```
--instance-type m3.xlarge --instance-count 3 --use-default-roles
```

## 向 Spark 提交工作

要向集群提交工作，请使用步骤在您的EMR集群上运行spark-submit脚本。使用中的addJobFlowSteps方法添加步骤 [AmazonElasticMapReduceClient](#)：

```
AWSCredentials credentials = new BasicAWSCredentials(accessKey, secretKey);
AmazonElasticMapReduceClient emr = new AmazonElasticMapReduceClient(credentials);
StepFactory stepFactory = new StepFactory();
AddJobFlowStepsRequest req = new AddJobFlowStepsRequest();
req.withJobFlowId("j-1K48XXXXXXHCB");

List<StepConfig> stepConfigs = new ArrayList<StepConfig>();

StepConfig sparkStep = new StepConfig()
    .withName("Spark Step")
    .withActionOnFailure("CONTINUE")
    .withHadoopJarStep(stepFactory.newScriptRunnerStep("/home/hadoop/spark/bin/spark-submit", "--class", "org.apache.spark.examples.SparkPi", "/home/hadoop/spark/lib/spark-examples-1.3.1-hadoop2.4.0.jar", "10"));

stepConfigs.add(sparkStep);
req.withSteps(stepConfigs);
AddJobFlowStepsResult result = emr.addJobFlowSteps(req);
```

## 覆盖 Spark 默认配置设置

建议您为不同的应用程序覆盖 Spark 默认配置值。您可以在提交应用程序时使用步骤完成此操作 (实质上是向 spark-submit 传递选项)。例如，您可能需要通过更改 spark.executor.memory 来更改为执行者进程分配的内存。您可以为 --executor-memory 开关提供与下类似的参数：

```
/home/hadoop/spark/bin/spark-submit --executor-memory 1g --class
org.apache.spark.examples.SparkPi /home/hadoop/spark/lib/spark-examples*.jar 10
```

同样地，您也可以调节 --executor-cores 和 --driver-memory。在步骤中，您可以向步骤提供以下参数：

```
--executor-memory 1g --class org.apache.spark.examples.SparkPi /home/hadoop/spark/lib/
spark-examples*.jar 10
```

您还可以使用 `--conf` 选项调节没有内置开关的设置。有关可调节的其他设置的更多信息，请参阅 Apache Spark 文档中的[动态加载 Spark 属性](#)主题。

## S3 DistCp 实用程序与早期AMI版本的 Amazon 的区别 EMR

### 亚马逊支持DistCp 的 S3 版本 EMR

Amazon DistCp 版本支持以下 S3 EMR AMI 版本。1.0.7 之后的 S3 DistCp 版本可以直接在集群上找到。使用JAR中的/home/hadoop/lib来了解最新功能。

版本	描述	发行日期
1.0.8	添加 <code>--appendToLastFile</code> 、 <code>--requirePreviousManifest</code> 和 <code>--storageClass</code> 选项。	2014 年 1 月 3 日
1.0.7	添加了 <code>--s3ServerSideEncryption</code> 选项。	2013 年 5 月 2 日
1.0.6	添加了 <code>--s3Endpoint</code> 选项。	2012 年 8 月 6 日
1.0.5	提高了指定要运行哪个 S3 DistCp 版本的能力。	2012 年 6 月 27 日
1.0.4	改进了 <code>--deleteOnSuccess</code> 选项。	2012 年 6 月 19 日
1.0.3	添加了对于 <code>--numberOfFiles</code> 和 <code>--startingIndex</code> 选项的支持。	2012 年 6 月 12 日
1.0.2	使用组时改进了文件命名。	2012 年 6 月 6 日
1.0.1	S3 的初始版本DistCp。	2012 年 1 月 19 日

## 向集群添加 S3 DistCp 复制步骤

要向正在运行的集群添加 S3 DistCp 复制步骤，请键入以下命令，替换 `j-3GYXXXXXX9I0K` 使用您的集群 ID，然后替换 `mybucket` 使用您的 Amazon S3 存储桶名称。

### Note

为了便于读取，包含 Linux 行继续符 (\)。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (^)。

```
aws emr add-steps --cluster-id j-3GYXXXXXX9I0K \
--steps Type=CUSTOM_JAR,Name="S3DistCp step",Jar=/home/hadoop/lib/emr-s3distcp-1.0.jar,
\
Args=["--s3Endpoint,s3-eu-west-1.amazonaws.com",\
"--src,s3://mybucket/logs/j-3GYXXXXXX9I0J/node/",\
"--dest,hdfs://output",\
"--srcPattern,.*[a-zA-Z,]+" ]
```

## Example 将 Amazon CloudFront 日志加载到 HDFS

此示例 HDFS 通过向正在运行的集群添加步骤来加载 Amazon CloudFront 日志。在此过程中，它将压缩格式从 Gzip (CloudFront 默认) 更改为 LZO。这很有用，因为使用压缩的数据 LZO 可以在解压缩时分成多个地图，因此您不必像使用 Gzip 那样等到压缩完成。当您使用 Amazon 分析数据时，这样可以提高性能 EMR。此示例还通过以下方式提高性能：使用在 `--groupBy` 选项中指定的正则表达式，将给定小时内的所有日志组合成为单个文件。与处理许多 GZIP 压缩的小文件相比，Amazon EMR 集群在处理少量大型 LZO 压缩文件时效率更高。要拆分 LZO 文件，必须对其进行索引并使用 `hadoop-lzo` 第三方库。

要将 Amazon CloudFront 日志加载到 HDFS，请键入以下命令，替换 `j-3GYXXXXXX9I0K` 使用您的集群 ID，然后替换 `mybucket` 使用您的 Amazon S3 存储桶名称。

### Note

为了便于读取，包含 Linux 行继续符 (\)。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (^)。

```
aws emr add-steps --cluster-id j-3GYXXXXXX9I0K \
```

```
--steps Type=CUSTOM_JAR,Name="S3DistCp step",Jar=/home/hadoop/lib/emr-s3distcp-1.0.jar,\
\  
Args=["--src,s3://mybucket/cf","--dest,hdfs:///local",\  
"--groupBy,. *XABCD12345678.([0-9]+-[0-9]+-[0-9]+-[0-9]+). *",\  
"--targetSize,128",\  
"--outputCodec,lzo","--deleteOnSuccess"]
```

假设前面的示例是在以下 CloudFront 日志文件上运行的。

```
s3://amzn-s3-demo-bucket1/cf/XABCD12345678.2012-02-23-01.HLUS3JKx.gz  
s3://amzn-s3-demo-bucket1/cf/XABCD12345678.2012-02-23-01.I9CNAZrg.gz  
s3://amzn-s3-demo-bucket1/cf/XABCD12345678.2012-02-23-02.YRRwERSA.gz  
s3://amzn-s3-demo-bucket1/cf/XABCD12345678.2012-02-23-02.dshVLXFE.gz  
s3://amzn-s3-demo-bucket1/cf/XABCD12345678.2012-02-23-02.LpLfuShd.gz
```

S3 将文件 DistCp 复制、连接和压缩到以下两个文件中，其中文件名由正则表达式的匹配确定。

```
hdfs:///local/2012-02-23-01.lzo  
hdfs:///local/2012-02-23-02.lzo
```



## 新增功能

本页介绍最新版本的 Amazon EMR 7.x、6.x 和 5.x 中提供的更改和功能。

这些发行说明也可在[亚马逊 EMR 7.2.0](#)、[亚马逊 EMR 6.15.0](#) 和 [Amazon EMR 5.36.2](#) 页面上找到，以及每个版本的应用程序版本、组件版本和可用配置分类。

- 有关早期发行版的发布说明，请参阅 [Amazon 发行说明EMR档案](#)。
- 要在亚马逊新EMR版本发布时获取更新，请订阅[亚马逊EMR发行说明的RSS提要](#)。

### Note

更高版本的 Amazon EMR 使用 AWS 签名版本 4 (SigV4) 来验证向 Amazon S3 发出的请求。我们建议您使用支持的 Amazon EMR 版本，SigV4这样您就可以访问新的 S3 存储桶并避免工作负载中断。有关更多信息以及支持的 Amazon EMR 版本列表SigV4，请参阅[Amazon EMR 和 AWS Signature 版本 4](#)。

## 亚马逊 EMR 7.2.0 ( 7.x 系列的最新版本 )

Amazon 的新EMR版本将在几天内在不同地区上市，从最初发布日期的第一个地区开始。在此期间，您所在区域可能无法提供最新发行版。

以下发行说明包含有关亚马逊 7.2.0 EMR 版本的信息。更改与 7.2.0 有关。

### 新功能

- 应用程序升级 – Amazon EMR 7.2.0 application upgrades include Iceberg 1.5.0-amzn-0 and Delta 3.1.0。
- 亚马逊EMR增加了支持，因此您可以将其他应用程序（例如 Flink 和 Hive）与 HBase Amazon S3 Express One Zone 存储类一起使用。
- 此版本增加了读取已恢复对象的功能，因此您可以使用S3A协议从 S3 位置读取 Glacier 对象。此功能适用于 Spark、Flink 和 Hive。
- 具有托管扩展功能的节点标签-根据市场类型或节点类型在实例上使用节点标签，以改进 Amazon 的自动扩展EMR。有关更多信息，请参阅[在 Amazon 中使用托管扩展EMR](#)。

## 已知问题

- EMRStudio 不支持 Python 3.11。

## 更改、增强功能和解决的问题

- 此版本修复了内部步骤清理操作期间可能出现的死锁问题。当步骤在EMR集群上完成时，此操作会管理这些步骤的生命周期。此问题会影响关键的 Amazon EMR 操作，例如分步操作和扩展。
- 此版本解决了具有某些预先存在的日志文件的自定义AMIs集群可能导致 Amazon EMR 日志管理守护程序失败的问题。
- Amazon EMR 7.2.0 将负责集群管理和监控活动的亚马逊EMR守护程序从 AWS SDK v1 升级到 v2。
- 当您启动带有最新补丁版本（亚马逊 EMR 5.36 或更高版本、6.6 或更高版本、7.0 或更高版本）的集群时，亚马逊将EMR使用最新的亚马逊 Linux 2023 或亚马逊 Linux 2 版本作为默认亚马逊 Linux 2 版本。EMR AMI有关更多信息，请参阅[使用亚马逊的默认 Amazon Linux AMI EMR](#)。

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2023.5.2 6.1.96-102.177.amzn2023 240708.1		2024年7月8日	美国东部（弗吉尼亚北部）、美国东部（俄亥俄州）、美国西部（加利福尼亚北部）、美国西部（俄勒冈）、欧洲（斯德哥尔摩）、欧洲（米兰）、欧洲（西班牙）、欧洲（法兰克福）、欧洲（苏黎世）、欧洲（伦敦）、欧洲（巴黎）、亚太地区（香港）、亚太地区（孟买）、亚太地区（海得拉巴）、亚

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、加拿大西部 (卡尔加里)、AWS GovCloud (美国西部)、AWS GovCloud (美国东部)、中国 (北京)、中国 (宁夏)

## 亚马逊 EMR 6.15.0 ( 6.x 系列的最新版本 )

Amazon 的新EMR版本将在几天内在不同地区上市，从最初发布日期的第一个地区开始。在此期间，您所在区域可能无法提供最新发行版。

以下发行说明包含 Amazon 6.15. EMR 0 版本的信息。更改与 6.14.0 有关。有关发布时间表的更多信息，请参阅 [6.15.0 更改日志](#)。

## 新功能

- 应用程序升级 – Amazon EMR 6.15.0 application upgrades include Apache Hadoop 3.3.6, Apache Hudi 0.14.0-amzn-0, Iceberg 1.4.0-amzn-0, and Trino 426.
- 在@@ [EMR集群上运行的启动速度更快 EC2](#) — 现在，在集EC2群上启动 Amaz EMR on 的速度最多可提高 35%。通过这一改进，大多数客户可以在 5 分钟或更短时间内启动集群。
- [CodeWhisperer 适用于 EMR Studio](#) — 您现在可以在编写代码时 CodeWhisperer 使用亚马逊和 Amazon EMR Studio 获取实时推荐 JupyterLab。CodeWhisperer 可以完成您的注释、完成单行代码、提出 line-by-line 建议并生成完整格式的函数。
- [使用 Flink 可以缩短任务重启时间](#) — 在 Amazon EMR 6.15.0 及更高版本中，Apache Flink 可以使用多种新机制来缩短任务恢复或扩展操作期间的任务重启时间。这样可以优化执行图的恢复和重启速度，从而提高作业稳定性。
- [开放表格式的表级和细粒度访问控制](#) — 在 Amazon EMR 6.15.0 及更高版本中，当您在访问 GI AWS ue 数据目录中数据的集群上在 Amazon EMR 上EC2运行 Spark 作业时，您可以使用 AWS Lake Formation 对基于 Hudi、Iceberg 或 Delta Lake 的表应用表、行、列和单元格级别的权限。
- Hadoop 升级 — 亚马逊 EMR 6.15.0 包括 Apache Hadoop 升级到 3.3.6 版。Hadoop 3.3.6是部署亚马逊 EMR 6.15时的最新版本，由Apache于2023年6月发布。亚马逊之前发布的版本EMR ( 6.9.0 至 6.14.x ) 使用了 Hadoop 3.3.3。

此次升级包括数百项改进和修复，以及一些功能，包括可重新配置的数据节点参数、在所有实时数据节点上启动批量重新配置操作的DFSAdmin选项，以及允许大量搜索的读者指定多个读取范围API的向量。Hadoop 3.3.6 还增加了对其预写日志 (WAL) 的支持HDFSAPIs和语义，因此HBase可以在其他存储系统实现上运行。有关更多信息，请参阅 Apache Hadoop 文档中的 [3.3.4](#)、[3.3.5](#) 和 [3.3.6](#) 版本的变更日志。

- 支持 Java ，版本 2 ——如果应用程序支持 v2，亚马逊 EMR 6.15.0 应用程序可以用 AWS SDK于 Java 版本 [1.12.569](#) 或 [2.20](#). 160。AWS SDK AWS SDK适用于 Java 2.x 的版本是对 1.x 版本代码库的重大改写。它基于 Java 8+ 构建，并增加了几个请求次数较多的功能。其中包括对非阻塞 I/O 的支持，以及在运行时插入其他HTTP实现的能力。有关更多信息，包括从 Java v1 到 v2 SDK 的迁移指南，请参阅[AWS SDK适用于 Java 的版本 2](#) 指南。

## 更改、增强功能和解决的问题

- 为了提高EMR集群的高可用性，此版本允许连接到使用IPv6终端节点的本地主机上的 Amazon EMR 守护程序。
- 此版本启用 TLS 1.2，以便在您的高可用性 ZooKeeper 集群的所有主节点上与预配置进行通信。

- 此版本改进了对主节点上维护的 ZooKeeper 事务日志文件的管理，以最大限度地减少日志文件越界增长并中断集群操作的情况。
- 此版本使高可用EMR性集群的节点内通信更具弹性。此改进减小了引导操作失败或集群启动失败的几率。
- Amazon EMR 6.15.0 中的 Tez 引入了配置，您可以指定这些配置来异步打开 Tez 分组拆分中的输入拆分。当单个 Tez 分组拆分中有大量输入拆分时，这样可以提高读取查询的性能。有关更多信息，请参阅 [Tez 异步拆分打开](#)。
- 当您启动带有最新补丁版本（亚马逊 EMR 5.36 或更高版本、6.6 或更高版本、7.0 或更高版本）的集群时，亚马逊将EMR使用最新的亚马逊 Linux 2023 或亚马逊 Linux 2 版本作为默认亚马逊 Linux 2 版本。EMR AMI有关更多信息，请参阅[使用亚马逊的默认 Amazon Linux AMI EMR](#)。

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 709.1	4.14.348	2024 年 7 月 23 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪) (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、AWS GovCloud (美国西部)、AWS GovCloud (美国东部)、(美国东部)、中国 (北京)、中国 (宁夏)、亚洲太平洋 (海得拉巴)、中东 (UAE)、欧洲 (西班牙)、欧洲 (苏黎世)、亚太地区 (墨尔本)、以色列 (特拉维夫)、加拿大西部 (卡尔加里)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 223.0	4.14.336	2024 年 3 月 8 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、中国 ( 北

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 131.0	4.14.336	2024年2月14日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、中国 ( 北

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024124.0	4.14.336	2024 年 2 月 7 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、中国 (北

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 109.0	4.14.334	2024 年 1 月 24 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、中国 (北

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 218.0	4.14.330	2024 年 1 月 2 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )



OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 206.0	4.14.330	2023 年 12 月 22 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023116.0	4.14.328	2023 年 12 月 11 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 101.0	4.14.327	2023 年 11 月 13 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			(美国东部)、中国(北京)、中国(宁夏)

## 亚马逊 EMR 5.36.2 ( 5.x 系列的最新版本 )

Amazon 的新EMR版本将在几天内在不同地区上市，从最初发布日期的第一个地区开始。在此期间，您所在区域可能无法提供最新发行版。

以下发行说明包含 Amazon EMR 版本 5.36.2 的信息。变化是相对于 5.36.1 的。有关发布时间表的信息，请参阅[更改日志](#)。

### 更改、增强功能和解决的问题

- 此版本改进了集群扩容逻辑，因此 Amazon EMR 不会将核心节点缩小到低于集群的HDFS重复因子设置。这一改进满足了数据冗余要求，并降低了扩展操作可能停滞的机会。
- 此版本为运行 Presto 或 Trino 的集群扩展工作流程添加了新的重试机制。这一改进降低了由于单个调整大小操作失败而导致集群调整无限期运行的风险。它还可以提高集群利用率，因为您的集群可以更快地向上和向下扩展。
- 修复了以下问题：当 Amazon EMR 优雅地停用核心节点时，集群扩容操作可能会停滞不前，核心节点在完全停用之前就会变为运行不正常。
- 当 Amazon EMR 重启单个节点时，可提高具有多个主节点的高可用性集群中节点的稳定性。
- 通过 Amazon 在亚马逊EC2上EMR运行，优化日志管理。因此，您可能会看到集群日志的存储成本略有降低。
- 改进了对主节点上维护的 ZooKeeper 事务日志文件的管理，以最大限度地减少日志文件越界增长并中断集群操作的情况。
- 修复了一个罕见的错误，该错误可能导致具有多个主节点的高可用性集群因无法与 Yarn ResourceManager 通信而发生故障。

- 当您启动带有最新补丁版本 ( 亚马逊 EMR 5.36 或更高版本、6.6 或更高版本、7.0 或更高版本 ) 的集群时，亚马逊将EMR使用最新的亚马逊 Linux 2023 或亚马逊 Linux 2 版本作为默认亚马逊 Linux 2 版本。EMR AMI有关更多信息，请参阅[使用亚马逊的默认 Amazon Linux AMI EMR](#)。

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 709.1	4.14.348	2024 年 7 月 23 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、AWS GovCloud ( 美国



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			东部 )、( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )、亚洲太平洋 ( 海得拉巴 )、中东 ( UAE )、欧洲 ( 西班牙 )、欧洲 ( 苏黎世 )、亚太地区 ( 墨尔本 )、以色列 ( 特拉维夫 )、加拿大西部 ( 卡尔加里 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024503.0	4.14.343	xxxxxx, 2024	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			中国 (北京)、中国 (宁夏)

## Amazon EMR 和 AWS Signature 版本 4

亚马逊EMR发布的版本使用 AWS 签名版本 4 (SigV4) 来验证向 Amazon S3 发出的请求。2020 年 6 月 24 日之后在 Amazon S3 中创建的存储桶不支持由签名版本 2 ( SigV2 ) 签名的请求。2020 年 6 月 24 日或之前创建的存储桶将继续支持 SigV2。我们建议您迁移到支持的 Amazon EMR 版本，SigV4这样您就可以访问新的 S3 存储桶并避免工作负载中断。

如果您使用亚马逊附带的应用程序，EMR例如 Apache Spark、Apache Hive 和 Presto，则无需更改应用程序代码即可使用。SigV4如果您使用未包含在 Amazon 中的自定义应用程序EMR，则可能需要更新代码才能使用SigV4。有关更多信息，请参阅《Amazon S3 用户指南》中的[从签名版本 2 转向签名版本 4](#)。

以下 Amazon EMR 版本支持SigV4:emr-4.7.4, emr-4.8.5, emr-4.9.6, emr-4.10.1, emr-5.1.1, emr-5.2.3, emr-5.3.2, emr-5.4.1, emr-5.5.4, emr-5.6.1, emr-5.7.1, emr-5.8.3, emr-5.9.1, emr-5.10.1, emr-5.11.4, emr-5.12.3, emr-5.13.1, emr-5.14.2, emr-5.15.1, emr-5.16.1, emr-5.17.2, emr-5.18.1, emr-5.19.1, emr-5.20.1, emr-5.21.2, and emr-5.22.0 and higher. 所有的 6.x 和 7.x 发行版都支持 SigV4。

## 缓解方法 CVE -2021-44228

### Note

对于亚马逊 6.9.0 及更高EMR版本，亚马逊EMR安装的所有使用 Log4j 库的组件都使用 Log4j 版本 2.17.1 或更高版本。

亚马逊EMR正在运行 EC2

在处理来自不可信来源的输入时，[CVE-2021-44228](#) 中讨论的问题与 2.0.0 和 2.14.1 之间的 Apache Log4j 核心版本有关。亚马逊在5.34.0之前的 EMR 5.x版本和 EMR 6.x版本中启动的亚马逊EMR集群包括Apache Hive、Flink、Presto和Trino等开源框架，它们使用这些版本的Apache Log EMR 4j。HUDI 但是，许多客户使用安装在其 Amazon EMR 集群上的开源框架来处理 and 记录来自不可信来源的输入。

我们建议您应用“适用于 Log4j CVE -2021-44228 的亚马逊 EMR Bootstrap 操作解决方案”，如下一节所述。该解决方案还解决了 CVE -2021-45046的问题。

### Note

亚马逊的引导操作脚本EMR已于2022年9月7日更新，增加了针对Oozie的增量错误修复和改进。如果您使用 Oozie，则应应用下一节中描述的更新后的 Amazon EMR 引导操作解决方案。

## Amazon EMR on EKS

如果您在默认配置下使用 [Amaz EMR on EKS](#)，则不会受到 CVE -2021-44228 中所述问题的影响，也不必应用本节中描述的解决方案。[适用于 Log4j 的亚马逊EMR引导操作解决方案-2021-44228 和 -2021-45046 CVE CVE](#)对于亚马逊EMR on EKS，Spark 的亚马逊EMR运行时使用 Apache Log4j 版本 1.2.17。EMR在上EKS使用 Amazon 时，不应将log4j.appender组件的默认设置更改为log。

## 适用于 Log4j 的亚马逊EMR引导操作解决方案-2021-44228 和 -2021-45046 CVE CVE

此解决方案提供了必须应用于您的 Amazon EMR 集群的 Amazon EMR 引导操作。对于每个 Amazon EMR 版本，您都可以在下面找到一个指向引导操作脚本的链接。要应用此引导操作，您应完成以下步骤：

1. 将与您的 Amazon EMR 版本对应的脚本复制到您的本地 S3 存储桶中 AWS 账户。请确保您使用的是特定于您的 Amazon EMR 版本的引导脚本。
2. 按照[EMR文档](#)中描述的说明，为您的EMR集群设置引导操作，以运行复制到您的 S3 存储桶的脚本。如果您为EMR集群配置了其他引导操作脚本，请确保将此脚本设置为要执行的第一个引导操作脚本。
3. 终止现有EMR集群，并使用引导操作脚本启动新集群。AWS 建议您在测试环境中测试引导脚本并验证应用程序，然后再将其应用于生产环境。如果您不使用EMR次要版本的最新修订版（例如 6.3.0），则必须使用最新的修订版（例如 6.3.1），然后应用上面讨论的解决方案。

## CVE-2021-44228 &amp; CVE -2021-45046-适用于亚马逊发布的引导脚本 EMR

亚马逊EMR版本号	脚本位置	脚本发布日期
6.5.0	s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-6.5.0-v2.sh	2022 年 3 月 24 日
6.4.0	s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-6.4.0-v2.sh	2022 年 3 月 24 日
6.3.1	s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-6.3.1-v2.sh	2022 年 3 月 24 日
6.2.1	s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-6.2.1-v2.sh	2022 年 3 月 24 日
6.1.1	s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-6.1.1-v2.sh	2021 年 12 月 14 日
6.0.1	s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-6.0.1-v2.sh	2021 年 12 月 14 日
5.34.0	s3://elasticmapreduce/ bootstrap-actions/	2021 年 12 月 12 日

亚马逊EMR版本号	脚本位置	脚本发布日期
	<code>log4j/patch-log4j-emr-5.34.0-v2.sh</code>	
5.33.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j-emr-5.33.1-v2.sh</code>	2021 年 12 月 12 日
5.32.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j-emr-5.32.1-v2.sh</code>	2021 年 12 月 13 日
5.31.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j-emr-5.31.1-v2.sh</code>	2021 年 12 月 13 日
5.30.2	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j-emr-5.30.2-v2.sh</code>	2021 年 12 月 14 日
5.29.0	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j-emr-5.29.0-v2.sh</code>	2021 年 12 月 14 日
5.28.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j-emr-5.28.1-v2.sh</code>	2021 年 12 月 15 日

亚马逊EMR版本号	脚本位置	脚本发布日期
5.27.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.27.1-v2.sh</code>	2021 年 12 月 15 日
5.26.0	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.26.0-v2.sh</code>	2021 年 12 月 15 日
5.25.0	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.25.0-v2.sh</code>	2021 年 12 月 15 日
5.24.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.24.1-v2.sh</code>	2021 年 12 月 15 日
5.23.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.23.1-v2.sh</code>	2021 年 12 月 15 日
5.22.0	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.22.0-v2.sh</code>	2021 年 12 月 15 日
5.21.2	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.21.2-v2.sh</code>	2021 年 12 月 15 日

亚马逊EMR版本号	脚本位置	脚本发布日期
5.20.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.20.1-v2.sh</code>	2021 年 12 月 15 日
5.19.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.19.1-v2.sh</code>	2021 年 12 月 15 日
5.18.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.18.1-v2.sh</code>	2021 年 12 月 15 日
5.17.2	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.17.2-v2.sh</code>	2021 年 12 月 15 日
5.16.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.16.1-v2.sh</code>	2021 年 12 月 15 日
5.15.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.15.1-v2.sh</code>	2021 年 12 月 15 日
5.14.2	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.14.2-v2.sh</code>	2021 年 12 月 15 日



亚马逊EMR版本号	脚本位置	脚本发布日期
5.13.1	<pre>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.13.1-v2.sh</pre>	2021 年 12 月 15 日
5.12.3	<pre>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.12.3-v2.sh</pre>	2021 年 12 月 15 日
5.11.4	<pre>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.11.4-v2.sh</pre>	2021 年 12 月 15 日
5.10.1	<pre>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.10.1-v2.sh</pre>	2021 年 12 月 15 日
5.9.1	<pre>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.9.1-v2.sh</pre>	2021 年 12 月 15 日
5.8.3	<pre>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.8.3-v2.sh</pre>	2021 年 12 月 15 日
5.7.1	<pre>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.7.1-v2.sh</pre>	2021 年 12 月 15 日

EMR发布版本	截至 2021 年 12 月的最新修订
6.3.0	6.3.1
6.2.0	6.2.1
6.1.0	6.1.1
6.0.0	6.0.1
5.33.0	5.33.1
5.32.0	5.32.1
5.31.0	5.31.1
5.30.0 或 5.30.1	5.30.2
5.28.0	5.28.1
5.27.0	5.27.1
5.24.0	5.24.1
5.23.0	5.23.1
5.21.0 或 5.21.1	5.21.2
5.20.0	5.20.1
5.19.0	5.19.1
5.18.0	5.18.1
5.17.0 或 5.17.1	5.17.2
5.16.0	5.16.1
5.15.0	5.15.1
5.14.0 或 5.14.1	5.14.2

EMR发布版本	截至 2021 年 12 月的最新修订
5.13.0	5.13.1
5.12.0、5.12.1、5.12.2	5.12.3
5.11.0、5.11.1、5.11.2、5.11.3	5.11.4
5.9.0	5.9.1
5.8.0、5.8.1、5.8.2	5.8.3
5.7.0	5.7.1

## 常见问题

- 版本@@ 超过 EMR 5 的EMR版本是否会受CVE到 -2021-44228 的影响？

不是。EMR版本 5 之前的EMR版本使用低于 2.0 的 Log4j 版本。

- 这个解决方案是否解决了 CVE -2021-45046 ？

是的，此解决方案还解决了 CVE-2021-45046 的问题。

- 该解决方案能否处理我在EMR集群上安装的自定义应用程序？

引导脚本仅更新由EMR安装的JAR文件。如果您通过引导操作、提交给EMR集群的步骤、使用自定义 Amazon Linux AMI 或任何其他机制在集群上安装和运行自定义应用程序和JAR文件，请与您的应用程序供应商合作，确定您的自定义应用程序是否受到 CVE -2021-44228 的影响，并确定适当的解决方案。

- 我应该如何处理开启的[自定义 docker 镜像](#)？EMREKS

如果您使用自定义 [docker 镜像](#)向亚马逊EMR添加自定义应用程序，或者EKS使用EKSwith自定义应用程序文件向亚马逊EMR提交任务，请与应用程序供应商合作，确定您的自定义应用程序是否受到 CVE -2021-44228 的影响，并确定适当的解决方案。

- 引导脚本如何缓解-2021-44228 和 CVE -2021-45046 中描述的问题？CVE

bootstrap 脚本通过添加一组新的指令来更新EMR启动指令。这些新指令删除了安装的所有开源框架通过 Log4j 使用的 JndiLookup 类文件。EMR这遵循了 [Apache 发布的](#)用于解决 Log4j 问题的建议。

- 是否有使用 Log4j 版本 2.17.1 或更高版本的更新？EMR

EMR5.34 之前的 5 个版本和 6.5 版本之前的 EMR 6 个版本使用与最新版本的 Log4j 不兼容的开源框架的旧版本。如果您继续使用这些版本，我们建议您应用引导操作来缓解中CVEs讨论的问题。在 EMR 5 版本 5.34 和 EMR 6 版本 6.5 之后，使用 Log4j 1.x 和 Log4j 2.x 的应用程序将分别升级为使用 Log4j 1.2.17 ( 或更高版本 ) 和 Log4j 2.17.1 ( 或更高版本 ) ，并且不需要使用上面提供的引导操作来缓解问题。CVE

- EMR发布会受CVE到 -2021-45105 的影响吗？

亚马逊EMR使用默认配置安装EMR的应用程序不受 CVE -2021-45105 的影响。在亚马逊安装的应用程序中EMR，只有 Apache Hive 使用 Apache Log4j 进行[上下文查询](#)，并且它不使用允许处理不当输入数据的方式使用非默认模式布局。

- Amazon 是否EMR受到以下任何CVE披露的影响？

下表包含与 Log4j 相关的列表CVEs，并注明了每个列表是否会影响 CVE Amazon。EMR仅当 Amazon EMR 使用默认配置安装应用程序时，此表中的信息才适用。

CVE	影响 EMR	注意
CVE-2022-23302	否	亚马逊EMR没有设置 Log4j JMSSink
CVE-2022-23305	否	亚马逊EMR没有设置 Log4j JDBCAppender
CVE-2022-23307	否	亚马逊EMR没有设置 Log4j Chainsaw
CVE-2020-9493	否	亚马逊EMR没有设置 Log4j Chainsaw
CVE-2021-44832	否	亚马逊EMR没有JDBCAppender使用连接字符串设置 Log4j JNDI
CVE-2021-4104	否	亚马逊EMR不使用 Log4j JMSAppender

CVE	影响 EMR	注意
CVE-2020-9488	否	Amazon 安装的应用程序EMR 不使用 Log4j SMTPAppender
CVE-2019-17571	否	Ama EMR zon 禁止公开访问集群，并且不会启动 SocketServer
CVE-2019-17531	否	我们建议您升级到最新的 Amazon EMR 发行版本。亚马逊 EMR 5.33.0 及更高版本使用 jackson-databind 2.6.7.4 或更高版本，6.1.0 及更高版本使用 jackson-databind 2.10.0 或更高版本。EMR这些版本的 jackson-databind 不受影响。CVE

## Amazon 发行说明EMR档案

下方提供了所有 Amazon EMR 版本的发行说明。有关每个发行版的全面版本信息，请参阅 [亚马逊 EMR 6.x 发布版本](#)、[亚马逊 EMR 5.x 发布版本](#) 和 [亚马逊 EMR 4.x 发布版本](#)。

要在亚马逊新EMR版本发布时获取更新，请订阅[亚马逊EMR发行说明的RSS提要](#)。

### 发行版 6.14.0

以下发行说明包含有关亚马逊 6.14.0 EMR 版本的信息。更改与 6.13.0 有关。有关发布时间表的更多信息，请参阅 [6.14.0 更改日志](#)。

#### 新功能

- Amazon EMR 6.14.0 supports Apache Spark 3.4.1, Apache Spark RAPIDS 23.06.0-amzn-2, Flink 1.17.1, Iceberg 1.3.1, and Trino 422.
- 对于您使用 [Amazon EMR 6.14.0 及更高版本创建的集群](#)，[ap-southeast-3亚太地区（雅加达）地区](#)现已推出 [Amazon EMR 托管扩展](#)。

## 更改、增强功能和解决的问题

- 6.14.0 版本优化了日志管理，亚马逊EMR在亚马逊上运行。EC2因此，您可能会看到集群日志的存储成本略有降低。
- 6.14.0 版本改进了扩展工作流程，以考虑不同的核心实例，这些实例的 Ama EBS zon 卷大小差异很大。此改进仅适用于核心节点；任务节点的缩减操作不受影响。
- 6.14.0 版本改进了 Amazon 与开源应用程序EMR交互的方式，例如。Apache Hadoop YARN ResourceManager and HDFS NameNode此改进降低了集群扩展导致操作延迟的风险，并减少了由于开源应用程序连接问题导致的启动故障。
- 6.14.0 发行版优化了集群启动时的应用程序安装。这缩短了某些 Amazon EMR 应用程序组合的集群启动时间。
- 6.14.0 版本修复了一个问题，即在自定义域中运行的集群遇到核心或任务节点重启时，集群缩小操作可能会停滞的问题。VPC
- 当您启动带有最新补丁版本（亚马逊 EMR 5.36 或更高版本、6.6 或更高版本、7.0 或更高版本）的集群时，亚马逊将EMR使用最新的亚马逊 Linux 2023 或亚马逊 Linux 2 版本作为默认亚马逊 Linux 2 版本。EMR AMI有关更多信息，请参阅[使用亚马逊的默认 Amazon Linux AMI EMR](#)。

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 709.1	4.14.348	2024 年 7 月 23 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪) (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、AWS GovCloud (美国西部)、AWS GovCloud (美国东部)、(美国东部)、中国 (北京)、中国 (宁夏)、亚洲太平洋 (海得拉巴)、中东 (UAE)、欧洲 (西班牙)、欧洲 (苏黎世)、亚太地区 (墨尔本)、以色列 (特拉维夫)、加拿大西部 (卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 223.0	4.14.336	2024 年 3 月 8 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、中国 ( 北



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 131.0	4.14.336	2024年2月14日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、中国 ( 北

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024124.0	4.14.336	2024 年 2 月 7 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、中国 ( 北

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 109.0	4.14.334	2024 年 1 月 24 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、中国 (北

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 218.0	4.14.330	2024 年 1 月 2 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 206.0	4.14.330	2023 年 12 月 22 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023116.0	4.14.328	2023 年 12 月 11 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 101.0	4.14.327	2023 年 11 月 17 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )
2.0.2023 906.0	4.14.322	2023 年 9 月 11 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )

## 发行版 6.13.0

以下发行说明包含有关亚马逊 6.13.0 EMR 版本的信息。更改与 6.12.0 有关。有关发布时间表的更多信息，请参阅 [6.13.0 更改日志](#)。

### 新功能

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

### 更改、增强功能和解决的问题

- 6.13.0 版本改进了 Amazon EMR 日志管理守护程序，确保在发出集群终止命令时，所有日志均按固定节奏上传到 Amazon S3。这有助于更快地终止集群。
- 6.13.0 版本增强了 Amazon EMR 日志管理功能，可确保将所有日志文件一致且及时地上传到 Amazon S3。这尤其有利于长时间运行的EMR集群。
- 当您启动带有最新补丁版本（亚马逊 EMR 5.36 或更高版本、6.6 或更高版本、7.0 或更高版本）的集群时，亚马逊将EMR使用最新的亚马逊 Linux 2023 或亚马逊 Linux 2 版本作为默认亚马逊 Linux 2 版本。EMR AMI有关更多信息，请参阅[使用亚马逊的默认 Amazon Linux AMI EMR](#)。

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 709.1	4.14.348	2024 年 7 月 23 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地



OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			区 ( 香港 ) 、 亚太地区 ( 孟买 ) 、 亚太地区 ( 东京 ) 、 亚太地区 ( 首尔 ) 、 亚太地区 ( 大阪 ) 、 亚太地区 ( 新加坡 ) 、 亚太地区 ( 新加坡 ) 、 亚太地区 ( 新加坡 ) 、 亚太地区 ( 孟买 ) 、 亚太地区 ( 东京 ) 、 亚太地区 ( 首尔 ) 、 亚太地区 ( 大阪 ) ( 悉尼 ) 、 亚太地区 ( 雅加达 ) 、 非洲 ( 开普敦 ) 、 南美洲 ( 圣保罗 ) 、 中东 ( 巴林 ) 、 加拿大 ( 中部 ) 、 AWS GovCloud ( 美国西部 ) 、 AWS GovCloud ( 美国东部 ) 、 ( 美国东部 ) 、 中国 ( 北京 ) 、 中国 ( 宁夏 ) 、 亚洲太平洋 ( 海得拉巴 ) 、 中东 ( UAE ) 、 欧洲 ( 西班牙 ) 、 欧洲 ( 苏黎世 ) 、 亚太地区 ( 墨尔本 ) 、 以色列 ( 特拉维夫 ) 、 加拿大西部 ( 卡尔加里 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 223.0	4.14.336	2024 年 3 月 8 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、中国 ( 北

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 131.0	4.14.336	2024年2月14日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、中国 ( 北

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024124.0	4.14.336	2024 年 2 月 7 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、中国 ( 北

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 109.0	4.14.334	2024 年 1 月 24 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、中国 ( 北



OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 218.0	4.14.330	2024 年 1 月 2 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 206.0	4.14.330	2023 年 12 月 22 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023116.0	4.14.328	2023 年 12 月 11 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 101.0	4.14.327	2023 年 11 月 16 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023020.1	4.14.326	2023 年 11 月 7 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023012.1	4.14.326	2023 年 10 月 26 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 926.0	4.14.322	2023 年 10 月 19 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 906.0	4.14.322	2023 年 10 月 4 日	(美国东部)、中国(北京)、中国(宁夏)  美国东部(弗吉尼亚北部)、美国东部(俄亥俄州)、美国西部(加利福尼亚北部)、美国西部(俄勒冈)、欧洲(斯德哥尔摩)、欧洲(米兰)、欧洲(西班牙)、欧洲(法兰克福)、欧洲(苏黎世)、欧洲(伦敦)、欧洲(巴黎)、亚太地区(香港)、亚太地区(孟买)、亚太地区(海得拉巴)、亚太地区(东京)、亚太地区(东京)、亚太地区(首尔)、亚太地区(大阪)、亚太地区(新加坡)、亚太地区(悉尼)、亚太地区(雅加达)、亚太地区(墨尔本)、非洲(开普敦)、南美洲(圣保罗)、中东(巴林)、中东( )、中东(UAE)、加拿大(中部)、以色列(特拉维夫)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.20230808.0	4.14.320	2023 年 8 月 24 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )



## 版本 6.12.0

以下发行说明包含有关亚马逊 6.12.0 EMR 版本的信息。更改与 6.11.0 有关。有关发布时间表的更多信息，请参阅 [6.12.0 更改日志](#)。

### 新功能

- Amazon EMR 6.12.0 supports Apache Spark 3.4.0, Apache Spark RAPIDS 23.06.0-amzn-0, CUDA 11.8.0, Apache Hudi 0.13.1-amzn-0, Apache Iceberg 1.3.0-amzn-0, Trino 414, and PrestoDB 0.281.
- 亚马逊EMR发布了 6.12.0 及更高版本的支持，与 Apache Livy、Apache Hive 到 HiveServer 2 (HS2)、Trino、Presto 和 Hue LDAP 集成。你还可以在使用 6.12.0 或更高版本的EMR集群上安装 Apache Spark 和 Apache Hadoop，然后对其进行配置以供使用。LDAP有关更多信息，请参阅[使用 Active Directory 或LDAP服务器进行亚马逊身份验证EMR](#)。

### 更改、增强功能和解决的问题

- 亚马逊EMR版本 6.12.0 及更高版本为 Flink 提供 Java 11 运行时支持。有关更多信息，请参阅[将 Flink 配置为使用 Java 11 运行](#)。
- 6.12.0 版本为运行 Presto 或 Trino 的集群的集群扩展工作流程添加了新的重试机制。EMR这一改进降低了由于单个调整大小操作失败而导致集群大小调整无限期停滞的风险。它还可以提高集群利用率，因为您的集群可以更快地向上和向下扩展。
- 6.12.0 版本修复了一个问题，即当处于正常停用状态的核心节点在完全停用之前出于任何原因变得运行不正常时，集群的缩减操作可能会停滞不前。
- 6.12.0 版本改进了集群缩小逻辑，因此您的集群不会尝试将核心节点缩小到低于集群HDFS重复因子设置的范围。这符合您的数据冗余要求，并减少了扩展操作可能停滞的机会。
- 6.12.0 版本EMR通过提高记录实例状态变化的速度，提高了 Amazon 运行状况监控服务的性能和效率。这一改进降低了运行多个自定义客户端工具或第三方应用程序的集群节点性能下降的机会。
- 6.12.0 版本提高了 Amazon 集群日志管理守护程序的性能。EMR因此，对于以高并发度运行步骤的 EMR集群，性能下降的可能性较小。
- 在 Amazon 6.12.0 EMR 版本中，日志管理守护程序已升级，可以识别所有正在使用本地实例存储上的打开文件句柄的日志，以及相关的进程。此次升级可确保 Amazon 在日志存档到 Amazon S3 后 EMR正确删除文件并回收存储空间。
- 6.12.0 版本包括日志管理进程守护程序增强功能，可删除本地集群文件系统中空的、未使用的步骤目录。过多的空目录会降低 Amazon EMR 守护程序的性能并导致磁盘过度使用。

- 6.12.0 版本允许对YARN时间轴服务器日志进行日志轮换。这样可以最大限度地减少磁盘过度使用情况，特别是对于长时间运行的集群。
- 在 Amazon EMR 6.10.0 及更高版本中，默认根卷大小已增加到 15 GB。早期版本的默认根卷大小为 10 GB。
- 当您启动带有最新补丁版本（亚马逊 EMR 5.36 或更高版本、6.6 或更高版本、7.0 或更高版本）的集群时，亚马逊将EMR使用最新的亚马逊 Linux 2023 或亚马逊 Linux 2 版本作为默认亚马逊 Linux 2 版本。EMR AMI有关更多信息，请参阅[使用亚马逊的默认 Amazon Linux AMI EMR](#)。

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 709.1	4.14.348	2024 年 7 月 23 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、AWS GovCloud ( 美国东部 )、( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )、亚洲太平洋 ( 海得拉巴 )、中东 ( UAE )、欧洲 ( 西班牙 )、欧洲 ( 苏黎世 )、亚太地区 ( 墨尔本 )、以色列 ( 特拉维夫 )、加拿大西部 ( 卡尔加里 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 223.0	4.14.336	2024 年 3 月 8 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、中国 ( 北

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 131.0	4.14.336	2024年2月14日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、中国 ( 北

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024124.0	4.14.336	2024 年 2 月 7 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、中国 ( 北



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 109.0	4.14.334	2024 年 1 月 24 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、中国 (北

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 218.0	4.14.330	2024 年 1 月 2 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 206.0	4.14.330	2023 年 12 月 22 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023116.0	4.14.328	2023 年 12 月 11 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 101.0	4.14.327	2023 年 11 月 16 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023020.1	4.14.326	2023 年 11 月 7 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023012.1	4.14.326	2023 年 10 月 26 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 926.0	4.14.322	2023 年 10 月 19 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 906.0	4.14.322	2023 年 10 月 4 日	( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )  美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.20230822.0	4.14.322	2023 年 8 月 30 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 808.0	4.14.320	2023 年 8 月 24 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 727.0	4.14.320	2023 年 8 月 14 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 719.0	4.14.320	2023 年 8 月 2 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 628.0	4.14.318	2023 年 7 月 12 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 (巴林)、中东 (UAE)、加拿大 (中部)

## 版本 6.11.1

以下发行说明包含有关亚马逊 6.11.1 EMR 版本的信息。更改与 6.11.0 有关。有关发布时间表的更多信息，请参阅 [6.11.1 更改日志](#)。

## 更改、增强功能和解决的问题

- 由于锁争用，如果在尝试停用节点的同时添加或移除节点，则该节点可能会陷入死锁。结果，Hadoop 资源管理器 (YARN) 变得无响应，并会影响所有传入和当前正在运行的容器。
- 此版本包括一项更改，允许高可用性集群在重启后从故障状态中恢复。
- 此版本包含针对 Hue 和的安全补丁HBase。
- 此版本修复了在 Spark with Amazon 上运行工作负载的集群EMR可能会在静默中收到使用contains、startsWithendsWith、和like的错误结果的问题。当您对在 Amazon EMR Hive3 Metastore 服务器中包含元数据的分区字段使用表达式时，就会出现此问题 ()。HMS
- 此版本修复了当没有用户定义的函数时，Glue 端的限制问题 () UDF。
- 此版本修复了一个问题，即节点日志聚合服务在日志推送器可以将容器日志推送到 S3 以防YARN停用时将其推送到 S3 之前，将其删除。
- 此版本修复了 Hadoop FairShare p 启用节点标签时调度器指标的问题。
- 此版本修复了您在 spark-defaults.conf 中为 spark.yarn.heterogeneousExecutors.enabled 配置设置默认 true 值时影响 Spark 性能的问题。
- 此版本修复了 Reduce Task 无法读取随机数据的问题。该问题因内存损坏错误导致 Hive 查询失败。
- 此版本为运行 Presto 或 Trino 的集群的集群扩展工作流程添加了新的重试机制。EMR这一改进降低了由于单个调整大小操作失败而导致集群大小调整无限期停滞的风险。它还可以提高集群利用率，因为您的集群可以更快地向上和向下扩展。
- 此版本改进了集群缩容逻辑，因此您的集群不会尝试将核心节点缩小到低于集群HDFS重复因子设置的范围。这符合您的数据冗余要求，并减少了扩展操作可能停滞的机会。
- 日志管理进程守护程序已升级，可以识别本地实例存储中所有包含打开文件句柄的使用中的日志，以及相关的进程。此次升级可确保 Amazon 在日志存档到 Amazon S3 后EMR正确删除文件并回收存储空间。
- 此版本包括日志管理进程守护程序增强功能，可删除本地集群文件系统中空的、未使用的步骤目录。过多的空目录会降低 Amazon EMR 守护程序的性能并导致磁盘过度使用。
- 当您启动带有最新补丁版本 ( 亚马逊 EMR 5.36 或更高版本、6.6 或更高版本、7.0 或更高版本 ) 的集群时，亚马逊将EMR使用最新的亚马逊 Linux 2023 或亚马逊 Linux 2 版本作为默认亚马逊 Linux 2 版本。EMR AMI有关更多信息，请参阅[使用亚马逊的默认 Amazon Linux AMI EMR](#)。

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 709.1	4.14.348	2024 年 7 月 23 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、AWS GovCloud (美国西部)、AWS GovCloud (美国



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			东部 )、( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )、亚洲太平洋 ( 海得拉巴 )、中东 ( UAE )、欧洲 ( 西班牙 )、欧洲 ( 苏黎世 )、亚太地区 ( 墨尔本 )、以色列 ( 特拉维夫 )、加拿大西部 ( 卡尔加里 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 223.0	4.14.336	2024 年 3 月 8 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、中国 ( 北

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 131.0	4.14.336	2024年2月14日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、中国 ( 北

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024124.0	4.14.336	2024 年 2 月 7 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、中国 (北

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			京 )、中国 ( 宁夏 )、加拿大西部 ( 卡尔加里 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 109.0	4.14.334	2024 年 1 月 24 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、中国 ( 北



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			京)、中国(宁夏)、加拿大西部(卡尔加里)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 218.0	4.14.330	2024 年 1 月 2 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 206.0	4.14.330	2023 年 12 月 22 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023116.0	4.14.328	2023 年 12 月 11 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 101.0	4.14.327	2023 年 11 月 16 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023020.1	4.14.326	2023 年 11 月 7 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023012.1	4.14.326	2023 年 10 月 26 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 926.0	4.14.322	2023 年 10 月 19 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 906.0	4.14.322	2023 年 10 月 4 日	(美国东部)、中国(北京)、中国(宁夏)  美国东部(弗吉尼亚北部)、美国东部(俄亥俄州)、美国西部(加利福尼亚北部)、美国西部(俄勒冈)、欧洲(斯德哥尔摩)、欧洲(米兰)、欧洲(西班牙)、欧洲(法兰克福)、欧洲(苏黎世)、欧洲(伦敦)、欧洲(巴黎)、亚太地区(香港)、亚太地区(孟买)、亚太地区(海得拉巴)、亚太地区(东京)、亚太地区(东京)、亚太地区(首尔)、亚太地区(大阪)、亚太地区(新加坡)、亚太地区(悉尼)、亚太地区(雅加达)、亚太地区(墨尔本)、非洲(开普敦)、南美洲(圣保罗)、中东(巴林)、中东( )、中东(UAE)、加拿大(中部)、以色列(特拉维夫)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 822.0	4.14.322	2023 年 8 月 30 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )



OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 808.0	4.14.320	2023 年 8 月 24 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 727.0	4.14.320	2023 年 8 月 14 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 米兰 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )

## 版本 6.11.0

以下发行说明包含有关亚马逊 6.11. EMR 0 版本的信息。更改与 6.10.0 有关。有关发布时间表的信息，请参阅[更改日志](#)。

## 新功能

- 亚马逊 EMR 6.11.0 支持 Apache Spark 3.3.2-amzn-0、Apache Spark 23.02.0-amzn-0、11 CUDA .8.0、Apache Hudi 0.13.0-amzn-0、Apache Iceberg RAPIDS 1.2.0-amzn-0、Trino 410-amzn-0 和 PrestoDB 0.279-amzn-0。

## 更改、增强功能和解决的问题

- 在亚马逊 EMR 6.11.0 中，DynamoDB 连接器已升级到 5.0.0 版。5.0.0 版本使用 AWS SDK for Java 2.x 之前的版本使用的是 AWS SDK for Java 1.x。由于此次升级，我们强烈建议您在将 DynamoDB 连接器与 Amazon 6.11 配合使用之前，先测试您的代码。EMR
- EMR 亚马逊 6.11.0 的 DynamoDB 连接器调用 DynamoDB 服务时，它将使用您为该属性提供的区域值。dynamodb.endpoint 我们建议您在配置 dynamodb.endpoint 时也配置 dynamodb.region，并且两个属性都以相同的 AWS 区域为目标。如果您使用 dynamodb.endpoint 但未进行配置 dynamodb.region，则适用于 Amazon 6.11.0 的 DynamoDB 连接器将返回无效的区域异常，并尝试协调来自 EC2 亚马逊实例元数据服务的信息 (IMDS)。如果连接器无法从中检索区域 IMDS，则默认为美国东部 (弗吉尼亚北部) (us-east-1)。以下错误是您未正确配置该 dynamodb.region 属性时可能会遇到的无效区域异常的示例：

```
software.amazon.awssdk.services.dynamodb.model.DynamoDbException: Credential should be scoped to a valid region.
```

有关受 AWS SDK for Java 升级到 2.x 影响的类的更多信息，请参阅 Amazon-EMR DynamoDB 连接器存储库中的 [AWS SDK for Java 从 1.x 升级到 2.x \(#175\)](#) 提交。GitHub
- 此版本修复了在执行列重命名操作后使用 Delta Lake 在 Amazon S3 中存储 Delta 表数据时列数据变为 NULL 的问题。有关 Delta Lake 中此实验性功能的更多信息，请参阅《Delta Lake User Guide》中的 [Column rename operation](#)。
- 6.11.0 版本修复了通过从具有多个主节点的集群中复制一个主节点来创建边缘节点时可能出现的问题。复制的边缘节点可能会导致缩减操作的延迟，或者导致主节点的内存使用率过高。有关如何创建边缘节点以及与 EMR 集群通信的更多信息，请参阅存储库中的 [Edge Node Create](#) aws-samples or。GitHub
- 6.11.0 版本改进了亚马逊用于在重启后 EMR 将 Amazon 卷 EBS 重新安装到实例的自动化流程。
- 6.11.0 版本修复了导致亚马逊向亚马逊 EMR 发布的 Hadoop 指标间歇性出现差距的问题。CloudWatch
- 6.11.0 版本修复了 EMR 群集中的一个问题，即由于磁盘过度使用而导致对包含群集节点排除列表的 YARN 配置文件的更新中断。不完整的更新阻碍了未来对集群的缩减操作。此版本可确保您的集群保持正常运行，并确保扩展操作按预期进行。

- 在 Amazon EMR 6.10.0 及更高版本中，默认根卷大小已增加到 15 GB。早期版本的默认根卷大小为 10 GB。
- Hadoop 3.3.3 在 YARN ([YARN-9608](#)) 中引入了一项更改，即在应用程序完成之前，容器运行的节点一直处于停用状态。此更改可确保如随机数据等本地数据不会丢失，并且您无需重新运行作业。无论是否启用托管扩展，这种方法还可能导致集群的资源利用不足。

在亚马逊EMR版本 6.11.0 及更高版本以及 6.8.1、6.9.1 和 6.10.1 中，将的值设置 `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` 为 `in` 以 `false` 解决此问题。 `yarn-site.xml`

虽然该修复程序解决了 YARN -9608 引入的问题，但由于启用了托管扩展的集群上的洗牌数据丢失，它可能会导致 Hive 作业失败。在此版本中，我们还通过设置 Hive `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-shuffle-data` 工作负载来降低这种风险。此配置仅适用于亚马逊 6.11.0 及更高EMR版本。

- 当您启动带有最新补丁版本（亚马逊 EMR 5.36 或更高版本、6.6 或更高版本、7.0 或更高版本）的集群时，亚马逊将EMR使用最新的亚马逊 Linux 2023 或亚马逊 Linux 2 版本作为默认亚马逊 Linux 2 版本。EMR AMI有关更多信息，请参阅[使用亚马逊的默认 Amazon Linux AMI EMR](#)。

#### Note

此版本不再获得自动AMI更新，因为又有 1 个补丁版本接替了该版本。补丁版本以第二位小数点后的数字 (6.8.1) 表示。要查看您是否使用的是最新补丁版本，请查看[发布指南](#)中的可用版本，或者在控制台中创建集群时查看 Amazon EMR 版本下拉列表，或者使用[ListReleaseLabels](#)API或[list-release-labels](#)CLI操作。要获取有关新版本的更新，请订阅“[最新消息?](#)”上的RSS提要 页面。

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 709.1	4.14.348	2024 年 7 月 23 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			利福尼亚北部)、美国西部(俄勒冈)、欧洲(斯德哥尔摩)、欧洲(米兰)、欧洲(法兰克福)、欧洲(爱尔兰)、欧洲(伦敦)、欧洲(巴黎)、亚太地区(香港)、亚太地区(孟买)、亚太地区(东京)、亚太地区(首尔)、亚太地区(大阪)、亚太地区(新加坡)、亚太地区(新加坡)、亚太地区(新加坡)、亚太地区(孟买)、亚太地区(东京)、亚太地区(首尔)、亚太地区(大阪)、亚太地区(悉尼)、亚太地区(雅加达)、非洲(开普敦)、南美洲(圣保罗)、中东(巴林)、加拿大(中部)、AWS GovCloud(美国西部)、AWS GovCloud(美国东部)、(美国东部)、中国(北京)、中国(宁夏)、亚洲太平洋(海得拉巴)、中东(UAE)、欧洲(西班牙)、欧洲

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			( 苏黎世 )、亚太地区 ( 墨尔本 )、以色列 ( 特拉维夫 )、加拿大西部 ( 卡尔加里 )
2.0.2023 808.0	4.14.320	2023 年 8 月 24 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 米兰 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 727.0	4.14.320	2023 年 8 月 14 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 719.0	4.14.320	2023 年 8 月 2 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 628.0	4.14.318	2023 年 7 月 12 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、加拿大 ( 中部 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( 巴林 )、中东 ( UAE )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 612.0	4.14.314	2023 年 6 月 23 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、加拿大 (中部)、欧洲 (斯德哥尔摩)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (米兰)、欧洲 (西班牙)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 (巴林)、中东 (UAE)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 504.1	4.14.313	2023 年 5 月 16 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 (巴林)、中东 (UAE)、加拿大 (中部)

## 版本 6.10.0

以下发行说明包含有关亚马逊 6.10. EMR 0 版本的信息。更改与 6.9.0 有关。有关发布时间表的信息，请参阅[更改日志](#)。

## 新功能

- 亚马逊 EMR 6.10.0 支持 Apache Spark 3.1、Apache Spark RAPIDS 22.12.0、CUDA 11.8.0、Apache Hudi 0.12.2-amzn-0、Apache Iceberg 1.1.0-amzn-0、Trino 403 和 PrestoDB 0.278.1。
- 亚马逊 EMR 6.10.0 包含原生 Trino-Hudi 连接器，可提供对 Hudi 表中数据的读取权限。您可以使用 `trino-cli --catalog hudi` 激活连接器，并使用 `trino-connector-hudi` 配置连接器以满足您的要求。与 Amazon 的原生集成 EMR 意味着您不再需要使用 `trino-connector-hive` 来查询 Hudi 表。有关新连接器支持的配置列表，请参阅 Trino 文档的 [Hudi connector](#) 页面。
- 亚马逊 EMR 发布的 6.10.0 及更高版本支持 Apache Zeppelin 与 Apache Flink 集成。请参阅在 [Amazon EMR 中通过 Zeppelin 使用 Flink 作业](#) 了解更多信息。

## 已知问题

- Hadoop 3.3.3 在 YARN ([YARN-9608](#)) 中引入了一项更改，即在应用程序完成之前，容器运行的节点一直处于停用状态。此更改可确保如随机数据等本地数据不会丢失，并且您无需重新运行作业。无论是否启用托管扩展，这种方法还可能导致集群的资源利用不足。

要在 Amazon EMR 6.10.0 中解决此问题，您可以将的值设置为 `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications`。false `yarn-site.xml` 在亚马逊 EMR 版本 6.11.0 及更高版本以及 6.8.1、6.9.1 和 6.10.1 中，配置 `false` 默认设置为以解决此问题。

## 更改、增强功能和解决的问题

- 亚马逊 EMR 6.10.0 消除了对 [Apache Spark Amazon Redshift 集成的](#) 依赖，并自动将所需的 Spark-Redshift 相关的 jar 添加到 `Spark:`、和的 `执行器类` 路径中。`minimal-json.jar spark-redshift.jar spark-avro.jar RedshiftJDBC.jar`
- 6.10.0 版本改进了集群上的日志管理守护程序，可以监视集群中的其他日志文件夹。EMR 这一改进最大限度地减少了磁盘过度使用情况。
- 6.10.0 版本在集群上日志管理进程守护程序停止后会自动重启该守护程序。这一改进降低了由于磁盘过度使用而导致节点出现运行状况不佳的风险。
- Amazon EMR 6.10.0 支持用于 EMRFS 用户映射的区域终端节点。
- 在 Amazon EMR 6.10.0 及更高版本中，默认根卷大小已增加到 15 GB。早期版本的默认根卷大小为 10 GB。

- 6.10.0 版本修复了当所有剩余的 Spark 执行器都在使用资源管理器的停用主机上时，导致 Spark 作业停滞的问题。YARN
- 在 Amazon EMR 6.0 到 6.9.x 中，具有动态分区和 BY 或 BY 子句ORDER的INSERT查询将始终有SORT两个缩减器。此问题是由OSS更改 [HIVE-20703](#) 引起的，该更改将动态排序分区优化置于基于成本的决策之下。如果您的工作负载不需要对动态分区进行排序，建议将 `hive.optimize.sort.dynamic.partition.threshold` 属性设置为 `-1` 以禁用新功能并获得计算正确的 Reducer 数量。此问题已在 OSS Hive 中作为 [HIVE-22269](#) 的一部分修复，并已在亚马逊 6.10.0 中修复。EMR
- 当您启动带有最新补丁版本（亚马逊 EMR 5.36 或更高版本、6.6 或更高版本、7.0 或更高版本）的集群时，亚马逊将EMR使用最新的亚马逊 Linux 2023 或亚马逊 Linux 2 版本作为默认亚马逊 Linux 2 版本。EMR AMI有关更多信息，请参阅[使用亚马逊的默认 Amazon Linux AMI EMR](#)。

### Note

此版本不再获得自动AMI更新，因为又有 1 个补丁版本接替了该版本。补丁版本以第二位小数点后的数字 (6.8.1) 表示。要查看您是否使用的是最新补丁版本，请查看[发布指南](#)中的可用版本，或者在控制台中创建集群时查看 Amazon EMR 版本下拉列表，或者使用[ListReleaseLabels](#)API或[list-release-labels](#)CLI操作。要获取有关新版本的更新，请订阅“[最新消息?](#)”上的RSS提要页面。

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 709.1	4.14.348	2024 年 7 月 23 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、AWS GovCloud ( 美国东部 )、( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )、亚洲太平洋 ( 海得拉巴 )、中东 ( UAE )、欧洲 ( 西班牙 )、欧洲 ( 苏黎世 )、亚太地区 ( 墨尔本 )、以色列 ( 特拉维夫 )、加拿大西部 ( 卡尔加里 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.20230808.0	4.14.320	2023 年 8 月 24 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、欧洲地区 (斯德哥尔摩)、欧洲地区 (米兰)、欧洲地区 (法兰克福)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、以色列 (特拉维夫)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 727.0	4.14.320	2023 年 8 月 14 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 719.0	4.14.320	2023 年 8 月 2 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 628.0	4.14.318	2023 年 7 月 12 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、加拿大 (中部)、欧洲 (斯德哥尔摩)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (米兰)、欧洲 (西班牙)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 (巴林)、中东 (UAE)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 612.0	4.14.314	2023 年 6 月 23 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、加拿大 (中部)、欧洲 (斯德哥尔摩)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (米兰)、欧洲 (西班牙)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 (巴林)、中东 (UAE)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 504.1	4.14.313	2023 年 5 月 16 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、加拿大 ( 中部 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( 巴林 )、中东 ( UAE )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 418.0	4.14.311	2023 年 5 月 3 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、加拿大 (中部)、欧洲 (斯德哥尔摩)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (米兰)、欧洲 (西班牙)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 (巴林)、中东 (UAE)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 404.1	4.14.311	2023 年 4 月 18 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、加拿大 ( 中部 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、欧洲 ( 法兰克福 )、欧洲 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 首尔 )、亚太地区 ( 首尔 )、亚太地区 ( 首尔 )、亚太地区 ( 首尔 )、亚太地区 ( 首尔 )、亚太地区 ( 香港 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( UAE )
2.0.2023 404.0	4.14.311	2023 年 4 月 10 日	美国东部 ( 弗吉尼亚州北部 )、欧洲地区 ( 巴黎 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 320.0	4.14.309	2023 年 3 月 30 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、加拿大 (中部)、欧洲 (斯德哥尔摩)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、欧洲 (法兰克福)、欧洲 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (首尔)、亚太地区 (首尔)、亚太地区 (首尔)、亚太地区 (首尔)、亚太地区 (首尔)、亚太地区 (香港)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 (UAE)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023070	4.14.304	2023 年 2 月 22 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、加拿大 (中部)、欧洲 (斯德哥尔摩)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、欧洲 (法兰克福)、欧洲 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (首尔)、亚太地区 (首尔)、亚太地区 (首尔)、亚太地区 (首尔)、亚太地区 (香港)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 (UAE)

## 发行版 6.9.0

以下发行说明包含 Amazon 6.9. EMR 0 版本的信息。更改与亚马逊 6.8.0 EMR 版本有关。有关发布时间表的信息，请参阅[更改日志](#)。



## 新功能

- 亚马逊EMR版本 6.9.0 支持 Apache Spark RAPIDS 22.08.0、Apache Hudi 0.12.1、Apache Iceberg 0.14.1、Trino 398 和 Tez 0.10.2。
- 亚马逊 6.9.0 EMR 版本包括一个新的开源应用程序 2.1.0。 [Delta Lake](#)
- 适用于 Apache Spark 的 Amazon Redshift 集成包含在亚马逊 6.9.0 及更高EMR版本中。本地集成之前是一种开源工具，现在是 Spark 连接器，您可以将其用于构建 Apache Spark 应用程序，这些应用程序可在 Amazon Redshift 和 Amazon Redshift Serverless 中读取和写入数据。有关更多信息，请参阅 [使用亚马逊 Redshift 与 Apache Spark 的集成与亚马逊 EMR](#)。
- Amazon EMR 版本 6.9.0 增加了对集群缩小期间将日志存档到 Amazon S3 的支持。之前，您只能在集群终止期间将日志文件存档到 Amazon S3。这项新功能可确保即使在节点终止后，集群上生成的日志文件仍保留在 Amazon S3 上。有关更多信息，请参阅[配置集群日志记录和调试](#)。
- 为了支持长时间运行的查询，Trino 现在包括容错执行机制。容错执行通过重试失败的查询或其组件任务来减少查询失败。有关更多信息，请参阅 [Trino 中的容错执行](#)。
- 你可以使用亚马逊EMR上的 Apache Flink 来统一BATCH和STREAM处理 Apache Hive 表或任何 Flink 表源的元数据，例如 Iceberg、Kinesis 或 Kafka。你可以使用、或 Amazon 将 AWS Glue 数据目录指定为 Flink 的 AWS Management Console元数据库。AWS CLI EMR API有关更多信息，请参阅 [在亚马逊中配置 Flink EMR](#)。
- 现在，你可以在装有 Amazon Studio 的EMR集群上为亚马逊上的 Apache Spark、Apache Hive 和 Presto 查询指定 AWS Identity and Access Management (IAM) 运行时角色和 AWS Lake Formation 基于访问控制EC2的权限。 SageMaker 有关更多信息，请参阅 [为 Amazon EMR 步骤配置运行时角色](#)。

## 已知问题

- 对于亚马逊EMR版本 6.9.0，Trino 不适用于启用 Apache Ranger 的集群。如果您需要将 Trino 与 Ranger 结合使用，请联系 [AWS Support](#)。
- 如果您使用适用于 Apache Spark 的 Amazon Redshift 集成，并且具有 Parquet 格式的时间、timetz、时间戳或 timestamptz（精度为微秒），连接器会将时间值舍入为最接近的毫秒值。解决方法是使用文本卸载格式 unload\_s3\_format 参数。
- 当你使用带有 Hive 分区位置格式的 Spark 来读取 Amazon S3 中的数据，并在亚马逊EMR版本 5.30.0 到 5.36.0 和 6.2.0 到 6.9.0 上运行 Spark 时，你可能会遇到一个问题，导致集群无法正确读取数据。如果您的分区具有以下所有特征，会发生这种情况：
  - 从同一个表扫描两个或多个分区。

- 至少有一个分区目录路径是至少一个其他分区目录路径的前缀，例如，`s3://bucket/table/p=a` 是 `s3://bucket/table/p=a b` 的前缀。
- 另一个分区目录中前缀之后的第一个字符的值为 UTF-8，小于该字符 (U+002F)。例如，在 `s3://bucket/table/p=a b` 中，`a` 和 `b` 之间出现的空格字符 (U+0020) 就属于此类。请注意，还有其他 14 个非控制字符：`!"#$%&'()*+,-`。有关更多信息，请参阅 [UTF-8 编码表和 Unicode 字符](#)。

解决方法是在 `spark-defaults` 分类中将

`spark.sql.sources.fastS3PartitionDiscovery.enabled` 配置设置为 `false`。

- 从 Amazon SageMaker Studio 到亚马逊EMR集群的连接可能会间歇性失败，并显示 403 禁止的响应代码。当在集群上设置IAM角色的时间超过 60 秒时，就会发生此错误。解决方法是，您可以安装 Amazon EMR 补丁以启用重试并将超时时间延长到最少 300 秒。启动集群时，按照以下步骤应用引导操作。

1. 从以下 Amazon S3 URIs 下载引导脚本和RPM文件。

```
s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/gcsc/replace-rpms.sh
s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/gcsc/emr-secret-agent-1.18.0-SNAPSHOT20221121212949.noarch.rpm
```

2. 将上一步中的文件上传到您自己的 Amazon S3 存储桶中。存储桶必须与您计划启动集群的 AWS 区域 位置相同。
3. 在启动EMR集群时包括以下引导操作。Replace ( 替换 ) `bootstrap_URI` 以及 `RPM_URI` 以及 URIs来自亚马逊 S3 的相应内容。

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

- 在 Amazon 5.36.0 和 6.6.0 到 6.9.0 EMR 版本中，由于 Log4 SecretAgent j2 属性中的文件名模式配置不正确，RecordServer服务组件可能会丢失日志数据。错误的配置导致组件每天只生成一个日志文件。当应用轮换策略时，它会重写现有文件，而不是按预期生成新的日志文件。应变方法是使用引导操作每小时生成一次日志文件，并在文件名中附加一个自动增量的整数来处理轮换。

对于 Amazon EMR 6.0 到 6.9.0 版本，请在启动集群时使用以下引导操作。

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

对于 Amazon EMR 5.36.0，请在启动集群时使用以下引导操作。

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

- Apache Flink 提供原生 S3 FileSystem 和 Hadoop FileSystem 连接器，允许应用程序创建 FileSink 并将数据写入 Amazon S3。这 FileSink 会失败，但有以下两个例外情况之一。

```
java.lang.UnsupportedOperationException: Recoverable writers on Hadoop are only supported for HDFS
```

```
Caused by: java.lang.NoSuchMethodError:
  org.apache.hadoop.io.retry.RetryPolicies.retryOtherThanRemoteAndSaslException(Lorg/
  apache/hadoop/io/retry/RetryPolicy;Ljava/util/Map;)Lorg/apache/hadoop/io/retry/
  RetryPolicy;
                                     at
  org.apache.hadoop.yarn.client.RMProxy.createRetryPolicy(RMProxy.java:302) ~[hadoop-
  yarn-common-3.3.3-amzn-0.jar:?]
```

要解决这个问题，你可以安装一个 Amazon EMR 补丁，该补丁修复了 Flink 中的上述问题。要在启动集群时应用引导操作，请完成以下步骤。

1. 将 flink-rpm 下载到 Amazon S3 存储桶中。你的RPM路是s3://*DOC-EXAMPLE-BUCKET*/rpms/flink/。
2. 使用以下URI命令从 Amazon S3 下载引导脚本和RPM文件。*regionName*替换为您计划启动集群 AWS 区域 的位置。

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

3. Hadoop 3.3.3 在 YARN ([YARN-9608](#)) 中引入了一项更改，即在应用程序完成之前，容器运行的节点一直处于停用状态。此更改可确保如随机数据等本地数据不会丢失，并且您无需重新运行作业。在 Amazon EMR 6.8.0 和 6.9.0 中，这种方法还可能导致启用或未启用托管扩展的集群上的资源利用不足。

在 A [mazon EMR 6.10.0](#) 中，有一种解决方法可以将的 `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` 值设置为 `in`。false `yarn-site.xml` 在亚马逊EMR版本 6.11.0 及更高版本以及 6.8.1、6.9.1 和 6.10.1 中，配置 `false` 默认设置为以解决此问题。

## 更改、增强和解决的问题

- 对于亚马逊 6.9.0 及更高EMR版本，亚马逊EMR安装的所有使用 Log4j 库的组件都使用 Log4j 版本 2.17.1 或更高版本。
- 当您在EMR亚马逊 6.0、6.7.0 和 6.8.0 版本上将 DynamoDB 连接器与 Spark 配合使用时，即使输入拆分引用非空数据，所有从表中读取的内容都会返回空结果。亚马逊 6.9.0 EMR 版本修复了这个问题。
- 当使用 Spark 读取数据时，亚马逊 EMR 6.9.0 增加了对基于 Lake Formation 的 Apache Hudi 访问控制的有限支持。SQL支持使用 Spark 的SELECT查询，SQL并且仅限于列级访问控制。有关更多信息，请参阅 [Hudi 和 Lake Formation](#)。
- 当您使用 Amazon EMR 6.9.0 创建启用了[节点标签](#)的 Hadoop 集群时，[YARN指标](#)会API返回所有分区的汇总信息，而不是默认分区。有关更多信息，请参阅 [YARN-11414](#)。
- 在亚马逊6.9.0 EMR 版本中，我们已将Trino更新到使用Java 17的398版本。之前支持的亚马逊 EMR 6.8.0 版 Trino 版本是在 Java 11 上运行的 Trino 388。有关此变更的更多信息，请参阅 Trino 博客上的 [Trino updates to Java 17](#)。
- 此版本修复了 Apache BigTop 和 Amaz EMR on EC2 集群启动序列之间的时间序列不匹配问题。当系统尝试同时执行两个或多个操作而不是按正确的顺序执行它们时，就会发生这种计时序列不匹配。因此，某些集群配置会遇到实例启动超时和较慢的集群启动时间。
- 当您启动带有最新补丁版本（亚马逊 EMR 5.36 或更高版本、6.6 或更高版本、7.0 或更高版本）的集群时，亚马逊将EMR使用最新的亚马逊 Linux 2023 或亚马逊 Linux 2 版本作为默认亚马逊 Linux 2 版本。EMR AMI有关更多信息，请参阅[使用亚马逊的默认 Amazon Linux AMI EMR](#)。

### Note

此版本不再获得自动AMI更新，因为又有 1 个补丁版本接替了该版本。补丁版本以第二位小数点后的数字 (6.8.**1**) 表示。要查看您是否使用的是最新补丁版本，请查看[发布指南](#)中的可用版本，或者在控制台中创建集群时查看 Amazon EMR 版本下拉列表，或者使用[ListReleaseLabels](#)API或[list-release-labels](#)CLI操作。要获取有关新版本的更新，请订阅“[最新消息？](#)”上的RSS提要 页面。

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 709.1	4.14.348	2024 年 7 月 23 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、AWS GovCloud (美国西部)、AWS GovCloud (美国

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			东部 )、( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )、亚洲太平洋 ( 海得拉巴 )、中东 ( UAE )、欧洲 ( 西班牙 )、欧洲 ( 苏黎世 )、亚太地区 ( 墨尔本 )、以色列 ( 特拉维夫 )、加拿大西部 ( 卡尔加里 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.20230808.0	4.14.320	2023 年 8 月 24 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、欧洲地区 (斯德哥尔摩)、欧洲地区 (米兰)、欧洲地区 (法兰克福)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、以色列 (特拉维夫)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 727.0	4.14.320	2023 年 8 月 14 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、亚太地区 (墨尔本)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 ( )、中东 (UAE)、加拿大 (中部)、以色列 (特拉维夫)



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 719.0	4.14.320	2023 年 8 月 2 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中部 )、以色列 ( 特拉维夫 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 628.0	4.14.318	2023 年 7 月 12 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 612.0	4.14.314	2023 年 6 月 23 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 504.1	4.14.313	2023 年 5 月 16 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 418.0	4.14.311	2023 年 5 月 3 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 404.1	4.14.311	2023 年 4 月 18 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)
2.0.2023 404.0	4.14.311	2023 年 4 月 10 日	美国东部 (弗吉尼亚州北部)、欧洲地区 (巴黎)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 320.0	4.14.309	2023 年 3 月 30 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.202307.0	4.14.305	2023 年 3 月 15 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)



OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 207.0	4.14.304	2023 年 2 月 22 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022 210.1	4.14.301	2023 年 1 月 12 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022 103.3	4.14.296	2022 年 12 月 5 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

## 发行版 6.8.0

以下发行说明包含 Amazon 6.8.0 EMR 版本的信息。更改与 6.7.0 有关。

### 新功能

- Amaz EMR on steps 功能现在支持 Apache Livy 终端节点和JDBC/ODBC客户端。有关更多信息，请参阅 [为 Amazon EMR 步骤配置运行时角色](#)。

- 亚马逊EMR版本6.8.0附带了Apache HBase 版本2.4.12。在此HBase版本中，您可以存档和删除HBase表。Amazon S3 存档过程会将所有表文件重命名为存档目录。这一过程成本高昂且时间较长。现在，您可以跳过存档过程，快速删除大型表。有关更多信息，请参阅 [使用外HBase壳](#)。

## 已知问题

- Hadoop 3.3.3 在 YARN ([YARN-9608](#)) 中引入了一项更改，即在应用程序完成之前，容器运行的节点一直处于停用状态。此更改可确保如随机数据等本地数据不会丢失，并且您无需重新运行作业。在 Amazon EMR 6.8.0 和 6.9.0 中，这种方法还可能导致启用或未启用托管扩展的集群上的资源利用不足。

在 [Amazon EMR 6.10.0](#) 中，有一种解决方法可以将 `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` 值设置为 `in`。 `false` `yarn-site.xml` 在亚马逊EMR版本 6.11.0 及更高版本以及 6.8.1、6.9.1 和 6.10.1 中，配置 `false` 默认设置为以解决此问题。

## 更改、增强和解决的问题

- 当亚马逊EMR发布 6.5.0、6.6.0 或 6.7.0 通过 Apache Spark shell 读取 Apache Phoenix 表时，亚马逊制作了一个。EMR NoSuchMethodError 亚马逊EMR版本 6.8.0 修复了这个问题。
- 亚马逊EMR版本6.8.0附带 [Apache Hudi 0.11.1](#)；但是，EMR亚马逊6.8.0集群也与Hudi 0.12.0的开源集群兼容。 `hudi-spark3.3-bundle_2.12`
- 亚马逊EMR版本 6.8.0 附带了 Apache Spark 3.0。此 Spark 发行版使用 Apache Log4j 2 和 `log4j2.properties` 文件，在 Spark 进程中配置 Log4j。如果您在集群中使用 Spark 或使用自定义配置参数创建EMR集群，并且想要升级到 Amazon EMR 版本 6.8.0，则必须迁移到 Apache Log4j 2 的新 `spark-log4j2` 配置分类和密钥格式。有关更多信息，请参阅 [从 Apache Log4j 1.x 迁移到 Log4j 2.x](#)。
- 当您启动带有最新补丁版本（亚马逊 EMR 5.36 或更高版本、6.6 或更高版本、7.0 或更高版本）的集群时，亚马逊将EMR使用最新的亚马逊 Linux 2023 或亚马逊 Linux 2 版本作为默认亚马逊 Linux 2 版本。EMR AMI有关更多信息，请参阅 [使用亚马逊的默认 Amazon Linux AMI EMR](#)。

### Note

此版本不再获得自动AMI更新，因为又有 1 个补丁版本接替了该版本。补丁版本以第二位小数点后的数字 (`6.8.1`) 表示。要查看您是否使用的是最新补丁版本，请查看 [发布指南](#) 中的可用版本，或者在控制台中创建集群时查看 Amazon EMR 版本下拉列表，或者使

用[ListReleaseLabels](#)API或[list-release-labels](#)CLI操作。要获取有关新版本的更新，请订阅“[最新消息？](#)”上的RSS提要 页面。

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 709.1	4.14.348	2024 年 7 月 23 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、AWS GovCloud (美国西部)、AWS GovCloud (美国

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			东部)、( 美国东部)、中国( 北京)、中国( 宁夏)、亚洲太平洋( 海得拉巴)、中东( UAE)、欧洲( 西班牙)、欧洲( 苏黎世)、亚太地区( 墨尔本)、以色列( 特拉维夫)
2.0.2023 808.0	4.14.320	2023 年 8 月 24 日	美国东部( 弗吉尼亚州北部)、美国东部( 俄亥俄州)、美国西部( 北加利福尼亚)、美国西部( 俄勒冈州)、亚太地区( 香港)、亚太地区( 孟买)、亚太地区( 东京)、亚太地区( 首尔)、亚太地区( 大阪)、亚太地区( 新加坡)、亚太地区( 悉尼)、亚太地区( 雅加达)、亚太地区( 墨尔本)、非洲( 开普敦)、南美洲( 圣保罗)、中东( 巴林)、加拿大( 中部)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 727.0	4.14.320	2023 年 8 月 14 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 米兰 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、



OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 719.0	4.14.320	2023 年 8 月 2 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、亚太地区 ( 墨尔本 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( )、中东 ( UAE )、加拿大 ( 中央 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 628.0	4.14.318	2023 年 7 月 12 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 612.0	4.14.314	2023 年 6 月 23 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 504.1	4.14.313	2023 年 5 月 16 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 418.0	4.14.311	2023 年 5 月 3 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 404.1	4.14.311	2023 年 4 月 18 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)
2.0.2023 404.0	4.14.311	2023 年 4 月 10 日	美国东部 (弗吉尼亚州北部)、欧洲地区 (巴黎)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 320.0	4.14.309	2023 年 3 月 30 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.202307.0	4.14.305	2023 年 3 月 15 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 207.0	4.14.304	2023 年 2 月 22 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 119.1	4.14.301	2023 年 2 月 3 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022 4.14.301 210.1		2023 年 12 月 22 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022 103.3	4.14.296	2022 年 12 月 5 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022 004.0	4.14.294	2022 年 11 月 2 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022 912.1	4.14.291	2022 年 9 月 6 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

## 已知问题

- 当您在EMR亚马逊 6.0、6.7.0 和 6.8.0 版本上将 DynamoDB 连接器与 Spark 配合使用时，即使输入拆分引用非空数据，所有从表中读取的内容都会返回空结果。这是因为 Spark 3.2.0 将 `spark.hadoopRDD.ignoreEmptySplits` 默认设置为 `true`。解决方法是将 `spark.hadoopRDD.ignoreEmptySplits` 显式设置为 `false`。亚马逊 6.9.0 EMR 版本修复了这个问题。

- 当你使用带有 Hive 分区位置格式的 Spark 来读取 Amazon S3 中的数据，并在亚马逊EMR版本 5.30.0 到 5.36.0 和 6.2.0 到 6.9.0 上运行 Spark 时，你可能会遇到一个问题，导致集群无法正确读取数据。如果您的分区具有以下所有特征，会发生这种情况：
  - 从同一个表扫描两个或多个分区。
  - 至少有一个分区目录路径是至少一个其他分区目录路径的前缀，例如，s3://bucket/table/p=a 是 s3://bucket/table/p=a b 的前缀。
  - 另一个分区目录中前缀之后的第一个字符的值为 UTF -8，小于该/字符 (U+002F)。例如，在 s3://bucket/table/p=a b 中，a 和 b 之间出现的空格字符 (U+0020) 就属于此类。请注意，还有其他 14 个非控制字符：! "\$ % & ' ( ) \* + , - 。有关更多信息，请参阅 [UTF-8 编码表和 Unicode 字符](#)。

解决方法是在 spark-defaults 分类中将

spark.sql.sources.fastS3PartitionDiscovery.enabled 配置设置为 false。

- 在 Amazon 5.36.0 和 6.6.0 到 6.9.0 EMR 版本中，由于 Log4 SecretAgent j2 属性中的文件名模式配置不正确，RecordServer 服务组件可能会丢失日志数据。错误的配置导致组件每天只生成一个日志文件。当应用轮换策略时，它会重写现有文件，而不是按预期生成新的日志文件。应变方法是使用引导操作每小时生成一次日志文件，并在文件名中附加一个自动增量的整数来处理轮换。

对于 Amazon EMR 6.0 到 6.9.0 版本，请在启动集群时使用以下引导操作。

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

对于 Amazon EMR 5.36.0，请在启动集群时使用以下引导操作。

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

有关发布时间表的更多信息，请参阅[更改日志](#)。

## 发行版 6.7.0

以下发行说明包含 Amazon 6.7. EMR 0 版本的信息。更改与 6.6.0 有关。

首次发布日期：2022 年 7 月 15 日

## 新功能

- 亚马逊EMR现在支持 Apache Spark 3.2.1、Apache Hive 3.1.3、0.11 Hudi、PrestoDB 0.272 和 Trino 0.378。
- 通过集群上的 Amazon EMR 步骤 ( Spark、Hive ) 支持基于IAM角色和湖泊组成的访问控制。EC2
- 在启用 Apache Ranger 的集群上支持 Apache Spark 数据定义语句。现在，这包括支持 Trino 应用程序在启用 Apache Ranger 的集群上读取和写入 Apache Hive 元数据。有关更多信息，请参阅在亚马逊[上使用 Trino 和 Apache Ranger 启用联合治理](#)。EMR
- 当您启动带有最新补丁版本 ( 亚马逊 EMR 5.36 或更高版本、6.6 或更高版本、7.0 或更高版本 ) 的集群时，亚马逊将EMR使用最新的亚马逊 Linux 2023 或亚马逊 Linux 2 版本作为默认亚马逊 Linux 2 版本。EMR AMI有关更多信息，请参阅[使用亚马逊的默认 Amazon Linux AMI EMR](#)。

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 709.1	4.14.348	2024 年 7 月 23 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、AWS GovCloud ( 美国东部 )、( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )、亚洲太平洋 ( 海得拉巴 )、中东 ( UAE )、欧洲 ( 西班牙 )、欧洲 ( 苏黎世 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 223.0	4.14.336	2024 年 3 月 8 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 131.0	4.14.336	2024年2月14日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024124.0	4.14.336	2024 年 2 月 7 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			中国(北京)、中国(宁夏)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 109.0	4.14.334	2024 年 1 月 24 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 218.0	4.14.330	2024 年 1 月 2 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			中国(北京)、中国(宁夏)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 206.0	4.14.330	2023 年 12 月 22 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			中国(北京)、中国(宁夏)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023116.0	4.14.328	2023 年 12 月 11 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 101.0	4.14.327	2023 年 11 月 16 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023020.1	4.14.326	2023 年 11 月 7 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023012.1	4.14.326	2023 年 10 月 26 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			中国(北京)、中国(宁夏)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 926.0	4.14.322	2023 年 10 月 19 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 906.0	4.14.322	2023 年 10 月 4 日	中国 (北京)、中国 (宁夏)  美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、欧洲地区 (斯德哥尔摩)、欧洲地区 (米兰)、欧洲地区 (法兰克福)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.20230822.0	4.14.322	2023 年 8 月 30 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、欧洲地区 (斯德哥尔摩)、欧洲地区 (米兰)、欧洲地区 (法兰克福)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)



OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 808.0	4.14.320	2023 年 8 月 24 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、欧洲地区 (斯德哥尔摩)、欧洲地区 (米兰)、欧洲地区 (法兰克福)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 727.0	4.14.320	2023 年 8 月 14 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 米兰 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 719.0	4.14.320	2023 年 8 月 2 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 西班牙 )、欧洲 ( 法兰克福 )、欧洲 ( 苏黎世 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 海得拉巴 )、亚太地区 ( 东京 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、中东 ( 巴林 )、中东 ( UAE )、加拿大 ( 中部 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 628.0	4.14.318	2023 年 7 月 12 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 612.0	4.14.314	2023 年 6 月 23 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 504.1	4.14.313	2023 年 5 月 16 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 418.0	4.14.311	2023 年 5 月 3 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 404.1	4.14.311	2023 年 4 月 18 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)
2.0.2023 404.0	4.14.311	2023 年 4 月 10 日	美国东部 (弗吉尼亚州北部)、欧洲地区 (巴黎)



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 320.0	4.14.309	2023 年 3 月 30 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 307.0	4.14.305	2023 年 3 月 15 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.202307.0	4.14.304	2023 年 2 月 22 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 119.1	4.14.301	2023 年 2 月 3 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022 210.1	4.14.301	2023 年 12 月 22 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022 103.3	4.14.296	2022 年 12 月 5 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022004.0	4.14.294	2022 年 11 月 2 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022 912.1	4.14.291	2022 年 10 月 7 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)
2.0.2022 719.0	4.14.287	2022 年 8 月 10 日	us-west-1 , eu-west-3 , eu-north-1 , ap-south-1 , me-south-1



OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022 606.1	4.14.281	2022 年 7 月 15 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

## 已知问题

- 当亚马逊EMR版本 6.5.0、6.6.0 或 6.7.0 通过 Apache Spark 外壳读取 Apache Phoenix 表时，会发生这种情况NoSuchMethodError是因为亚马逊使用了错误的。EMR Hbase.compat.version亚马逊EMR版本 6.8.0 修复了这个问题。
- 当您在EMR亚马逊 6.0、6.7.0 和 6.8.0 版本上将 DynamoDB 连接器与 Spark 配合使用时，即使输入拆分引用非空数据，所有从表中读取的内容都会返回空结果。这是因为 Spark

3.2.0 将 `spark.hadoopRDD.ignoreEmptySplits` 默认设置为 `true`。解决方法是将 `spark.hadoopRDD.ignoreEmptySplits` 显式设置为 `false`。亚马逊 6.9.0 EMR 版本修复了这个问题。

- 当你使用带有 Hive 分区位置格式的 Spark 来读取 Amazon S3 中的数据，并在亚马逊 EMR 版本 5.30.0 到 5.36.0 和 6.2.0 到 6.9.0 上运行 Spark 时，你可能会遇到一个问题，导致集群无法正确读取数据。如果您的分区具有以下所有特征，会发生这种情况：
  - 从同一个表扫描两个或多个分区。
  - 至少有一个分区目录路径是至少一个其他分区目录路径的前缀，例如，`s3://bucket/table/p=a` 是 `s3://bucket/table/p=a b` 的前缀。
  - 另一个分区目录中前缀之后的第一个字符的值为 UTF -8，小于该/字符 (U+002F)。例如，在 `s3://bucket/table/p=a b` 中，`a` 和 `b` 之间出现的空格字符 (U+0020) 就属于此类。请注意，还有其他 14 个非控制字符：`!"#$%&'()*+,-.`。有关更多信息，请参阅 [UTF-8 编码表和 Unicode 字符](#)。

解决方法是在 `spark-defaults` 分类中将 `spark.sql.sources.fastS3PartitionDiscovery.enabled` 配置设置为 `false`。

- 在 Amazon 5.36.0 和 6.6.0 到 6.9.0 EMR 版本中，由于 Log4 SecretAgent j2 属性中的文件名模式配置不正确，RecordServer 服务组件可能会丢失日志数据。错误的配置导致组件每天只生成一个日志文件。当应用轮换策略时，它会重写现有文件，而不是按预期生成新的日志文件。应变方法是使用引导操作每小时生成一次日志文件，并在文件名中附加一个自动增量的整数来处理轮换。

对于 Amazon EMR 6.0 到 6.9.0 版本，请在启动集群时使用以下引导操作。

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

对于 Amazon EMR 5.36.0，请在启动集群时使用以下引导操作。

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

- 在 Amazon EMR 6.7 或更低版本上运行的集群 `GetClusterSessionCredentialsAPI` 不支持。

## 发行版 6.6.0

以下发行说明包含 Amazon 6.6. EMR 0 版本的信息。更改与 6.5.0 有关。

首次发布日期：2022 年 5 月 9 日

文档更新日期：2022 年 6 月 15 日

## 新功能

- 亚马逊 EMR 6.6 现在支持 Apache Spark 3.2、Apache Spark RAPIDS 22.02、CUDA 11、Apache Hudi 0.10.1、Apache Iceberg 0.13、Trino 0.367 和 PrestoDB 0.267。
- 当您启动带有最新补丁版本（亚马逊 EMR 5.36 或更高版本、6.6 或更高版本、7.0 或更高版本）的集群时，亚马逊将EMR使用最新的亚马逊 Linux 2023 或亚马逊 Linux 2 版本作为默认亚马逊 Linux 2 版本。EMR AMI有关更多信息，请参阅[使用亚马逊的默认 Amazon Linux AMI EMR](#)。

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 709.1	4.14.348	2024 年 7 月 23 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 223.0	4.14.336	2024 年 3 月 8 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			中国(北京)、中国(宁夏)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 131.0	4.14.336	2024年2月14日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			中国(北京)、中国(宁夏)



OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024124.0	4.14.336	2024 年 2 月 7 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			中国(北京)、中国(宁夏)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2024 109.0	4.14.334	2024 年 1 月 24 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			中国(北京)、中国(宁夏)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 218.0	4.14.330	2024 年 1 月 2 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 206.0	4.14.330	2023 年 12 月 22 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			中国(北京)、中国(宁夏)



OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023116.0	4.14.328	2023 年 12 月 11 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			中国(北京)、中国(宁夏)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 101.0	4.14.327	2023 年 11 月 16 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			中国 ( 北京 )、中国 ( 宁夏 )

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023020.1	4.14.326	2023 年 11 月 7 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
			中国(北京)、中国(宁夏)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023012.1	4.14.326	2023 年 10 月 26 日	美国东部 ( 弗吉尼亚北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 加利福尼亚北部 )、美国西部 ( 俄勒冈 )、欧洲 ( 斯德哥尔摩 )、欧洲 ( 米兰 )、欧洲 ( 法兰克福 )、欧洲 ( 爱尔兰 )、欧洲 ( 伦敦 )、欧洲 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 新加坡 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 ) ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )、AWS GovCloud ( 美国西部 )、( 美国东部 )、AWS GovCloud ( 美国东部 )、

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
			中国 ( 北京 )、中国 ( 宁夏 )



OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 926.0	4.14.322	2023 年 10 月 19 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (法兰克福)、欧洲 (爱尔兰)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (新加坡)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)、AWS GovCloud (美国西部)、(美国东部)、AWS GovCloud (美国东部)、

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 906.0	4.14.322	2023 年 10 月 4 日	中国 (北京)、中国 (宁夏)  美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、欧洲地区 (斯德哥尔摩)、欧洲地区 (米兰)、欧洲地区 (法兰克福)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 822.0	4.14.322	2023 年 8 月 30 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 米兰 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、亚太地区 ( 雅加达 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )、加拿大 ( 中部 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 808.0	4.14.320	2023 年 8 月 24 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、欧洲地区 (斯德哥尔摩)、欧洲地区 (米兰)、欧洲地区 (法兰克福)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 727.0	4.14.320	2023 年 8 月 14 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、欧洲地区 (斯德哥尔摩)、欧洲地区 (米兰)、欧洲地区 (法兰克福)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、加拿大 (中部)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 719.0	4.14.320	2023 年 8 月 2 日	美国东部 (弗吉尼亚北部)、美国东部 (俄亥俄州)、美国西部 (加利福尼亚北部)、美国西部 (俄勒冈)、欧洲 (斯德哥尔摩)、欧洲 (米兰)、欧洲 (西班牙)、欧洲 (法兰克福)、欧洲 (苏黎世)、欧洲 (伦敦)、欧洲 (巴黎)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (海得拉巴)、亚太地区 (东京)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、亚太地区 (雅加达)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)、中东 (巴林)、中东 (UAE)、加拿大 (中部)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 628.0	4.14.318	2023 年 7 月 12 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 612.0	4.14.314	2023 年 6 月 23 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)



OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 504.1	4.14.313	2023 年 5 月 16 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 418.0	4.14.311	2023 年 5 月 3 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 404.1	4.14.311	2023 年 4 月 18 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)
2.0.2023 404.0	4.14.311	2023 年 4 月 10 日	美国东部 (弗吉尼亚州北部)、欧洲地区 (巴黎)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 320.0	4.14.309	2023 年 3 月 30 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 307.0	4.14.305	2023 年 3 月 15 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.202307.0	4.14.304	2023 年 2 月 22 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2023 119.1	4.14.301	2023 年 2 月 3 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022 210.1	4.14.301	2023 年 12 月 22 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )



OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022 103.3	4.14.296	2022 年 12 月 5 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label ( 亚马逊 Linux 版本 )	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022004.0	4.14.294	2022 年 11 月 2 日	美国东部 ( 弗吉尼亚州北部 )、美国东部 ( 俄亥俄州 )、美国西部 ( 北加利福尼亚 )、美国西部 ( 俄勒冈州 )、加拿大 ( 中部 )、欧洲地区 ( 斯德哥尔摩 )、欧洲地区 ( 爱尔兰 )、欧洲地区 ( 伦敦 )、欧洲地区 ( 巴黎 )、欧洲地区 ( 法兰克福 )、欧洲地区 ( 米兰 )、亚太地区 ( 香港 )、亚太地区 ( 孟买 )、亚太地区 ( 雅加达 )、亚太地区 ( 东京 )、亚太地区 ( 首尔 )、亚太地区 ( 大阪 )、亚太地区 ( 新加坡 )、亚太地区 ( 悉尼 )、非洲 ( 开普敦 )、南美洲 ( 圣保罗 )、中东 ( 巴林 )

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022 912.1	4.14.291	2022 年 10 月 7 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)
2.0.2022 805.0	4.14.287	2022 年 8 月 30 日	us-west-1

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022 719.0	4.14.287	2022 年 8 月 10 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022 426.0	4.14.281	2022 年 6 月 10 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

OsRelease Label (亚马逊 Linux 版本)	Amazon Linux 内核版本	可用日期	支持的区域
2.0.2022 406.1	4.14.275	2022 年 5 月 2 日	美国东部 (弗吉尼亚州北部)、美国东部 (俄亥俄州)、美国西部 (北加利福尼亚)、美国西部 (俄勒冈州)、加拿大 (中部)、欧洲地区 (斯德哥尔摩)、欧洲地区 (爱尔兰)、欧洲地区 (伦敦)、欧洲地区 (巴黎)、欧洲地区 (法兰克福)、欧洲地区 (米兰)、亚太地区 (香港)、亚太地区 (孟买)、亚太地区 (雅加达)、亚太地区 (东京)、亚太地区 (首尔)、亚太地区 (大阪)、亚太地区 (新加坡)、亚太地区 (悉尼)、非洲 (开普敦)、南美洲 (圣保罗)、中东 (巴林)

- [在 Amazon EMR 6.6 及更高版本中，使用 Log4j 1.x 和 Log4j 2.x 的应用程序将分别升级为使用 Log4j 1.2.17 \(或更高版本\) 和 Log4j 2.17.1 \(或更高版本\)，并且不需要使用提供的引导操作来缓解问题。](#) CVE
- [托管扩展] Spark shuffle 数据托管扩展优化——对于 Amazon EMR 5.34.0 及更高版本以及 6.4.0 及更高EMR版本，托管扩展现在支持 Spark 洗牌数据 (Spark 跨分区重新分配以执行特定操作的数据)。有关洗牌操作的更多信息，请参阅《[亚马逊EMREMR管理指南](#)》和《[Spark 编程指南](#)》EMR 中的“[在亚马逊使用托管扩展](#)”。

- 从亚马逊 EMR 5.32.0 和 6.5.0 开始，Apache Spark 的动态执行器大小设置默认处于启用状态。要打开或关闭此功能，您可以使用 `spark.yarn.heterogeneousExecutors.enabled` 配置参数。

## 更改、增强和解决的问题

- 对于使用EMR默认AMI选项且仅安装常见应用程序（例如 Apache Hadoop、Apache Spark 和 Apache Hive）的集群，亚马逊平均将集群启动时间EMR缩短多达 80 秒。

## 已知问题

- 当亚马逊EMR版本 6.5.0、6.6.0 或 6.7.0 通过 Apache Spark 外壳读取 Apache Phoenix 表时，会发生这种情况NoSuchMethodError是因为亚马逊使用了错误的。EMR Hbase.compat.version亚马逊EMR版本 6.8.0 修复了这个问题。
- 当您在EMR亚马逊 6.0、6.7.0 和 6.8.0 版本上将 DynamoDB 连接器与 Spark 配合使用时，即使输入拆分引用非空数据，所有从表中读取的内容都会返回空结果。这是因为 Spark 3.2.0 将 `spark.hadoopRDD.ignoreEmptySplits` 默认设置为 `true`。解决方法是将 `spark.hadoopRDD.ignoreEmptySplits` 显式设置为 `false`。亚马逊 6.9.0 EMR 版本修复了这个问题。
- 在 Trino 长期运行的集群上，Amazon EMR 6.0 在 Trino `jvm.config` 中启用了垃圾收集日志参数，以便从垃圾收集日志中获得更好的见解。此更改会将许多垃圾回收日志附加到 `launcher.log (/var/log/trino/launcher.log)` 文件。如果您在 Amazon EMR 6.6.0 中运行 Trino 集群，则由于附加的日志，在集群运行了几天后，您可能会遇到节点磁盘空间耗尽的情况。

此问题的解决方法是在为 Amazon 6.0 创建或克隆集群时，将以下脚本作为引导操作运行，以禁用 `jvm.config` 中的垃圾收集日志参数。EMR

```
#!/bin/bash
set -ex
PRESTO_PUPPET_DIR='/var/aws/emr/bigtop-deploy/puppet/modules/trino'
sudo bash -c "sed -i '/-Xlog/d' ${PRESTO_PUPPET_DIR}/templates/jvm.config"
```

- 当你使用带有 Hive 分区位置格式的 Spark 来读取 Amazon S3 中的数据，并在亚马逊EMR版本 5.30.0 到 5.36.0 和 6.2.0 到 6.9.0 上运行 Spark 时，你可能会遇到一个问题，导致集群无法正确读取数据。如果您的分区具有以下所有特征，会发生这种情况：
  - 从同一个表扫描两个或多个分区。

- 至少有一个分区目录路径是至少一个其他分区目录路径的前缀，例如，`s3://bucket/table/p=a` 是 `s3://bucket/table/p=a b` 的前缀。
- 另一个分区目录中前缀之后的第一个字符的值为 UTF-8，小于该字符 (U+002F)。例如，在 `s3://bucket/table/p=a b` 中，`a` 和 `b` 之间出现的空格字符 (U+0020) 就属于此类。请注意，还有其他 14 个非控制字符：`!"#$%&'()*+,-`。有关更多信息，请参阅 [UTF-8 编码表和 Unicode 字符](#)。

解决方法是在 `spark-defaults` 分类中将

`spark.sql.sources.fastS3PartitionDiscovery.enabled` 配置设置为 `false`。

- 在 Amazon 5.36.0 和 6.6.0 到 6.9.0 EMR 版本中，由于 Log4j SecretAgent j2 属性中的文件名模式配置不正确，RecordServer 服务组件可能会丢失日志数据。错误的配置导致组件每天只生成一个日志文件。当应用轮换策略时，它会重写现有文件，而不是按预期生成新的日志文件。应变方法是使用引导操作每小时生成一次日志文件，并在文件名中附加一个自动增量的整数来处理轮换。

对于 Amazon EMR 6.0 到 6.9.0 版本，请在启动集群时使用以下引导操作。

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

对于 Amazon EMR 5.36.0，请在启动集群时使用以下引导操作。

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

## 发行版 5.35.0

这是亚马逊 5.35.0 EMR 版本的发行说明。

以下发行说明包含 Amazon EMR 版本 5.35.0 的信息。更改与 5.34.0 有关。

首次发布日期: 2022 年 3 月 30 日

### 新功能

- 使用 Log4j 1.x 和 Log4j 2.x 的亚马逊 EMR 版本 5.35 应用程序已分别升级为使用 Log4j 1.2.17 (或更高版本) 和 Log4j 2.17.1 (或更高版本)，并且不需要使用引导操作来缓解先前版本中的问题。CVE 请参阅 [缓解方法 CVE-2021-44228](#)。



## 更改、增强和解决的问题

## Flink 更改

更改类型	描述
升级	<ul style="list-style-type: none"> <li>• 将 Flink 版本更新到 1.14.2。</li> <li>• log4j 升级到 2.17.1。</li> </ul>

## Hadoop 更改

更改类型	描述
自 5.34.0 起的 Hadoop 开源向后移植 EMR	<ul style="list-style-type: none"> <li>• <a href="#">YARN-10438</a> : containerId 在 ClientRMService #中处理空值 () getContainerReport</li> <li>• <a href="#">YARN-7266</a> : 时间轴服务器事件处理程序线程已锁定</li> <li>• <a href="#">YARN-10438</a> : 如果 RollingLevelDb 文件损坏或丢失，则无法启动 ATS 1.5</li> <li>• <a href="#">HADOOP-13500</a> : 同步配置属性对象的小版本</li> <li>• <a href="#">YARN-10651</a> : 在 in 中 CapacityScheduler 坠毁。NPE AbstractYarnScheduler updateNodeResource()</li> <li>• <a href="#">HDFS-12221</a> : 替换中的 xerces XmlEditsVisitor</li> <li>• <a href="#">HDFS-16410</a> : <a href="#">解析中的 X ml 不安全</a> OfflineEditsXMLLoader</li> </ul>
Hadoop 更改和修复	<ul style="list-style-type: none"> <li>• 中KMS使用的 Tomcat , HttpFS 已升级到 8.5.75</li> <li>• 在 FileSystemOptimizedCommitter V2 中 , 成功标记写在创建提交者时定义的 commitJob 输出路径中。由于任务级别 commitJob 的输出路径可能不同 , 因此已将路径更正为使用清单文件中定义的路径。对于 Hive 作业 , 这会</li> </ul>

更改类型	描述
	导致在执行诸如动态分区或UNIONALL之类的操作时正确写入成功标记。

## Hive 更改

更改类型	描述
Hive 已升级到开源 <a href="#">版本 2.3.9</a> ，包括这些修复 JIRA	<ul style="list-style-type: none"> <li>• <a href="#">HIVE-17155</a>: HiveConf .java 中的 findConfFile () 在 conf 路径中存在一些问题</li> <li>• <a href="#">HIVE-24797</a> : 在解析 Avro 架构时禁用验证默认值</li> <li>• <a href="#">HIVE-21563</a> : 通过禁用 Once 来提高 Table# 性能 getEmptyTable registerAllFunctions</li> <li>• <a href="#">HIVE-18147</a> : 使用 java.net 时，测试可能会失败。BindException: 地址已在使用中</li> <li>• <a href="#">HIVE-24608</a> : 在 Hive 2.3.x 的客户端中切换回 get_table HMS</li> <li>• <a href="#">HIVE-211200</a> : 矢量化——日期列抛出 java.lang. UnsupportedOperationException 用于实木复合地板</li> <li>• <a href="#">HIVE-19228</a> : 移除 commons-httpclient 3.x 用法</li> </ul>
自 5.34.0 起的 Hive 开源向后移植 EMR	<ul style="list-style-type: none"> <li>• <a href="#">HIVE-19990</a> : 在连接条件下使用间隔文字进行查询失败</li> <li>• <a href="#">HIVE-25824</a> : 将 branch-2.3 升级到 log4j 2.17.0</li> <li>• <a href="#">TEZ-4062</a> : 任务完成后应中止推测尝试调度</li> <li>• <a href="#">TEZ-4108</a> : NullPointerException 在投机执行期间竞赛条件</li> <li>• <a href="#">TEZ-3918</a> : 设置 tez.task.log.level 不起作用</li> </ul>
Hive 升级和修复	<ul style="list-style-type: none"> <li>• 将 Log4j 版本升级到 2.17.1</li> </ul>

更改类型	描述
	<ul style="list-style-type: none"> <li>将ORC版本升级到 1.4.3</li> <li>修复了由于处罚线程而导致的死锁 ShuffleScheduler</li> </ul>
新功能	<ul style="list-style-type: none"> <li>添加了在 AM 日志中打印 Hive 查询的功能 默认情况下，将禁用该功能。标记/配置：<code>tez.am.emr.print.hive.query.in.log</code>。状态（默认）：<code>FALSE</code>。</li> </ul>

## Oozie 更改

更改类型	描述
Oozie 自 5.34.0 起的开源向后移植 EMR	<ul style="list-style-type: none"> <li><a href="#">OOZIE-3652</a>：Oozie 启动器应在出现目录列表时重试目录列表 NoSuchFileException</li> </ul>

## Pig 更改

更改类型	描述
升级	<ul style="list-style-type: none"> <li>log4j 升级到 1.2.17。</li> </ul>

## 已知问题

- 当你使用带有 Hive 分区位置格式的 Spark 来读取 Amazon S3 中的数据，并在亚马逊EMR版本 5.30.0 到 5.36.0 和 6.2.0 到 6.9.0 上运行 Spark 时，你可能会遇到一个问题，导致集群无法正确读取数据。如果您的分区具有以下所有特征，会发生这种情况：
  - 从同一个表扫描两个或多个分区。
  - 至少有一个分区目录路径是至少一个其他分区目录路径的前缀，例如，`s3://bucket/table/p=a` 是 `s3://bucket/table/p=a b` 的前缀。
  - 另一个分区目录中前缀之后的第一个字符的值为 UTF -8，小于该/字符 (U+002F)。例如，在 `s3://bucket/table/p=a b` 中，`a` 和 `b` 之间出现的空格字符 (U+0020) 就属于此类。请注意，还有其他 14 个非控制字符：`!"#$%&'()*+,-.`。有关更多信息，请参阅 [UTF-8 编码表和 Unicode 字符](#)。

解决方法是在 `spark-defaults` 分类中将 `spark.sql.sources.fastS3PartitionDiscovery.enabled` 配置设置为 `false`。

## 发行版 5.34.0

以下发行说明包含有关亚马逊 5.34.0 EMR 版本的信息。更改与 5.33.1 有关。

首次发布日期：2022 年 1 月 20 日

发布更新日期：2022 年 3 月 21 日

### 新功能

- [托管扩展] Spark shuffle 数据托管扩展优化——对于 Amazon EMR 5.34.0 及更高版本以及 6.4.0 及更高EMR版本，托管扩展现在支持 Spark 洗牌数据（Spark 跨分区重新分配以执行特定操作的数据）。有关洗牌操作的更多信息，请参阅《[亚马逊EMREMR管理指南](#)》和《[Spark 编程指南](#)》EMR 中的“[在亚马逊使用托管扩展](#)”。
- [Hudi] 简化了 Hudi 配置的改进。预设情况下禁用乐观并发控制。

### 更改、增强和解决的问题

- 此版本旨在修复 Amazon EMR Scaling 无法成功扩展/缩小集群或导致应用程序故障时出现的问题。
- 以前，在多主集群上手动重启资源管理器会导致 Amaz EMR on 集群上的守护程序（例如 Zookeeper）在 Zookeeper znode 文件中重新加载所有先前已停用或丢失的节点。在某些情况下，这会导致超出默认限制。Amazon EMR 现已从 Zookeeper 文件中删除已停用或丢失超过一小时的节点记录，并且内部限制也已提高。
- 修复了 Amaz EMR on 集群上守护程序运行状况检查活动（例如收集YARN节点状态和HDFS节点状态）时，对利用率很高的大型集群的扩展请求失败的问题。之所以发生这种情况，是因为集群上的守护程序无法将节点的运行状况数据传送到内部 Ama EMR zon 组件。
- 改进了EMR集群上的守护程序，可在重复使用 IP 地址时正确跟踪节点状态，从而提高扩展操作期间的可靠性。
- [SPARK-29683](#)。修复了集群缩减期间出现任务失败的问题，因为 Spark 假定所有可用节点都被拒绝列出。
- [YARN-9011](#)。修复了集群尝试向上或向下扩展时，由于YARN停用过程中的争用条件而导致任务失败的问题。

- 通过确保 Amaz EMR on 集群上守护程序和 YARN /之间的节点状态始终保持一致，修复了集群扩展期间步骤或任务失败的问题。HDFS
- 修复了启用了 Kerberos 身份验证的 Amazon 集群的EMR集群操作（例如缩减和步骤提交）失败的问题。这是因为 Amaz EMR on 集群上守护程序没有续订 Kerberos 票证，而要与主节点上YARN运行的 HDFS /进行安全通信，则需要续订 Kerberos 票证。
- Zeppelin 已升级到版本 0.10.0。
- Livy 修复 - 已升级到 0.7.1
- Spark 性能改进-在 5.34.0 中重写某些 Spark 配置值时，异构执行器将被禁用。EMR
- 默认情况下HDFS，Web 和 HTTPFS 服务器处于禁用状态。您可以HDFS使用 Hadoop 配置重新启用 Web。dfs.webhdfs.enabledHTTPFS 服务器可以通过使用 `sudo systemctl start hadoop-httpfs` 启动。

## 已知问题

- 与 Livy 用户模仿一起使用的 Amazon N EMR otebooks 功能不起作用，因为 HTTPFS 在默认情况下处于禁用状态。在这种情况下，EMR笔记本无法连接到启用了 Livy 模拟的集群。解决方法是在使用将EMR笔记本连接到集群之前启动 HTTPFS 服务器。`sudo systemctl start hadoop-httpfs`
- Hue 查询在 Amazon EMR 6.4.0 中不起作用，因为 Apache Hadoop httpFS 服务器默认处于禁用状态。要在亚马逊 EMR 6.4.0 上使用 Hue，要么使用亚马逊步骤在亚马逊EMR主节点上手动启动 HttpFS 服务器`sudo systemctl start hadoop-httpfs`，要么使用[亚马逊步骤](#)。EMR
- 与 Livy 用户模仿一起使用的 Amazon N EMR otebooks 功能不起作用，因为 HTTPFS 在默认情况下处于禁用状态。在这种情况下，EMR笔记本无法连接到启用了 Livy 模拟的集群。解决方法是在使用将EMR笔记本连接到集群之前启动 HTTPFS 服务器。`sudo systemctl start hadoop-httpfs`
- 当你使用带有 Hive 分区位置格式的 Spark 来读取 Amazon S3 中的数据，并在亚马逊EMR版本 5.30.0 到 5.36.0 和 6.2.0 到 6.9.0 上运行 Spark 时，你可能会遇到一个问题，导致集群无法正确读取数据。如果您的分区具有以下所有特征，会发生这种情况：
  - 从同一个表扫描两个或多个分区。
  - 至少有一个分区目录路径是至少一个其他分区目录路径的前缀，例如，`s3://bucket/table/p=a` 是 `s3://bucket/table/p=a b` 的前缀。
  - 另一个分区目录中前缀之后的第一个字符的值为 UTF -8，小于该/字符 (U+002F)。例如，在 `s3://bucket/table/p=a b` 中，`a` 和 `b` 之间出现的空格字符 (U+0020) 就属于此类。请注意，

还有其他 14 个非控制字符：`!"#$%&'()*+,-`。有关更多信息，请参阅 [UTF-8 编码表和 Unicode 字符](#)。

解决方法是在 `spark-defaults` 分类中将 `spark.sql.sources.fastS3PartitionDiscovery.enabled` 配置设置为 `false`。

## 发行版 6.5.0

以下发行说明包含 Amazon 6.5.0 EMR 版本的信息。更改与 6.4.0 有关。

首次发布日期：2022 年 1 月 20 日

发布更新日期：2022 年 3 月 21 日

### 新功能

- [托管扩展] Spark shuffle 数据托管扩展优化——对于 Amazon EMR 5.34.0 及更高版本以及 6.4.0 及更高EMR版本，托管扩展现在支持 Spark 洗牌数据（Spark 跨分区重新分配以执行特定操作的数据）。有关洗牌操作的更多信息，请参阅《[亚马逊EMREMR管理指南](#)》和《[Spark 编程指南](#)》EMR 中的“[在亚马逊使用托管扩展](#)”。
- 从亚马逊 EMR 5.32.0 和 6.5.0 开始，Apache Spark 的动态执行器大小设置默认处于启用状态。要打开或关闭此功能，您可以使用 `spark.yarn.heterogeneousExecutors.enabled` 配置参数。
- 支持 Apache Iceberg 开放表格式，用于大型分析数据集。
- 支持 ranger-trino-plugin 2.0.1-amzn-1
- 支持 toree 0.5.0

### 更改、增强和解决的问题

- 亚马逊 EMR 6.5 发行版现在支持 Apache Iceberg 0.12.0，并通过适用于 Apache Spark 的亚马逊 Runtime、适用于 Presto 的亚马逊 EMR Runtime 和 Apache Hive 的EMR亚马逊 Runtime 提供了运行时改进。EMR
- [Apache Iceberg](#) 是一种适用于 Amazon S3 中大型数据集的开放表格式，它提供了对大型表、原子提交、并发写入和SQL兼容表演化的快速查询性能。在 EMR 6.5 版本中，您可以使用带有 Iceberg 表格式的 Apache Spark 3.1.2。
- Apache Hudi 0.9 增加了 Spark SQL DDL 和支援。DML这允许您仅SQL使用语句创建、更新插入 Hudi 表。Apache Hudi 0.9 还包括查询端和写入器端的性能改进。

- Amazon EMR Runtime for Apache Hive 通过删除暂存操作期间的重命名操作提高了 Apache Hive 在 Amazon S3 上的性能，并提高了用于修复表的元数据仓检查 (MSCK) 命令的性能。

## 已知问题

- 当亚马逊EMR版本 6.5.0、6.6.0 或 6.7.0 通过 Apache Spark 外壳读取 Apache Phoenix 表时，会发生这种情况NoSuchMethodError是因为亚马逊使用了错误的。EMR Hbase.compat.version亚马逊EMR版本 6.8.0 修复了这个问题。
- 高可用性 ( HA ) 的 Hbase 捆绑集群无法使用默认卷大小和实例类型进行预置。此问题的变通解决方法是增加根卷大小。
- 要将 Spark 操作与 Apache Oozie 一起使用，必须将以下配置添加到 Oozie workflow.xml 文件中。否则，Oozie 启动的 Spark 执行程序的路径中EMRFS将缺少几个关键库，例如 Hadoop 和。

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

- 当你使用带有 Hive 分区位置格式的 Spark 来读取 Amazon S3 中的数据，并在亚马逊EMR版本 5.30.0 到 5.36.0 和 6.2.0 到 6.9.0 上运行 Spark 时，你可能会遇到一个问题，导致集群无法正确读取数据。如果您的分区具有以下所有特征，会发生这种情况：
  - 从同一个表扫描两个或多个分区。
  - 至少有一个分区目录路径是至少一个其他分区目录路径的前缀，例如，s3://bucket/table/p=a 是 s3://bucket/table/p=a b 的前缀。
  - 另一个分区目录中前缀之后的第一个字符的值为 UTF -8，小于该/字符 (U+002F)。例如，在 s3://bucket/table/p=a b 中，a 和 b 之间出现的空格字符 (U+0020) 就属于此类。请注意，还有其他 14 个非控制字符：!"#\$%&'()\*+,-。有关更多信息，请参阅 [UTF-8 编码表和 Unicode 字符](#)。

解决方法是在 spark-defaults 分类中将

spark.sql.sources.fastS3PartitionDiscovery.enabled 配置设置为 false。

## 发行版 6.4.0

以下发行说明包含有关亚马逊 6.4.0 EMR 版本的信息。更改与 6.3.0 有关。

首次发布日期：2021 年 9 月 20 日

发布更新日期：2022 年 3 月 21 日

## 受支持的应用程序

- AWS SDK for Java 版本 1.12.31
- CloudWatch 水槽版本 2.2.0
- DynamoDB 连接器 4.16.0
- EMRFS版本 2.47.0
- 亚马逊 EMR Goodies 版本 3.2.0
- 亚马逊 EMR Kinesis Connector 版本 3.5.0
- 亚马逊EMR录音服务器版本 2.1.0
- 亚马逊EMR脚本版本 2.5.0
- Flink 1.13.1
- Ganglia 3.7.2
- AWS Glue Hive Metastore 客户端版本 3.3.0
- Hadoop 3.2.1-amzn-4
- HBase版本 2.4.4-amzn-0
- HBase-operator-tools 1.1.0
- HCatalog版本 3.1.2-amzn-5
- Hive 3.1.2-amzn-5
- Hudi 0.8.0-amzn-0
- Hue 4.9.0
- Java JDK 版本 Corretto-8.302.08.1 ( 版本 1.8.0\_302-b08 )
- JupyterHub 版本 1.4.1
- Livy 0.7.1-incubating
- MXNet版本 1.8.0
- Oozie 5.2.1
- Phoenix 5.1.2
- Pig 0.17.0
- Presto 0.254.1-amzn-0
- Trino 359
- Apache RangerKMS ( 多主机透明加密 ) 版本 2.0.0
- ranger-plugins 2.0.1-amzn-0



- ranger-s3-plugin 1.2.0
- SageMaker Spark SDK 版本 1.4.1
- Scala 版本 2.12.10 ( 打开 JDK 64 位服务器虚拟机 , Java 1.8.0\_282 )
- Spark 3.1.2-amzn-0
- spark-rapids 0.4.1
- Sqoop 1.4.7
- TensorFlow 版本 2.4.1
- tez 0.9.2
- Zeppelin 0.9.0
- Zookeeper 3.5.7
- 连接器和驱动程序 : DynamoDB 连接器 4.16.0

## 新功能

- [托管扩展] Spark shuffle 数据托管扩展优化 ——对于 Amazon EMR 5.34.0 及更高版本以及 6.4.0 及更高EMR版本，托管扩展现在支持 Spark 洗牌数据 ( Spark 跨分区重新分配以执行特定操作的数据 )。有关洗牌操作的更多信息，请参阅《[亚马逊EMREMR管理指南](#)》和《[Spark 编程指南](#)》EMR 中的“[在亚马逊使用托管扩展](#)”。
- 在支持 Apache Ranger 的EMR亚马逊集群上，你可以使用、和使用 Apache SQL Spark 将数据插入或更新 Apache Hive 元数据仓表。INSERT INTO INSERT OVERWRITE ALTER TABLEALTERTABLE与 Spark 一起使用时SQL，分区位置必须是表位置的子目录。Amazon 目前EMR不支持在分区位置与表位置不同的分区中插入数据。
- Presto SQL 已[更名为 Trino](#)。
- Hive : SELECT获取到LIMIT子句中提到的记录数后，立即停止查询执行，从而加快带LIMIT子句的简单查询的执行速度。简单SELECT查询是指没有 BY/b GROUP y 子句ORDER的查询或没有 reducer 阶段的查询。例如，SELECT \* from <TABLE> WHERE <Condition> LIMIT <Number>。

## Hudi 并发控制

- Hudi 现在支持 Optimistic Concurrency Control (OCC)，它可以与写入操作 ( 如UPSERT和 ) 一起使用，INSERT以允许多个写入器对同一 Hudi 表进行更改。这是文件级的OCC，因此，如果任意两个提交 ( 或写入者 ) 的更改不冲突，则可以写入同一个表。有关更多信息，请参阅 [Hudi 并发性控制](#)。
- 亚马逊EMR集群安装了 Zookeeper，可以将其用作锁定提供程序。OCC为了便于使用此功能，Amazon EMR 集群已预先配置了以下属性：

```
hoodie.write.lock.provider=org.apache.hudi.client.transaction.lock.ZookeeperBasedLockProvider
hoodie.write.lock.zookeeper.url=<EMR Zookeeper URL>
hoodie.write.lock.zookeeper.port=<EMR Zookeeper Port>
hoodie.write.lock.zookeeper.base_path=/hudi
```

要启用OCC，您需要使用其 Hudi 任务选项或使用 Amazon 配置在集群级别配置以下属性：EMRAPI

```
hoodie.write.concurrency.mode=optimistic_concurrency_control
hoodie.cleaner.policy.failed.writes=LAZY (Performs cleaning of failed writes lazily
instead of inline with every write)
hoodie.write.lock.zookeeper.lock_key=<Key to uniquely identify the Hudi table> (Table
Name is a good option)
```

Hudi 监控：亚马逊 CloudWatch 集成以报告 Hudi 指标

- 亚马逊EMR支持向亚马逊 CloudWatch发布 Hudi 指标。通过设置以下所需配置来启用：

```
hoodie.metrics.on=true
hoodie.metrics.reporter.type=CLOUDWATCH
```

- 以下是您可以更改的可选 Hudi 配置：

设置	描述	值
hoodie.metrics.cloudwatch.report.period.seconds	向 Amazon 报告指标的频率（以秒为单位）CloudWatch	默认值为 60s，这对于亚马逊提供的默认一分钟分辨率来说没问题 CloudWatch
hoodie.metrics.cloudwatch.metric.prefix	要添加到每个指标名称的前缀	默认值为空（无前缀）
hoodie.metrics.cloudwatch.namespace	发布指标时使用的 Amazon CloudWatch 命名空间	默认值为 Hudi
hoodie.metrics.cloudwatch.maxDatumsPer请求	向 Amazon 发出的单次请求中包含的最大数据数量 CloudWatch	默认值为 20，这与亚马逊的 CloudWatch 默认值相同

## Amazon EMR Hudi 配置支持和改进

- 客户现在可以利用EMR配置API和重新配置功能在集群级别配置 Hudi 配置。与 Spark 和 Hive 等其他应用程序一样，通过 `/etc/hudi/CONF/hudi-defaults.conf` 引入了基于文件的新配置支持。EMR配置一些默认值以改善用户体验：
  - 已配置 `hoodie.datasource.hive_sync.jdbcurl` 为群集 Hive 服务器URL，不再需要指定。这在在 Spark 集群模式下运行作业时特别有用，您以前必须在此模式下指定 A EMR mazon 主 IP。
  - HBase 特定配置，这对于在 Hudi 中使用HBase索引很有用。
  - Zookeeper 锁定提供程序的特定配置，如并发控制下所述，这使得使用乐观并发控制 () 变得更加容易。OCC
- 还引入了其他更改，以减少需要通过的配置数量，并在可能的情况下自动推断：
  - 该 `partitionBy` 关键字可用于指定分区列。
  - 启用 Hive Sync 时，不再强制通过 `HIVE_TABLE_OPT_KEY`，`HIVE_PARTITION_FIELDS_OPT_KEY`，`HIVE_PARTITION_EXTRACTOR_CLASS_OPT_KEY`。这些值可以根据 Hudi 表名称和分区字段推断出来。
  - `KEYGENERATOR_CLASS_OPT_KEY`不强制通过，可以从更简单的 `SimpleKeyGenerator`和 `ComplexKeyGenerator`情况下推断。

## Hudi 注意事项

- Hudi 不支持在 Hive 中用于读取时合并 (MoR) 和 Bootstrap 表格中的矢量化执行。例如，当 `hive.vectorized.execution.enabled` 设置为 `true` 时，Hudi 实时表的 `count(*)` 失败。作为解决方法，您可以通过将 `hive.vectorized.execution.enabled` 设置为 `false` 禁用矢量化读入。
- 多写作器支持与 Hudi 引导启动功能不兼容。
- Flink Streamer 和 Flink SQL 是此版本中的实验性功能。建议不要在生产部署中使用这些功能。

## 更改、增强功能和解决的问题

此版本旨在修复 Amazon EMR Scaling 无法成功扩展/缩小集群或导致应用程序故障时出现的问题。

- 以前，在多主集群上手动重启资源管理器会导致 Amazon EMR on 集群上的守护程序（例如 Zookeeper）在 Zookeeper znode 文件中重新加载所有先前已停用或丢失的节点。在某些情况下，这会导致超出默认限制。Amazon EMR 现已从 Zookeeper 文件中删除已停用或丢失超过一小时的节点记录，并且内部限制也已提高。
- 修复了 Amazon EMR on 集群上守护程序运行状况检查活动（例如收集 YARN 节点状态和 HDFS 节点状态）时，对利用率很高的大型集群的扩展请求失败的问题。之所以发生这种情况，是因为集群上的守护程序无法将节点的运行状况数据传送到内部 Amazon EMR zon 组件。
- 改进了 EMR 集群上的守护程序，可在重复使用 IP 地址时正确跟踪节点状态，从而提高扩展操作期间的可靠性。
- [SPARK-29683](#)。修复了集群缩减期间出现任务失败的问题，因为 Spark 假定所有可用节点都被拒绝列出。
- [YARN-9011](#)。修复了集群尝试向上或向下扩展时，由于 YARN 停用过程中的争用条件而导致任务失败的问题。
- 通过确保 Amazon EMR on 集群上守护程序和 YARN 之间的节点状态始终保持一致，修复了集群扩展期间步骤或任务失败的问题。HDFS
- 修复了启用了 Kerberos 身份验证的 Amazon 集群的 EMR 集群操作（例如缩减和步骤提交）失败的问题。这是因为 Amazon EMR on 集群上守护程序没有续订 Kerberos 票证，而要与主节点上 YARN 运行的 HDFS 进行安全通信，则需要续订 Kerberos 票证。
- 配置群集以修复 Apache T YARN timeline Server 版本 1 和 1.5 的性能问题

Apache T YARN timeline Server 版本 1 和 1.5 可能会导致非常活跃的大型 EMR 集群出现性能问题 `yarn.resourcemanager.system-metrics-publisher.enabled=true`，尤其是在使用 Amazon EMR 的默认设置时。开源 YARN 时间轴服务器 v2 解决了与时间 YARN 轴服务器可扩展性相关的性能问题。

此问题的其他解决方法包括：

- 配置 `yarn.resourceManager system-metrics-publisher yarn-site.xml` 中的 `.enabled=false`。
- 如下所述，在创建群集时启用此问题的修复程序。

以下 Amazon EMR 版本包含针对此 YARN 时间轴服务器性能问题的修复程序。

EMR5.30.2、5.31.1、5.32.1、5.33.1、5.34.x、6.0.1、6.1.1、6.2.1、6.3.1、6.3.1、6.3.1、6.4.x

要在上述任何指定的 Amazon EMR 版本上启用修复，请在使用 [aws emr create-cluster](#) 命令参数传 `true` 入的配置 JSON 文件中将这些属性设置为：`--configurations file:///./configurations.json`。或者使用 [重新配置控制台 UI](#) 启用修复程序。

配置 .json 文件内容的示例：

```
[
  {
    "Classification": "yarn-site",
    "Properties": {
      "yarn.resourcemanager.system-metrics-publisher.timeline-server-v1.enable-batch":
        "true",
      "yarn.resourcemanager.system-metrics-publisher.enabled": "true"
    },
    "Configurations": []
  }
]
```

- 默认情况下HDFS，Web 和 HTTPFS 服务器处于禁用状态。您可以HDFS使用 Hadoop 配置重新启用 Web。dfs.webhdfs.enabledHTTPFS 服务器可以通过使用 `sudo systemctl start hadoop-httpfs` 启动。
- HTTPS现在，亚马逊 Linux 存储库已默认启用。如果您使用 Amazon S3 VPCE 策略来限制对特定存储桶的访问，则必须将新的 Amazon Linux 存储桶ARNarn:aws:s3:::amazonlinux-2-repos-\$region/\*添加到您的策略中（\$region替换为终端节点所在的区域）。有关更多信息，请在 AWS 讨论论坛中查看此主题。[公告：Amazon Linux 2 现在支持HTTPS在连接软件包存储库时使用。](#)
- Hive：通过为最后一个作业启用暂存目录，可以提高写入查询性能。HDFS最终任务的临时数据是写入的，HDFS而不是 Amazon S3，而且由于数据是从HDFS最终表位置 (Amazon S3) 移动到最终表位置 (Amazon S3)，而不是在 Amazon S3 设备之间移动，因此性能得到了改善。
- Hive：使用 Glue 元存储分区修剪，查询编译时间最多可缩短 2.5 倍。
- 默认情况下，当 Hive 将内置UDFs内容传递给 Hive Metastore 服务器时，只有内置内容的子集会传递给 Glue Metastore，因UDFs为 Glue 仅支持有限的表达式运算符。如果您设置hive.glue.partition.pruning.client=true，则所有分区修剪发生在客户端。如果您设置hive.glue.partition.pruning.server=true，则所有分区修剪发生在服务器端。

## 已知问题

- Hue 查询在 Amazon EMR 6.4.0 中不起作用，因为 Apache Hadoop httpFS 服务器默认处于禁用状态。要在亚马逊 EMR 6.4.0 上使用 Hue，要么使用亚马逊步骤在亚马逊EMR主节点上手动启动 HttpFS 服务器`sudo systemctl start hadoop-httpfs`，要么使用[亚马逊步骤](#)。EMR

- 与 Livy 用户模仿一起使用的 Amazon N EMR otebooks 功能不起作用，因为 HTTPFS 在默认情况下处于禁用状态。在这种情况下，EMR笔记本无法连接到启用了 Livy 模拟的集群。解决方法是在使用将EMR笔记本连接到集群之前启动 HTTPFS 服务器。`sudo systemctl start hadoop-httpfs`
- 在亚马逊 6.4.0 EMR 版本中，Phoenix 不支持 Phoenix 连接器组件。
- 要将 Spark 操作与 Apache Oozie 一起使用，必须将以下配置添加到 Oozie workflow.xml 文件中。否则，Oozie 启动的 Spark 执行程序的类路径中EMRFS将缺少几个关键库，例如 Hadoop 和。

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

- 当你使用带有 Hive 分区位置格式的 Spark 来读取 Amazon S3 中的数据，并在亚马逊EMR版本 5.30.0 到 5.36.0 和 6.2.0 到 6.9.0 上运行 Spark 时，你可能会遇到一个问题，导致集群无法正确读取数据。如果您的分区具有以下所有特征，会发生这种情况：
  - 从同一个表扫描两个或多个分区。
  - 至少有一个分区目录路径是至少一个其他分区目录路径的前缀，例如，`s3://bucket/table/p=a` 是 `s3://bucket/table/p=a b` 的前缀。
  - 另一个分区目录中前缀之后的第一个字符的值为 UTF -8，小于该/字符 (U+002F)。例如，在 `s3://bucket/table/p=a b` 中，`a` 和 `b` 之间出现的空格字符 (U+0020) 就属于此类。请注意，还有其他 14 个非控制字符：`!"#$%&'()*+,-`。有关更多信息，请参阅 [UTF-8 编码表和 Unicode 字符](#)。

解决方法是在 `spark-defaults` 分类中将

`spark.sql.sources.fastS3PartitionDiscovery.enabled` 配置设置为 `false`。

## 发行版 5.32.0

以下发行说明包含有关亚马逊 5.32.0 EMR 版本的信息。更改与 5.31.0 有关。

首次发布日期：2021 年 1 月 8 日

### 升级

- 已将 Amazon Glue 连接器升级到 1.14.0
- 已将 Amazon SageMaker Spark 升级 SDK 到 1.4.1 版
- 已升级 AWS SDK for Java 到 1.11.890 版本
- 升级后的 EMR DynamoDB 连接器版本 4.16.0

- 已升级EMRFS到 2.45.0 版
- 已将EMR日志分析指标升级到 1.18.0 版
- 已将EMR MetricsAndEventsApiGateway 客户端升级到 1.5.0 版
- 已将 R EMR ecord Server 升级到 1.8.0 版
- 已EMR将 S3 Dist CP 升级到版本 2.17.0
- 已将 S EMR ecret Agent 升级到 1.7.0 版
- 已将 Flink 升级到版本 1.11.2
- 已将 Hadoop 升级到版本 2.10.1-amzn-0
- 已将 Hive 升级到版本 2.3.7-amzn-3
- 已将 Hue 升级到版本 4.8.0
- 已将 Mxnet 升级到版本 1.7.0
- 已将 OpenCV 升级到版本 4.4.0
- 已将 Presto 升级到版本 0.240.1-amzn-0
- 已将 Spark 升级到版本 2.4.7-amzn-0
- 已升级 TensorFlow 到 2.3.1 版

## 更改、增强功能和解决的问题

- 此版本旨在修复 Amazon EMR Scaling 无法成功扩展/缩小集群或导致应用程序故障时出现的问题。
- 修复了 Amaz EMR on 集群上守护程序运行状况检查活动 ( 例如收集YARN节点状态和HDFS节点状态 ) 时, 对利用率很高的大型集群的扩展请求失败的问题。之所以发生这种情况, 是因为集群上的守护程序无法将节点的运行状况数据传送到内部 Ama EMR zon 组件。
- 改进了EMR集群上的守护程序, 可在重复使用 IP 地址时正确跟踪节点状态, 从而提高扩展操作期间的可靠性。
- [SPARK-29683](#)。修复了集群缩减期间出现任务失败的问题, 因为 Spark 假定所有可用节点都被拒绝列出。
- [YARN-9011](#)。修复了集群尝试向上或向下扩展时, 由于YARN停用过程中的争用条件而导致任务失败的问题。
- 通过确保 Amaz EMR on 集群上守护程序和 YARN /之间的节点状态始终保持一致, 修复了集群扩展期间步骤或任务失败的问题。HDFS

- 修复了启用了 Kerberos 身份验证的 Amazon 集群的EMR集群操作（例如缩减和步骤提交）失败的问题。这是因为 Amaz EMR on 集群上守护程序没有续订 Kerberos 票证，而要与主节点上YARN运行的 HDFS /进行安全通信，则需要续订 Kerberos 票证。
- 较新的Amazon EMR 版本通过降低亚马逊较旧AL2版本的“最大打开文件数”限制来解决这个问题。EMR。亚马逊EMR发布的 5.30.1、5.30.2、5.31.1、5.32.1、5.32.1、6.0.1、6.1.1、6.2.1、5.33.0、6.3.0及更高版本现在包含一个永久修复程序，其中包含一个具有更高的“最大打开文件数”设置的永久修复。
- 升级了组件版本。
- 有关组件版本的列表，请参阅本指南中的[关于 Amazon EMR 版本](#)。

## 新功能

- 从亚马逊 EMR 5.32.0 和 6.5.0 开始，Apache Spark 的动态执行器大小设置默认处于启用状态。要打开或关闭此功能，您可以使用 `spark.yarn.heterogeneousExecutors.enabled` 配置参数。
- 实例元数据服务 (IMDS) V2 支持状态：Amazon EMR 5.23.1、5.27.1 和 5.32 或更高版本的组件用于所有调用。IMDSv2 IMDS对于应用程序代码中的IMDS调用，您可以同时使用IMDSv1和IMDSv2，或者将配置IMDS为仅IMDSv2用于提高安全性。对于其他 5.x EMR 版本，禁用IMDSv1会导致集群启动失败。
- 从 Amazon EMR 5.32.0 开始，您可以启动与 Apache Ranger 原生集成的集群。Apache Ranger 是一个开源框架，可跨 Hadoop 平台启用、监控和管理全面的数据安全。有关更多信息，请参阅 [Apache Ranger](#)。通过原生集成，您可以自带Apache Ranger来在亚马逊上实施精细的数据访问控制。EMR请参阅亚马逊EMR发布指南中的[将亚马逊EMR与 Apache Ranger 集成](#)。
- 亚马逊EMR版本 5.32.0 支持亚马逊EMR。EKS如需详细了解如何开始使用 EMR onEKS，请参阅 [Amazon EMR on EKS](#)。
- 亚马逊EMR版本 5.32.0 支持亚马逊 EMR Studio（预览版）。有关 EMR Studio 入门的更多详细信息，请参阅 [Amazon EMR Studio（预览版）](#)。
- 限定范围的托管策略：为了与 AWS 最佳实践保持一致，Amazon 引入EMR了 v2 EMR 范围的默认托管策略，以取代即将弃用的策略。参见 [Amazon EMR 托管策略](#)。

## 已知问题

- 对于 Amazon EMR 6.3.0 和 6.2.0 私有子网集群，您无法访问 Ganglia 网页用户界面。您将收到“access denied (403)”错误。其他网络UIs，例如 Spark、Hue JupyterHub、Zeppelin、Livy 和 Tez，都正常运行。公有子网集群上的 Ganglia Web UI 访问也正常工作。要解决该问题，请在具有



`sudo systemctl restart httpd` 的主节点上重新启动 `httpd` 服务。此问题已在 Amazon EMR 6.4.0 中修复。

- 降低旧版本的“最大打开文件数”限制 AL2 [在新版本中已修复]。亚马逊EMR版本：  
`emr-5.30.x`、`emr-5.31.0`、`emr-5.32.0`、`emr-6.0.0`、`emr-6.1.0` 和 `emr-6.2.0` 基于较旧版本的 Linux of Amazon 2 ( )AL2，当使用默认版本创建亚马逊集群时，“最大打开文件数”的用户限制较低。EMR AMI 亚马逊EMR发布的版本为 5.30.1、5.30.2、5.31.1、5.32.1、6.0.1、6.1.1、6.2.1、5.33.0、6.3.0 及更高版本，包括具有更高的“最大打开文件数”设置的永久修复程序。如果使用打开文件数限制较低的发行版，会在提交 Spark 任务时导致“Too many open files”（打开的文件过多）错误。在受影响的版本中，亚马逊EMR默认AMI的“最大打开文件数”的 `ulimit` 设置为 4096，低于 Linux 2 中的 65536 个文件限制。latest Amazon AMI Spark 驱动程序和执行程序尝试打开超过 4096 个文件时，“打开的最大文件数”的较低 `ulimit` 设置会导致 Spark 任务失败。为了解决这个问题，Amazon EMR 提供了一个引导操作 (BA) 脚本，可以在创建集群时调整 `ulimit` 设置。

如果您使用的是无法永久修复此问题的旧EMR版 Amazon，则可以使用以下解决方法将实例控制器 `ulimit` 明确设置为最多 65536 个文件。

从命令行显式设置 `ulimit`

1. 编辑 `/etc/systemd/system/instance-controller.service`，将以下参数添加到 Service (服务) 部分。

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. 重启 InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

使用引导操作 (BA) 设置 `ulimit`

您还可以在创建集群时使用引导操作 ( BA ) 脚本将实例控制器 `ulimit` 配置为 65536 个文件。

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
```

```

EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload

```

### ⚠ Important

EMR运行 Amazon Linux 或 Amazon Linux 2 Amazon 机器映像 (AMIs) 的集群使用默认 Amazon Linux 行为，不会自动下载和安装需要重启的重要关键内核更新。这与运行默认 Amazon Linux 的其他亚马逊EC2实例的行为相同AMI。如果在 Amazon EMR 版本发布后需要重启的新 Amazon Linux 软件CUDA更新（例如内核和更新）可用，则默认运行的EMR集群实例AMI不会自动下载和安装这些更新。NVIDIA要获取内核更新，您可以[自定义您的亚马逊EMRAMI以使用最新的亚马逊 Linux AMI](#)。

- 该 GovCloud 区域目前不支持控制台支持创建指定 AWS Ranger 集成选项的安全配置。可以使用来完成安全配置CLI。请参阅 [《Amazon EMR 管理指南》中的创建EMR安全配置](#)。
- 在使用 Amazon EMR 5.31.0 AtRestEncryption 或 5.32.0 的集群上启用或HDFS加密时，Hive 查询会导致以下运行时异常。

```

TaskAttempt 3 failed, info=[Error: Error while running task ( failure ) :
attempt_1604112648850_0001_1_01_000000_3:java.lang.RuntimeException:
java.lang.RuntimeException: Hive Runtime Error while closing
operators: java.io.IOException: java.util.ServiceConfigurationError:
org.apache.hadoop.security.token.TokenIdentifier: Provider
org.apache.hadoop.hbase.security.token.AuthenticationTokenIdentifier not found

```

- 当你使用带有 Hive 分区位置格式的 Spark 来读取 Amazon S3 中的数据，并在亚马逊EMR版本 5.30.0 到 5.36.0 和 6.2.0 到 6.9.0 上运行 Spark 时，你可能会遇到一个问题，导致集群无法正确读取数据。如果您的分区具有以下所有特征，会发生这种情况：
  - 从同一个表扫描两个或多个分区。

- 至少有一个分区目录路径是至少一个其他分区目录路径的前缀，例如，`s3://bucket/table/p=a` 是 `s3://bucket/table/p=a b` 的前缀。
- 另一个分区目录中前缀之后的第一个字符的值为 UTF -8，小于该/字符 (U+002F)。例如，在 `s3://bucket/table/p=a b` 中，`a` 和 `b` 之间出现的空格字符 (U+0020) 就属于此类。请注意，还有其他 14 个非控制字符：`!"#$%&'()*+,-`。有关更多信息，请参阅 [UTF-8 编码表和 Unicode 字符](#)。

解决方法是在 `spark-defaults` 分类中将

`spark.sql.sources.fastS3PartitionDiscovery.enabled` 配置设置为 `false`。

## 发行版 6.2.0

以下发行说明包含有关亚马逊 6.2.0 EMR 版本的信息。更改与 6.1.0 有关。

首次发布日期：2020 年 12 月 9 日

上次更新日期：2021 年 10 月 4 日

受支持的应用程序

- AWS SDK for Java 版本 1.11.828
- emr-record-server 版本 1.7.0
- Flink 1.11.2
- Ganglia 3.7.2
- Hadoop 3.2.1-amzn-1
- HBase版本 2.2.6-amzn-0
- HBase-operator-tools 1.0.0
- HCatalog版本 3.1.2-amzn-0
- Hive 3.1.2-amzn-3
- Hudi 0.6.0-amzn-1
- Hue 4.8.0
- JupyterHub 版本 1.1.0
- Livy 0.7.0
- MXNet版本 1.7.0

- Oozie 5.2.0
- Phoenix 5.0.0
- Pig 0.17.0
- Presto 0.238.3-amzn-1
- Presto SQL 版本 343
- Spark 3.0.1-amzn-0
- spark-rapids 0.2.0
- TensorFlow 版本 2.3.1
- Zeppelin 0.9.0-preview1
- Zookeeper 3.4.14
- 连接器和驱动程序 : DynamoDB 连接器 4.16.0

## 新功能

- HBase: 删除了提交阶段的重命名并添加了持续HFile跟踪。请参阅 Amazon EMR 发布指南中的[持续HFile跟踪](#)。
- HBase: 向后移植[创建一个在压缩时强制缓存块的配置](#)。
- PrestoDB : 改进了动态分区修剪。基于规则的连接重新排序对未分区数据运行。
- 限定范围的托管策略 : 为了与 AWS 最佳实践保持一致 , Amazon 引入EMR了 v2 EMR 范围的默认托管策略 , 以取代即将弃用的策略。参见 [Amazon EMR 托管政策](#)。
- 实例元数据服务 (IMDS) V2 支持状态 : 对于 Amazon EMR 6.2 或更高版本 , Amazon EMR 组件IMDSv2用于所有IMDS调用。对于应用程序代码中的IMDS调用 , 您可以同时使用IMDSv1和IMDSv2 , 或者将配置IMDS为仅IMDSv2用于提高安全性。如果您IMDSv1在早期的 Amazon EMR 6.x 版本中禁用 , 则会导致集群启动失败。

## 更改、增强功能和解决的问题

- 此版本旨在修复 Amazon EMR Scaling 无法成功扩展/缩小集群或导致应用程序故障时出现的问题。
- 修复了 Amaz EMR on 集群上守护程序运行状况检查活动 ( 例如收集YARN节点状态和HDFS节点状态 ) 时 , 对利用率很高的大型集群的扩展请求失败的问题。之所以发生这种情况 , 是因为集群上的守护程序无法将节点的运行状况数据传送到内部 Ama EMR zon 组件。
- 改进了EMR集群上的守护程序 , 可在重复使用 IP 地址时正确跟踪节点状态 , 从而提高扩展操作期间的可靠性。

- [SPARK-29683](#)。修复了集群缩减期间出现任务失败的问题，因为 Spark 假定所有可用节点都被拒绝列出。
- [YARN-9011](#)。修复了集群尝试向上或向下扩展时，由于YARN停用过程中的争用条件而导致任务失败的问题。
- 通过确保 Amaz EMR on 集群上守护程序和 YARN /之间的节点状态始终保持一致，修复了集群扩展期间步骤或任务失败的问题。HDFS
- 修复了启用了 Kerberos 身份验证的 Amazon 集群的EMR集群操作（例如缩减和步骤提交）失败的问题。这是因为 Amaz EMR on 集群上守护程序没有续订 Kerberos 票证，而要与主节点上YARN运行的 HDFS /进行安全通信，则需要续订 Kerberos 票证。
- 较新的Amazon EMR 版本通过降低亚马逊较旧AL2版本的“最大打开文件数”限制来解决这个问题 EMR。亚马逊EMR发布的 5.30.1、5.30.2、5.31.1、5.32.1、5.32.1、6.0.1、6.1.1、6.2.1、5.33.0、6.3.0及更高版本现在包含一个永久修复程序，其中包含一个具有更高的“最大打开文件数”设置的永久修复。
- Spark：改进了 Spark 运行时的性能。

## 已知问题

- 亚马逊 EMR 6.2 在 6.2.0 版本中对 /etc/cron libinstance-controller-java .d/ 文件设置的权限不正确。EMR当文件的权限应为 644 (-rw-r--r--) 时，它们为 645 (-rw-r--r-x)。因此，Amazon EMR 版本 6.2 不记录实例状态日志，并且 /emr/instance-logs 目录为空。此问题已在 Amazon EMR 6.3.0 及更高版本中修复。

要解决此问题，请在集群启动时将以下脚本作为引导操作运行。

```
#!/bin/bash
sudo chmod 644 /etc/cron.d/libinstance-controller-java
```

- 对于 Amazon EMR 6.2.0 和 6.3.0 私有子网集群，您无法访问 Ganglia 网页用户界面。您将收到“access denied (403)”错误。其他网络UIs，例如 Spark、Hue JupyterHub、Zeppelin、Livy 和 Tez，都正常运行。公有子网集群上的 Ganglia Web UI 访问也正常工作。要解决该问题，请在具有 sudo systemctl restart httpd 的主节点上重新启动 httpd 服务。此问题已在 Amazon EMR 6.4.0 中修复。
- 亚马逊 EMR 6.2.0 中存在一个问题，即 httpd 持续失败，导致 Ganglia 不可用。您会收到“cannot connect to the server (无法连接到服务器)”错误。要修复已在运行但存在此问题的群集，SSH请

访问群集主节点，然后将该行 Listen 80 添加到 `httpd.conf` 位于的文件中 `/etc/httpd/conf/httpd.conf`。此问题已在 Amazon EMR 6.3.0 中修复。

- HTTPD 使用安全配置时，在 EMR 6.2.0 集群上失败。因此，Ganglia Web 应用程序用户界面不可用。要访问 Ganglia Web 应用程序用户界面，请将 Listen 80 添加到群集主节点上的 `/etc/httpd/conf/httpd.conf` 文件中。有关连接到集群的信息，请参阅[使用连接到主节点SSH](#)。

EMR 当您使用安全配置时，笔记本也无法与 EMR 6.2.0 集群建立连接。笔记本将无法列出内核和提交 Spark 任务。我们建议您 EMR 改用其他版本的 Amazon 的 EMR 笔记本电脑。

- 降低旧版本的“最大打开文件数”限制 AL2 [在新版本中已修复]。亚马逊 EMR 版本：  
emr-5.30.x、emr-5.31.0、emr-5.32.0、emr-6.0.0、emr-6.1.0 和 emr-6.2.0 基于较旧版本的 Linux of Amazon 2 () AL2，当使用默认版本创建亚马逊集群时，“最大打开文件数”的用户限制较低。EMR AMI 亚马逊 EMR 发布的版本为 5.30.1、5.30.2、5.31.1、5.32.1、6.0.1、6.1.1、6.2.1、5.33.0、6.3.0 及更高版本，包括具有更高的“最大打开文件数”设置的永久修复程序。如果使用打开文件数限制较低的发行版，会在提交 Spark 任务时导致“Too many open files”（打开的文件过多）错误。在受影响的版本中，亚马逊 EMR 默认 AMI 的“最大打开文件数”的 `ulimit` 设置为 4096，低于 Linux 2 中的 65536 个文件限制。latest Amazon AMI Spark 驱动程序和执行程序尝试打开超过 4096 个文件时，“打开的最大文件数”的较低 `ulimit` 设置会导致 Spark 任务失败。为了解决这个问题，Amazon EMR 提供了一个引导操作 (BA) 脚本，可以在创建集群时调整 `ulimit` 设置。

如果您使用的是无法永久修复此问题的旧 EMR 版 Amazon，则可以使用以下解决方法将实例控制器 `ulimit` 明确设置为最多 65536 个文件。

从命令行显式设置 `ulimit`

1. 编辑 `/etc/systemd/system/instance-controller.service`，将以下参数添加到 Service (服务) 部分。

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. 重启 InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

使用引导操作 (BA) 设置 `ulimit`

您还可以在创建集群时使用引导操作 ( BA ) 脚本将实例控制器 ulimit 配置为 65536 个文件。

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

• **⚠ Important**

Amazon EMR 6.1.0 和 6.2.0 包含一个性能问题，可能会严重影响 Hudi 的所有插入、更新和删除操作。如果您计划在亚马逊 EMR 6.1.0 或 6.2.0 中使用 Hudi，则应联系 AWS 支持人员获取已修补的 Hudi。RPM

• **⚠ Important**

EMR运行 Amazon Linux 或 Amazon Linux 2 Amazon 机器映像 (AMIs) 的集群使用默认 Amazon Linux 行为，不会自动下载和安装需要重启的重要关键内核更新。这与运行默认 Amazon Linux 的其他亚马逊EC2实例的行为相同AMI。如果在 Amazon EMR 版本发布后需要重启的新 Amazon Linux 软件CUDA更新（例如内核和更新）可用，则默认运行的EMR集群实例AMI不会自动下载和安装这些更新。NVIDIA要获取内核更新，您可以[自定义您的亚马逊EMRAMI以使用最新的亚马逊 Linux AMI](#)。

- 亚马逊 EMR 6.2.0 Maven 工件尚未发布。它们将与 Amazon 的未来版本一起发布EMR。
- 使用 HBase storefile 系统表进行持续HFile跟踪不支持HBase区域复制功能。有关HBase区域复制的更多信息，请参阅[时间轴一致的高可用读取](#)。

- 亚马逊 EMR 6.x 和 EMR 5.x Hive 存储桶版本的区别

EMR5.x 使用 OOS Apache Hive 2，而在 EMR 6.x 中使用 OOS Apache Hive 3。开源 Hive2 使用分桶版本 1，而开源 Hive3 使用分桶版本 2。Hive 2 (EMR5.x) 和 Hive 3 (EMR6.x) 之间的这种存储分区版本差异意味着 Hive 存储分区哈希功能不同。请参见以下示例。

下表分别是在 EMR 6.x 和 EMR 5.x 中创建的示例。

```
-- Using following LOCATION in EMR 6.x
CREATE TABLE test_bucketing (id INT, desc STRING)
PARTITIONED BY (day STRING)
CLUSTERED BY(id) INTO 128 BUCKETS
LOCATION 's3://your-own-s3-bucket/emr-6-bucketing/';

-- Using following LOCATION in EMR 5.x
LOCATION 's3://your-own-s3-bucket/emr-5-bucketing/';
```

在 EMR 6.x 和 EMR 5.x 中插入相同的数据。

```
INSERT INTO test_bucketing PARTITION (day='01') VALUES(66, 'some_data');
INSERT INTO test_bucketing PARTITION (day='01') VALUES(200, 'some_data');
```

检查 S3 位置会显示存储桶文件名不同，因为哈希函数在 EMR 6.x (Hive 3) 和 EMR 5.x (Hive 2) 之间有所不同。

```
[hadoop@ip-10-0-0-122 ~]$ aws s3 ls s3://your-own-s3-bucket/emr-6-bucketing/day=01/
2020-10-21 20:35:16          13 000025_0
2020-10-21 20:35:22          14 000121_0
[hadoop@ip-10-0-0-122 ~]$ aws s3 ls s3://your-own-s3-bucket/emr-5-bucketing/day=01/
2020-10-21 20:32:07          13 000066_0
2020-10-21 20:32:51          14 000072_0
```

您还可以通过 EMR 6.x 的 Hive CLI 中运行以下命令来查看版本差异。请注意，它将返回分桶版本 2。

```
hive> DESCRIBE FORMATTED test_bucketing;
...
Table Parameters:
  bucketing_version      2
...
```



- 具有多个主节点的集群和 Kerberos 身份验证中的已知问题

如果您在 Amazon 5.20.0 及更高EMR版本中运行具有多个主节点和 Kerberos 身份验证的集群，则在集群运行一段时间后，集群操作可能会遇到问题，例如缩小规模或步骤提交。具体时间段取决于您定义的 Kerberos 票证有效期。缩减问题会影响您提交的自动缩减和显式缩减请求。其它集群操作也可能受到影响。

解决办法：

- SSH以hadoop用户身份访问具有多个主节点的EMR群集的主节点。
- 运行以下命令，为 hadoop 用户续订 Kerberos 票证。

```
kinit -kt <keytab_file> <principal>
```

通常情况下，keytab 文件位于 /etc/hadoop.keytab，而 principal 为 hadoop/<hostname>@<REALM> 格式。

#### Note

此解决方法将在 Kerberos 票证有效期内生效。默认情况下，此持续时间为 10 个小时，但可以通过 Kerberos 设置进行配置。Kerberos 票证过期后，您必须重新运行上述命令。

- 当你使用带有 Hive 分区位置格式的 Spark 来读取 Amazon S3 中的数据，并在亚马逊EMR版本 5.30.0 到 5.36.0 和 6.2.0 到 6.9.0 上运行 Spark 时，你可能会遇到一个问题，导致集群无法正确读取数据。如果您的分区具有以下所有特征，会发生这种情况：
  - 从同一个表扫描两个或多个分区。
  - 至少有一个分区目录路径是至少一个其他分区目录路径的前缀，例如，s3://bucket/table/p=a 是 s3://bucket/table/p=a b 的前缀。
  - 另一个分区目录中前缀之后的第一个字符的值为 UTF -8，小于该/字符 (U+002F)。例如，在 s3://bucket/table/p=a b 中，a 和 b 之间出现的空格字符 (U+0020) 就属于此类。请注意，还有其他 14 个非控制字符：! "\$ % & ' ( ) \* + , -。有关更多信息，请参阅 [UTF-8 编码表和 Unicode 字符](#)。

解决方法是在 spark-defaults 分类中将

spark.sql.sources.fastS3PartitionDiscovery.enabled 配置设置为 false。

## 发行版 5.31.0

以下发行说明包含有关亚马逊 5.31.0 EMR 版本的信息。更改与 5.30.1 有关。

首次发布日期：2020 年 10 月 9 日

上次更新日期：2020 年 10 月 15 日

### 升级

- 已将 Amazon Glue 连接器升级到版本 1.13.0
- 已将 Amazon SageMaker Spark 升级SDK到 1.4.0 版
- 已将 Amazon Kinesis 连接器升级到版本 3.5.9
- 已升级 AWS SDK for Java 到 1.11.852 版本
- 已将 Bigtop-tomcat 升级到版本 8.5.56
- 已EMR将 FS 升级到 2.43.0 版
- 已将EMR MetricsAndEventsApiGateway 客户端升级到 1.4.0 版
- 已EMR将 S3 Dist CP 升级到版本 2.15.0
- 已EMR将 S3 Select 升级到 1.6.0 版
- 已将 Flink 升级到版本 1.11.0
- 已将 Hadoop 升级到版本 2.10.0
- 已将 Hive 升级到版本 2.3.7
- 已将 Hudi 升级到版本 0.6.0
- 已将 Hue 升级到版本 4.7.1
- 已升级 JupyterHub 到 1.1.0 版
- 已将 Mxnet 升级到版本 1.6.0
- 已将 OpenCV 升级到版本 4.3.0
- 已将 Presto 升级到版本 0.238.3
- 已升级 TensorFlow 到 2.1.0 版

### 更改、增强功能和解决的问题

- 此版本旨在修复 Amazon EMR Scaling 无法成功扩展/缩小集群或导致应用程序故障时出现的问题。

- 修复了 Amaz EMR on 集群上守护程序运行状况检查活动（例如收集YARN节点状态和HDFS节点状态）时，对利用率很高的大型集群的扩展请求失败的问题。之所以发生这种情况，是因为集群上的守护程序无法将节点的运行状况数据传送到内部 Ama EMR zon 组件。
- 改进了EMR集群上的守护程序，可在重复使用 IP 地址时正确跟踪节点状态，从而提高扩展操作期间的可靠性。
- [SPARK-29683](#)。修复了集群缩减期间出现任务失败的问题，因为 Spark 假定所有可用节点都被拒绝列出。
- [YARN-9011](#)。修复了集群尝试向上或向下扩展时，由于YARN停用过程中的争用条件而导致任务失败的问题。
- 通过确保 Amaz EMR on 集群上守护程序和 YARN /之间的节点状态始终保持一致，修复了集群扩展期间步骤或任务失败的问题。HDFS
- 修复了启用了 Kerberos 身份验证的 Amazon 集群的EMR集群操作（例如缩减和步骤提交）失败的问题。这是因为 Amaz EMR on 集群上守护程序没有续订 Kerberos 票证，而要与主节点上YARN运行的 HDFS /进行安全通信，则需要续订 Kerberos 票证。
- 较新的Amazon EMR 版本通过降低亚马逊较旧AL2版本的“最大打开文件数”限制来解决这个问题 EMR。亚马逊EMR发布的  
5.30.1、5.30.2、5.31.1、5.32.1、5.32.1、6.0.1、6.1.1、6.2.1、5.33.0、6.3.0及更高版本现在包含一个永久修复程序，其中包含一个具有更高的“最大打开文件数”设置的永久修复。
- 亚马逊 5.31.0 及更高EMR版本支持 [Hive 列统计信息](#)。
- 升级了组件版本。
- EMRFS亚马逊 EMR 5.31.0 中的 S3EC V2 Support。在 S3 Java SDK 版本 1.11.837 及更高版本中，引入了加密客户端版本 2 (S3EC V2)，其中包含各种安全增强功能。有关更多信息，请参阅下列内容：
  - S3 博客文章：[更新至 Amazon S3 加密客户端](#)。
  - AWS SDK for Java 开发者指南：[将加密和解密客户端迁移到 V2](#)。
  - EMR管理指南：[Amazon S3 客户端加密](#)。

加密客户端 V1 仍可在中使用，SDK以实现向后兼容。

## 新功能

- 降低旧版本的“最大打开文件数”限制 AL2 [在新版本中已修复]。亚马逊EMR版本：  
emr-5.30.x、emr-5.31.0、emr-5.32.0、emr-6.0.0、emr-6.1.0 和 emr-6.2.0 基于较旧版本的 Linux ofAmazon 2 ()AL2，当使用默认版本创建亚马逊集群时，“最大打开文件数”的用户限制较低。EMR

AMI亚马逊EMR发布的版本为5.30.1、5.30.2、5.31.1、5.32.1、6.0.1、6.1.1、6.2.1、5.33.0、6.3.0及更高版本，包括具有更高的“最大打开文件数”设置的永久修复。如果使用打开文件数限制较低的发行版，会在提交 Spark 任务时导致“Too many open files”（打开的文件过多）错误。在受影响的版本中，亚马逊EMR默认AMI的“最大打开文件数”的ulimit设置为4096，低于Linux 2中的65536个文件限制。latestAmazon AMISpark 驱动程序和执行程序尝试打开超过 4096 个文件时，“打开的最大文件数”的较低 ulimit 设置会导致 Spark 任务失败。为了解决这个问题，Amazon EMR 提供了一个引导操作 (BA) 脚本，可以在创建集群时调整ulimit设置。

如果您使用的是无法永久修复此问题的旧EMR版 Amazon，则可以使用以下解决方法将实例控制器 ulimit 明确设置为最多 65536 个文件。

### 从命令行显式设置 ulimit

1. 编辑 `/etc/systemd/system/instance-controller.service`，将以下参数添加到 Service (服务) 部分。

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. 重启 InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

### 使用引导操作 (BA) 设置 ulimit

您还可以在创建集群时使用引导操作 ( BA ) 脚本将实例控制器 ulimit 配置为 65536 个文件。

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
```

```
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- 使用 Amazon EMR 5.31.0，您可以启动与 Lake Formation 集成的集群。这种集成为 Glue 数据目录中的数据库和表提供了精细的列级数据筛选。AWS 还支持从企业身份系统对 EMR 笔记本或 Apache Zeppelin 进行联合单点登录。有关更多信息，请参阅《[亚马逊EMR管理指南](#)》[AWS Lake Formation中的亚马逊EMR与集成](#)。

Amazon EMR with Lake Formation 目前在 16 个 AWS 地区上市：美国东部（俄亥俄州和弗吉尼亚北部）、美国西部（加利福尼亚北部和俄勒冈州）、亚太地区（孟买、首尔、新加坡、悉尼和东京）、加拿大（中部）、欧洲（法兰克福、爱尔兰、伦敦、巴黎和斯德哥尔摩）、南美洲（圣保罗）。

## 已知问题

- 具有多个主节点的集群和 Kerberos 身份验证中的已知问题

如果您在 Amazon 5.20.0 及更高 EMR 版本中运行具有多个主节点和 Kerberos 身份验证的集群，则在集群运行一段时间后，集群操作可能会遇到问题，例如缩小规模或步骤提交。具体时间段取决于您定义的 Kerberos 票证有效期。缩减问题会影响您提交的自动缩减和显式缩减请求。其它集群操作也可能受到影响。

解决办法：

- SSH 以 hadoop 用户身份访问具有多个主节点的 EMR 群集的主节点。
- 运行以下命令，为 hadoop 用户续订 Kerberos 票证。

```
kinit -kt <keytab_file> <principal>
```

通常情况下，keytab 文件位于 `/etc/hadoop.keytab`，而 principal 为 `hadoop/<hostname>@<REALM>` 格式。

**Note**

此解决方法将在 Kerberos 票证有效期内生效。默认情况下，此持续时间为 10 个小时，但可以通过 Kerberos 设置进行配置。Kerberos 票证过期后，您必须重新运行上述命令。

- 在使用 Amazon EMR 5.31.0 AtRestEncryption 或 5.32.0 的集群上启用或HDFS加密时，Hive 查询会导致以下运行时异常。

```
TaskAttempt 3 failed, info=[Error: Error while running task ( failure ) :
attempt_1604112648850_0001_1_01_000000_3:java.lang.RuntimeException:
java.lang.RuntimeException: Hive Runtime Error while closing
operators: java.io.IOException: java.util.ServiceConfigurationError:
org.apache.hadoop.security.token.TokenIdentifier: Provider
org.apache.hadoop.hbase.security.token.AuthenticationTokenIdentifier not found
```

- 当你使用带有 Hive 分区位置格式的 Spark 来读取 Amazon S3 中的数据，并在亚马逊EMR版本 5.30.0 到 5.36.0 和 6.2.0 到 6.9.0 上运行 Spark 时，你可能会遇到一个问题，导致集群无法正确读取数据。如果您的分区具有以下所有特征，会发生这种情况：
  - 从同一个表扫描两个或多个分区。
  - 至少有一个分区目录路径是至少一个其他分区目录路径的前缀，例如，s3://bucket/table/p=a 是 s3://bucket/table/p=a b 的前缀。
  - 另一个分区目录中前缀之后的第一个字符的值为 UTF -8，小于该/字符 (U+002F)。例如，在 s3://bucket/table/p=a b 中，a 和 b 之间出现的空格字符 (U+0020) 就属于此类。请注意，还有其他 14 个非控制字符：!"#\$%&'()\*+,-。有关更多信息，请参阅 [UTF-8 编码表和 Unicode 字符](#)。

解决方法是在 spark-defaults 分类中将

spark.sql.sources.fastS3PartitionDiscovery.enabled 配置设置为 false。

## 发行版 6.1.0

以下发行说明包含有关亚马逊 6.1.0 EMR 版本的信息。更改与 6.0.0 有关。

首次发布日期：2020 年 9 月 4 日

上次更新日期：2020 年 10 月 15 日

## 受支持的应用程序

- AWS SDK for Java 版本 1.11.828
- Flink 1.11.0
- Ganglia 3.7.2
- Hadoop 3.2.1-amzn-1
- HBase版本 2.2.5
- HBase-operator-tools 1.0.0
- HCatalog版本 3.1.2-amzn-0
- Hive 3.1.2-amzn-1
- Hudi 0.5.2-incubating
- Hue 4.7.1
- JupyterHub 版本 1.1.0
- Livy 0.7.0
- MXNet版本 1.6.0
- Oozie 5.2.0
- Phoenix 5.0.0
- Presto 0.232
- Presto SQL 版本 338
- Spark 3.0.0-amzn-0
- TensorFlow 版本 2.1.0
- Zeppelin 0.9.0-preview1
- Zookeeper 3.4.14
- 连接器和驱动程序 : DynamoDB 连接器 4.14.0

## 新功能

- ARM从亚马逊EMR版本 5.30.0 和亚马逊EMR 6.1.0 版开始支持实例类型。
- 从 Amazon 6.1.0 和 5.30.0 EMR 版本开始，支持 M6g 通用实例类型。有关更多信息，请参阅 Amazon EMR 管理指南中的[支持的实例类型](#)。
- 从 Amazon EMR 版本 5.23.0 开始，支持EC2置放群组功能作为多个主节点集群的选项。目前，置放群组功能仅支持主节点类型，并将 SPREAD 策略应用于这些主节点。SPREAD 策略将一小组实例

放置在单独的基础硬件上，以防止发生硬件故障时出现多个主节点丢失的问题。有关更多信息，请参阅《Amazon EMR 管理指南》中的[与EC2置放群组EMR集成](#)。

- 托管扩展 — 在 Amazon 6.1.0 EMR 版本中，您可以启用 Amazon EMR 托管扩展，根据工作负载自动增加或减少集群中的实例或单元数量。Amazon 会持续评估集群指标，以做出扩展决策，从而优化集群的成本和速度。托管扩展也适用于亚马逊 5.30.0 及更高 EMR 版本，6.0.0 除外。有关更多信息，请参阅 Amazon EMR 管理指南中的[扩展集群资源](#)。
- 6.1.0 支持 Presto SQL 版本 338。EMR 有关更多信息，请参阅 [Presto](#)。
  - 只有 EMR 6.1.0 及更高版本支持 Presto SQL，EMR 6.0.0 或 5.x 不支持。
  - 可以继续使用应用程序名称 Presto 在集群上安装 PrestoDB。要在集群上安装 Presto，请使用应用程序名称 PrestoSQL。
  - 您可以安装 PrestoDB 或 Presto，但不能在单个集群上同时安装两者。如果在尝试创建集群时同时指定了 PrestoDB 和 Presto，则会发生验证错误，集群创建请求将失败。
  - 单主集群和多主集群都支持 Presto。在多主集群上，需要外部 Hive 元数据仓库才能运行 Presto 或 PrestoDB。请参见[具有多个主节点的EMR集群中支持的应用程序](#)。
- ECR Apache Hadoop 和带有 Docker 的 Apache Spark 支持自动身份验证：Spark 用户可以使用 Docker Hub 和亚马逊弹性容器注册表（亚马逊 ECR）中的 Docker 镜像来定义环境和库依赖关系。

使用 [Amazon EMR 6.x 配置 Docker 并使用 Docker 运行 Spark 应用程序](#)。

- EMR 支持 Apache Hive ACID 事务：Amazon EMR 6.1.0 增加了对 Hive ACID 事务的支持，因此它符合数据库的 ACID 属性。借助此功能，您可以使用 Amazon Simple Storage Service (Amazon S3) 中的数据在 Hive 托管表中运行 INSERT、UPDATE、DELETE 和 MERGE 操作。对于流媒体摄取、数据重述、使用批量更新和缓慢更改维度等用例来说，MERGE 是一项关键功能。有关更多信息，包括配置示例和用例，请参阅[亚马逊EMR支持 Apache Hive ACID 事务](#)。

## 更改、增强功能和解决的问题

- 此版本旨在修复 Amazon EMR Scaling 无法成功扩展/缩小集群或导致应用程序故障时出现的问题。
- 修复了 Amazon EMR 集群上守护程序运行状况检查活动（例如收集 YARN 节点状态和 HDFS 节点状态）时，对利用率很高的大型集群的扩展请求失败的问题。之所以发生这种情况，是因为集群上的守护程序无法将节点的运行状况数据传送到内部 Amazon EMR 组件。
- 改进了 EMR 集群上的守护程序，可在重复使用 IP 地址时正确跟踪节点状态，从而提高扩展操作期间的可靠性。
- [SPARK-29683](#)。修复了集群缩减期间出现任务失败的问题，因为 Spark 假定所有可用节点都被拒绝列出。



- [YARN-9011](#)。修复了集群尝试向上或向下扩展时，由于YARN停用过程中的争用条件而导致任务失败的问题。
- 通过确保 Amaz EMR on 集群上守护程序和 YARN /之间的节点状态始终保持一致，修复了集群扩展期间步骤或任务失败的问题。HDFS
- 修复了启用了 Kerberos 身份验证的 Amazon 集群的EMR集群操作（例如缩减和步骤提交）失败的问题。这是因为 Amaz EMR on 集群上守护程序没有续订 Kerberos 票证，而要与主节点上YARN运行的 HDFS /进行安全通信，则需要续订 Kerberos 票证。
- 较新的Amazon EMR 版本通过降低亚马逊较旧AL2版本的“最大打开文件数”限制来解决这个问题 EMR。亚马逊EMR发布的 5.30.1、5.30.2、5.31.1、5.32.1、5.32.1、6.0.1、6.1.1、6.2.1、5.33.0、6.3.0及更高版本现在包含一个永久修复程序，其中包含一个具有更高的“最大打开文件数”设置的永久修复。
- EMR6.0.0 不支持 Apache Flink，但在 6.1.0 和 Flink 1.11.0 上EMR支持 Apache Flink。这是首个正式支持 Hadoop 3 的 Fink 版本。请参阅 [Apache Flink 1.11.0 发布公告](#)。
- Ganglia 已从默认 EMR 6.1.0 软件包捆绑包中移除。

## 已知问题

- 降低旧版本的“最大打开文件数”限制 AL2 [在新版本中已修复]。亚马逊EMR版本：  
emr-5.30.x、emr-5.31.0、emr-5.32.0、emr-6.0.0、emr-6.1.0 和 emr-6.2.0 基于较旧版本的 Linux ofAmazon 2 ()AL2，当使用默认版本创建亚马逊集群时，“最大打开文件数”的用户限制较低。EMR AMI亚马逊EMR发布的版本为5.30.1、5.30.2、5.31.1、5.32.1、6.0.1、6.1.1、6.2.1、5.33.0、6.3.0 及更高版本，包括具有更高的“最大打开文件数”设置的永久修复。如果使用打开文件数限制较低的发行版，会在提交 Spark 任务时导致“Too many open files”（打开的文件过多）错误。在受影响的版本中，亚马逊EMR默认AMI“最大打开文件数”的ulimit设置为4096，低于Linux 2中的65536个文件限制。latestAmazon AMISpark 驱动程序和执行程序尝试打开超过 4096 个文件时，“打开的最大文件数”的较低 ulimit 设置会导致 Spark 任务失败。为了解决这个问题，Amazon EMR 提供了一个引导操作 (BA) 脚本，可以在创建集群时调整ulimit设置。

如果您使用的是无法永久修复此问题的旧EMR版 Amazon，则可以使用以下解决方法将实例控制器 ulimit 明确设置为最多 65536 个文件。

从命令行显式设置 ulimit

1. 编辑 `/etc/systemd/system/instance-controller.service`，将以下参数添加到 Service (服务) 部分。

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

## 2. 重启 InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

### 使用引导操作 (BA) 设置 ulimit

您还可以在创建集群时使用引导操作 ( BA ) 脚本将实例控制器 ulimit 配置为 65536 个文件。

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

#### Important

Amazon EMR 6.1.0 和 6.2.0 包含一个性能问题，可能会严重影响 Hudi 的所有插入、更新和删除操作。如果您计划在亚马逊 EMR 6.1.0 或 6.2.0 中使用 Hudi，则应联系 AWS 支持人员获取已修补的 Hudi。RPM

- 如果您使用 `spark.driver.extraJavaOptions` 和设置自定义垃圾收集配置 `spark.executor.extraJavaOptions`，则由于垃圾收集配置冲突，将导致 EMR 6.1 版的驱动程序/执行程序启动失败。在 6.1.0 EMR 版本中，您应该使用属性 `spark.driver.defaultJavaOptions` 为驱动程序和执行程序指定自定义 Spark 垃圾收集配

置，改为使用。spark.executor.defaultJavaOptions在 [Apache Spark 运行时环境](#)和在[亚马逊 EMR 6.1.0 上配置 Spark 垃圾回收](#)中阅读更多内容。

- 在 Oozie 中使用 Pig ( 以及在 Hue 中，因为 Hue 使用 Oozie 操作来运行 Pig 脚本 ) 会生成一个错误，即无法加载 native-lzo 库。此错误消息是信息性的，不会阻止 Pig 运行。
- Hudi 并发支持：目前 Hudi 不支持并发写入单个 Hudi 表。此外，Hudi 会回滚处于运行状态的写入器所做的所有更改后再允许新写入器启动。并发写入可能会干扰此机制并引入竞争条件，这会导致数据损坏。您应确保作为数据处理工作流程的一部分，任何时候都只有一个 Hudi 写入器对 Hudi 表进行操作。Hudi 支持多个并发读取器对同一 Hudi 表进行操作。
- 具有多个主节点的集群和 Kerberos 身份验证中的已知问题

如果您在 Amazon 5.20.0 及更高EMR版本中运行具有多个主节点和 Kerberos 身份验证的集群，则在集群运行一段时间后，集群操作可能会遇到问题，例如缩小规模或步骤提交。具体时间段取决于您定义的 Kerberos 票证有效期。缩减问题会影响您提交的自动缩减和显式缩减请求。其它集群操作也可能受到影响。

解决办法：

- SSH以hadoop用户身份访问具有多个主节点的EMR群集的主节点。
- 运行以下命令，为 hadoop 用户续订 Kerberos 票证。

```
kinit -kt <keytab_file> <principal>
```

通常情况下，keytab 文件位于 /etc/hadoop.keytab，而 principal 为 hadoop/  
<hostname>@<REALM> 格式。

#### Note

此解决方法将在 Kerberos 票证有效期内生效。默认情况下，此持续时间为 10 个小时，但可以通过 Kerberos 设置进行配置。Kerberos 票证过期后，您必须重新运行上述命令。

- Amazon EMR 6.1.0 中存在一个影响运行 Presto 的集群的问题。在较长时间 ( 天 ) 后，集群可能会引发错误，例如“su: failed to execute /bin/bash: Resource temporarily unavailable”或“shell request failed on channel 0”。此问题是由亚马逊内部EMR进程 (InstanceController) 引起的，该进程生成了太多的轻量级进程 (LWP)，最终导致 Hadoop 用户超出其 nproc 限制。这可以阻止用户打开其它进程。此问题的解决方案是升级到 EMR 6.2.0。

## 发行版 6.0.0

以下发行说明包含 Amazon EMR 版本 6.0.0 的信息。

首次发布日期：2020 年 3 月 10 日

### 受支持的应用程序

- AWS SDK for Java 版本 1.11.711
- Ganglia 3.7.2
- Hadoop 3.2.1
- HBase版本 2.2.3
- HCatalog版本 3.1.2
- Hive 3.1.2
- Hudi 0.5.0-incubating
- Hue 4.4.0
- JupyterHub 版本 1.0.0
- Livy 0.6.0
- MXNet版本 1.5.1
- Oozie 5.1.0
- Phoenix 5.0.0
- Presto 0.230
- Spark 2.4.4
- TensorFlow 版本 1.14.0
- 齐柏林飞艇版本 0.9.0-SNAPSHOT
- Zookeeper 3.4.14
- 连接器和驱动程序：DynamoDB 连接器 4.14.0

#### Note

Flink、Sqoop、Pig 和 Mahout 在亚马逊版本 6.0.0 中不可用。EMR

## 新功能

- YARN Docker Runtime Support——诸如 Spark 作业之类的YARN应用程序现在可以在 Docker 容器的环境中运行。这使您可以轻松地在 Docker 映像中定义依赖关系，而无需在 Amazon EMR 集群上安装自定义库。有关更多信息，请参阅[使用 Amazon EMR 版本 6.0.0 配置 Docker 集成和使用 Docker 运行 Spark 应用程序](#)。
- Hive LLAP Support-Hive 现在支持LLAP执行模式，以提高查询性能。有关更多信息，请参阅[使用 Hive。LLAP](#)

## 更改、增强功能和解决的问题

- 此版本旨在修复 Amazon EMR Scaling 无法成功扩展/缩小集群或导致应用程序故障时出现的问题。
- 修复了 Amazon EMR on 集群上守护程序运行状况检查活动（例如收集YARN节点状态和HDFS节点状态）时，对利用率很高的大型集群的扩展请求失败的问题。之所以发生这种情况，是因为集群上的守护程序无法将节点的运行状况数据传送到内部 Amazon EMR 组件。
- 改进了EMR集群上的守护程序，可在重复使用 IP 地址时正确跟踪节点状态，从而提高扩展操作期间的可靠性。
- [SPARK-29683](#)。修复了集群缩减期间出现任务失败的问题，因为 Spark 假定所有可用节点都被拒绝列出。
- [YARN-9011](#)。修复了集群尝试向上或向下扩展时，由于YARN停用过程中的争用条件而导致任务失败的问题。
- 通过确保 Amazon EMR on 集群上守护程序和 YARN 之间的节点状态始终保持一致，修复了集群扩展期间步骤或任务失败的问题。HDFS
- 修复了启用了 Kerberos 身份验证的 Amazon 集群的EMR集群操作（例如缩减和步骤提交）失败的问题。这是因为 Amazon EMR on 集群上守护程序没有续订 Kerberos 票证，而要与主节点上YARN运行的 HDFS 进行安全通信，则需要续订 Kerberos 票证。
- 较新的Amazon EMR 版本通过降低亚马逊较旧AL2版本的“最大打开文件数”限制来解决这个问题。亚马逊EMR发布的 5.30.1、5.30.2、5.31.1、5.32.1、5.32.1、6.0.1、6.1.1、6.2.1、5.33.0、6.3.0及更高版本现在包含一个永久修复程序，其中包含一个具有更高的“最大打开文件数”设置的永久修复。
- Amazon Linux
  - 亚马逊 Linux 2 是 EMR 6.x 版本系列的操作系统。
  - systemd用于服务管理，而不是upstart使用的 inAmazon Linux 1。
- Java 开发套件 (JDK)

- C JDK orretto 8 是 6.x 版本系列EMR的JDK默认版本。
- Scala
  - Scala 2.12 与 Apache Spark 和 Apache Livy 一起使用。
- Python 3
  - Python 3 现在是中默认版本的 Python EMR。
- YARN节点标签
  - 从 Amazon EMR 6.x 版本系列开始，默认情况下YARN节点标签功能处于禁用状态。默认情况下，应用程序主进程可以在核心节点和任务节点上运行。您可以通过配置以下属性来启用YARN节点标签功能：`yarn.node-labels.enabled`和`yarn.node-labels.am.default-node-label-expression`。有关更多信息，请参阅[了解主节点、核心节点和任务节点](#)。

## 已知问题

- 降低旧版本的“最大打开文件数”限制 AL2 [在新版本中已修复]。亚马逊EMR版本：`emr-5.30.x`、`emr-5.31.0`、`emr-5.32.0`、`emr-6.0.0`、`emr-6.1.0` 和 `emr-6.2.0` 基于较旧版本的 Linux of Amazon 2 ()AL2，当使用默认版本创建亚马逊集群时，“最大打开文件数”的用户限制较低。EMR AMI亚马逊EMR发布的版本为5.30.1、5.30.2、5.31.1、5.32.1、6.0.1、6.1.1、6.2.1、5.33.0、6.3.0 及更高版本，包括具有更高的“最大打开文件数”设置的永久修复。如果使用打开文件数限制较低的发行版，会在提交 Spark 任务时导致“Too many open files”（打开的文件过多）错误。在受影响的版本中，亚马逊EMR默认AMI的“最大打开文件数”的`ulimit`设置为4096，低于Linux 2中的65536个文件限制。latest Amazon AMI Spark 驱动程序和执行程序尝试打开超过 4096 个文件时，“打开的最大文件数”的较低 `ulimit` 设置会导致 Spark 任务失败。为了解决这个问题，Amazon EMR 提供了一个引导操作 (BA) 脚本，可以在创建集群时调整`ulimit`设置。

如果您使用的是无法永久修复此问题的旧EMR版 Amazon，则可以使用以下解决方法将实例控制器 `ulimit` 明确设置为最多 65536 个文件。

### 从命令行显式设置 `ulimit`

1. 编辑 `/etc/systemd/system/instance-controller.service`，将以下参数添加到 Service (服务) 部分。

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. 重启 InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

### 使用引导操作 (BA) 设置 ulimit

您还可以在创建集群时使用引导操作 ( BA ) 脚本将实例控制器 ulimit 配置为 65536 个文件。

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- Spark 交互式外壳 PySpark，包括 SparkR 和 spark-shell，不支持将 Docker 与其他库一起使用。
- 要在亚马逊 6.0.0 EMR 版本中使用 Python 3，必须添加PATH到。yarn.nodemanager.env-whitelist
- 当您使用 AWS Glue 数据目录作为 Hive 的元存储时，不支持 Live Long 和 Process (LLAP) 功能。
- 使用集成 Spark 和 Docker 的 Amazon EMR 6.0.0 时，您需要在集群中配置具有相同实例类型和相同数量的EBS卷的实例，以避免在提交 Docker 运行时的 Spark 任务时失败。
- 在亚马逊 EMR 6.0.0 中，亚马逊 S3 HBase 上的存储模式受到 [HBASE-24](#) 286. 问题的影响。HBase 使用现有 S3 数据创建集群时，master 无法初始化。
- 具有多个主节点的集群和 Kerberos 身份验证中的已知问题

如果您在 Amazon 5.20.0 及更高EMR版本中运行具有多个主节点和 Kerberos 身份验证的集群，则在集群运行一段时间后，集群操作可能会遇到问题，例如缩小规模或步骤提交。具体时间段取决于您

定义的 Kerberos 票证有效期。缩减问题会影响您提交的自动缩减和显式缩减请求。其它集群操作也可能受到影响。

解决办法：

- SSH以hadoop用户身份访问具有多个主节点的EMR群集的主节点。
- 运行以下命令，为 hadoop 用户续订 Kerberos 票证。

```
kinit -kt <keytab_file> <principal>
```

通常情况下，keytab 文件位于 /etc/hadoop.keytab，而 principal 为 hadoop/<hostname>@<REALM> 格式。

#### Note

此解决方法将在 Kerberos 票证有效期内生效。默认情况下，此持续时间为 10 个小时，但可以通过 Kerberos 设置进行配置。Kerberos 票证过期后，您必须重新运行上述命令。

## 发行版 5.30.1

以下发行说明包含有关亚马逊 5.30.1 EMR 版本的信息。更改与 5.30.0 有关。

首次发布日期：2020 年 6 月 30 日

上次更新时间：2020 年 8 月 24 日

更改、增强功能和解决的问题

- 较新的 Amazon EMR 版本通过降低亚马逊较旧 AL2 版本的“最大打开文件数”限制来解决这个问题。亚马逊 EMR 发布的 5.30.1、5.30.2、5.31.1、5.32.1、5.32.1、6.0.1、6.1.1、6.2.1、5.33.0、6.3.0 及更高版本现在包含一个永久修复程序，其中包含一个具有更高的“最大打开文件数”设置的永久修复。
- 修复了实例控制器进程生成无限量进程的问题。
- 修复了以下问题：Hue 无法运行 Hive 查询并显示“database is locked ( 数据库已锁定 )”消息、阻止执行查询的问题。
- 修复了 Spark 问题，允许在 EMR 集群上同时运行更多任务。
- 修复了一个 Jupyter notebook 问题，该问题会导致 Jupyter 服务器中出现“too many files open error ( 打开过多文件错误 )”。



- 修复了集群启动时间的问题。

## 新功能

- Tez 用户界面和YARN时间轴服务器永久应用程序接口适用于亚马逊 6.x EMR 版本和 5.30.1 及EMR 更高版本。通过单击链接访问永久应用程序历史记录，无需通过SSH连接设置 Web 代理，即可快速访问作业历史记录。活动和已终止集群的日志将在应用程序结束后保留 30 天。有关更多信息，请参阅 Amazon EMR 管理指南中的[查看永久应用程序用户界面](#)。
- EMR笔记本执行APIs可以通过脚本或命令行执行EMR笔记本。无需 AWS 控制台即可启动、停止、列出和描述EMR笔记本执行的功能，使您能够以编程方式控制EMR笔记本电脑。借助参数化笔记本单元，您可以将不同的参数值传递给笔记本，而无需为每组新参数值创建笔记本副本。参见[EMR API 操作](#)。有关示例代码，请参阅[以编程方式执行 EMR Notebook 的示例命令](#)。

## 已知问题

- 降低旧版本的“最大打开文件数”限制 AL2 [在新版本中已修复]。亚马逊EMR版本：  
emr-5.30.x、emr-5.31.0、emr-5.32.0、emr-6.0.0、emr-6.1.0 和 emr-6.2.0 基于较旧版本的 Linux ofAmazon 2 ()AL2，当使用默认版本创建亚马逊集群时，“最大打开文件数”的用户限制较低。EMR AMI亚马逊EMR发布的版本为5.30.1、5.30.2、5.31.1、5.32.1、6.0.1、6.1.1、6.2.1、5.33.0、6.3.0 及更高版本，包括具有更高的“最大打开文件数”设置的永久修复。如果使用打开文件数限制较低的发行版，会在提交 Spark 任务时导致“Too many open files”（打开的文件过多）错误。在受影响的版本中，亚马逊EMR默认AMI“最大打开文件数”的ulimit设置为4096，低于Linux 2中的65536个文件限制。latestAmazon AMISpark 驱动程序和执行程序尝试打开超过 4096 个文件时，“打开的最大文件数”的较低 ulimit 设置会导致 Spark 任务失败。为了解决这个问题，Amazon EMR 提供了一个引导操作 (BA) 脚本，可以在创建集群时调整ulimit设置。

如果您使用的是无法永久修复此问题的旧EMR版 Amazon，则可以使用以下解决方法将实例控制器 ulimit 明确设置为最多 65536 个文件。

### 从命令行显式设置 ulimit

1. 编辑 `/etc/systemd/system/instance-controller.service`，将以下参数添加到 Service (服务) 部分。

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. 重启 InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

## 使用引导操作 (BA) 设置 ulimit

您还可以在创建集群时使用引导操作 ( BA ) 脚本将实例控制器 ulimit 配置为 65536 个文件。

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- EMR笔记本电脑

在 5.30.1 EMR 版本中，允许您在集群主节点上安装内核和其他 Python 库的功能默认处于禁用状态。有关此功能的更多信息，请参阅[在集群主节点上安装内核和 Python 库](#)。

要启动此功能，请执行以下操作：

1. 确保附加到 Notebook 服务角色的权限EMR策略允许执行以下操作：

```
elasticmapreduce:ListSteps
```

有关更多信息，请参阅[EMR笔记本的服务角色](#)。

2. 使用在 AWS CLI 集群上运行设置 EMR Notebook 的步骤，如以下示例所示。Replace ( 替换 ) `us-east-1` 与您的集群所在的区域相同。有关更多信息，请参阅[使用 AWS CLI向集群中添加步骤](#)。

```
aws emr add-steps --cluster-id MyClusterID --steps
  Type=CUSTOM_JAR,Name=EMRNotebooksSetup,ActionOnFailure=CONTINUE,Jar=s3://us-
east-1.elasticmapreduce/libs/script-runner/script-runner.jar,Args=["s3://
  awssupportdatasvcs.com/bootstrap-actions/EMRNotebooksSetup/emr-notebooks-setup.sh"]
```

- 托管扩展

在未安装 Presto 的 5.30.0 和 5.30.1 的集群上进行托管扩展操作可能会导致应用程序故障或导致统一的实例组或实例集处于 ARRESTED 状态，尤其是在缩减操作之后快速执行扩展操作时。

解决方法是，即使您的任务不需要 Presto，也可以在创建包含亚马逊EMR版本 5.30.0 和 5.30.1 的集群时选择 Presto 作为要安装的应用程序。

- 具有多个主节点的集群和 Kerberos 身份验证中的已知问题


如果您在 Amazon 5.20.0 及更高EMR版本中运行具有多个主节点和 Kerberos 身份验证的集群，则在集群运行一段时间后，集群操作可能会遇到问题，例如缩小规模或步骤提交。具体时间段取决于您定义的 Kerberos 票证有效期。缩减问题会影响您提交的自动缩减和显式缩减请求。其它集群操作也可能受到影响。

解决办法：

- SSH以hadoop用户身份访问具有多个主节点的EMR群集的主节点。
- 运行以下命令，为 hadoop 用户续订 Kerberos 票证。

```
kinit -kt <keytab_file> <principal>
```

通常情况下，keytab 文件位于 `/etc/hadoop.keytab`，而 principal 为 `hadoop/<hostname>@<REALM>` 格式。

 Note

此解决方法将在 Kerberos 票证有效期内生效。默认情况下，此持续时间为 10 个小时，但可以通过 Kerberos 设置进行配置。Kerberos 票证过期后，您必须重新运行上述命令。

- 当你使用带有 Hive 分区位置格式的 Spark 来读取 Amazon S3 中的数据，并在亚马逊EMR版本 5.30.0 到 5.36.0 和 6.2.0 到 6.9.0 上运行 Spark 时，你可能会遇到一个问题，导致集群无法正确读取数据。如果您的分区具有以下所有特征，会发生这种情况：
  - 从同一个表扫描两个或多个分区。
  - 至少有一个分区目录路径是至少一个其他分区目录路径的前缀，例如，s3://bucket/table/p=a 是 s3://bucket/table/p=a b 的前缀。
  - 另一个分区目录中前缀之后的第一个字符的值为 UTF -8，小于该/字符 (U+002F)。例如，在 s3://bucket/table/p=a b 中，a 和 b 之间出现的空格字符 (U+0020) 就属于此类。请注意，还有其他 14 个非控制字符：! "\$ % & ' ( ) \* + , - 。有关更多信息，请参阅 [UTF-8 编码表和 Unicode 字符](#)。

解决方法是在 spark-defaults 分类中将 spark.sql.sources.fastS3PartitionDiscovery.enabled 配置设置为 false。

## 发行版 5.30.0

以下发行说明包含 Amazon EMR 版本 5.30.0 的信息。更改与 5.29.0 有关。

首次发布日期：2020 年 5 月 13 日

上次更新日期：2020 年 6 月 25 日

### 升级

- 已升级 AWS SDK for Java 到 1.11.759 版本
- 已将 Amazon SageMaker Spark 升级 SDK 到 1.3.0 版
- 已将 R EMR ecord Server 升级到 1.6.0 版
- 已将 Flink 升级到版本 1.10.0
- 已将 Ganglia 升级到版本 3.7.2
- 已升级 HBase 到 1.4.13 版
- 已将 Hudi 升级到版本 0.5.2-incubating
- 已将 Hue 升级到版本 4.6.0
- 已升级 JupyterHub 到 1.1.0 版
- 已将升级 Livy 到版本 0.7.0-incubating
- 已将 Oozie 升级到版本 5.2.0

- 已将 Presto 升级到版本 0.232
- 已将 Spark 升级到版本 2.4.5
- 升级后的连接器和驱动程序：亚马逊 Glue Connector 1.12.0；亚马逊 Kinesis Connector 3.5.0；DynamoDB 连接器 4.14.0 EMR

## 新功能

- EMR笔记本-与使用 5.30.0 创建的EMR集群一起使用时，EMR笔记本内核在集群上运行。这可以提高笔记本的性能，并允许您安装和自定义内核。您还可以在集群主节点上安装 Python 库。有关更多信息，请参阅《EMR管理指南》中的[安装和使用内核和库](#)。
- 托管扩展 — 在 Amazon 5.30.0 及更高EMR版本中，您可以启用EMR托管扩展，从而根据工作负载自动增加或减少集群中的实例或单元数量。Amazon 会EMR持续评估集群指标，以做出扩展决策，从而优化集群的成本和速度。有关更多信息，请参阅 Amazon EMR 管理指南中的[扩展集群资源](#)。
- 加密存储在 Amazon S3 中的日志文件 — 在 Amazon 5.30.0 及更高EMR版本中，您可以使用 AWS KMS 客户托管密钥加密存储在 Amazon S3 中的日志文件。有关更多信息，请参阅《亚马逊EMR管理指南》中的加密存储在 Amazon [S3 中的日志文件](#)。
- 亚马逊 Linux 2 支持 — 在 5.30.0 及更高EMR版本中，EMR uses Amazon Linux 2 操作系统。新的自定义AMIs ( Amazon 机器映像 ) 必须基于 the Amazon Linux 2 AMI。有关更多信息，请参阅[使用自定义AMI](#)。
- Presto Graceful Auto Scale — 使用 5.30.0 的EMR集群可以设置自动缩放超时时间，让 Presto 任务有时间在节点停用之前完成运行。有关更多信息，请参阅 [使用采用 Graceful Decommission 的 Presto 自动扩展配置](#)。
- 使用新的分配策略选项创建舰队实例 — 5.12.1 及更高EMR版本中提供了新的分配策略选项。它加快了集群预置、提高了 Spot 分配的准确性并减少了竞价型实例中断。需要更新非默认EMR服务角色。请查看[配置实例集](#)。
- sudo systemctl stop 和 sudo systemctl 启动命令 — 在 5.30.0 及更高EMR版本中，use Amazon Linux 2 操作系统使用和命令来重启服务。EMR sudo systemctl stop sudo systemctl start有关更多信息，请参阅[如何在 Amazon 中重启服务EMR？](#)。

## 更改、增强功能和解决的问题

- EMR默认情况下，版本 5.30.0 不安装 Ganglia。您可以在创建集群时明确选择 Ganglia 进行安装。
- Spark 性能优化。
- Presto 性能优化。

- Python 3 是亚马逊 5.30.0 及更高EMR版本的默认版本。
- 用于私有子网中服务访问的默认托管安全组已使用新规则进行更新。如果使用自定义安全组进行服务访问，则必须包含与默认托管安全组相同的规则。有关更多信息，请参阅 [EMRAmazon 服务访问托管安全组 \(私有子网\)](#)。如果您为 Amazon 使用自定义服务角色EMR，则必须向授予权限，`ec2:describeSecurityGroups`这样EMR才能验证安全组的创建是否正确。如果您使用 `EMR_DefaultRole`，则此权限已包含在默认托管式策略中。

## 已知问题

- 降低旧版本的“最大打开文件数”限制 AL2 [在新版本中已修复]。亚马逊EMR版本：`emr-5.30.x`、`emr-5.31.0`、`emr-5.32.0`、`emr-6.0.0`、`emr-6.1.0` 和 `emr-6.2.0` 基于较旧版本的 Linux ofAmazon 2 ()AL2，当使用默认版本创建亚马逊集群时，“最大打开文件数”的用户限制较低。EMR AMI亚马逊EMR发布的版本为5.30.1、5.30.2、5.31.1、5.32.1、6.0.1、6.1.1、6.2.1、5.33.0、6.3.0 及更高版本，包括具有更高的“最大打开文件数”设置的永久修复。如果使用打开文件数限制较低的发行版，会在提交 Spark 任务时导致“Too many open files”（打开的文件过多）错误。在受影响的版本中，亚马逊EMR默认AMI“最大打开文件数”的

如果您使用的是无法永久修复此问题的旧EMR版 Amazon，则可以使用以下解决方法将实例控制器 ulimit 明确设置为最多 65536 个文件。

### 从命令行显式设置 ulimit

1. 编辑 `/etc/systemd/system/instance-controller.service`，将以下参数添加到 Service (服务) 部分。

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. 重启 InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

### 使用引导操作 (BA) 设置 ulimit

您还可以在创建集群时使用引导操作 ( BA ) 脚本将实例控制器 ulimit 配置为 65536 个文件。

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- 托管扩展

在未安装 Presto 的 5.30.0 和 5.30.1 的集群上进行托管扩展操作可能会导致应用程序故障或导致统一的实例组或实例集处于 ARRESTED 状态，尤其是在缩减操作之后快速执行扩展操作时。

解决方法是，即使您的任务不需要 Presto，也可以在创建包含亚马逊EMR版本 5.30.0 和 5.30.1 的集群时选择 Presto 作为要安装的应用程序。

- 具有多个主节点的集群和 Kerberos 身份验证中的已知问题

如果您在 Amazon 5.20.0 及更高EMR版本中运行具有多个主节点和 Kerberos 身份验证的集群，则在集群运行一段时间后，集群操作可能会遇到问题，例如缩小规模或步骤提交。具体时间段取决于您定义的 Kerberos 票证有效期。缩减问题会影响您提交的自动缩减和显式缩减请求。其它集群操作也可能受到影响。

解决办法：

- SSH以hadoop用户身份访问具有多个主节点的EMR群集的主节点。
- 运行以下命令，为 hadoop 用户续订 Kerberos 票证。

```
kinit -kt <keytab_file> <principal>
```

通常情况下，keytab 文件位于 `/etc/hadoop.keytab`，而 principal 为 `hadoop/<hostname>@<REALM>` 格式。

#### Note

此解决方法将在 Kerberos 票证有效期内生效。默认情况下，此持续时间为 10 个小时，但可以通过 Kerberos 设置进行配置。Kerberos 票证过期后，您必须重新运行上述命令。

- Hue 4.6.0 的默认数据库引擎是 SQLite，当您尝试将 Hue 与外部数据库配合使用时，这会导致出现问题。若要解决此问题，请在您的 `hue-ini` 配置分类中将 `engine` 设置为 `mysql`。此问题已在亚马逊 EMR 版本 5.30.1 中修复。
- 当你使用带有 Hive 分区位置格式的 Spark 来读取 Amazon S3 中的数据，并在亚马逊 EMR 版本 5.30.0 到 5.36.0 和 6.2.0 到 6.9.0 上运行 Spark 时，你可能会遇到一个问题，导致集群无法正确读取数据。如果您的分区具有以下所有特征，会发生这种情况：
  - 从同一个表扫描两个或多个分区。
  - 至少有一个分区目录路径是至少一个其他分区目录路径的前缀，例如，`s3://bucket/table/p=a` 是 `s3://bucket/table/p=a b` 的前缀。
  - 另一个分区目录中前缀之后的第一个字符的值为 UTF-8，小于该/字符 (U+002F)。例如，在 `s3://bucket/table/p=a b` 中，`a` 和 `b` 之间出现的空格字符 (U+0020) 就属于此类。请注意，还有其他 14 个非控制字符：`!"#$%&'()*+,-`。有关更多信息，请参阅 [UTF-8 编码表和 Unicode 字符](#)。

解决方法是在 `spark-defaults` 分类中将 `spark.sql.sources.fastS3PartitionDiscovery.enabled` 配置设置为 `false`。

## 发行版 5.29.0

以下发行说明包含有关亚马逊 5.29.0 EMR 版本的信息。更改与 5.28.1 有关。

首次发布日期：2020 年 1 月 17 日

### 升级

- 已升级 AWS SDK for Java 到 1.11.682 版本
- 已将 Hive 升级到版本 2.3.6
- 已将 Flink 升级到版本 1.9.1



- 已将 EMRFS 升级到版本 2.38.0
- 已将 EMR DynamoDB 连接器升级到版本 4.13.0

### 更改、增强功能和解决的问题

- Spark
  - Spark 性能优化。
- EMRFS
  - 将管理指南更新为 `emrfs-site.xml` 默认设置以实现了一致视图。

### 已知问题

- 具有多个主节点的集群和 Kerberos 身份验证中的已知问题

如果您在 Amazon 5.20.0 及更高EMR版本中运行具有多个主节点和 Kerberos 身份验证的集群，则在集群运行一段时间后，集群操作可能会遇到问题，例如缩小规模或步骤提交。具体时间段取决于您定义的 Kerberos 票证有效期。缩减问题会影响您提交的自动缩减和显式缩减请求。其它集群操作也可能受到影响。

#### 解决办法：

- SSH以hadoop用户身份访问具有多个主节点的EMR群集的主节点。
- 运行以下命令，为 hadoop 用户续订 Kerberos 票证。

```
kinit -kt <keytab_file> <principal>
```

通常情况下，keytab 文件位于 `/etc/hadoop.keytab`，而 principal 为 `hadoop/<hostname>@<REALM>` 格式。

#### Note

此解决方法将在 Kerberos 票证有效期内生效。默认情况下，此持续时间为 10 个小时，但可以通过 Kerberos 设置进行配置。Kerberos 票证过期后，您必须重新运行上述命令。

## 版本 5.28.1

以下发行说明包含有关亚马逊 5.28.1 EMR 版本的信息。更改与 5.28.0 有关。

首次发布日期：2020 年 1 月 10 日

### 更改、增强功能和解决的问题

- Spark
  - 修复了 Spark 兼容性问题。
- CloudWatch 指标
  - 修复了在具有多个主节点的EMR集群上发布亚马逊 CloudWatch 指标的问题。
- 已禁用日志消息
  - 已禁用假日志消息“...using old version (<4.5.8) of Apache http client”（使用低于版本 4.5.8 的 Apache http 客户端）。

### 已知问题

- 具有多个主节点的集群和 Kerberos 身份验证中的已知问题

如果您在 Amazon 5.20.0 及更高EMR版本中运行具有多个主节点和 Kerberos 身份验证的集群，则在集群运行一段时间后，集群操作可能会遇到问题，例如缩小规模或步骤提交。具体时间段取决于您定义的 Kerberos 票证有效期。缩减问题会影响您提交的自动缩减和显式缩减请求。其它集群操作也可能受到影响。

#### 解决办法：

- SSH以hadoop用户身份访问具有多个主节点的EMR群集的主节点。
- 运行以下命令，为 hadoop 用户续订 Kerberos 票证。

```
kinit -kt <keytab_file> <principal>
```

通常情况下，keytab 文件位于 `/etc/hadoop.keytab`，而 principal 为 `hadoop/<hostname>@<REALM>` 格式。

**Note**

此解决方法将在 Kerberos 票证有效期内生效。默认情况下，此持续时间为 10 个小时，但可以通过 Kerberos 设置进行配置。Kerberos 票证过期后，您必须重新运行上述命令。

## 版本 5.28.0

以下发行说明包含有关亚马逊 5.28. EMR 0 版本的信息。更改与 5.27.0 有关。

首次发布日期：2019 年 11 月 12 日

### 升级

- 已将 Flink 升级到版本 1.9.0
- 已将 Hive 升级到版本 2.3.6
- 已升级MXNet到 1.5.1 版
- 已将 Phoenix 升级到版本 4.14.3
- 已将 Presto 升级到版本 0.227
- 已将 Zeppelin 升级到版本 0.8.2

### 新功能

- [Apache Hudi](#) 现在EMR可供亚马逊在你创建集群时安装。有关更多信息，请参阅 [Hudi](#)。
- (2019 年 11 月 25 日) 您现在可以选择并行运行多个步骤以提高集群利用率并节省成本。您还可以取消待处理和正在运行的步骤。有关更多信息，请参阅[使用 AWS CLI 和控制台处理步骤](#)。
- (2019 年 12 月 3 日) 您现在可以在上创建和运行EMR集群 AWS Outposts。AWS Outposts 在本地设施中启用原生 AWS 服务、基础设施和运营模式。在 AWS Outposts 环境中，您可以使用与 AWS 云端相同的 AWS APIs工具和基础架构。有关更多信息，请参阅[上的EMR集群 AWS Outposts](#)。
- (2020 年 3 月 11 日) 从 Amazon EMR 版本 5.28.0 开始，您可以在本地区域子网上创建和运行亚马逊EMR集群，作为支持 AWS 本地区域的 AWS 区域的逻辑扩展。本地区域使得 Amazon EMR 功能和部分 AWS 服务（例如计算和存储服务）可以位于离用户更近的地方，从而为本地运行的应用程序提供极低的延迟访问。有关可用的 Local Zones 列表，请参阅 [AWS Local Zones](#)。有关访问可用 AWS 本地区域的信息，请参阅[区域、可用区和本地区域](#)。

Local Zones 目前不支持亚马逊EMR笔记本电脑，也不支持EMR使用接口VPC终端节点 (AWS PrivateLink) 直接连接到亚马逊。

## 更改、增强功能和解决的问题

- 扩展了对高可用性集群的应用程序支持
  - 有关更多信息，请参阅 [Amazon EMR 管理指南中的具有多个主节点的EMR集群中支持的应用程序](#)。
- Spark
  - 性能优化
- Hive
  - 性能优化
- Presto
  - 性能优化

## 已知问题

- 具有多个主节点的集群和 Kerberos 身份验证中的已知问题

如果您在 Amazon 5.20.0 及更高EMR版本中运行具有多个主节点和 Kerberos 身份验证的集群，则在集群运行一段时间后，集群操作可能会遇到问题，例如缩小规模或步骤提交。具体时间段取决于您定义的 Kerberos 票证有效期。缩减问题会影响您提交的自动缩减和显式缩减请求。其它集群操作也可能受到影响。

### 解决办法：

- SSH以hadoop用户身份访问具有多个主节点的EMR群集的主节点。
- 运行以下命令，为 hadoop 用户续订 Kerberos 票证。

```
kinit -kt <keytab_file> <principal>
```

通常情况下，keytab 文件位于 `/etc/hadoop.keytab`，而 principal 为 `hadoop/<hostname>@<REALM>` 格式。

**Note**

此解决方法将在 Kerberos 票证有效期内生效。默认情况下，此持续时间为 10 个小时，但可以通过 Kerberos 设置进行配置。Kerberos 票证过期后，您必须重新运行上述命令。

## 版本 5.27.0

以下发行说明包含有关亚马逊 5.27. EMR 0 版本的信息。更改与 5.26.0 有关。

首次发布日期：2019 年 9 月 23 日

### 升级

- AWS SDK for Java 1.11.615
- Flink 1.8.1
- JupyterHub 1.0.0
- Spark 2.4.4
- Tensorflow 1.14.0
- 连接器和驱动程序：
  - DynamoDB 连接器 4.12.0

### 新功能

- (2019 年 10 月 24 日) 所有亚马逊EMR版本均提供EMR笔记本电脑的以下新功能。
  - 现在，您可以将 Git 存储库与EMR笔记本关联，将您的笔记本存储在版本控制的环境中。您可以通过远程 Git 存储库与同行共享代码，并重复使用现有的 Jupyter notebook。有关更多信息，请参阅《亚马逊EMR管理指南》中的“将 Git 存储库与亚马逊EMR[笔记本关联](#)”。
  - [nbdime 实用程序](#)现已在EMR笔记本中提供，用于简化比较和合并笔记本电脑。
  - EMR现在支持笔记本电脑 JupyterLab。JupyterLab 是一个基于 Web 的交互式开发环境，与 Jupyter 笔记本电脑完全兼容。现在，你可以选择在 Jupyter 笔记本编辑器 JupyterLab 或 Jupyter 笔记本编辑器中打开笔记本。
- (2019 年 10 月 30 日) 在 Amazon 5.25.0 及更高EMR版本中，您可以从集群摘要页面或控制台的“应用程序历史记录”选项卡连接到 Spark 历史服务器用户界面。您无需通过SSH连接设置 Web 代理，而是可以快速访问 Spark 历史服务器用户界面来查看应用程序指标并访问活动集群和已终止集

群的相关日志文件。有关更多信息，请参阅 Amazon EMR 管理指南中的[集群外访问永久应用程序用户界面](#)。

## 更改、增强功能和解决的问题

- 带有多个主节点的 Amazon EMR 集群
  - 您可以在具有多个主节点的 Amazon EMR 集群上安装和运行 Flink。有关更多信息，请参阅[支持的应用程序和功能](#)。
  - 您可以在具有多个主节点的 Amazon EMR 集群上配置HDFS透明加密。有关更多信息，请参阅在[具有多个主节点的EMR集群上进行HDFS透明加密](#)。
  - 现在，您可以修改在具有多个主节点的 Amazon EMR 集群上运行的应用程序的配置。有关更多信息，请参阅[为运行的集群中的实例组提供配置](#)。
- Amazon EMR-DynamoDB 连接器
  - Amazon EMR-DynamoDB Connector 现在支持以下 DynamoDB 数据类型：布尔值、列表、地图、项目、空。有关更多信息，请参阅[设置 Hive 表以运行 Hive 命令](#)。

## 已知问题

- 具有多个主节点的集群和 Kerberos 身份验证中的已知问题

如果您在 Amazon 5.20.0 及更高EMR版本中运行具有多个主节点和 Kerberos 身份验证的集群，则在集群运行一段时间后，集群操作可能会遇到问题，例如缩小规模或步骤提交。具体时间段取决于您定义的 Kerberos 票证有效期。缩减问题会影响您提交的自动缩减和显式缩减请求。其它集群操作也可能受到影响。

### 解决办法：

- SSH以hadoop用户身份访问具有多个主节点的EMR群集的主节点。
- 运行以下命令，为 hadoop 用户续订 Kerberos 票证。

```
kinit -kt <keytab_file> <principal>
```

通常情况下，keytab 文件位于 /etc/hadoop.keytab，而 principal 为 hadoop/<hostname>@<REALM> 格式。

**Note**

此解决方法将在 Kerberos 票证有效期内生效。默认情况下，此持续时间为 10 个小时，但可以通过 Kerberos 设置进行配置。Kerberos 票证过期后，您必须重新运行上述命令。

## 版本 5.26.0

以下发行说明包含有关亚马逊 5.26. EMR 0 版本的信息。更改与 5.25.0 有关。

首次发布日期：2019 年 8 月 8 日

上次更新日期：2019 年 8 月 19 日

### 升级

- AWS SDK for Java 1.11.595
- HBase1.4.10
- Phoenix 4.14.2
- 连接器和驱动程序：
  - DynamoDB 连接器 4.11.0
  - MariaDB 连接器 2.4.2
  - 亚马逊 Redshift JDBC Driver 1.2.32.1056

### 新功能

- (测试版) 使用亚马逊 EMR 5.26.0，您可以启动与 Lake Formation 集成的集群。这种集成提供了对 Glue 数据目录中的数据库和表的精细列级访问。AWS 它还支持从企业身份系统对 EMR 笔记本或 Apache Zeppelin 进行联合单点登录。有关更多信息，请参阅[将 Amazon EMR 与 AWS Lake Formation \(测试版\) 集成](#)。
- (2019 年 8 月 19 日) 所有支持安全组的亚马逊 EMR 版本现已推出亚马逊 EMR 封锁公开访问权限。屏蔽公共访问是适用于每个 AWS 地区的账户范围的设置。除非将端口指定为例外，否则当任何与集群关联的安全组都有允许某个端口上来自 IPv4 0.0.0.0/0 或 IPv6::/0 (公共访问) 的入站流量的规则时，阻止公共访问会阻止集群启动。默认情况下，端口 22 是一个例外。有关更多信息，请参阅[《亚马逊 EMR 管理指南》中的“使用 Amazon EMR 阻止公共访问”](#)。

## 更改、增强功能和解决的问题

### • EMR笔记本电脑

- 在 EMR 5.26.0 及更高版本中，EMR笔记本除了默认的 Python 库外，还支持笔记本电脑范围的 Python 库。无需重新创建集群或重新将笔记本附加到集群，您即可从笔记本编辑器中安装笔记本范围的库。笔记本范围的库是在 Python 虚拟环境中创建的，因此适用于当前笔记本会话。这使得您可以隔离笔记本依赖项。有关更多信息，请参阅 Amazon EMR 管理指南中的[使用笔记本范围的库](#)。

### • EMRFS

- 您可以通过将设置 `fs.s3.consistent.metadata.etag.validation.enabled` 为来启用 ETag 验证功能（测试版）`true`。借助此功能，ETags 可以 EMRFS 使用 Amazon S3 验证正在读取的对象是否为最新可用版本。此功能对于 Amazon S3 上的文件被覆盖，同时保留相同名称的 `read-after-update` 用例非常有用。此 ETag 验证功能目前不适用于 S3 Select。有关更多信息，请参阅[配置统一视图](#)。

### • Spark

- 现在默认启用了以下优化：DISTINCT 之前的动态分区修剪、JOIN 后续查询的 SQL 计划统计推断的改进 INTERSECT、标量子 DISTINCT 查询的扁平化、优化的联接重新排序以及布隆过滤器联接。有关更多信息，请参阅[优化 Spark 性能](#)。
- 改进了排序合并连接的整个阶段代码生成。
- 改进了查询片段和子查询重用。
- 改进了 Spark 启动时的预分配执行程序。
- 连接的较小侧包含广播提示时，不再应用 Bloom 筛选条件连接。

### • Tez

- 已解决 Tez 中存在的问题。Tez UI 现在可以在具有多个主节点的 Amazon EMR 集群上运行。

## 已知问题

- 改进的“排序合并连接的整个阶段代码生成”功能在启用后会增加内存压力。此优化可提高性能，但如果 `spark.yarn.executor.memoryOverheadFactor` 未调整，不能提供足够的内存，则会导致任务重试或失败。要禁用此功能，请将 `spark.sql.sortMergeJoinExec.extendedCodegen.enabled` 设置为 `false`。
- 具有多个主节点的集群和 Kerberos 身份验证中的已知问题

如果您在 Amazon 5.20.0 及更高 EMR 版本中运行具有多个主节点和 Kerberos 身份验证的集群，则在集群运行一段时间后，集群操作可能会遇到问题，例如缩小规模或步骤提交。具体时间段取决于您



定义的 Kerberos 票证有效期。缩减问题会影响您提交的自动缩减和显式缩减请求。其它集群操作也可能受到影响。

解决办法：

- SSH以hadoop用户身份访问具有多个主节点的EMR群集的主节点。
- 运行以下命令，为 hadoop 用户续订 Kerberos 票证。

```
kinit -kt <keytab_file> <principal>
```

通常情况下，keytab 文件位于 `/etc/hadoop.keytab`，而 principal 为 `hadoop/<hostname>@<REALM>` 格式。

#### Note

此解决方法将在 Kerberos 票证有效期内生效。默认情况下，此持续时间为 10 个小时，但可以通过 Kerberos 设置进行配置。Kerberos 票证过期后，您必须重新运行上述命令。

## 版本 5.25.0

以下发行说明包含有关亚马逊 5.25. EMR 0 版本的信息。更改与 5.24.1 有关。

首次发布日期：2019 年 7 月 17 日

上次更新日期：2019 年 10 月 30 日

亚马逊 EMR 5.25.0

升级

- AWS SDK for Java 1.11.566
- Hive 2.3.5
- Presto 0.220
- Spark 2.4.3
- TensorFlow 1.13.1
- Tez 0.9.2
- Zookeeper 3.4.14

## 新功能

- ( 2019 年 10 月 30 日 ) 从 Amazon EMR 版本 5.25.0 开始，您可以从集群摘要页面或控制台的“应用程序历史记录”选项卡连接到 Spark 历史服务器用户界面。您无需通过SSH连接设置 Web 代理，而是可以快速访问 Spark 历史服务器用户界面来查看应用程序指标并访问活动集群和已终止集群的相关日志文件。有关更多信息，请参阅 Amazon EMR 管理指南中的[集群外访问永久应用程序用户界面](#)。

## 更改、增强功能和解决的问题

- Spark
  - 通过使用 Bloom 筛选条件预筛选输入，提高了某些连接的性能。默认情况下，优化处于禁用状态，但可以通过以下方式启用：将 Spark 配置参数 `spark.sql.bloomFilterJoin.enabled` 设置为 `true`。
  - 改进了按字符串类型列分组的性能。
  - 改进了未HBase安装的集群的 R4 实例类型的默认 Spark 执行器内存和内核配置。
  - 解决了动态分区修剪功能之前存在的一个问题，即修剪的表必须位于联接的左侧。
  - DISTINCT在INTERSECT优化之前进行了改进，适用于其他涉及别名的情况。
  - 改进了JOIN后续DISTINCT查询的SQL计划统计数据推断。默认情况下，该改进处于禁用状态，但可以通过以下方式启用：将 Spark 配置参数 `spark.sql.statsImprovements.enabled` 设置为 `true`。此优化是“Distinct before Intersect”功能所需的，将 `spark.sql.optimizer.distinctBeforeIntersect.enabled` 设置为 `true` 时将自动启用。
  - 根据表格大小和筛选条件优化了联接顺序。默认情况下，该优化处于禁用状态，但可以通过以下方式启用：将 Spark 配置参数 `spark.sql.optimizer.sizeBasedJoinReorder.enabled` 设置为 `true`。

有关更多信息，请参阅[优化 Spark 性能](#)。

- EMRFS
  - 现在，默认情况下`fs.s3.buckets.create.enabled`，该EMRFS设置处于禁用状态。通过测试，我们发现禁用此设置可提高性能并可防止意外创建 S3 存储桶。如果您的应用程序需使用此功能，则可以通过以下方式启用：将 `emrfs-site` 配置分类中的 `fs.s3.buckets.create.enabled` 设置为 `true`。有关更多信息，请参阅[在创建集群时提供配置](#)。
- 安全配置中的本地磁盘加密和 S3 加密改进 ( 2019 年 8 月 5 日 )

- 在安全配置设置中将 Amazon S3 加密设置与本地磁盘加密设置分开。
- 添加了在 5.24.0 及更高版本中启用EBS加密的选项。选择此选项后，除了存储卷之外，还会加密根设备卷。以前的版本需要使用自定义AMI来加密根设备卷。
- 有关更多信息，请参阅 Amazon EMR 管理指南中的[加密选项](#)。

## 已知问题

- 具有多个主节点的集群和 Kerberos 身份验证中的已知问题

如果您在 Amazon 5.20.0 及更高EMR版本中运行具有多个主节点和 Kerberos 身份验证的集群，则在集群运行一段时间后，集群操作可能会遇到问题，例如缩小规模或步骤提交。具体时间段取决于您定义的 Kerberos 票证有效期。缩减问题会影响您提交的自动缩减和显式缩减请求。其它集群操作也可能受到影响。

解决办法：

- SSH以hadoop用户身份访问具有多个主节点的EMR群集的主节点。
- 运行以下命令，为 hadoop 用户续订 Kerberos 票证。

```
kinit -kt <keytab_file> <principal>
```

通常情况下，keytab 文件位于 `/etc/hadoop.keytab`，而 principal 为 `hadoop/<hostname>@<REALM>` 格式。

### Note

此解决方法将在 Kerberos 票证有效期内生效。默认情况下，此持续时间为 10 个小时，但可以通过 Kerberos 设置进行配置。Kerberos 票证过期后，您必须重新运行上述命令。

## 版本 5.24.1

以下发行说明包含有关亚马逊 5.24.1 EMR 版本的信息。更改与 5.24.0 有关。

首次发布日期：2019 年 6 月 26 日

## 更改、增强功能和解决的问题

- 更新了亚马逊的默认 Amazon Linux AMI，EMR使其包含重要的 Linux 内核安全更新，包括 TCPSACK拒绝服务问题 ([AWS-2019-005](#))。

## 已知问题

- 具有多个主节点的集群和 Kerberos 身份验证中的已知问题

如果您在 Amazon 5.20.0 及更高EMR版本中运行具有多个主节点和 Kerberos 身份验证的集群，则在集群运行一段时间后，集群操作可能会遇到问题，例如缩小规模或步骤提交。具体时间段取决于您定义的 Kerberos 票证有效期。缩减问题会影响您提交的自动缩减和显式缩减请求。其它集群操作也可能受到影响。

### 解决办法：

- SSH以hadoop用户身份访问具有多个主节点的EMR群集的主节点。
- 运行以下命令，为 hadoop 用户续订 Kerberos 票证。

```
kinit -kt <keytab_file> <principal>
```

通常情况下，keytab 文件位于 `/etc/hadoop.keytab`，而 principal 为 `hadoop/<hostname>@<REALM>` 格式。

### Note

此解决方法将在 Kerberos 票证有效期内生效。默认情况下，此持续时间为 10 个小时，但可以通过 Kerberos 设置进行配置。Kerberos 票证过期后，您必须重新运行上述命令。

## 版本 5.24.0

以下发行说明包含有关亚马逊 5.24. EMR 0 版本的信息。更改与 5.23.0 有关。

首次发布日期：2019 年 6 月 11 日

上次更新时间：2019 年 8 月 5 日

## 升级

- Flink 1.8.0
- Hue 4.4.0
- JupyterHub 0.9.6
- Livy 0.6.0
- MxNet 1.4.0
- Presto 0.219
- Spark 2.4.2
- AWS SDK for Java 1.11.546
- 连接器和驱动程序：
  - DynamoDB 连接器 4.9.0
  - MariaDB 连接器 2.4.1
  - 亚马逊 Redshift JDBC Driver 1.2.27.1051

## 更改、增强功能和解决的问题

- Spark
  - 添加了对动态修剪分区的优化。默认情况下禁用优化。要启用该优化，请将 Spark 参数 `spark.sql.dynamicPartitionPruning.enabled` 设置为 `true`。
  - 改进了 INTERSECT 查询的性能。默认情况下禁用此优化。要启用该优化，请将 Spark 参数 `spark.sql.optimizer.distinctBeforeIntersect.enabled` 设置为 `true`。
  - 添加了对展平标量子查询的优化，可使用相同关系进行聚合。默认情况下禁用优化。要启用该优化，请将 Spark 参数 `spark.sql.optimizer.flattenScalarSubqueriesWithAggregates.enabled` 设置为 `true`。
  - 改进了整个阶段代码生成。

有关更多信息，请参阅[优化 Spark 性能](#)。

- 安全配置中的本地磁盘加密和 S3 加密改进 ( 2019 年 8 月 5 日 )
  - 在安全配置设置中将 Amazon S3 加密设置与本地磁盘加密设置分开。
  - 添加了启用EBS加密的选项。选择此选项后，除了存储卷之外，还会加密根设备卷。以前的版本需要使用自定义AMI来加密根设备卷。

- 有关更多信息，请参阅 Amazon EMR 管理指南中的[加密选项](#)。

## 已知问题

- 具有多个主节点的集群和 Kerberos 身份验证中的已知问题

如果您在 Amazon 5.20.0 及更高EMR版本中运行具有多个主节点和 Kerberos 身份验证的集群，则在集群运行一段时间后，集群操作可能会遇到问题，例如缩小规模或步骤提交。具体时间段取决于您定义的 Kerberos 票证有效期。缩减问题会影响您提交的自动缩减和显式缩减请求。其它集群操作也可能受到影响。

### 解决办法：

- SSH以hadoop用户身份访问具有多个主节点的EMR群集的主节点。
- 运行以下命令，为 hadoop 用户续订 Kerberos 票证。

```
kinit -kt <keytab_file> <principal>
```

通常情况下，keytab 文件位于 /etc/hadoop.keytab，而 principal 为 hadoop/  
<hostname>@<REALM> 格式。

### Note

此解决方法将在 Kerberos 票证有效期内生效。默认情况下，此持续时间为 10 个小时，但可以通过 Kerberos 设置进行配置。Kerberos 票证过期后，您必须重新运行上述命令。

## 版本 5.23.0

以下发行说明包含有关亚马逊 5.23. EMR 0 版本的信息。更改与 5.22.0 有关。

首次发布日期：2019 年 4 月 1 日

上次更新时间：2019 年 4 月 30 日

## 升级

- AWS SDK for Java 1.11.519

## 新功能

- (2019年4月30日) 在 Amazon EMR 5.23.0 及更高版本中，您可以启动包含三个主节点的集群，以支持YARN资源管理器、Spark HDFS NameNode、Hive 和 Ganglia 等应用程序的高可用性。使用此功能，主节点不再发生潜在的单点故障。如果其中一个主节点出现故障，Amazon EMR 会自动故障转移到备用主节点，并将故障主节点替换为具有相同配置和引导操作的新主节点。有关更多信息，请参阅[计划和配置主节点](#)。

## 已知问题

- Tez 用户界面 (已在亚马逊EMR版本 5.26.0 中修复)

Tez UI 不适用于具有多个主节点的EMR集群。

- Hue (已在亚马逊 5.24. EMR 0 版本中修复)

- 在亚马逊上运行的 Hue EMR 不支持 Solr。从 Amazon 5.20.0 EMR 版本开始，配置错误问题会导致 Solr 被启用，并显示类似于以下内容的无害错误消息：

```
Solr server could not be contacted properly:
HTTPConnectionPool('host=ip-xx-xx-xx-xx.ec2.internal',
port=1978): Max retries exceeded with url: /solr/admin/info/
system?user.name=hue&doAs=administrator&wt=json (Caused by
NewConnectionError(': Failed to establish a new connection: [Errno 111]
Connection refused',))
```

要防止显示 Solr 错误消息：

1. 使用连接到主节点命令行SSH。
2. 使用文本编辑器打开 hue.ini 文件。例如：

```
sudo vim /etc/hue/conf/hue.ini
```

3. 搜索术语 appblacklist，并将该行修改为以下内容：

```
appblacklist = search
```

4. 保存更改并重新启动 Hue，如以下示例所示：

```
sudo stop hue; sudo start hue
```

- 具有多个主节点的集群和 Kerberos 身份验证中的已知问题

如果您在 Amazon 5.20.0 及更高EMR版本中运行具有多个主节点和 Kerberos 身份验证的集群，则在集群运行一段时间后，集群操作可能会遇到问题，例如缩小规模或步骤提交。具体时间段取决于您定义的 Kerberos 票证有效期。缩减问题会影响您提交的自动缩减和显式缩减请求。其它集群操作也可能受到影响。

解决办法：

- SSH以hadoop用户身份访问具有多个主节点的EMR群集的主节点。
- 运行以下命令，为 hadoop 用户续订 Kerberos 票证。

```
kinit -kt <keytab_file> <principal>
```

通常情况下，keytab 文件位于 /etc/hadoop.keytab，而 principal 为 hadoop/<hostname>@<REALM> 格式。

#### Note

此解决方法将在 Kerberos 票证有效期内生效。默认情况下，此持续时间为 10 个小时，但可以通过 Kerberos 设置进行配置。Kerberos 票证过期后，您必须重新运行上述命令。

## 版本 5.22.0

以下发行说明包含有关亚马逊 5.22. EMR 0 版本的信息。更改与 5.21.0 有关。

#### Important

从亚马逊EMR版本 5.22.0 开始，亚马逊仅EMR使用 AWS 签名版本 4 来验证向亚马逊 S3 发出的请求。早期的 Amazon EMR 版本在某些情况下使用 AWS 签名版本 2，除非发行说明中注明仅使用签名版本 4。有关更多信息，请参阅《[亚马逊简单存储服务开发者AWS 指南](#)》中的[对请求进行身份验证 \( AWS 签名版本 4 \)](#)和[对请求进行身份验证 \( 签名版本 2 \)](#)。

首次发布日期：2019 年 3 月 20 日

升级

- Flink 1.7.1



- HBase1.4.9
- Oozie 5.1.0
- Phoenix 4.14.1
- Zeppelin 0.8.1
- 连接器和驱动程序：
  - DynamoDB 连接器 4.8.0
  - MariaDB 连接器 2.2.6
  - 亚马逊 Redshift JDBC Driver 1.2.20.1043

## 新功能

- 修改了EBS仅限存储空间的EC2实例类型的默认EBS配置。当您使用 Amazon EMR 版本 5.22.0 及更高版本创建集群时，默认EBS存储量会根据实例的大小而增加。此外，我们将增加的存储空间分配给多个卷，从而提高了IOPS性能。如果要使用不同的EBS实例存储配置，可以在创建EMR集群或向现有集群添加节点时指定该配置。有关默认为每种实例类型分配的存储量和卷数量的更多信息，请参阅 Amazon EMR 管理指南中的[实例默认EBS存储](#)。

## 更改、增强功能和解决的问题

- Spark
  - 在 Spark 上YARN引入了一个新的配置属性 `spark.yarn.executor.memoryOverheadFactor`。此属性的值是一个缩放系数，它将内存开销值设置为执行程序内存的百分比，最小为 384 MB。如果内存开销设置为使用 `spark.yarn.executor.memoryOverhead`，则此属性不发挥任何作用。默认值为 0.1875，表示 18.75%。与Spark内部设置的10%默认值相比，Amazon的默认值在YARN容器中为执行程序的内存开销EMR留出了更多的空间。根据经验，Amazon的EMR默认值为18.75%，表明-DS基准测试中与内存相关的故障较少。TPC
  - 向后移植了 [SPARK-26316](#) 以提高性能。
- 在亚马逊EMR版本 5.19.0、5.20.0 和 5.21.0 中，YARN节点标签存储在目录中。HDFS在某些情况下，这会导致核心节点启动延迟，然后导致集群超时和启动失败。从 Amazon EMR 5.22.0 开始，此问题已得到解决。YARN节点标签存储在每个群集节点的本地磁盘上，避免依赖于HDFS。

## 已知问题

- Hue (已在亚马逊 5.24. EMR 0 版本中修复)

- 在亚马逊上运行的 Hue EMR 不支持 Solr。从 Amazon 5.20.0 EMR 版本开始，配置错误问题会导致 Solr 被启用，并显示类似于以下内容的无害错误消息：

```
Solr server could not be contacted properly:
HTTPConnectionPool('host=ip-xx-xx-xx-xx.ec2.internal',
port=1978): Max retries exceeded with url: /solr/admin/info/
system?user.name=hue&doAs=administrator&wt=json (Caused by
NewConnectionError(': Failed to establish a new connection: [Errno 111]
Connection refused',))
```

要防止显示 Solr 错误消息：

1. 使用连接到主节点命令行SSH。
2. 使用文本编辑器打开 hue.ini 文件。例如：

```
sudo vim /etc/hue/conf/hue.ini
```

3. 搜索术语 appblacklist，并将该行修改为以下内容：

```
appblacklist = search
```

4. 保存更改并重新启动 Hue，如以下示例所示：

```
sudo stop hue; sudo start hue
```

- 具有多个主节点的集群和 Kerberos 身份验证中的已知问题

如果您在 Amazon 5.20.0 及更高EMR版本中运行具有多个主节点和 Kerberos 身份验证的集群，则在集群运行一段时间后，集群操作可能会遇到问题，例如缩小规模或步骤提交。具体时间段取决于您定义的 Kerberos 票证有效期。缩减问题会影响您提交的自动缩减和显式缩减请求。其它集群操作也可能受到影响。

解决办法：

- SSH以hadoop用户身份访问具有多个主节点的EMR群集的主节点。
- 运行以下命令，为 hadoop 用户续订 Kerberos 票证。

```
kinit -kt <keytab file> <principal>
```

通常情况下，keytab 文件位于 `/etc/hadoop.keytab`，而 principal 为 `hadoop/<hostname>@<REALM>` 格式。

#### Note

此解决方法将在 Kerberos 票证有效期内生效。默认情况下，此持续时间为 10 个小时，但可以通过 Kerberos 设置进行配置。Kerberos 票证过期后，您必须重新运行上述命令。

## 发布版本 5.21.1

以下发行说明包含有关亚马逊 5.21.1 EMR 版本的信息。更改与 5.21.0 有关。

首次发布日期：2019 年 7 月 18 日

### 更改、增强功能和解决的问题

- 更新了亚马逊的默认 Amazon L AMI linux，EMR使其包含重要的 Linux 内核安全更新，包括 TCPSACK拒绝服务问题 ([AWS-2019-005](#))。

### 已知问题

- 具有多个主节点的集群和 Kerberos 身份验证中的已知问题

如果您在 Amazon 5.20.0 及更高EMR版本中运行具有多个主节点和 Kerberos 身份验证的集群，则在集群运行一段时间后，集群操作可能会遇到问题，例如缩小规模或步骤提交。具体时间段取决于您定义的 Kerberos 票证有效期。缩减问题会影响您提交的自动缩减和显式缩减请求。其它集群操作也可能受到影响。

### 解决办法：

- SSH以hadoop用户身份访问具有多个主节点的EMR群集的主节点。
- 运行以下命令，为 hadoop 用户续订 Kerberos 票证。

```
kinit -kt <keytab_file> <principal>
```

通常情况下，keytab 文件位于 `/etc/hadoop.keytab`，而 principal 为 `hadoop/<hostname>@<REALM>` 格式。

**Note**

此解决方法将在 Kerberos 票证有效期内生效。默认情况下，此持续时间为 10 个小时，但可以通过 Kerberos 设置进行配置。Kerberos 票证过期后，您必须重新运行上述命令。

## 版本 5.21.0

以下发行说明包含有关亚马逊 5.21. EMR 0 版本的信息。更改与 5.20.0 有关。

首次发布日期：2019 年 2 月 18 日

上次更新时间：2019 年 4 月 3 日

### 升级

- Flink 1.7.0
- Presto 0.215
- AWS SDK for Java 1.11.479

### 新功能

- ( 2019 年 4 月 3 日 ) 在 Amazon 5.21.0 及更高 EMR 版本中，您可以覆盖集群配置，并为正在运行的集群中的每个实例组指定其他配置分类。您可以使用 Amazon EMR 控制台、AWS Command Line Interface (AWS CLI) 或 AWS SDK。有关更多信息，请参阅[为运行的集群中的实例组提供配置](#)。

### 更改、增强功能和解决的问题

- Zeppelin
  - 向后移植 [ZEPPELIN-3878](#)。

### 已知问题

- Hue ( 已在亚马逊 5.24. EMR 0 版本中修复 )
  - 在亚马逊上运行的 Hue EMR 不支持 Solr。从 Amazon 5.20.0 EMR 版本开始，配置错误问题会导致 Solr 被启用，并显示类似于以下内容的无害错误消息：

```
Solr server could not be contacted properly:
HTTPConnectionPool('host=ip-xx-xx-xx-xx.ec2.internal',
port=1978): Max retries exceeded with url: /solr/admin/info/
system?user.name=hue&doAs=administrator&wt=json (Caused by
NewConnectionError(': Failed to establish a new connection: [Errno 111]
Connection refused',))
```

要防止显示 Solr 错误消息:

1. 使用连接到主节点命令行SSH。
2. 使用文本编辑器打开 hue.ini 文件。例如：

```
sudo vim /etc/hue/conf/hue.ini
```

3. 搜索术语 appblacklist，并将该行修改为以下内容：

```
appblacklist = search
```

4. 保存更改并重新启动 Hue，如以下示例所示：

```
sudo stop hue; sudo start hue
```

- Tez

- 此问题已在亚马逊 EMR 5.22.0 中修复。

当你通过 `http://` 连接到 Tez 用户界面时 `MasterDNS:8080/tez-ui` 通过SSH连接到群集主节点，出现错误“适配器操作失败-时间轴服务器 () ATS 遥不可及。要么已关闭，要么CORS出现“未启用”，要么任务意外显示 N/A。

这是由于 Tez UI 使用 `localhost` 而不是主节点的主机名向 YARN 时间轴服务器发出请求所致。解决方法：将脚本作为引导操作或步骤运行。脚本更新 Tez `configs.env` 文件中的主机名。有关更多信息以及脚本的位置信息，请参阅[引导说明](#)。

- 在亚马逊EMR版本 5.19.0、5.20.0 和 5.21.0 中，YARN节点标签存储在目录中。HDFS在某些情况下，这会导致核心节点启动延迟，然后导致集群超时和启动失败。从 Amazon EMR 5.22.0 开始，此问题已得到解决。YARN节点标签存储在每个群集节点的本地磁盘上，避免依赖于HDFS。
- 具有多个主节点的集群和 Kerberos 身份验证中的已知问题

如果您在 Amazon 5.20.0 及更高EMR版本中运行具有多个主节点和 Kerberos 身份验证的集群，则在集群运行一段时间后，集群操作可能会遇到问题，例如缩小规模或步骤提交。具体时间段取决于您

定义的 Kerberos 票证有效期。缩减问题会影响您提交的自动缩减和显式缩减请求。其它集群操作也可能受到影响。

解决办法：

- SSH以hadoop用户身份访问具有多个主节点的EMR群集的主节点。
- 运行以下命令，为 hadoop 用户续订 Kerberos 票证。

```
kinit -kt <keytab_file> <principal>
```

通常情况下，keytab 文件位于 /etc/hadoop.keytab，而 principal 为 hadoop/<hostname>@<REALM> 格式。

#### Note

此解决方法将在 Kerberos 票证有效期内生效。默认情况下，此持续时间为 10 个小时，但可以通过 Kerberos 设置进行配置。Kerberos 票证过期后，您必须重新运行上述命令。

## 版本 5.20.0

以下发行说明包含有关亚马逊 5.20. EMR 0 版本的信息。更改与 5.19.0 有关。

首次发布日期：2018 年 12 月 18 日

上次更新时间：2019 年 1 月 22 日

升级

- Flink 1.6.2
- HBase1.4.8
- Hive 2.3.4
- Hue 4.3.0
- MXNet1.3.1
- Presto 0.214
- Spark 2.4.0
- TensorFlow 1.12.0

- Tez 0.9.1
- AWS SDK for Java 1.11.461

## 新功能

- (2019年1月22日) 亚马逊中的 Kerberos EMR 已得到改进，支持从外部对委托人进行身份验证。KDC这可以集中管理委托人，因为多个集群可以共享一个外部KDC集群。此外，外部KDC用户可与 Active Directory 域建立跨领域信任。这使得所有集群可以从 Active Directory 对委托人进行身份验证。有关更多信息，请参阅《亚马逊EMR管理指南》中的[“使用 Kerberos 身份验证”](#)。

## 更改、增强功能和解决的问题

- 亚马逊的默认 Amazon AMI Linux EMR
  - Python 3 软件包已从 Python 3.4 升级到 3.6。
- 经过 EMRFS S3 优化的提交者
  - EMRFS S3 优化的提交器现在默认处于启用状态，这提高了写入性能。有关更多信息，请参阅[使用 EMRFS S3 优化的提交器](#)。
- Hive
  - 向后移植 [HIVE-16686](#)。
- 集成 Spark 和 Hive 的 Glue
  - 在 EMR 5.20.0 或更高版本中，当使用 Glue Data Catalog 作为元数据存储时，Spark 和 Hive 会自动启用并行分区修剪。此更改通过并行执行多个请求来检索分区，显著缩短查询计划时间。可同时执行的分段总数介于 1 到 10 之间。默认值为 5，这是建议的设置。您可以通过以下方式更改该值：指定 hive-site 配置分类中的属性 `aws.glue.partition.num.segments`。如果发生节流，则可以通过将值更改为 1 来关闭此功能。有关更多信息，请参阅[AWS Glue 分段结构](#)。

## 已知问题

- Hue (已在亚马逊 5.24. EMR 0 版本中修复)
  - 在亚马逊上运行的 Hue EMR 不支持 Solr。从 Amazon 5.20.0 EMR 版本开始，配置错误问题会导致 Solr 被启用，并显示类似于以下内容的无害错误消息：

```
Solr server could not be contacted properly:
HTTPConnectionPool('host=ip-xx-xx-xx-xx.ec2.internal',
```

```
port=1978): Max retries exceeded with url: /solr/admin/info/system?user.name=hue&doAs=administrator&wt=json (Caused by NewConnectionError(': Failed to establish a new connection: [Errno 111] Connection refused',))
```

要防止显示 Solr 错误消息:

1. 使用连接到主节点命令行SSH。
2. 使用文本编辑器打开 hue.ini 文件。例如：

```
sudo vim /etc/hue/conf/hue.ini
```

3. 搜索术语 appblacklist，并将该行修改为以下内容：

```
appblacklist = search
```

4. 保存更改并重新启动 Hue，如以下示例所示：

```
sudo stop hue; sudo start hue
```

- Tez

- 此问题已在亚马逊 EMR 5.22.0 中修复。

当你通过 `http://` 连接到 Tez 用户界面时 `MasterDNS:8080/tez-ui` 通过SSH连接到群集主节点，出现错误“适配器操作失败-时间轴服务器 () ATS 遥不可及。要么已关闭，要么CORS出现“未启用”，要么任务意外显示 N/A。

这是由于 Tez UI 使用 `localhost` 而不是主节点的主机名向 YARN 时间轴服务器发出请求所致。解决方法：将脚本作为引导操作或步骤运行。脚本更新 Tez `configs.env` 文件中的主机名。有关更多信息以及脚本的位置信息，请参阅[引导说明](#)。

- 在亚马逊EMR版本 5.19.0、5.20.0 和 5.21.0 中，YARN节点标签存储在目录中。HDFS在某些情况下，这会导致核心节点启动延迟，然后导致集群超时和启动失败。从 Amazon EMR 5.22.0 开始，此问题已得到解决。YARN节点标签存储在每个群集节点的本地磁盘上，避免依赖于HDFS。
- 具有多个主节点的集群和 Kerberos 身份验证中的已知问题

如果您在 Amazon 5.20.0 及更高EMR版本中运行具有多个主节点和 Kerberos 身份验证的集群，则在集群运行一段时间后，集群操作可能会遇到问题，例如缩小规模或步骤提交。具体时间段取决于您定义的 Kerberos 票证有效期。缩减问题会影响您提交的自动缩减和显式缩减请求。其它集群操作也可能受到影响。



### 解决办法：

- SSH以hadoop用户身份访问具有多个主节点的EMR群集的主节点。
- 运行以下命令，为 hadoop 用户续订 Kerberos 票证。

```
kinit -kt <keytab_file> <principal>
```

通常情况下，keytab 文件位于 /etc/hadoop.keytab，而 principal 为 hadoop/<hostname>@<REALM> 格式。

#### Note

此解决方法将在 Kerberos 票证有效期内生效。默认情况下，此持续时间为 10 个小时，但可以通过 Kerberos 设置进行配置。Kerberos 票证过期后，您必须重新运行上述命令。

## 版本 5.19.0

以下发行说明包含有关亚马逊 5.19. EMR 0 版本的信息。更改与 5.18.0 有关。

首次发布日期：2018 年 11 月 7 日

上次更新时间：2018 年 11 月 19 日

### 升级

- Hadoop 2.8.5
- Flink 1.6.1
- JupyterHub 0.9.4
- MXNet1.3.0
- Presto 0.212
- TensorFlow 1.11.0
- Zookeeper 3.4.13
- AWS SDK for Java 1.11.433

## 新功能

- (2018年11月19日) EMR笔记本是一个基于 Jupyter Notebook 的托管环境。它支持 Spark PySpark、Spark R 和 Scala 的 Spark 魔法内核。SQLEMR笔记本可以与使用 Amazon EMR 版本 5.18.0 及更高版本创建的集群一起使用。有关更多信息，请参阅 Amazon EMR 管理指南中的[使用 EMR笔记本](#)。
- 使用 Spark 和编写 Parquet 文件时，可以使用 EMRFS S3 优化的提交器。EMRFS此提交程序改进了写入性能。有关更多信息，请参阅[使用 EMRFS S3 优化的提交器](#)。

## 更改、增强功能和解决的问题

- YARN
  - 修改了限制应用程序主进程在核心节点上运行的逻辑。此功能现在使用YARN节点标签功能和yarn-site和capacity-scheduler配置分类中的属性。有关信息，请参阅。<https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-plan-instances-guidelines.html#emr-plan-spot-YARN>。
- 亚马逊的默认 Amazon AMI Linux EMR
  - 默认情况下，不再安装 ruby18、php56 和 gcc48。如果需要，可以使用 yum 安装它们。
  - 默认情况下，不再安装 aws-sdk ruby gem。如果需要，可以使用 gem install aws-sdk 进行安装。此外，还可以安装特定组件。例如，gem install aws-sdk-s3。

## 已知问题

- EMR笔记本-在某些情况下，当多个笔记本编辑器处于打开状态时，笔记本编辑器可能无法连接到集群。如果发生这种情况，请清除浏览器 Cookie，然后重新打开笔记本编辑器。
- CloudWatch ContainerPending 指标和自动缩放 — (在 5.20.0 中已修复) Amazon EMR 可能会发出负值。ContainerPending如果在自动伸缩规则中使用 ContainerPending，自动伸缩的行为方式可能会不符合预期。请避免在自动伸缩中使用 ContainerPending。
- 在亚马逊EMR版本 5.19.0、5.20.0 和 5.21.0 中，YARN节点标签存储在目录中。HDFS在某些情况下，这会导致核心节点启动延迟，然后导致集群超时和启动失败。从 Amazon EMR 5.22.0 开始，此问题已得到解决。YARN节点标签存储在每个群集节点的本地磁盘上，避免依赖于HDFS。

## 版本 5.18.0

以下发行说明包含有关亚马逊 5.18.0 EMR 版本的信息。更改与 5.17.0 有关。

首次发布日期：2018 年 10 月 24 日

## 升级

- Flink 1.6.0
- HBase1.4.7
- Presto 0.210
- Spark 2.3.2
- Zeppelin 0.8.0

## 新功能

- 从 Amazon EMR 5.18.0 开始，您可以使用 Amazon 构EMR件存储库针对特定亚马逊版本中提供的库和依赖项的确切版本构建任务代码。EMR有关更多信息，请参阅 [使用 Amazon EMR 项目存储库检查依赖项](#)。

## 更改、增强功能和解决的问题

- Hive
  - 添加了对 S3 Select 的支持。有关更多信息，请参阅[将 S3 Select 与 Hive 结合使用以提高查询性能](#)。
- Presto
  - 添加了对 [S3 Select](#) Pushdown 的支持。有关更多信息，请参阅[使用 S3 Select Pushdown 搭配 Presto 提高性能](#)。
- Spark
  - Spark 的默认 log4j 配置已更改为 Spark Streaming 任务每小时的滚动容器日志。这有助于防止删除长时间运行的 Spark Streaming 任务的日志。

## 发布版本 5.17.1

以下发行说明包含有关亚马逊 5.17.1 EMR 版本的信息。更改与 5.17.0 有关。

首次发布日期：2019 年 7 月 18 日

## 更改、增强功能和解决的问题

- 更新了亚马逊的默认 Amazon L AMI linux，EMR使其包含重要的 Linux 内核安全更新，包括 TCPSACK拒绝服务问题 ([AWS-2019-005](#))。

## 版本 5.17.0

以下发行说明包含有关亚马逊 5.17. EMR 0 版本的信息。更改与 5.16.0 有关。

首次发布日期：2018 年 8 月 30 日

### 升级

- Flink 1.5.2
- HBase1.4.6
- Presto 0.206

### 新功能

- 添加了对 Tensorflow 的支持。有关更多信息，请参阅[TensorFlow](#)。

## 更改、增强功能和解决的问题

- JupyterHub
  - Amazon S3 中添加了对笔记本持久性的支持。有关更多信息，请参阅[在 Amazon S3 中配置笔记本的持久性](#)。
- Spark
  - 添加了对 [S3 Select](#) 的支持。有关更多信息，请参阅 [将 S3 Select 与 Spark 结合使用以提高查询性能](#)。
- 解决了亚马逊EMR版本 5.14.0、5.15.0 或 5.16.0 中的 Cloudwatch 指标和自动扩展功能的问题。

### 已知问题

- 创建使用 Kerberos 的集群时，如果安装了 Livy，Livy 将失败，并显示未启用简单身份验证的错误。重新启动 Livy 服务器可解决此问题。解决方法是在集群创建过程中添加一个在主节点上运行 `sudo restart livy-server` 的步骤。

- 如果你使用AMI基于亚马逊 Linux 的自定义 Amazon LinuxAMI，创建日期为 2018-08-11，Oozie 服务器将无法启动。如果你使用 Oozie，请AMI根据创建日期不同的亚马逊 Linux AMI ID 创建自定义账号。您可以使用以下 AWS CLI 命令返回所有 HVM Amazon Linux AMIs 版本为 2018.03 的图片 IDs列表以及发布日期，以便您可以选择合适的 Amazon Linux AMI 作为基础。MyRegion 替换为您的地区标识符，例如 us-west-2。

```
aws ec2 --region MyRegion describe-images --owner amazon --query 'Images[?Name!=`null`][[?starts_with(Name, `amzn-ami-hvm-2018.03`) == `true`]. [CreationDate,ImageId,Name]' --output text | sort -rk1
```

## 版本 5.16.0

以下发行说明包含有关亚马逊 5.16. EMR 0 版本的信息。更改与 5.15.0 有关。

首次发布日期：2018 年 7 月 19 日

### 升级

- Hadoop 2.8.4
- Flink 1.5.0
- Livy 0.5.0
- MXNet1.2.0
- Phoenix 4.14.0
- Presto 0.203
- Spark 2.3.1
- AWS SDK for Java 1.11.336
- CUDA9.2
- Redshift JDBC Driver 1.2.15.1025

### 更改、增强功能和解决的问题

- HBase
  - [向后移植 HBASE -20723](#)
- Presto

- 配置更改为支持LDAP身份验证。有关更多信息，请参阅 [在亚马逊上使用 Presto 的LDAP身份验证 EMR](#)。
- Spark
  - [Apache Spark 版本 2.3.1](#)，从亚马逊EMR版本 5.16.0 开始推出，地址CVE为 [-2018-8024](#) 和 [-2018-1334](#)。CVE建议您将 Spark 的早期版本迁移到 Spark 2.3.1 版本或更高版本。

## 已知问题

- 此发行版不支持 c1.medium 或 m1.small 实例类型。使用这些实例类型的集群将无法启动。解决方法：指定其它实例类型或使用其它发行版。
- 创建使用 Kerberos 的集群时，如果安装了 Livy，Livy 将失败，并显示未启用简单身份验证的错误。重新启动 Livy 服务器可解决此问题。解决方法是在集群创建过程中添加一个在主节点上运行 `sudo restart livy-server` 的步骤。
- 在主节点重启或实例控制器重新启动后，将不会收集 CloudWatch 指标，并且自动扩展功能将在 Amazon EMR 版本 5.14.0、5.15.0 或 5.16.0 中不可用。此问题已在亚马逊 EMR 5.17.0 中修复。

## 版本 5.15.0

以下发行说明包含有关亚马逊 5.15. EMR 0 版本的信息。更改与 5.14.0 有关。

首次发布日期：2018 年 6 月 21 日

### 升级

- 已升级HBase到 1.4.4
- 已将 Hive 升级到 2.3.3
- 已将 Hue 升级到 4.2.0
- 已将 Oozie 升级到 5.0.0
- 已将 Zookeeper 升级到 3.4.12
- 已升级 AWS SDK到 1.11.333

### 更改、增强功能和解决的问题

- Hive
  - [向后移植 HIVE -18069](#)

- Hue
  - 更新了 Hue，启用 Kerberos 后可以使用 Livy 正确地进行身份验证。在亚马逊上使用 Kerberos 时，现在支持 Livy。EMR
- JupyterHub
  - 已更新，以 JupyterHub 便 Amazon 默认EMR安装LDAP客户端库。
  - 修复了生成自签名凭证的脚本中的错误。

## 已知问题

- 此发行版不支持 c1.medium 或 m1.small 实例类型。使用这些实例类型的集群将无法启动。解决方法：指定其它实例类型或使用其它发行版。
- 在主节点重启或实例控制器重新启动后，将不会收集 CloudWatch 指标，并且自动扩展功能将在 Amazon EMR 版本 5.14.0、5.15.0 或 5.16.0 中不可用。此问题已在亚马逊 EMR 5.17.0 中修复。

## 版本 5.14.1

以下发行说明包含有关亚马逊 5.14.1 EMR 版本的信息。更改与 5.14.0 有关。

首次发布日期：2018 年 10 月 17 日

更新了 Amazon AMI 的默认设置EMR以解决潜在的安全漏洞。

## 版本 5.14.0

以下发行说明包含有关亚马逊 5.14. EMR 0 版本的信息。更改与 5.13.0 有关。

首次发布日期：2018 年 6 月 4 日

## 升级

- 已将 Apache Flink 升级到 1.4.2
- 将 Apache 升级MXnet到 1.1.0
- 已将 Apache Sqoop 升级到 1.4.7

## 新功能

- 增加了 JupyterHub 支持。有关更多信息，请参阅 [JupyterHub](#)。

## 更改、增强功能和解决的问题

- EMRFS
  - 对 Amazon S3 的请求中的 userAgent 字符串已更新为包含调用委托人的用户和组信息。这可以与 AWS CloudTrail 日志一起使用，以实现更全面的请求跟踪。
- HBase
  - 包括 [HBASE-20447](#)，它解决了可能导致缓存问题，尤其是在拆分区域的情况下。
- MXnet
  - 新增了 OpenCV 库。
- Spark
  - 当 Spark 使用将 Parquet 文件写入 Amazon S3 位置时 EMRFS，FileOutputCommitter 算法已更新为使用版本 2 而不是版本 1。这将减少重命名的数量，从而提高应用程序性能。此更改不会影响：
    - Spark 以外的应用程序。
    - 写入其他文件系统的应用程序，例如 HDFS（仍使用版本 1 FileOutputCommitter）。
    - 使用其他输出格式（例如文本或 csv）且已经使用 EMRFS 直接写入的应用程序。

## 已知问题

- JupyterHub
  - 不支持在创建集群时使用配置分类来设置 JupyterHub 和单个 Jupyter 笔记本。手动编辑每个用户的 jupyterhub\_config.py 文件和 jupyter\_notebook\_config.py 文件。有关更多信息，请参阅 [正在配置 JupyterHub](#)。
  - JupyterHub 无法在私有子网内的群集上启动，失败并显示消息 `Error: ENOENT: no such file or directory, open '/etc/jupyter/conf/server.crt'`。这由生成自签名凭证的脚本中的错误所致。使用以下解决方法生成自签名凭证。在连接到主节点时执行所有命令。
    1. 将凭证生成脚本从容器复制到主节点：

```
sudo docker cp jupyterhub:/tmp/gen_self_signed_cert.sh ./
```

2. 使用文本编辑器更改第 23 行，将公有主机名更改为本地主机名，如下所示：

```
local hostname=$(curl -s $EC2_METADATA_SERVICE_URI/local-hostname)
```

## 3. 运行脚本，生成自签名凭证：



```
sudo bash ./gen_self_signed_cert.sh
```

4. 将脚本生成的凭证文件移至 `/etc/jupyter/conf/` 目录：

```
sudo mv /tmp/server.crt /tmp/server.key /etc/jupyter/conf/
```

您可以通过 `tail` 该 `jupyter.log` 文件来验证是否已 JupyterHub 重新启动并且正在返回 200 响应码。例如：

```
tail -f /var/log/jupyter/jupyter.log
```

该命令应返回与以下示例类似的响应：

```
# [I 2018-06-14 18:56:51.356 JupyterHub app:1581] JupyterHub is now running at
https://:9443/
# 19:01:51.359 - info: [ConfigProxy] 200 GET /api/routes
```

- 在主节点重启或实例控制器重新启动后，将不会收集 CloudWatch 指标，并且自动扩展功能将在 Amazon EMR 版本 5.14.0、5.15.0 或 5.16.0 中不可用。此问题已在亚马逊 EMR 5.17.0 中修复。

## 版本 5.13.0

以下发行说明包含有关亚马逊 5.13.0 EMR 版本的信息。更改与 5.12.0 有关。

### 升级

- 已将 Spark 升级到 2.3.0
- 已升级 HBase 到 1.4.2
- 已将 Presto 升级到 0.194
- 已升级 AWS SDK for Java 到 1.11.297

### 更改、增强功能和解决的问题

- Hive
  - 向后移植 [HIVE-15436](#)。增强了 HiveAPIs，仅返回视图。

## 已知问题

- MXNet目前没有 OpenCV 库。

## 版本 5.12.2

以下发行说明包含有关亚马逊 5.12.2 EMR 版本的信息。更改与 5.12.1 有关。

首次发布日期：2018年 8 月 29 日

更改、增强功能和解决的问题

- 此版本解决了潜在的安全漏洞。

## 版本 5.12.1

以下发行说明包含有关亚马逊 5.12.1 EMR 版本的信息。更改与 5.12.0 有关。

首次发布日期：2018 年 3 月 29 日

更改、增强功能和解决的问题

- 更新了亚马逊 Linux 版的亚马逊 defaultAmazon Linux AMI 内核EMR以修复潜在漏洞。

## 版本 5.12.0

以下发行说明包含有关亚马逊 5.12.0 EMR 版本的信息。更改与 5.11.1 有关。

升级

- AWS SDK适用于 Java 1.11.238 ⇒ 1.11.267。有关更多信息，请参阅适用于 [Java AWS SDK 的 Change Log](#) on GitHub。
- Hadoop 2.7.3 升级到 2.8.3。有关更多信息，请参阅 [Apache Hadoop 发行版](#)。
- Fink 1.3.2 升级到 1.4.0。有关详细信息，请参阅 [Apache Flink 1.4.0 版本公告](#)。
- HBase1.3.1 ⇒ 1.4.0。有关更多信息，请参阅[HBase发布公告](#)。
- Hue 4.0.1 升级到 4.1.0。有关更多信息，请参阅[发布说明](#)。
- MxNet 0.12.0 ⇒ 1.0.0。有关更多信息，请参阅“[MXNet更改日志](#)” GitHub。
- Presto 0.187 升级到 0.188。有关更多信息，请参阅[发布说明](#)。

## 更改、增强功能和解决的问题

- Hadoop
  - `yarn.resourcemanager.decommissioning.timeout` 属性已更改为 `yarn.resourcemanager.nodemanager-graceful-decommission-timeout-secs`。您可以使用此属性自定义集群缩减。有关更多信息，请参阅 Amazon EMR 管理指南中的[集群扩容](#)。
  - Hadoop 在 `cp` (复制) 命令中 CLI 添加了该 `-d` 选项，该命令指定了直接复制。可以使用它来避免创建中间 `.COPYING` 文件，这加快了在 Amazon S3 之间复制数据的速度。有关更多信息，请参阅[HADOOP-12384](#)。
- Pig
  - 添加了 `pig-env` 配置分类，这简化了 Pig 环境属性的配置。有关更多信息，请参阅[配置应用程序](#)。
- Presto
  - 新增 `presto-connector-redshift` 配置分类，您可以将其用于配置 Presto `redshift.properties` 配置文件中的值。有关更多信息，请参阅 Presto 文档中[Redshift 连接器](#)以及[配置应用程序](#)。
  - EMRFS 已添加对 Presto 的支持，这是默认配置。亚马逊早期 EMR 版本使用的是 `PrestoS3FileSystem`，这是唯一的选择。有关更多信息，请参阅[EMRFS 和 Presto FileSystem S3 配置](#)。

### Note

如果您使用亚马逊 EMR 版本 5.12.0 在 Amazon S3 中查询基础数据，则可能会出现 Presto 错误。这是因为 Presto 无法从 `emrfs-site.xml` 提取配置分类值。解决方法是在 `usr/lib/presto/plugin/hive-hadoop2/` 下创建一个 `emrfs` 子目录，并在 `usr/lib/presto/plugin/hive-hadoop2/emrfs` 中创建一个指向现有 `/usr/share/aws/emr/emrfs/conf/emrfs-site.xml` 文件的符号链接。然后重新启动 `presto-server` 进程 ( 首先执行 `sudo presto-server stop`，然后执行 `sudo presto-server start` )。

- Spark
  - 向后移植的[SPARK-22036 : BigDecimal 乘法有时](#)会返回空值。

## 已知问题

- MXNet 不包括 OpenCV 库。

- SparkR 不适用于使用自定义创建的集群，因为默认情况下 AMI R 未安装在群集节点上。

## 发布版本 5.11.3

以下发行说明包含有关亚马逊 5.11.3 EMR 版本的信息。更改与 5.11.2 有关。

首次发布日期：2019 年 7 月 18 日

更改、增强功能和解决的问题

- 更新了亚马逊的默认 Amazon L AMI inux，EMR使其包含重要的 Linux 内核安全更新，包括 TCPSACK拒绝服务问题 ([AWS-2019-005](#))。

## 版本 5.11.2

以下发行说明包含有关亚马逊 5.11.2 EMR 版本的信息。更改与 5.11.1 有关。

首次发布日期：2018年 8 月 29 日

更改、增强功能和解决的问题

- 此版本解决了潜在的安全漏洞。

## 版本 5.11.1

以下发行说明包含 Amazon EMR 版本 5.11.1 版本的信息。更改与亚马逊 EMR 5.11.0 版本有关。

首次发布日期：2018 年 1 月 22 日

更改、增强功能和解决的问题

- 更新了亚马逊版 Linux 的亚马逊 L defaultAmazon inux AMI 内核，EMR以解决与推测执行相关的漏洞 ( CVE-2017-5715、-2017-5753 和 -2017 CVE -5754 )。CVE有关更多信息，请参阅 <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>。

## 已知问题

- MXNet不包括 OpenCV 库。

- 默认情况下，Hive 2.3.2 设置 `hive.compute.query.using.stats=true`。这会导致查询从现有统计数据而不是直接从数据中获取数据，这可能会造成混淆。例如，如果您有一个包含 `hive.compute.query.using.stats=true` 的表并且将新文件上载到表 LOCATION，则在该表上运行 `SELECT COUNT(*)` 查询会返回来自统计数据的计数，而不是选择已添加的行。

作为解决方法，请使用 `ANALYZE TABLE` 命令收集新的统计数据，或者设置 `hive.compute.query.using.stats=false`。有关更多信息，请参阅 Apache Hive 文档中的 [Hive 中的统计数据](#)。

## 版本 5.11.0

以下发行说明包含 Amazon EMR 版本 5.11.0 版本的信息。更改与亚马逊 EMR 5.10.0 版本有关。

### 升级

以下应用程序和组件已在此版本中进行升级以包含以下版本。

- Hive 2.3.2
- Spark 2.2.1
- SDK适用于 Java 1.11.238

### 新功能

- Spark
  - 增加了 `spark.decommissioning.timeout.threshold` 设置，这将改进使用竞价型实例时的 Spark 停用行为。有关更多信息，请参阅 [配置节点停用行为](#)。
  - 在 Spark 中添加了 `aws-sagemaker-spark-sdk` 组件，它安装了 Amazon SageMaker Spark 以及与 Spark 与 [亚马逊](#) 集成的相关依赖项 SageMaker。您可以使用 Amazon SageMaker Spark 通过亚马逊 SageMaker 阶段构建 Spark 机器学习 (ML) 管道。有关更多信息，请参阅《亚马逊 SageMaker 开发者指南》中的 [SageMaker Spark 自述文件](#) GitHub 和 SageMaker在亚马逊[上使用 Apache Spark](#)。

### 已知问题

- MXNet不包括 OpenCV 库。

- 默认情况下，Hive 2.3.2 设置 `hive.compute.query.using.stats=true`。这会导致查询从现有统计数据而不是直接从数据中获取数据，这可能会造成混淆。例如，如果您有一个包含 `hive.compute.query.using.stats=true` 的表并且将新文件上载到表 LOCATION，则在该表上运行 `SELECT COUNT(*)` 查询会返回来自统计数据的计数，而不是选择已添加的行。

作为解决方法，请使用 `ANALYZE TABLE` 命令收集新的统计数据，或者设置 `hive.compute.query.using.stats=false`。有关更多信息，请参阅 Apache Hive 文档中的 [Hive 中的统计数据](#)。

## 版本 5.10.0

以下发行说明包含 Amazon EMR 版本 5.10.0 版本的信息。更改与 Amazon EMR 5.9.0 版本有关。

### 升级

以下应用程序和组件已在此版本中进行升级以包含以下版本。

- AWS SDK for Java 1.11.221
- Hive 2.3.1
- Presto 0.187

### 新功能

- 添加了对 Kerberos 身份验证的支持。有关更多信息，请参阅《亚马逊EMR管理指南》中的“[使用 Kerberos 身份验证](#)”
- 为向 Amazon S3 发EMRFS出的请求添加了对IAM角色的支持。有关更多信息，请参阅《亚马逊EMR管理指南》中的为向 Amazon S3 发出的EMRFS[请求配置IAM角色](#)。
- 增加了对GPU基于 P2 和 P3 的实例类型的支持。有关更多信息，请参阅[亚马逊 EC2 P2 实例](#)和[亚马逊 EC2 P3 实例](#)。NVIDIA默认情况下，驱动程序 384.81 和CUDA驱动程序 9.0.176 安装在这些实例类型上。
- 增加了对 [Apache MXNet](#) 的支持。

### 更改、增强功能和解决的问题

- Presto

- 增加了对使用 Glue 数据 AWS 目录作为默认 Hive 元数据仓的支持。有关更多信息，请参阅[Presto 与 AWS Glue 数据目录配合使用](#)。
- 增加了对[地理空间函数](#)的支持。
- 为联接添加了[溢出到磁盘](#)支持。
- 增加了对 [Redshift 连接器](#)的支持。
- Spark
  - 向后移植了 [SPARK-20640](#)，这使得使用和属性可以配置 rpc 超时和洗牌注册值的重试次数。`spark.shuffle.registration.timeout`  
`spark.shuffle.registration.maxAttempts`
  - 向后移植的 [SPARK-21549](#)，它更正了向非位置写入自定义 OutputFormat 内容时发生的错误。HDFS
- 已逆向移植 [Hadoop 13270](#)
- Numpy、Scipy 和 Matplotlib 库已从亚马逊基础库中删除。EMR AMI如果您的应用程序需要这些库，应用程序存储库中提供了它们，因此您可以通过引导操作使用 `yum install` 在所有节点上安装它们。
- Amazon EMR 基础AMI不再包含应用程序RPM包，因此这些RPM软件包不再存在于集群节点上。Custom AMIs 和 Amazon EMR 基础AMI现在引用 Amazon S3 中的RPM软件包存储库。
- 由于 Amazon 引入了按秒计费EC2，因此默认的缩减行为现在是任务完成时终止，而不是在实例时间终止。有关更多信息，请参阅[配置集群缩减](#)。

## 已知问题

- MXNet不包括 OpenCV 库。
- 默认情况下，Hive 2.3.1 设置 `hive.compute.query.using.stats=true`。这会导致查询从现有统计数据而不是直接从数据中获取数据，这可能会造成混淆。例如，如果您有一个包含 `hive.compute.query.using.stats=true` 的表并且将新文件上载到表 LOCATION，则在该表上运行 `SELECT COUNT(*)` 查询会返回来自统计数据的计数，而不是选择已添加的行。

作为解决方法，请使用 `ANALYZE TABLE` 命令收集新的统计数据，或者设置 `hive.compute.query.using.stats=false`。有关更多信息，请参阅 Apache Hive 文档中的[Hive 中的统计数据](#)。

## 版本 5.9.0

以下发行说明包含 Amazon 5.9.0 EMR 版本的信息。更改与 Amazon EMR 5.8.0 版本有关。

发布日期：2017 年 10 月 5 日

最近功能更新时间：2017 年 10 月 12 日

### 升级

以下应用程序和组件已在此版本中进行升级以包含以下版本。

- AWS SDK for Java 版本 1.11.183
- Flink 1.3.2
- Hue 4.0.1
- Pig 0.17.0
- Presto 0.184

### 新功能

- 添加了 Livy 支持 (0.4.0-incubating 版)。有关更多信息，请参阅[Apache Livy](#)。
- 添加了对 Hue Notebook for Spark 的支持。
- 增加了对 i3 系列亚马逊EC2实例的支持 (2017 年 10 月 12 日)。

### 更改、增强功能和解决的问题

- Spark
  - 添加了一组新功能，有助于确保 Spark 能够更为正常地处理因手动调整大小或自动扩展策略请求导致的节点终止。有关更多信息，请参阅 [配置节点停用行为](#)。
  - SSL用于区块传输服务的传输中加密，而不是 3DES，当使用带 AES-NI 的 Amazon EC2 实例类型时，这可以提高性能。
  - 向后移植 [SPARK-21494](#)。
- Zeppelin
  - 向后移植 [ZEPPELIN-2377](#)。
- HBase



- 添加了 [HBASE-18533](#) 补丁，它允许使用HBase BucketCache 配置分类为配置提供其他值。hbase-site
- Hue
  - 为 Hue 中的 Hive 查询编辑器添加了 Glue 数据目录支持。
  - 默认情况下，Hue 中的超级用户可以访问允许 Amazon EMR IAM 角色访问的所有文件。新建用户不会自动拥有对 Amazon S3 filebrowser 的访问权限，并且必须为其组启用 filebrowser.s3\_access 权限。
- 解决了导致无法访问使用 AWS Glue 数据目录创建的底层JSON数据的问题。

## 已知问题

- 当所有应用程序都安装完毕并且默认的 Amazon EBS 根卷大小未更改时，集群启动将失败。解决方法是，使用中的aws emr create-cluster命令 AWS CLI 并指定更大的--ebs-root-volume-size参数。
- 默认情况下，Hive 2.3.0 设置 hive.compute.query.using.stats=true。这会导致查询从现有统计数据而不是直接从数据中获取数据，这可能会造成混淆。例如，如果您有一个包含 hive.compute.query.using.stats=true 的表并且将新文件上载到表 LOCATION，则在该表上运行 SELECT COUNT(\*) 查询会返回来自统计数据的计数，而不是选择已添加的行。

作为解决方法，请使用 ANALYZE TABLE 命令收集新的统计数据，或者设置 hive.compute.query.using.stats=false。有关更多信息，请参阅 Apache Hive 文档中的 [Hive 中的统计数据](#)。

## 版本 5.8.2

以下发行说明包含有关亚马逊 5.8.2 EMR 版本的信息。更改与 5.8.1 有关。

首次发布日期：2018 年 3 月 29 日

更改、增强功能和解决的问题

- 更新了亚马逊 Linux 版的亚马逊 defaultAmazon Linux AMI 内核EMR以修复潜在漏洞。

## 版本 5.8.1

以下发行说明包含 Amazon EMR 版本 5.8.1 版本的信息。更改与 Amazon EMR 5.8.0 版本有关。

首次发布日期：2018 年 1 月 22 日

## 更改、增强功能和解决的问题

- 更新了亚马逊版 Linux 的亚马逊 L defaultAmazon inux AMI 内核，EMR以解决与推测执行相关的漏洞 ( CVE-2017-5715、-2017-5753 和 -2017 CVE -5754 )。CVE有关更多信息，请参阅 <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>。

## 版本 5.8.0

以下发行说明包含 Amazon EMR 版本 5.8.0 版本的信息。更改与 Amazon EMR 5.7.0 版本有关。

首次发布日期：2017 年 8 月 10 日

最近功能更新时间：2017 年 9 月 25 日

## 升级

以下应用程序和组件已在此版本中进行升级以包含以下版本：

- AWS SDK1.11.160
- Flink 1.3.1
- Hive 2.3.0。有关更多信息，请参阅 Apache Hive 站点上的[发布说明](#)。
- Spark 2.2.0。有关更多信息，请参阅 Apache Spark 站点上的[发布说明](#)。

## 新功能

- 添加了对查看应用程序历史记录的支持 (2017 年 9 月 25 日)。有关更多信息，请参阅《Amazon EMR 管理指南》中的[查看应用程序历史记录](#)。

## 更改、增强功能和解决的问题

- 与 AWS Glue 数据目录集成
  - 添加了 Hive 和 Spark SQL 使用 AWS Glue 数据目录作为 Hive 元数据存储的功能。有关更多信息，请参阅[使用 AWS Glue 数据目录作为 Hive 的元数据库](#) 和[使用 AWS Glue 数据目录作为 Spark 的元数据库 SQL](#)。

- 在集群详细信息中添加了应用程序历史记录，允许您查看YARN应用程序的历史数据和 Spark 应用程序的其他详细信息。有关更多信息，请参阅《Amazon EMR 管理指南》中的[查看应用程序历史记录](#)。
- Oozie
  - 向后移植 [OOZIE-2748](#)。
- Hue
  - [向后移植 HUE -5859](#)
- HBase
  - 添加了用于通过 Java 管理扩展 (JMX) 公开HBase主服务器启动时间的补丁 `丁getMasterInitializedTime`。
  - 添加了改进集群启动时间的补丁。

## 已知问题

- 当所有应用程序都安装完毕并且默认的 Amazon EBS 根卷大小未更改时，集群启动将失败。解决方法是，使用中的 `aws emr create-cluster` 命令 AWS CLI 并指定更大的 `--ebs-root-volume-size` 参数。
- 默认情况下，Hive 2.3.0 设置 `hive.compute.query.using.stats=true`。这会导致查询从现有统计数据而不是直接从数据中获取数据，这可能会造成混淆。例如，如果您有一个包含 `hive.compute.query.using.stats=true` 的表并且将新文件上载到表 LOCATION，则在该表上运行 `SELECT COUNT(*)` 查询会返回来自统计数据的计数，而不是选择已添加的行。

作为解决方法，请使用 `ANALYZE TABLE` 命令收集新的统计数据，或者设置 `hive.compute.query.using.stats=false`。有关更多信息，请参阅 Apache Hive 文档中的 [Hive 中的统计数据](#)。

- Spark – 在使用 Spark 时，`appusher` 进程守护程序存在文件处理程序泄漏问题，长时间运行的 Spark 任务在几个小时或几天后可能会出现此情况。要解决此问题，请连接到主节点并键入 `sudo /etc/init.d/appusher stop`。这将停止 `appusher` 守护程序，亚马逊EMR将自动重启该守护程序。
- 应用程序历史记录
  - 死 Spark 执行程序的历史数据不可用。
  - 应用程序历史记录对使用安全配置来启用传输中加密的集群不可用。

## 版本 5.7.0

以下发行说明包含 Amazon EMR 5.7.0 版本的信息。更改与 Amazon EMR 5.6.0 版本有关。

发布日期：2017 年 7 月 13 日

### 升级

- Flink 1.3.0
- Phoenix 4.11.0
- Zeppelin 0.7.2

### 新功能

- 增加了在创建集群AMI时指定自定义 Amazon Linux 的功能。有关更多信息，请参阅[使用自定义 AMI](#)。

### 更改、增强功能和解决的问题

- HBase
  - 增加了配置HBase只读副本集群的功能。请参阅[使用只读副本集群](#)。
  - 多个错误修复和增强功能
- Presto – 添加了配置 `node.properties` 的功能。
- YARN-增加了配置功能 `container-log4j.properties`
- Sqoop-向后移植的 [SQOOP-2880](#)，它引入了一个允许你设置 Sqoop 临时目录的参数。

## 版本 5.6.0

以下发行说明包含 Amazon EMR 5.6.0 版本的信息。更改与亚马逊 EMR 5.5.0 版本有关。

发布日期：2017 年 6 月 5 日

### 升级

- Flink 1.2.1
- HBase1.3.1

- Mahout 0.13.0。这是第一个在亚马逊 5.0 及更高版本中支持 Spark 2.x 的 Mahout EMR 版本。
- Spark 2.1.1

## 更改、增强功能和解决的问题

- Presto
  - 通过使用TLS安全配置启用传输中加密，增加了在 Presto 节点之间启用 SSL /安全通信的功能。有关更多信息，请参阅[传输中的数据加密](#)。
  - 已逆向移植 [Presto 7661](#)，它向 EXPLAIN ANALYZE 语句添加了 VERBOSE 选项，以报告有关查询计划的更详细、高低级别的统计数据。

## 版本 5.5.3

以下发行说明包含有关亚马逊 5.5.3 EMR 版本的信息。更改与 5.5.2 有关。

首次发布日期：2018年 8 月 29 日

### 更改、增强功能和解决的问题

- 此版本解决了潜在的安全漏洞。

## 版本 5.5.2

以下发行说明包含 Amazon EMR 版本 5.5.2 的信息。更改与 5.5.1 有关。

首次发布日期：2018 年 3 月 29 日

### 更改、增强功能和解决的问题

- 更新了亚马逊 Linux 版的亚马逊 defaultAmazon Linux AMI 内核EMR以修复潜在漏洞。

## 版本 5.5.1

以下发行说明包含有关 Amazon EMR 5.5.1 版本的信息。更改与亚马逊 EMR 5.5.0 版本有关。

首次发布日期：2018 年 1 月 22 日

## 更改、增强功能和解决的问题

- 更新了亚马逊版 Linux 的亚马逊 L defaultAmazon inux AMI 内核，EMR以解决与推测执行相关的漏洞 ( CVE-2017-5715、-2017-5753 和 -2017 CVE -5754 )。CVE有关更多信息，请参阅 <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>。

## 版本 5.5.0

以下发行说明包含有关 Amazon EMR 5.5.0 版本的信息。更改与亚马逊 EMR 5.4.0 版本有关。

发布日期：2017 年 4 月 26 日

## 升级

- Hue 3.12
- Presto 0.170
- Zeppelin 0.7.1
- ZooKeeper 3.4.10

## 更改、增强功能和解决的问题

- Spark
  - 向后移植的 Spark Patch ([SPARK-20115](#)) 已修复DAGScheduler，当外部洗牌服务不可用于 Spark 版本 2.1.0 ( 包含在本版本中 ) 时，可以重新计算所有丢失的随机播放块。
- Flink
  - Flink 现在使用 Scala 2.11 进行构建。如果您使用 Scala API 和库，我们建议您在项目中使用 Scala 2.11。
  - 解决了 HADOOP\_CONF\_DIR 和 YARN\_CONF\_DIR 默认值未正确设置，因此 start-scala-shell.sh 无法工作的问题。此外，还添加了使用 env.hadoop.conf.dir 或 env.yarn.conf.dir 配置类别中的 /etc/flink/conf/flink-conf.yaml 和 flink-conf 设置这些值的功能。
  - 引入了一个新的EMR特定于命令flink-scala-shell的封装器。start-scala-shell.sh我们建议使用此命令而不是 start-scala-shell。新命令可简化执行。例如，flink-scala-shell -n 2 将使用任务并行度 2 启动 Flink Scala Shell。

- 引入了一个新的EMR特定于命令flink-yarn-session的封装器。yarn-session.sh我们建议使用此命令而不是 yarn-session。新命令可简化执行。例如，flink-yarn-session -d -n 2 将使用两个任务管理器在分离状态下启动长时间运行的 Flink 会话。
- 已解决 [\(FLINK-6125\) Commons httpclient 在 Flink 1.2 中不再有阴影。](#)
- Presto
  - 增加了对LDAP身份验证的支持。在亚马逊上LDAP与 Presto 一起使用EMR需要启用 Presto 协调器的HTTPS访问权限 ( http-server.https.enabled=true中 )。config.properties有关配置的详细信息，请参阅 Presto 文档中的[LDAP身份验证](#)。
  - 增加了对 SHOW GRANTS 的支持。
- 亚马逊 EMR Base Linux AMI
  - 亚马逊EMR发布的版本现在基于亚马逊 Linux 2017.03。有关更多信息，请参阅[亚马逊 Linux AMI 2017.03 发行说明](#)。
  - 从亚马逊EMR基本的 Linux 镜像中移除了 Python 2.6。默认安装 Python 2.7 和 3.4。如果需要，您可以手动安装 Python 2.6。

## 版本 5.4.0

以下发行说明包含 Amazon EMR 5.4.0 版本的信息。更改与亚马逊 EMR 5.3.0 版本有关。

发布日期：2017 年 3 月 8 日

### 升级

此版本提供以下升级：

- 已升级到 Flink 1.2.0
- 已升级到 Hbase 1.3.0
- 已升级到 Phoenix 4.9.0

#### Note

如果您从亚马逊的早期版本升级EMR到亚马逊版本 5.4.0 或更高EMR版本并使用二级索引，请按照 A [pache Phoenix](#) 文档中的说明升级本地索引。Amazon 从hbase-site分类中EMR删除了所需的配置，但需要重新填充索引。支持在线和离线升级索引。在线升级为默认值，这意味着，在从版本 4.8.0 或更高版本的 Phoenix 客户端初始化时重新填充索引。要指定离

线升级，请在phoenix-site分类中将phoenix.client.localIndexUpgrade配置设置为 false，然后SSH将配置设置为要运行的主节点psql [zookeeper] -1。

- 已升级到 Presto 0.166
- 已升级到 Zeppelin 0.7.0

## 更改和增强功能

以下是对亚马逊EMR发布标签 emr-5.4.0 的版本所做的更改：

- 增加了对 r4 实例的支持。参见 [Amazon EC2 实例类型](#)。

## 版本 5.3.1

以下发行说明包含 Amazon EMR 5.3.1 版本的信息。更改与亚马逊 EMR 5.3.0 版本有关。

发布日期：2017 年 2 月 7 日

对向后移植齐柏林飞艇补丁进行了细微的更改，并更新了亚马逊的默认补丁。AMI EMR

## 版本 5.3.0

以下发行说明包含 Amazon EMR 5.3.0 版本的信息。更改与亚马逊 EMR 5.2.1 版本有关。

发布日期：2017 年 1 月 26 日

## 升级

此版本提供以下升级：

- 已升级到 Hive 2.1.1
- 已升级到 Hue 3.11.0
- 已升级到 Spark 2.1.0
- 已升级到 Oozie 4.3.0
- 已升级到 Flink 1.1.4

## 更改和增强功能

以下是对亚马逊EMR发布的版本标签 emr-5.3.0 所做的更改：



- Hue 新增补丁可使您使用 `interpreters_shown_on_wheel` 设置配置解释器在笔记本选择轮盘上最先显示的内容，而不受 `hue.ini` 文件中排序的限制。
- 新增 `hive-parquet-logging` 配置分类，您可以将其用于配置 Hive `parquet-logging.properties` 文件中的值。

## 版本 5.2.2

以下发行说明包含 Amazon EMR 5.2.2 版本的信息。更改与亚马逊 EMR 5.2.1 版本有关。

发布日期：2017 年 5 月 2 日

### 早期版本中已解决的已知问题

- 向后移植 [SPARK-194459](#)，它解决了从包含 `char/varchar` ORC 列的表中读取可能失败的问题。

## 版本 5.2.1

以下发行说明包含 Amazon EMR 5.2.1 版本的信息。更改与亚马逊 EMR 5.2.0 版本有关。

发布日期：2016 年 12 月 29 日

### 升级

此版本提供以下升级：

- 已升级到 Presto 0.157.1。有关更多信息，请参阅 Presto 文档中的 [Presto 发布说明](#)。
- 已升级到 Zookeeper 3.4.9。有关更多信息，请参阅 Apache ZooKeeper 文档中的 [ZooKeeper 发行说明](#)。

### 更改和增强功能

以下是对亚马逊EMR发布的版本标签 `emr-5.2.1` 所做的更改：

- 在亚马逊 4.8.3 及更高EMR版本中增加了对亚马逊 EC2 `m4.16xlarge` 实例类型的支持，不包括 5.0.0、5.0.3 和 5.2.0。
- 亚马逊EMR发布的版本现在基于亚马逊 Linux 2016.09。有关更多信息，请参阅 <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>。

- 现在，Flink 和YARN配置路径的位置是默认设置的/etc/default/flink，因为您无需设置环境变量FLINK\_CONF\_DIR，运行flink或yarn-session.sh驱动脚本即可启动 Flink 作业。HADOOP\_CONF\_DIR
- 增加了对 FlinkKinesisConsumer 类的支持。

## 早期版本中已解决的已知问题

- 修复了 Hadoop 中的一个问题，即由于在大型集群中复制和删除同一文件之间存在竞争，ReplicationMonitor 线程可能会长时间卡住。
- 修复了任务状态未成功更新时 ControlledJob # toString 失败并出现空指针异常 (NPE) 的问题。

## 版本 5.2.0

以下发行说明包含 Amazon EMR 5.2.0 版本的信息。更改与亚马逊 EMR 5.1.0 版本有关。

发布日期：2016 年 11 月 21 日

## 更改和增强功能

此版本中提供了以下更改和增强功能：

- 为添加了 Amazon S3 存储模式HBase。
- 允许您为 r HBase oodir 指定 Amazon S3 的位置。有关更多信息，请参阅 [Amaz HBase on S3](#)。

## 升级

此版本提供以下升级：

- 已升级到 Spark 2.0.2

## 早期版本中已解决的已知问题

- 修复了EBS仅限实例类型的 /mnt 限制为 2 TB 的问题。
- 修复了输出到相应的 .out 文件而不是常规 log4j 配置的 .log 文件 (每小时转动一次) 的 instance-controller 和 logpusher 日志的问题。.out 文件不会轮换，因此这最终将填满 /emr 分区。此问题仅影响硬件虚拟机 (HVM) 实例类型。

## 版本 5.1.0

以下发行说明包含 Amazon EMR 5.1.0 版本的信息。更改与 Amazon EMR 5.0.0 版本有关。

发布日期：2016 年 11 月 3 日

### 更改和增强功能

此版本中提供了以下更改和增强功能：

- 增加了对 Flink 1.1.3 的支持。
- Presto 已作为 Hue 的记事本部分中的选项添加。

### 升级

此版本提供以下升级：

- 已升级到 HBase 1.2.3
- 已升级到 Zeppelin 0.6.2

### 早期版本中已解决的已知问题

- 修复了 Amazon S3 上的 Tez 查询问题，ORC文件表现不如早期的 Amazon EMR 4.x 版本。

## 版本 5.0.3

以下发行说明包含 Amazon EMR 5.0.3 版本的信息。更改与 Amazon EMR 5.0.0 版本有关。

发布日期：2016 年 10 月 24 日

### 升级

此版本提供以下升级：

- 已升级到 Hadoop 2.7.3
- 已升级到 Presto 0.152.3，它包括对 Presto Web 界面的支持。可使用端口 8889 访问 Presto 协调器上的 Presto Web 界面。有关 Presto Web 界面的更多信息，请参阅 Presto 文档中的 [Web 界面](#)。
- 已升级到 Spark 2.0.1

- 亚马逊EMR发布的版本现在基于亚马逊 Linux 2016.09。有关更多信息，请参阅 <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>。

## 版本 5.0.0

发布日期：2016 年 7 月 27 日

### 升级

此版本提供以下升级：

- 已升级到 Hive 2.1
- 已升级到 Presto 0.150
- 已升级到 Spark 2.0
- 已升级到 Hue 3.10.0
- 已升级到 Pig 0.16.0
- 已升级到 Tez 0.8.4
- 已升级到 Zeppelin 0.6.1

### 更改和增强功能

以下是对亚马逊EMR发布的版本标签 emr-5.0.0 或更高版本所做的更改：

- 亚马逊EMR支持 Hive ( 版本 2.1 ) 和 Pig ( 版本 0.16.0 ) 的最新开源版本。如果您过去曾EMR在亚马逊上使用过 Hive 或 Pig，这可能会影响某些用例。有关更多信息，请参阅 [Hive](#) 和 [Pig](#)。
- Hive 和 Pig 的默认执行引擎现在是 Tez。要更改该设置，您可以在 hive-site 和 pig-properties 配置分类中分别编辑相应的值。
- 添加了增强型步骤调试功能，可让您查看步骤失败的根本原因 (如果服务可以确定原因)。有关更多信息，请参阅 Amazon EMR 管理指南中的 [增强型步骤调试](#)。
- 先前以“-Sandbox”结尾的应用程序不再拥有该后缀。这可能会中断您的自动化，例如，如果您使用脚本来启动具有这些应用程序的集群。下表显示了亚马逊 EMR 4.7.2 与亚马逊 EMR 5.0.0 中的应用程序名称。

## 应用程序名称更改

亚马逊 EMR 4.7.2	亚马逊 EMR 5.0.0
Oozie-Sandbox	Oozie
Presto-Sandbox	Presto
Sqoop-Sandbox	Sqoop
Zeppelin-Sandbox	Zeppelin
ZooKeeper-沙箱	ZooKeeper

- Spark 现在针对 Scala 2.11 进行编译。
- Java 8 现在是默认设置JVM。所有应用程序均使用 Java 8 runtime 运行。对任何应用程序的字节代码目标都没有进行更改。大多数应用程序继续运行 Java 7。
- Zeppelin 现在包括身份验证功能。有关更多信息，请参阅 [Zeppelin](#)。
- 添加了对安全配置的支持，这使您可以更轻松地创建和应用加密选项。有关更多信息，请参阅[数据加密](#)。

## 版本 4.9.5

以下发行说明包含有关亚马逊 4.9.5 EMR 版本的信息。更改与 4.9.4 有关。

首次发布日期：2018年 8 月 29 日

更改、增强功能和解决的问题

- HBase
  - 此版本解决了潜在的安全漏洞。

## 版本 4.9.4

以下发行说明包含有关亚马逊 4.9.4 EMR 版本的信息。更改与 4.9.3 有关。

首次发布日期：2018 年 3 月 29 日

## 更改、增强功能和解决的问题

- 更新了亚马逊 Linux 版的亚马逊 defaultAmazon Linux AMI 内核EMR以修复潜在漏洞。

## 版本 4.9.3

以下发行说明包含 Amazon EMR 4.9.3 版本的信息。更改与 Amazon EMR 4.9.2 版本有关。

首次发布日期：2018 年 1 月 22 日

## 更改、增强功能和解决的问题

- 更新了亚马逊版 Linux 的亚马逊 L defaultAmazon inux AMI 内核，EMR以解决与推测执行相关的漏洞 ( CVE-2017-5715、-2017-5753 和 -2017 CVE -5754 )。CVE有关更多信息，请参阅 <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>。

## 版本 4.9.2

以下发行说明包含 Amazon EMR 4.9.2 版本的信息。更改与 Amazon EMR 4.9.1 版本有关。

发布日期：2017 年 7 月 13 日

此版本略微进行了一些改动、错误修复和增强。

## 版本 4.9.1

以下发行说明包含 Amazon EMR 4.9.1 版本的信息。更改与 Amazon EMR 4.8.4 版本有关。

发布日期：2017 年 4 月 10 日

## 早期版本中已解决的已知问题

- [HIVE-9976 和- 10106 的向后移植 HIVE](#)
- 修复了大量节点 ( 大于 2,000 ) 和容器 ( 大于 5,000 ) 会导致内存不足错误的问题，例如："Exception in thread 'main' java.lang.OutOfMemoryError"。YARN

## 更改和增强功能

以下是对亚马逊EMR发布的版本标签 emr-4.9.1 所做的更改：

- 亚马逊EMR发布的版本现在基于亚马逊 Linux 2017.03。有关更多信息，请参阅 <https://aws.amazon.com/amazon-linux-ami/2017.03-release-notes/>。
- 从亚马逊EMR基本的 Linux 镜像中移除了 Python 2.6。如果需要，您可以手动安装 Python 2.6。

## 版本 4.8.4

以下发行说明包含 Amazon EMR 4.8.4 版本的信息。更改与 Amazon EMR 4.8.3 版本有关。

发布日期：2017 年 2 月 7 日

此版本略微进行了一些改动、错误修复和增强。

## 版本 4.8.3

以下发行说明包含 Amazon EMR 4.8.3 版本的信息。更改与亚马逊 EMR 4.8.2 版本有关。

发布日期：2016 年 12 月 29 日

## 升级

此版本提供以下升级：

- 已升级到 Presto 0.157.1。有关更多信息，请参阅 Presto 文档中的 [Presto 发布说明](#)。
- 已升级到 Spark 1.6.3。有关更多信息，请参阅 Apache Spark 文档中的 [Spark 发布说明](#)。
- 已升级到 ZooKeeper 3.4.9。有关更多信息，请参阅 Apache ZooKeeper 文档中的 [ZooKeeper 发行说明](#)。

## 更改和增强功能

以下是对亚马逊EMR发布的版本标签 emr-4.8.3 所做的更改：

- 在亚马逊 4.8.3 及更高EMR版本中增加了对亚马逊 EC2 m4.16xlarge 实例类型的支持，不包括 5.0.0、5.0.3 和 5.2.0。
- 亚马逊EMR发布的版本现在基于亚马逊 Linux 2016.09。有关更多信息，请参阅 <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>。

## 早期版本中已解决的已知问题

- 修复了 Hadoop 中的一个问题，即由于在大型集群中复制和删除同一文件之间存在竞争，ReplicationMonitor 线程可能会长时间卡住。
- 修复了任务状态未成功更新时 ControlledJob # toString 失败并出现空指针异常 (NPE) 的问题。

## 版本 4.8.2

以下发行说明包含 Amazon EMR 4.8.2 版本的信息。更改与 Amazon EMR 4.8.0 版本有关。

发布日期：2016 年 10 月 24 日

### 升级

此版本提供以下升级：

- 已升级到 Hadoop 2.7.3
- 已升级到 Presto 0.152.3，它包括对 Presto Web 界面的支持。可使用端口 8889 访问 Presto 协调器上的 Presto Web 界面。有关 Presto Web 界面的更多信息，请参阅 Presto 文档中的 [Web 界面](#)。
- 亚马逊EMR发布的版本现在基于亚马逊 Linux 2016.09。有关更多信息，请参阅 <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>。

## 版本 4.8.0

发布日期：2016 年 9 月 7 日

### 升级

此版本提供以下升级：

- 已升级到 HBase 1.2.2
- 已升级到 Presto-Sandbox 0.151
- 已升级到 Tez 0.8.4
- 已升级到 Zeppelin-Sandbox 0.6.1

### 更改和增强功能

以下是对亚马逊EMR发布的版本标签 emr-4.8.0 所做的更改：



- YARN修复了 ApplicationMaster 尝试清理因实例已终止而不再存在的容器的问题。
- 在 Oozie 示例中更正了 Hive-server2 URL 的 Hive2 动作。
- 添加了对其它 Presto 目录的支持。
- [向后移植的补丁：HIVE-8948、-12679、-13405、HIVE-3116、-12689 HIVE PHOENIX HADOOP](#)
- 添加了对安全配置的支持，这使您可以更轻松地创建和应用加密选项。有关更多信息，请参阅[数据加密](#)。

## 版本 4.7.2

以下发行说明包含亚马逊 EMR 4.7.2 的信息。

发布日期：2016 年 7 月 15 日

### 功能

此版本提供以下功能：

- 已升级到 Mahout 0.12.2
- 已升级到 Presto 0.148
- 已升级到 Spark 1.6.2
- 现在，您可以使用 a URI 作为参数来创建“以 AWSCredentialsProvider 供EMRFS使用”。有关更多信息，请参阅[AWSCredentialsProvider 为创建EMRFS](#)。
- EMRFS现在允许用户使用中的属性为其一致视图元数据配置自定义 DynamoDB 终端节点。fs.s3.consistent.dynamodb.endpoint emrfs-site.xml
- 在 /usr/bin 中添加了一个名为 spark-example 的脚本，它将包装 /usr/lib/spark/spark/bin/run-example，因此您可以直接运行示例。例如，要运行 Spark 发行版附带的 SparkPi 示例，可以spark-example SparkPi 100从命令行运行，也可以command-runner.jar将其用作中的一个步骤API。

### 早期版本中已解决的已知问题

- 修复了 Oozie 在安装 Spark 后拥有的 spark-assembly.jar 未位于正确位置 (这导致使用 Oozie 启动 Spark 应用程序失败) 的问题。
- 修复了容器中基于 Spark Log4J 的日志记录的问题。YARN

## 版本 4.7.1

发布日期：2016 年 6 月 10 日

### 早期版本中已解决的已知问题

- 修复了延长VPC使用私有子网启动的集群的启动时间的问题。该错误仅影响在 Amazon EMR 4.7.0 版本中启动的集群。
- 修复了在 Amazon EMR 4.7.0 版本中启动EMR的集群在亚马逊中未正确处理文件列表的问题。

## 版本 4.7.0

### Important

亚马逊 EMR 4.7.0 已被弃用。改用亚马逊 EMR 4.7.1 或更高版本。

发布日期：2016 年 6 月 2 日

### 功能

此版本提供以下功能：

- 已添加 Apache Phoenix 4.7.0
- 已添加 Apache Tez 0.8.3
- 已升级到 HBase 1.2.1
- 已升级到 Mahout 0.12.0
- 已升级到 Presto 0.147
- 已将升级 AWS SDK for Java 到 1.10.75
- 已从 `mapreduce.cluster.local.dir` 中的 `mapred-site.xml` 属性中删除最终标志以允许用户以本地模式运行 Pig。

### 集群上提供亚马逊 Redshift JDBC 驱动程序

亚马逊 Redshift JDBC 驱动程序现已包含在 `/usr/share/aws/redshift/jdbc` 中。`/usr/share/aws/redshift/jdbc/RedshiftJDBC41.jar` 是兼容 JDBC 4.1 的亚马逊 Redshift 驱动程序。

序，也是兼容 4. JDBC 0 的亚马逊 Redshift 驱动程序 `/usr/share/aws/redshift/jdbc/RedshiftJDBC4.jar`。有关更多信息，请参阅 Amazon Redshift 管理指南中的[配置 JDBC 连接](#)。

## Java 8

除了 Presto 之外，Open JDK 1.7 是所有应用程序的 JDK 默认设置。但是，Open JDK 1.7 和 1.8 都已安装。有关如何为应用程序设置 JAVA\_HOME 的信息，请参阅[配置应用程序以使用 Java 8](#)。

### 早期版本中已解决的已知问题

- 修复了 emr-4.6.0 中显著影响亚马逊 EMR 吞吐量优化 HDD (st1) EBS 卷性能的内核问题。
- 修复了如果在未选择 Hadoop 作为应用程序的情况下指定了任何 HDFS 加密区域，则集群会失败的问题。
- 将默认 HDFS 写入策略从更改 RoundRobin 为 AvailableSpaceVolumeChoosingPolicy。某些卷在 RoundRobin 配置中未得到正确利用，这导致核心节点出现故障且不可靠 HDFS。
- 修复了在 EMRFSLI 为一致视图创建默认 DynamoDB 元数据表时会导致异常的问题。
- 修复了在 EMRFS 多部分重命名和复制操作期间可能发生的死锁问题。
- 修复了 EMRFS 导致默认 CopyPart 大小为 5 MB 的问题。默认值现已相应地设置为 128 MB。
- 修复了与 Zeppelin upstart 配置有关的问题，此问题可能会阻止您停止服务。
- 修复了 Spark 和 Zeppelin 的问题，该问题使你无法使用该 `s3a://URI` 方案，`/usr/lib/hadoop/hadoop-aws.jar` 因为它们各自的职业路径中没有正确加载。
- 向后移植 [HUE-2484](#)。
- 向后移植了 Hue 3.9.0 中的[提交](#)（不 JIRA 存在），以修复浏览器示例的问题。HBase
- 向后移植 [HIVE-9073](#)。

## 版本 4.6.0

发布日期：2016 年 4 月 21 日

### 功能

此版本提供以下功能：

- 已添加 HBase 1.2.0
- 已添加 Zookeeper-Sandbox 3.4.8

- 已升级到 Presto-Sandbox 0.143
- 亚马逊EMR发布的版本现在基于亚马逊 Linux 2016.03.0。有关更多信息，请参阅 <https://aws.amazon.com/amazon-linux-ami/2016.03-release-notes/>。

## 影响吞吐量优化 HDD (st1) EBS 卷类型的问题

Linux 内核版本 4.2 及更高版本中的问题会严重影响吞吐量优化 HDD (st1) EBS 卷的EMR性能。此版本 (emr-4.6.0) 使用内核版本 4.4.5，因此会受到影响。因此，如果您想使用 st1 卷，我们建议不要使用 emr-4.6.0。EBS您可以将 emr-4.5.0 或之前的亚马逊EMR版本与 st1 一起使用，而不会产生任何影响。此外，我们将随将来版本一起提供修复程序。

## Python 默认值

现在，默认情况下已安装 Python 3.4，但 Python 2.7 将保留系统默认值。您可以使用引导操作将 Python 3.4 配置为系统默认值；您可以使用配置API在spark-env分类/usr/bin/python3.4中将PYSPARK\_PYTHON 设置为以影响所 PySpark使用的 Python 版本。

## Java 8

除了 Presto 之外，Open JDK 1.7 是所有应用程序的JDK默认设置。但是，Open JDK 1.7 和 1.8 都已安装。有关如何为应用程序设置 JAVA\_HOME 的信息，请参阅[配置应用程序以使用 Java 8](#)。

## 早期版本中已解决的已知问题

- 修复了应用程序预置有时会因生成的密码导致随机失败的问题。
- 之前，mysqld 已安装在所有节点上。现在，它仅安装在主实例上，而且仅在所选应用程序将 mysql-server 作为组件包含时安装。目前，以下应用程序包括该mysql-server组件：、Hive HCatalog、Hue、Presto-Sandbox 和 Sqoop-Sandbox。
- 从默认值 32 更改yarn.scheduler.maximum-allocation-vcores为 80，这修复了 emr-4.0 中引入的一个问题，该问题主要发生在 Spark 中，该问题主要发生在 Spark 中，核心实例类型是 YARN vcore 设置为 32 的少数大型实例类型之一；即 c4.8xlarge、cc2.8xlarge、hs1.8xlarge、i2.8xlarge、i2.8xlarge、i2.8xlarge、m2.4xlarge、d2.8xlarge，或 m4.10xlarge 受到了此问题的影响。maximizeResourceAllocation
- s3-dist-cp 现在用于所有 EMRFS Amazon S3 提名，不再暂存到临时目录。HDFS
- 修复了与针对客户端加密分段上载的异常处理有关的问题。
- 添加了允许用户更改 Amazon S3 存储类的选项。默认情况下，此设置为 STANDARD。emrfs-site 配置分类设置为 fs.s3.storageClass，可能的值为 STANDARD、STANDARD\_IA 和

REDUCED\_REDUNDANCY。有关存储类的更多信息，请参阅《Amazon Simple Storage Service 用户指南》中的[存储类](#)。

## 版本 4.5.0

发布日期：2016 年 4 月 4 日

### 功能

此版本提供以下功能：

- 已升级到 Spark 1.6.1
- 已升级到 Hadoop 2.7.2
- 已升级到 Presto 0.140
- 增加了对 Amazon S3 服务器端加密的 AWS KMS 支持。

### 早期版本中已解决的已知问题

- 修复了节点重启后 My SQL 和 Apache 服务器无法启动的问题。
- 修复IMPORT了存储在 Amazon S3 中的非分区表无法正常运行问题
- 修复了与 Presto 有关的问题，此问题导致在写入 Hive 表时要求暂存目录为 /mnt/tmp 而不是 /tmp。

## 版本 4.4.0

发布日期：2016 年 3 月 14 日

### 功能

此版本提供以下功能：

- 已添加 HCatalog 1.0.0
- 已添加 Sqoop-Sandbox 1.4.6
- 已升级到 Presto 0.136
- 已升级到 Zeppelin 0.5.6
- 已升级到 Mahout 0.11.1

- 默认情况下已启用 `dynamicResourceAllocation`。
- 已添加针对此版本的所有配置分类的表。有关更多信息，请参阅[配置应用程序](#)中的“配置分类”表。

## 早期版本中已解决的已知问题

- 修复了该`maximizeResourceAllocation`设置无法为YARN ApplicationMaster守护程序保留足够内存的问题。
- 修复了自定义时遇到的问题DNS。如果 `resolve.conf` 中的任何条目位于提供的自定义条目之前，则自定义条目不可解析。此行为受到集群的影响，VPC其中默认VPC名称服务器作为顶部条目插入`resolve.conf`。
- 修复了默认 Python 已移至版本 2.7 且未为该版本安装 boto 的问题。
- 修复了YARN容器和 Spark 应用程序生成唯一的 Ganglia 循环数据库 (rrd) 文件的问题，该文件会导致连接到实例的第一个磁盘已满。由于此修复，YARN容器级别的指标已被禁用，Spark 应用程序级别的指标已被禁用。
- 修复了导致日志推送程序中删除所有空日志文件夹的问题。结果是 Hive 无法登录CLI，因为日志推送器正在删除其下的空`user`文件夹。`/var/log/hive`
- 修复了影响 Hive 导入的问题，此问题影响分区并导致导入期间出现错误。
- 修复了EMRFS和 `s3-dist-cp` 无法正确处理包含句点的存储桶名称的问题。
- 更改了中的行为，EMRFS以便在启用版本控制的存储桶中不会连续创建\_`$folder`标记文件，这可能有助于提高启用版本控制的存储桶的性能。
- 更改了中的EMRFS行为，使其不使用指令文件，除非启用了客户端加密。如果您要在使用客户端加密时删除说明文件，可将 `emrfs-site.xml` 属性 `fs.s3.cse.cryptoStorageMode.deleteInstructionFiles.enabled` 设置为 `true`。
- 更改了YARN日志聚合，将聚合目标处的日志保留两天。默认目标为集群的HDFS存储空间。如果您要更改此持续时间，请在创建集群时使用 `yarn.log-aggregation.retain-seconds` 配置分类来更改 `yarn-site` 的值。与往常一样，您可以在创建集群时使用 `log-uri` 参数将应用程序日志保存到 Amazon S3。

## 已应用的修补程序

此版本中包含了来自开源项目的以下修补程序：

- [HIVE-9655](#)
- [HIVE-9183](#)

- [HADOOP-12810](#)

## 版本 4.3.0

发布日期：2016 年 1 月 19 日

### 功能

此版本提供以下功能：

- 已升级到 Hadoop 2.7.1
- 已升级到 Spark 1.6.0
- 已将 Ganglia 升级到 3.7.2
- 已将 Presto 升级到 0.130

Amazon 对 `spark.dynamicAllocation.enabled` 何时设置为 true EMR 进行了一些更改；默认情况下设置为 false。如果设置为 true，则会影响由 `maximizeResourceAllocation` 设置设定的默认值：

- 若 `spark.dynamicAllocation.enabled` 设为 true，则 `spark.executor.instances` 将不被 `maximizeResourceAllocation` 设置。
- 目前，`spark.driver.memory` 设置根据集群中的实例类型进行配置，与 `spark.executors.memory` 设置的方式类似。但是，由于 Spark 驱动程序应用程序可以在主实例或其中一个核心实例上运行（例如，分别在 YARN 客户端和集群模式下），因此该 `spark.driver.memory` 设置是根据这两个实例组之间较小实例类型的实例类型进行设置的。
- 现在，该 `spark.default.parallelism` 设置设置为 YARN 容器可用 CPU 内核数量的两倍。在上一版本中，这是该值的一半。
- 对为 Spark YARN 进程预留的内存开销的计算进行了调整，使其更加准确，从而使 Spark 可用的内存总量（即 `spark.executor.memory`）略有增加。

### 早期版本中已解决的已知问题

- YARN 默认情况下，日志聚合现在处于启用状态。
- 修复了启用日志聚合后无法将日志推送到集群的 Amazon S3 YARN 日志存储桶的问题。
- YARN 现在，在所有节点类型中，容器大小的最小值为 32。

- 修复了与 Ganglia 有关的问题，此问题已导致大型集群中主节点上的磁盘 I/O 过多。
- 修复了在关闭集群时阻止应用程序日志推送至 Amazon S3 的问题。
- 修复了导致EMRFSCLI某些命令失败的问题。
- 修复了齐柏林飞艇无法在底层加载依赖项的问题。 SparkContext
- 修复了因发出尝试添加实例的调整大小命令导致的问题。
- 修复了 Hive 中出现的 AS 对 Amazon S3 SELECT 进行过多列表调用的问题。 CREATE TABLE
- 修复了在安装 Hue、Oozie 和 Ganglia 时无法正常预置大型集群的问题。
- 修复了 s3-dist-cp 中的问题，此问题导致即使在失败并出现错误的情况下仍将返回零退出代码。

## 已应用的修补程序

此版本中包含了来自开源项目的以下修补程序：

- [OOZIE-2402](#)
- [HIVE-12502](#)
- [HIVE-10631](#)
- [HIVE-12213](#)
- [HIVE-10559](#)
- [HIVE-12715](#)
- [HIVE-10685](#)

## 版本 4.2.0

发布日期：2015 年 11 月 18 日

### 功能

此版本提供以下功能：

- 已添加 Ganglia 支持
- 已升级到 Spark 1.5.2
- 已升级到 Presto 0.125
- 已将 Oozie 升级到 4.2.0
- 已将 Zeppelin 升级到 0.5.5



- 已将升级 AWS SDK for Java 到 1.10.27

## 早期版本中已解决的已知问题

- 修复了未使用默认元数据表名称的问题。EMRFS CLI
- 修复了在 Amazon S3 中使用ORC由支持的表时遇到的问题。
- 修复了遇到的 Python 版本在 Spark 配置中不匹配的问题。
- 修复了由于中的群集问题而无法报告YARN节点状态DNS的问题VPC。
- 修复了节点YARN停用时遇到的问题，该问题导致应用程序挂起或无法调度新应用程序。
- 修复了集群以 TIMED OUT \_ 状态终止时遇到的问题STARTING。
- 修复了在其他版本中包含 EMRFS Scala 依赖项时遇到的问题。Scala 依赖项已被删除。

## 配置应用程序

您可以提供配置对象来覆盖应用程序的默认配置。您可以使用速记语法来提供配置，也可以在文件中引用配置对象。JSON配置对象包含分类、属性和可选的嵌套配置。属性对应于您想要更改的应用程序设置。您可以在单个JSON对象中为多个应用程序指定多个分类。

### Warning

Ama EMR zon Describe 和 List API 操作以纯文本形式发出自定义和可配置的设置，这些设置用作亚马逊EMR任务流程的一部分。要在这些设置中提供密码等敏感信息，请参阅将在 [AWS Secrets Manager](#) 中存储敏感配置数据。

可用的配置分类因 Amazon EMR 发行版本而异。有关特定发行版中所支持配置分类的列表，请参阅 [关于 Amazon EMR 发行](#) 下有关该发行版的页面。

以下是配置列表的示例JSON文件。

```
[
  {
    "Classification": "core-site",
    "Properties": {
      "hadoop.security.groups.cache.secs": "250"
    }
  },
  {
    "Classification": "mapred-site",
    "Properties": {
      "mapred.tasktracker.map.tasks.maximum": "2",
      "mapreduce.map.sort.spill.percent": "0.90",
      "mapreduce.tasktracker.reduce.tasks.maximum": "5"
    }
  }
]
```

配置分类通常可以映射到应用程序特定的配置文件。例如，hive-site 分类映射到 Hive 的 hive-site.xml 配置文件中的设置。此情况的一个例外是不再支持的引导操作 configure-daemons，它用于设置 --namenode-heap-size 等环境参数。与此类似的选项已归入 hadoop-env 和 yarn-env 分类，并具有自己的嵌套导出分类。如果任何分类以 env 结尾，请使用导出子分类。

另一个例外是 `s3get`，它用于将客户 `EncryptionMaterialsProvider` 对象放置在集群中的每个节点上来进行客户端加密。为了实现此目的，已向 `emrfs-site` 分类添加一个选项。

以下是一个 `hadoop-env` 分类示例。

```
[
  {
    "Classification": "hadoop-env",
    "Properties": {

    },
    "Configurations": [
      {
        "Classification": "export",
        "Properties": {
          "HADOOP_DATANODE_HEAPSIZE": "2048",
          "HADOOP_NAMENODE_OPTS": "-XX:GCTimeRatio=19"
        },
        "Configurations": [

        ]
      }
    ]
  }
]
```

以下是一个 `yarn-env` 分类示例。

```
[
  {
    "Classification": "yarn-env",
    "Properties": {

    },
    "Configurations": [
      {
        "Classification": "export",
        "Properties": {
          "YARN_RESOURCEMANAGER_OPTS": "-Xdebug -Xrunjdw:transport=dt_socket"
        },
        "Configurations": [

        ]
      }
    ]
  }
]
```

```

    }
  ]
}
]
```

以下设置不属于配置文件，但是 Amazon 可能会使用这些设置EMR来代表您配置多个设置。

由 Amazon 精心策划的设置 EMR

应用程序	发行版标注分类	有效属性	何时使用
Spark	spark	maximizeResourceAllocation	配置执行程序以利用每个节点的最大资源。

主题

- [在创建集群时配置应用程序](#)
- [在正在运行的集群中重新配置实例组](#)
- [在 AWS Secrets Manager中存储敏感配置数据](#)
- [配置应用程序来使用特定 Java 虚拟机](#)

## 在创建集群时配置应用程序

创建集群时，您可以使用 Amazon EMR 控制台、AWS Command Line Interface (AWS CLI) 或覆盖应用程序的默认配置 AWS SDK。

要覆盖应用程序的原定设置配置，您需要指定某个配置分类中的自定义值。配置分类对应于应用程序的配置XML文件，例如hive-site.xml。

配置分类因 Amazon EMR 发布版本而异。有关特定发行版中可用配置分类的列表，请参阅该发行版的详细信息页面。例如，[亚马逊EMR发布了 6.4.0](#)。

### 在创建集群时，在控制台中提供配置

要提供配置，请导航到创建集群页面并展开软件设置。然后，您可以使用控制台中以阴影文本JSON形式演示的速记语法直接输入配置。否则，您可以为带有JSONConfigurations对象的文件提供 Amazon S3 URI。

要为实例组提供配置，请在集群列表中选择集群，然后选择配置选项卡。在实例组配置表中，选择要编辑的实例组，然后选择重新配置。

## 创建集群 AWS CLI 时使用提供配置

您可以使用 `create-cluster` 通过提供存储在本地或 Amazon S3 中的 JSON 文件的路径来提供配置。以下示例假设您使用的是 Amazon 的默认角色 EMR 并且这些角色已经创建。如果您需要创建角色，请先运行 `aws emr create-default-roles`。

如果您的配置位于本地目录中，您可以使用示例命令。

```
aws emr create-cluster --use-default-roles --release-label emr-7.2.0 --applications
  Name=Hive \
--instance-type m5.xlarge --instance-count 3 --configurations file:///./
configurations.json
```

如果您的配置位于 Amazon S3 路径中，则需要设置以下解决方法，然后才能将 Amazon S3 路径传递给 `create-cluster` 命令。

```
#!/bin/sh
# Assume the ConfigurationS3Path is not public, and its present in the same AWS account
as the EMR cluster
ConfigurationS3Path="s3://my-bucket/config.json"
# Get a presigned HTTP URL for the s3Path
ConfigurationURL=`aws s3 presign $ConfigurationS3Path --expires-in 300`
# Fetch the presigned URL, and minify the JSON so that it spans only a single line
Configurations=`curl $ConfigurationURL | jq -c .`
aws emr create-cluster --use-default-roles --release-label emr-5.34.0 --instance-type
m5.xlarge --instance-count 2 --applications Name=Hadoop Name=Spark --configurations
$Configurations
```

## 创建集群 SDK 时使用 Java 提供配置

以下程序摘要说明如何使用 AWS SDK for Java 提供配置。

```
Application hive = new Application().withName("Hive");

Map<String,String> hiveProperties = new HashMap<String,String>();
hiveProperties.put("hive.join.emit.interval","1000");
hiveProperties.put("hive.merge.mapfiles","true");
```

```
Configuration myHiveConfig = new Configuration()
    .withClassification("hive-site")
    .withProperties(hiveProperties);

RunJobFlowRequest request = new RunJobFlowRequest()
    .withName("Create cluster with ReleaseLabel")
    .withReleaseLabel("emr-5.20.0")
    .withApplications(hive)
    .withConfigurations(myHiveConfig)
    .withServiceRole("EMR_DefaultRole")
    .withJobFlowRole("EMR_EC2_DefaultRole")
    .withInstances(new JobFlowInstancesConfig()
        .withEc2KeyName("myEc2Key")
        .withInstanceCount(3)
        .withKeepJobFlowAliveWhenNoSteps(true)
        .withMasterInstanceType("m4.large")
        .withSlaveInstanceType("m4.large")
    );
```

## 在正在运行的集群中重新配置实例组

在 Amazon 5.21.0 及更高EMR版本中，您可以重新配置集群应用程序，并为正在运行的集群中的每个实例组指定其他配置分类。为此，您可以使用 Amazon EMR 控制台、AWS Command Line Interface (AWS CLI) 或 AWS SDK。

当您在新 Amazon EMR 控制台中更新实例组的应用程序配置时，控制台会尝试将新配置与现有配置合并，以创建新的活动配置。在 Amazon EMR 无法合并配置的不寻常情况下，控制台会提醒您。

在您提交实例组的重新配置请求后，Amazon 会为新的配置规范EMR分配一个版本号。您可以通过查看 CloudWatch 事件来跟踪配置的版本号或实例组的状态。有关更多信息，请参阅[监控 CloudWatch 事件](#)。

### Note

您只能覆盖（而不能删除）集群创建过程中指定的集群配置。如果现有配置与您提供的文件之间存在差异，Amazon 会将EMR将手动修改的配置（例如您在使用SSH连接到集群时修改的配置）重置为指定实例组的集群默认值。

## 重新配置实例组时的注意事项

### 重新配置操作

当您使用亚马逊EMR控制台、AWS Command Line Interface (AWS CLI) 或提交重新配置请求时 AWS SDK，Amazon EMR 会检查现有的集群配置文件。如果现有配置与您提供的文件之间存在差异，Amazon EMR 会启动重新配置操作，重新启动某些应用程序，并将任何手动修改的配置（例如您在使用SSH连接到集群时修改的配置）重置为指定实例组的集群默认值。

#### Note

每次实例组重新配置期间，Amazon 都会EMR执行一些默认操作。这些默认操作可能与您所做的集群自定义冲突，并导致重新配置失败。如何对重新配置失败问题进行故障排查的相关信息，请参阅[对实例组重新配置问题进行故障排查](#)。

Amazon EMR 还会针对您在请求中指定的配置分类启动重新配置操作。有关这些操作的完整列表，请参阅您使用的Amazon EMR 版本的配置分类部分。例如，[6.2.0 配置分类](#)。

#### Note

《亚马逊EMR发行指南》仅列出了从亚马逊EMR版本 5.32.0 和 6.2.0 开始的重新配置操作。

### 服务中断

Amazon EMR 遵循滚动流程来重新配置任务和核心实例组中的实例。只能同时修改并重新启动实例组中的 10% 实例。该过程需要更长的时间才能完成，但降低了在运行的集群中发生潜在应用程序故障的可能性。

要在YARN重启期间运行YARN作业，您可以创建具有多个主节点的 Amazon EMR 集群，也可以在`yarn-site`配置分类`true`中设置为`yarn.resourcemanager.recovery.enabled`。有关使用多个主节点的更多信息，请参阅[高可用性YARN ResourceManager](#)。

### 应用程序验证

重新配置重启过程完成后，Amazon 会EMR检查集群上的每个应用程序是否都在运行。如果任何应用程序不可用，则整个重新配置操作失败。如果重新配置操作失败，Amazon EMR 会将配置参数恢复为先前的工作版本。

**Note**

为避免重新配置失败，我们建议您仅在计划使用的集群上安装应用程序。我们还建议您在提交重新配置请求之前，先确保所有集群应用程序均正在正常运行。

## 重新配置的类型

您可以以两种方式之一重新配置实例组重新配置实例组：

- **覆盖。**默认重新配置方法，也是亚马逊5.35.0和6.6.0之前EMR版本中唯一可用的方法。此重新配置方法不加区分地使用新提交的配置集覆盖任何集群上的文件。该方法会清除在重新API配置之外对配置文件所做的任何更改。
- **合并。**亚马逊EMR版本 5.35.0 和 6.6.0 及更高版本支持重新配置方法，但亚马逊EMR控制台除外，没有版本支持。此重新配置方法合并新提交的配置与已经存在于集群中的配置。此选项仅添加或修改您提交的新配置。它会保留现有的配置。

**Note**

亚马逊EMR继续覆盖一些必不可少的 Hadoop 配置，这些配置是确保服务正常运行所必需的。

## 限制

在重新配置正在运行的集群中的实例组时，请考虑以下限制：

- 非YARN应用程序可能会在重启期间失败或导致群集问题，尤其是在应用程序配置不正确的情况下。在重新启动过程之后，接近最大CPU内存和使用量的集群可能会遇到问题。对于主实例组而言，情况尤其如此。
- 当实例组调整大小时，您无法提交重新配置请求。如果在调整实例组大小时启动重新配置，在实例组完成大小调整后，才能启动重新配置，反之亦然。
- 重新配置实例组后，Amazon EMR 会重新启动应用程序以使新配置生效。如果在重新配置期间正在使用应用程序，则可能会出现作业失败或其它意外应用程序行为。
- 如果实例组的重新配置失败，Amazon EMR 会将配置参数恢复到之前的工作版本。如果恢复过程也失败，您必须提交新的 ModifyInstanceGroup 请求以从 SUSPENDED 状态中恢复实例组。
- Phoenix 配置分类的重新配置请求仅在亚马逊EMR版本 5.23.0 及更高版本中受支持，亚马逊EMR版本 5.21.0 或 5.22.0 不支持。



- 只有亚马逊版本 5.30.0 及更高版本支持HBase配置分类的重新配置请求，亚马逊EMR版本 5.23.0 到 5.29.0 不支持重新配置请求。EMR
- 只有在 Amazon 5.27.0 及更高EMR版本中，Amazon 才EMR支持在具有多个主节点的 Amazon EMR 集群上请求应用程序重新配置。
- 具有多个主节点的 Amazon EMR 集群不支持重新KMS配置hdfs-encryption-zones分类或任何 Hadoop 配置分类。
- Amazon EMR 目前不支持某些需要重新启动的容量计划程序的重新配置请求。YARN ResourceManager例如，您无法完全删除队列。

## 在控制台中重新配置实例组

### Note

Amazon EMR 控制台不支持合并类型重新配置。

1. 打开亚马逊EMR控制台，[网址为 https://console.aws.amazon.com/emr](https://console.aws.amazon.com/emr)
2. 在集群列表中，在 Name (名称) 下面选择要重新配置的活动集群。
3. 打开集群的集群详细信息页面，然后转到 Configurations (配置) 选项卡。
4. 在 Filter (筛选条件) 下拉列表中，选择要重新配置的实例组。
5. 在重新配置下拉菜单中，选择在表格中编辑或在JSON文件中编辑。
  - Edit in table (在表中编辑) – 在配置分类表中，编辑现有配置的属性和值，或者选择 Add configuration (添加配置) 来提供额外的配置分类。
  - 在JSON文件中编辑-直接在中输入配置JSON，或使用速记语法 (以阴影文本演示)。否则，请 URI为带有JSONConfigurations对象的文件提供 Amazon S3。

### Note

配置分类表中的 Source (源) 列表示是在您创建集群时提供配置，还是在您为该实例组指定额外的配置时提供配置。您可以编辑来自两个来源的实例组配置。您无法删除初始集群配置，但可以覆盖实例组的这些配置。

您还可以直接在表中添加或编辑嵌套的配置分类。例如，要提供 `hadoop-env` 的额外 `export` 子分类，请在表中添加一个 `hadoop.export` 配置分类。然后，为该分类提供特定的属性和值。

6. (可选) 选择 `Apply this configuration to all active instance groups` (将该配置应用于所有活动实例组)。
7. 保存更改。

## 使用重新配置实例组 CLI

使用 `modify-instance-groups` 命令为运行的集群中的一个实例组指定新配置。

### Note

在以下示例中，替换 `<j-2AL4XXXXXX5T9>` 使用您的集群 ID，然后替换 `<ig-1xxxxxxx9>` 使用您的实例组 ID。

### Example – 替换实例组的配置

以下示例引用了一个名 `instanceGroups.json` 为的配置JSON文件，该文件用于编辑实例组的 `YARNNodeManager` 磁盘运行状况检查器的属性。

1. 准备配置分类，并在运行命令的相同目录中将其保存为 `instanceGroups.json`。

```
[
  {
    "InstanceGroupId": "<ig-1xxxxxxx9>",
    "Configurations": [
      {
        "Classification": "yarn-site",
        "Properties": {
          "yarn.nodemanager.disk-health-checker.enable": "true",
          "yarn.nodemanager.disk-health-checker.max-disk-utilization-per-disk-percentage": "100.0"
        },
        "Configurations": []
      }
    ]
  }
]
```

```
    ]
  }
]
```

## 2. 运行以下命令。

```
aws emr modify-instance-groups --cluster-id <j-2AL4XXXXXX5T9> \
--instance-groups file://instanceGroups.json
```

### Example – 为实例组添加配置

如果要为实例组添加配置，还必须在新的 ModifyInstanceGroup 请求中包含以前为实例组指定的配置。否则，将删除以前指定的配置。

以下示例为YARN NodeManager 虚拟内存检查器添加了一个属性。该配置还包括先前为YARN NodeManager 磁盘运行状况检查器指定的值，因此这些值不会被覆盖。

## 1. 准备 instanceGroups.json 文件中具有的以下内容，并将其保存到您将在其中运行该命令的同一目录中。

```
[
  {
    "InstanceGroupId": "<ig-1xxxxxxx9>",
    "Configurations": [
      {
        "Classification": "yarn-site",
        "Properties": {
          "yarn.nodemanager.disk-health-checker.enable": "true",
          "yarn.nodemanager.disk-health-checker.max-disk-utilization-per-disk-percentage": "100.0",
          "yarn.nodemanager.vmem-check-enabled": "true",
          "yarn.nodemanager.vmem-pmem-ratio": "3.0"
        },
        "Configurations": []
      }
    ]
  }
]
```

## 2. 运行以下命令。

```
aws emr modify-instance-groups --cluster-id <j-2AL4XXXXXX5T9> \  
--instance-groups file://instanceGroups.json
```

### Example – 使用 Merge ( 合并 ) 类型重新配置将配置添加到实例组中

当您想要使用默认的覆盖重新配置方法添加配置时，您必须在新 ModifyInstanceGroup 请求中包括该实例组的所有以前指定的配置。否则，覆盖会删除您以前指定的配置。您不需要使用合并重新配置执行此操作。相反，您必须确保您的请求仅包括新配置。

以下示例为YARN NodeManager 虚拟内存检查器添加了一个属性。由于这是合并类型的重新配置，因此它不会覆盖先前为YARNNodeManager 磁盘运行状况检查器指定的值。

1. 准备 instanceGroups.json 文件中具有的以下内容，并将其保存到您将在其中运行该命令的同一目录中。

```
[  
  {"InstanceGroupId": "<ig-1xxxxxxx9>",  
   "ReconfigurationType": "MERGE",  
   "Configurations": [  
     {"Classification": "yarn-site",  
      "Properties": {  
        "yarn.nodemanager.vmem-check-enabled": "true",  
        "yarn.nodemanager.vmem-pmem-ratio": "3.0"  
      }  
    },  
     "Configurations": []  
   ]  
 }  
]
```

## 2. 运行以下命令。

```
aws emr modify-instance-groups --cluster-id <j-2AL4XXXXXX5T9> \  
--instance-groups file://instanceGroups.json
```

## Example – 删除实例组的配置

要删除实例组的配置，请提交新的重新配置请求以排除以前的配置。

### Note

您只能覆盖初始集群配置。您无法删除该配置。

例如，要从上一个示例中删除YARN NodeManager 磁盘运行状况检查器的配置，请提交instanceGroups.json包含以下内容的新配置。

```
[
  {
    "InstanceGroupId": "<ig-1xxxxxxx9>",
    "Configurations": [
      {
        "Classification": "yarn-site",
        "Properties": {
          "yarn.nodemanager.vmem-check-enabled": "true",
          "yarn.nodemanager.vmem-pmem-ratio": "3.0"
        },
        "Configurations": []
      }
    ]
  }
]
```

### Note

要删除上一重新配置请求中的所有配置，请提交包含一组空配置的重新配置请求。例如：

```
[
  {
    "InstanceGroupId": "<ig-1xxxxxxx9>",
    "Configurations": []
  }
]
```

## Example – 在一个请求中重新配置实例组并调整其大小

以下示例JSON演示了如何在同一个请求中重新配置实例组并调整其大小。

```
[
  {
    "InstanceGroupId": "<ig-1xxxxxxx9>",
    "InstanceCount": 5,
    "EC2InstanceIdsToTerminate": ["i-123"],
    "ForceShutdown": true,
    "ShrinkPolicy": {
      "DecommissionTimeout": 10,
      "InstanceResizePolicy": {
        "InstancesToTerminate": ["i-123"],
        "InstancesToProtect": ["i-345"],
        "InstanceTerminationTimeout": 20
      }
    },
    "Configurations": [
      {
        "Classification": "yarn-site",
        "Configurations": [],
        "Properties": {
          "yarn.nodemanager.disk-health-checker.enable": "true",
          "yarn.nodemanager.disk-health-checker.max-disk-utilization-per-disk-
percentage": "100.0"
        }
      }
    ]
  }
]
```

## 使用 Java 重新配置实例组 SDK

### Note

在以下示例中，替换 `<j-2AL4XXXXXX5T9>` 使用您的集群 ID，然后替换 `<ig-1xxxxxxx9>` 使用您的实例组 ID。

以下代码段使用 AWS SDK for Java 为实例组提供新配置。

```
AWSCredentials credentials = new BasicAWSCredentials("access-key", "secret-key");
AmazonElasticMapReduce emr = new AmazonElasticMapReduceClient(credentials);

Map<String,String> hiveProperties = new HashMap<String,String>();
hiveProperties.put("hive.join.emit.interval","1000");
hiveProperties.put("hive.merge.mapfiles","true");

Configuration configuration = new Configuration()
    .withClassification("hive-site")
    .withProperties(hiveProperties);

InstanceGroupModifyConfig igConfig = new InstanceGroupModifyConfig()
    .withInstanceId("<ig-1xxxxxxx9>")
    .withReconfigurationType("MERGE");
    .withConfigurations(configuration);

ModifyInstanceGroupsRequest migRequest = new ModifyInstanceGroupsRequest()
    .withClusterId("<j-2AL4XXXXXX5T9>")
    .withInstanceGroups(igConfig);

emr.modifyInstanceGroups(migRequest);
```

以下代码段通过提供一组空配置来删除以前为实例组指定的配置。

```
List<Configuration> configurations = new ArrayList<Configuration>();

InstanceGroupModifyConfig igConfig = new InstanceGroupModifyConfig()
    .withInstanceId("<ig-1xxxxxxx9>")
    .withConfigurations(configurations);

ModifyInstanceGroupsRequest migRequest = new ModifyInstanceGroupsRequest()
    .withClusterId("<j-2AL4XXXXXX5T9>")
    .withInstanceGroups(igConfig);

emr.modifyInstanceGroups(migRequest);
```

## 对实例组重新配置问题进行故障排查

如果实例组的重新配置过程失败，Amazon 会EMR恢复重新配置并使用 Amazon 事件记录失败消息。CloudWatch 此事件能够提供该重新配置失败的简短摘要。其会列出重新配置失败的实例以及相应的失败消息。下面是一个失败消息示例。

```
The reconfiguration operation for instance group ig-1xxxxxxx9 in Amazon EMR
cluster j-2AL4XXXXXX5T9 (ExampleClusterName)
failed at 2021-01-01 00:00 UTC and took 2 minutes to fail. Failed configuration version
is example12345.
Failure message: Instance i-xxxxxxx1, i-xxxxxxx2, i-xxxxxxx3 failed with message "This
is an example failure message".
```

要收集有关重新配置失败的更多数据，您可以检查节点预配置日志。对于收到类似以下消息时，这样做尤其有用。

```
i-xxxxxxx1 failed with message "Unable to complete transaction and some changes were
applied."
```

### On the node

通过连接到节点来访问节点预配置日志

1. 用于SSH连接到重新配置失败的节点。有关说明，请参阅《亚马逊 [Linux 实例EC2用户指南](#)》中的“[连接到您的 Linux 实例](#)”。
2. 导航到以下包含节点预置日志文件的目录。

```
/mnt/var/log/provision-node/
```

3. 打开 reports 子目录并搜索节点预置报告，从而进行重新配置。该reports目录按重新配置版本号、通用唯一标识符 (UUID)、Amazon EC2 实例 IP 地址和时间戳来组织日志。每个报告都是一个压缩YAML文件，其中包含有关重新配置过程的详细信息。

以下为报告文件名和路径的示例。

```
/reports/2/ca598xxx-cxxx-4xxx-bxxx-6dbxxxxxxxxxxx/ip-10-73-xxx-
xxx.ec2.internal/202104061715.yaml.gz
```

4. 您可以使用文件查看器（如以下示例中的 zless）来查看报告。



```
zless 202104061715.yaml.gz
```

## Amazon S3

使用 Amazon S3 来访问节点预配置日志

1. 登录 AWS Management Console 并打开 Amazon S3 控制台，网址为<https://console.aws.amazon.com/s3/>。
2. 打开您在配置集群时指定的 Amazon S3 存储桶，以便将日志文件存档。
3. 导航到以下包含节点预配置日志文件的文件夹：

```
DOC-EXAMPLE-BUCKET/elasticmapreduce/<cluster id>/node/<instance id>/provision-  
node/
```

4. 打开 reports 文件夹并搜索节点预配置报告，从而进行重新配置。该reports文件夹按重新配置版本号、通用唯一标识符 (UUID)、Amazon EC2 实例 IP 地址和时间戳来组织日志。每个报告都是一个压缩YAML文件，其中包含有关重新配置过程的详细信息。

以下为报告文件名和路径的示例。

```
/reports/2/ca598xxx-cxxx-4xxx-bxxx-6dbxxxxxxxxxx/ip-10-73-xxx-  
xxx.ec2.internal/202104061715.yaml.gz
```

5. 要查看日志文件，您可以将其作为文本文件从 Amazon S3 下载到本地计算机。有关说明，请参阅[下载对象](#)。

每个日志文件都包含关联重新配置の詳細预配置报告。要查找错误消息信息，您可以搜索报告的 err 日志级别。报告格式取决于您的集群上的 Amaz EMR on 版本。

以下示例显示了 5.32.0 和 6.2.0 之前的亚马逊EMR发行版本的错误信息。

```
- !ruby/object:Puppet::Util::Log  
  level: !ruby/sym err  
  tags:  
    - err  
  message: "Example detailed error message."  
  source: Puppet  
  time: 2021-01-01 00:00:00.000000 +00:00
```

亚马逊EMR发布的版本 5.32.0 和 6.2.0 及更高版本改为使用以下格式。

```
- level: err
  message: 'Example detailed error message.'
  source: Puppet
  tags:
  - err
  time: '2021-01-01 00:00:00.000000 +00:00'
  file:
  line:
```

## 在 AWS Secrets Manager 中存储敏感配置数据

Amazon 以明文形式EMR描述和列出API发出自定义配置数据的操作（例如DescribeCluster和ListInstanceGroups）。亚马逊与EMR集成，AWS Secrets Manager 因此您可以将数据存储在 Secrets Manager 中，并在配置ARN中使用该密钥。这样，您就不会以明文形式将敏感配置数据传递给 AmazonEMR，也不会将其暴露给外部。APIs如果您指明密钥值对包含存储在 Secrets Manager 中的密钥，则 Amazon 在向集群发送配置数据时会EMR检索此密钥。ARN当 Amazon 使用外部方式显示配置时APIs，它EMR不会发送注释。

### 创建密钥

要创建密钥，请遵循《AWS Secrets Manager 用户指南》中[创建 AWS Secrets Manager 密钥](#)的说明。在步骤 3 中，必须选择 Plaintext（明文）字段来输入您的敏感值。

请注意，虽然 Secrets Manager 允许密钥最多包含 65536 字节，但亚马逊将属性密钥（不包括注释）和检索到的密钥值的总长度EMR限制为 1024 个字符。

### 授予 Amazon 检索密钥的EMR权限

Amazon EMR 使用IAM服务角色为您预置和管理集群。Amazon 的服务角色EMR定义了亚马逊在预置资源和执行不在集群内运行的 Amazon EC2 实例环境中执行的服务级别任务EMR时允许执行的操作。有关服务角色的更多信息，请参阅 [Amazon 的服务角色EMR \(EMR角色\)](#) 和 [自定义IAM角色](#)。

要允许亚马逊EMR从 Secrets Manager 中检索密钥值，请在启动集群时向您的亚马逊EMR角色添加以下政策声明。

```
{
  "Sid": "AllowSecretsRetrieval",
  "Effect": "Allow",
```

```

    "Action": "secretsmanager:GetSecretValue",
    "Resource": [
      "arn:aws:secretsmanager:<region>:<aws-account-id>:secret:<secret-name>"
    ]
  }
}

```

如果您使用客户管理的密钥创建密钥 AWS KMS key，则还必须向 Amazon EMR 角色添加您使用的密钥的 `kms:Decrypt` 权限。有关更多信息，请参阅 AWS Secrets Manager 用户指南中的 [AWS Secrets Manager 的身份验证和访问控制](#)。

## 在配置分类中使用密钥

您可以向任何配置属性添加 `EMR.secret@` 注释，以表明其键值对包含存储在 Secrets Manager 中的密钥的。

以下示例说明如何在配置分类 ARN 中提供密钥：

```

{
  "Classification": "core-site",
  "Properties": {
    "presto.s3.access-key": "<sensitive-access-key>",
    "EMR.secret@presto.s3.secret-key": "arn:aws:secretsmanager:<region>:<aws-account-id>:secret:<secret-name>"
  }
}

```

当您创建集群并提交带注释的配置时，Amazon EMR 会验证配置属性。如果您的配置有效，Amazon 会从配置中删除注释并从 Secrets Manager 中检索密钥以创建实际配置，然后再将其应用于集群：

```

{
  "Classification": "core-site",
  "Properties": {
    "presto.s3.access-key": "<sensitive-access-key>",
    "presto.s3.secret-key": "<my-secret-key-retrieved-from-Secrets-Manager>"
  }
}

```

当您调用类似的操作时 `DescribeCluster`，Amazon EMR 返回集群上的当前应用程序配置。如果应用程序配置属性被标记为包含密钥 ARN，则该 `DescribeCluster` 调用返回的应用程序配置包含密钥值 ARN，而不是密钥值。这样可以确保密钥值仅在集群上可见：

```
{
  "Classification": "core-site",
  "Properties": {
    "presto.s3.access-key": "<sensitive-access-key>",
    "presto.s3.secret-key": "arn:aws:secretsmanager:<region>:<aws-account-id>:secret:<secret-name>"
  }
}
```

## 更新密钥值

每当附加的实例组启动、重新配置或调整大小时，Amazon 都会从带注释的配置中EMR检索密钥值。您可以使用 Secrets Manager 修改正在运行的集群的配置中使用的密钥的值。完成后，您可以向想要接收更新值的每个实例组提交重新配置请求。如需详细了解如何重新配置实例组以及重新配置时需要考虑的事项，请参阅 [在正在运行的集群中重新配置实例组](#)。

## 配置应用程序来使用特定 Java 虚拟机

Amazon EMR 版本具有不同的默认 Java 虚拟机 (JVM) 版本。本页说明了对不同版本和应用程序的 JVM 支持。

### 注意事项

有关应用程序支持的 Java 版本的信息，请参阅 [Amazon EMR 发行指南](#) 中的应用程序页面。

- Amazon EMR 仅支持在一个集群中运行一个运行时版本，不支持在同一集群上的不同运行时版本上运行不同的节点或应用程序。
- 对于亚马逊 EMR 7.x，支持 Java 17 的应用程序的默认 Java 虚拟机 (JVM) 为 Java 17，Apache Livy 除外。有关应用程序支持的JDK版本的更多信息，请参阅《Amazon 发行指南》中的相应EMR 发布页面。
- 从亚马逊 EMR 7.1.0 开始，Flink 支持并默认设置为 Java 17。要使用其他版本的 Java 运行时，请覆盖中的设置flink-conf。有关将 Flink 配置为使用 Java 8 或 Java 11 的更多信息，请参阅[将 Flink 配置为使用 Java 11 运行](#)。
- 对于亚马逊 EMR 5.x 和 6.x，默认 Java 虚拟机 (JVM) 是 Java 8。
  - 对于亚马逊EMR版本 6.12.0 及更高版本，某些应用程序还支持 Java 11 和 17。
  - 对于亚马逊6.9.0及更高EMR版本，Trino默认支持Java 17。有关 Trino 上 Java 17 的更多信息，请参阅 Trino 博客上的 [Trino updates to Java 17](#)。

在选择运行时系统版本时，请记住以下特定于应用程序的注意事项：

### 特定于应用程序的 Java 配置说明

应用程序	Java 配置说明
Spark	<p>要使用非默认 Java 版本运行 Spark，必须同时配置 Spark 和 Hadoop。有关示例，请参阅<a href="#">覆盖 JVM</a>。</p> <ul style="list-style-type: none"> <li>在 spark-env 中配置 JAVA_HOME 以更新主实例进程的 Java 运行时系统。例如，spark-submit、spark-shell 和 Spark 历史记录服务器。</li> <li>修改 Hadoop 配置以更新 Spark 执行器的 Java 运行时和 YARN ApplicationMaster</li> </ul>
火花 RAPIDS	你可以RAPIDS使用为 Spark 配置的 Java 版本运行。
Iceberg	您可以使用正在使用 Iceberg 的应用程序的已配置 Java 版本来运行 Iceberg。
Delta	您可以使用正在使用 Delta 的应用程序的已配置 Java 版本来运行 Delta。
Hudi	您可以使用正在使用 Hudi 的应用程序的已配置 Java 版本来运行 Hudi。
Hadoop	要更新JVM适用于 Hadoop 的，请进行修改。hadoop-env 有关示例，请参阅 <a href="#">覆盖 JVM</a> 。
Hive	要将 Hive 的 Java 版本设置为 11 或 17，请将 Hadoop JVM 设置配置为要使用的 Java 版本。
HBase	要更新 fo JVM rHBase，请修改hbase-env。默认情况下，除非您覆盖中的EMR设置，否则

应用程序	Java 配置说明
Flink	<p>Amazon 会HBaseJVM根据 Hadoop 的JVM配置进行设置。hbase-env 有关示例，请参阅<a href="#">覆盖 JVM</a>。</p> <p>要更新 Flink JVM 的，请进行修改flink-conf。默认情况下，除非您覆盖中的EMR设置，否则亚马逊会JVM根据 Hadoop 的JVM配置设置 Flink。flink-conf 有关更多信息，请参阅<a href="#">将 Flink 配置为使用 Java 11 运行</a>。</p>
Oozie	<p>要将 Oozie 配置为在 Java 11 或 17 上运行，请配置 Oozie Server、Oozie LauncherAM Launcher AM，然后更改客户端可执行文件和作业配置。您也可以将 EmbeddedOozieServer 配置为在 Java 17 上运行。有关更多信息，请参阅<a href="#">为 Oozie 配置 Java 版本</a>。</p>
Pig	<p>Pig 仅支持 Java 8。您无法在 Hadoop 中使用 Java 11 或 17，也不能在同一个集群上运行 Pig。</p>

## 覆盖 JVM

要覆盖亚马逊EMR版本的JVM设置（例如，在使用 Amazon 6.12.0 EMR 版本的集群中使用 Java 17），请为其环境分类提供该JAVA\_HOME设置，该设置`application-env`适用于除了 Flink 之外的所有应用程序。对于 Flink 来说，环境分类是 `flink-conf`。有关使用 Flink 配置 Java 运行时系统的步骤，请参阅[将 Flink 配置为使用 Java 11 运行](#)。

### 主题

- [使用 Apache Spark 覆盖该JVM设置](#)
- [使用 Apache 覆盖该JVM设置 HBase](#)
- [使用 Apache Hadoop 和 Hive 覆盖JVM设置](#)

## 使用 Apache Spark 覆盖该JVM设置

在 Amazon 6.12 及更高EMR版本中使用 Spark 时，如果您编写驱动程序以在集群模式下提交，则驱动程序使用 Java 8，但您可以将环境设置为执行程序使用 Java 11 或 17。而且，当您在 Amazon EMR 版本低于 5.x 的情况下使用 Spark，并编写驱动程序以在集群模式下提交时，驱动程序使用 Java 7。不过，您可以设置环境以确保执行程序使用 Java 8。

要覆盖 Spark JVM 的，我们建议您同时设置 Hadoop 和 Spark 分类。

```
{
  "Classification": "hadoop-env",
    "Configurations": [
      {
        "Classification": "export",
          "Configurations": [],
          "Properties": {
            "JAVA_HOME": "/usr/lib/jvm/java-1.8.0"
          }
      }
    ],
    "Properties": {}
  },
  {
    "Classification": "spark-env",
      "Configurations": [
        {
          "Classification": "export",
            "Configurations": [],
            "Properties": {
              "JAVA_HOME": "/usr/lib/jvm/java-1.8.0"
            }
        }
      ],
      "Properties": {}
    }
  }
```

## 使用 Apache 覆盖该JVM设置 HBase

HBase要配置为使用 Java 11，可以在启动集群时设置以下配置。

```
[
```

```
{
  "Classification": "hbase-env",
  "Configurations": [
    {
      "Classification": "export",
      "Configurations": [],
      "Properties": {
        "JAVA_HOME": "/usr/lib/jvm/jre-11"
      }
    }
  ],
  "Properties": {}
}
```

## 使用 Apache Hadoop 和 Hive 覆盖JVM设置

以下示例说明如何将 Hadoop 和 Hive 的版本设置为版本 17。JVM

```
[
  {
    "Classification": "hadoop-env",
    "Configurations": [
      {
        "Classification": "export",
        "Configurations": [],
        "Properties": {
          "JAVA_HOME": "/usr/lib/jvm/jre-17"
        }
      }
    ],
    "Properties": {}
  }
]
```

## 服务端口

以下是YARN和HDFS服务端口。这些设置反映 Hadoop 默认值。其它应用程序服务托管在默认端口上，除非另有指定。有关更多信息，请参阅应用程序的项目文档。



## YARN和的端口设置 HDFS

设置	主机名/端口
<code>fs.default.name</code>	默认 ( <code>hdfs://<i>emrDeterminedIP</i>:8020</code> )
<code>dfs.datanode.address</code>	默认 ( <code>0.0.0.0:50010</code> )
<code>dfs.datanode.http.address</code>	默认 ( <code>0.0.0.0:50075</code> )
<code>dfs.datanode.https.address</code>	默认 ( <code>0.0.0.0:50475</code> )
<code>dfs.datanode.ipc.address</code>	默认 ( <code>0.0.0.0:50020</code> )
<code>dfs.http.address</code>	默认 ( <code>0.0.0.0:50070</code> )
<code>dfs.https.address</code>	默认 ( <code>0.0.0.0:50470</code> )
<code>dfs.secondary.http.address</code>	默认 ( <code>0.0.0.0:50090</code> )
<code>yarn.nodemanager.address</code>	默认 ( <code>\${yarn.nodemanager.hostname}:0</code> )
<code>yarn.nodemanager.localizer.address</code>	默认 ( <code>\${yarn.nodemanager.hostname}:8040</code> )
<code>yarn.nodemanager.webapp.address</code>	默认 ( <code>\${yarn.nodemanager.hostname}:8042</code> )
<code>yarn.resourcemanager.address</code>	默认 ( <code>\${yarn.resourcemanager.hostname}:8032</code> )
<code>yarn.resourcemanager.admin.address</code>	默认 ( <code>\${yarn.resourcemanager.hostname}:8033</code> )
<code>yarn.resourcemanager.resource-tracker.address</code>	默认 ( <code>\${yarn.resourcemanager.hostname}:8031</code> )
<code>yarn.resourcemanager.scheduler.address</code>	默认 ( <code>\${yarn.resourcemanager.hostname}:8030</code> )

设置	主机名/端口
<code>yarn.resourcemanager.webapp.address</code>	默认 ( <code>\${yarn.resourcemanager.hostname}:8088</code> )
<code>yarn.web-proxy.address</code>	默认 (no-value)
<code>yarn.resourcemanager.hostname</code>	<i>emrDeterminedIP</i>

### Note

这个词 *emrDeterminedIP* 是由 Amazon EMR 控制平面生成的 IP 地址。在较新的版本中，已删除该约定，但 `yarn.resourcemanager.hostname` 和 `fs.default.name` 设置除外。

## 应用程序用户

应用程序以自己的用户身份运行进程。例如，Hive 以用户身份 JVMs 运行 hive、以用户身份 MapReduce JVMs 运行 mapred，依此类推。在以下进程状态示例中说明了这一点。

```

USER      PID %CPU %MEM    VSZ   RSS TTY      STAT START   TIME COMMAND
hive      6452  0.2  0.7 853684 218520 ?        S1   16:32   0:13 /usr/lib/jvm/
java-openjdk/bin/java -Xmx256m -Dhive.log.dir=/var/log/hive -Dhive.log.file=hive-
metastore.log -Dhive.log.threshold=INFO -Dhadoop.log.dir=/usr/lib/hadoop
hive      6557  0.2  0.6 849508 202396 ?        S1   16:32   0:09 /usr/lib/jvm/java-
openjdk/bin/java -Xmx256m -Dhive.log.dir=/var/log/hive -Dhive.log.file=hive-server2.log
-Dhive.log.threshold=INFO -Dhadoop.log.dir=/usr/lib/hadoop/l
hbase     6716  0.1  1.0 1755516 336600 ?        S1   Jun21   2:20 /usr/lib/jvm/java-
openjdk/bin/java -Dproc_master -XX:OnOutOfMemoryError=kill -9 %p -Xmx1024m -ea -XX:
+UseConcMarkSweepGC -XX:+CMSIncrementalMode -Dhbase.log.dir=/var/
hbase     6871  0.0  0.7 1672196 237648 ?        S1   Jun21   0:46 /usr/lib/jvm/java-
openjdk/bin/java -Dproc_thrift -XX:OnOutOfMemoryError=kill -9 %p -Xmx1024m -ea -XX:
+UseConcMarkSweepGC -XX:+CMSIncrementalMode -Dhbase.log.dir=/var/
hdfs     7491  0.4  1.0 1719476 309820 ?        S1   16:32   0:22 /usr/lib/jvm/java-
openjdk/bin/java -Dproc_namenode -Xmx1000m -Dhadoop.log.dir=/var/log/hadoop-hdfs -
Dhadoop.log.file=hadoop-hdfs-namenode-ip-10-71-203-213.log -Dhadoo
yarn     8524  0.1  0.6 1626164 211300 ?        S1   16:33   0:05 /usr/lib/jvm/java-
openjdk/bin/java -Dproc_proxyserver -Xmx1000m -Dhadoop.log.dir=/var/log/hadoop-yarn -
Dyarn.log.dir=/var/log/hadoop-yarn -Dhadoop.log.file=yarn-yarn-

```

```
yarn      8646  1.0  1.2 1876916 385308 ?      S1   16:33   0:46 /usr/lib/jvm/java-  
openjdk/bin/java -Dproc_resourcemanager -Xmx1000m -Dhadoop.log.dir=/var/log/hadoop-yarn  
-Dyarn.log.dir=/var/log/hadoop-yarn -Dhadoop.log.file=yarn-y  
mapred   9265  0.2  0.8 1666628 260484 ?      S1   16:33   0:12 /usr/lib/jvm/java-  
openjdk/bin/java -Dproc_historyserver -Xmx1000m -Dhadoop.log.dir=/usr/lib/hadoop/logs -  
Dhadoop.log.file=hadoop.log -Dhadoop.home.dir=/usr/lib/hadoop
```

## 使用 Amazon EMR 项目存储库检查依赖项

您可以使用 Amazon EMR 构件存储库构建针对特定 Amazon EMR 发行版 ( 从 Amazon EMR 发布版 5.18.0 开始 ) 附带的准确版本的库和依赖项的 Apache Hive 和 Apache Hadoop 任务代码。针对存储库中的 Amazon EMR 项目进行构建可确保针对其创建任务的库的版本是在集群上运行时提供的相同版本，从而帮助避免运行时类路径问题。目前，Amazon EMR 项目仅适用于 Maven 构建。

要访问项目存储库，请将存储库 URL 添加到 Maven 设置文件或特定项目的 pom.xml 配置文件。之后，您可以在项目配置中指定依赖项。对于依赖项版本，请为 [亚马逊 EMR 5.x 发布版本](#) 上所需的发行版使用 Component Versions (组件版本) 下列出的版本。例如，[the section called “5.36.2 组件版本”](#) 上提供了最新 Amazon EMR 发行版的组件版本。如果项目的构件未在 Component Versions (组件版本) 下列出，请指定为该发行版中的 Hive 和 Hadoop 列出的版本。例如，对于 Amazon EMR 发行版 5.18.0 中的 Hadoop 组件，版本为 2.8.4-amzn-1。

项目存储库 URL 具有以下语法：

```
https://s3-endpoint/region-ID-emr-artifacts/emr-release-label/repos/maven/
```

- *s3-endpoint* 是存储库区域的 Amazon Simple Storage Service ( Amazon S3 ) 终端节点，而 *region-ID* 是对应的区域。例如，s3.us-west-1.amazonaws.com 和 us-west-1。有关更多信息，请参阅《Amazon Web Services 一般参考》中的 Amazon S3 endpoints。由于区域之间的项目不存在差异，因此，您可以为开发环境指定最方便的区域。
- *emr-release-label* 是将运行您的代码的 Amazon EMR 集群的发布标签。发行版标注的格式是 emr-x.x.x，例如 emr-5.36.2。一个 EMR 版本系列可能包含多个版本。例如，如果您使用的是 EMR 发行版 5.24.1，请在构件存储库 URL 中使用 5.24 系列中的第一个 EMR 发行版标注，即 emr-5.24.0：

```
https://s3-endpoint/region-ID-emr-artifacts/emr-5.24.0/repos/maven/
```

### Example Maven pom.xml 的配置

以下 pom.xml 示例配置 Maven 项目以使用 us-west-1 中的项目存储库针对 emr-5.18.0 Apache Hadoop 和 Apache Hive 项目进行构建。由于快照版本在项目存储库中不可用，因此已在 pom.xml 中禁用快照。以下示例中的省略号 (...) 指示忽略其它配置参数。请勿将它们复制到 Maven 项目。

```
<project>
```

```
...
<repositories>
  ...
  <repository>
    <id>emr-5.18.0-artifacts</id>
    <name>EMR 5.18.0 Releases Repository</name>
    <releases>
      <enabled>true</enabled>
    </releases>
    <snapshots>
      <enabled>false</enabled>
    </snapshots>
    <url>https://s3.us-west-1.amazonaws.com/us-west-1-emr-artifacts/emr-5.18.0/repos/
maven/</url>
  </repository>
  ...
</repositories>
...
<dependencies>
  ...
  <dependency>
    <groupId>org.apache.hive</groupId>
    <artifactId>hive-exec</artifactId>
    <version>2.3.3-amzn-2</version>
  </dependency>
  <dependency>
    <groupId>org.apache.hadoop</groupId>
    <artifactId>hadoop-common</artifactId>
    <version>2.8.4-amzn-1</version>
  </dependency>
  ...
</dependencies>

</project>
```

## EMR文件系统 (EMRFS)

EMR文件系统 (EMRFS) 是所有亚马逊EMR集群用来EMR直接从亚马逊读取和写入常规文件到 Amazon S3 的实现。HDFSEMRFS提供了在 Amazon S3 中存储永久数据以便与 Hadoop 配合使用的便利，同时还提供数据加密等功能。

数据加密允许您加密EMRFS写入 Amazon S3 的对象，并EMRFS允许您使用 Amazon S3 中的加密对象。如果您使用的是 Amazon EMR 发行版 4.8.0 或更高版本，则可以使用安全配置为 Amazon S3 中的EMRFS对象设置加密以及其他加密设置。有关更多信息，请参阅[加密选项](#)。如果您使用的是早期版本的 AmazonEMR，则可以手动配置加密设置。有关更多信息，请参阅 [使用EMRFS属性指定 Amazon S3 加密](#)。

Amazon S3 为所有LIST操作和所有GET操作提供了很强的读后写一致性。PUT AWS 区域这意味着您使用写入的内容EMRFS就是从 Amazon S3 读取的内容，对性能没有影响。有关更多信息，请参阅 [Amazon S3 数据一致性模型](#)。

在使用 Amazon EMR 发行版 5.10.0 或更高版本时，您可以根据集群用户、群组或EMRFS数据在 Amazon S3 中的位置，使用不同的IAM角色处理对 Amazon S3 的EMRFS请求。有关更多信息，请参阅[为向 Amazon S3 发出的EMRFS请求配置IAM角色](#)。

### Warning

在为运行 Apache Spark 任务的亚马逊EMR集群开启推测执行功能之前，请查看以下信息。EMRFS包括 EMRFS S3 优化的提交器，该 OutputCommitter 实现针对使用时向 Amazon S3 写入文件进行了优化。EMRFS[如果您对将数据写入 Amazon S3 且不使用 EMRFS S3 优化提交程序的应用程序打开 Apache Spark 推测执行功能，则可能会遇到-10063 中描述的数据正确性问题](#)。SPARK如果您使用的是亚马逊版本 5.19 之前的亚马逊EMR版本，或者使用ORC和CSV等格式向 Amazon S3 写入文件，则可能会发生这种情况。EMRFS S3 优化的提交器不支持这些格式。有关使用 S3 优化提交器的完整要求列表，请参阅 [EMRFS S3 优化提交者的EMRFS要求](#)。

EMRFS直接写入通常在不支持 EMRFS S3 优化的提交器时使用，例如在编写以下内容时：

- Parquet 以外的输出格式，例如ORC或文本。
- 使用 Spark 的 Hadoop 文件。RDD API
- 使用 Hive SerDe 的实木复合地板。请参阅 [Hive 元存储 Parquet 表转换](#)。

EMRFS在以下情况下不使用直接写入：

- 启用 EMRFS S3 优化的提交器时。请参阅 [EMRFS S3 优化提交者的要求](#)。
- 在 `partitionOverwriteMode` 设置为动态的情况下写入动态分区时。
- 写入自定义分区位置（例如不符合 Hive 默认分区位置约定的位置）时。
- 使用除之外的文件系统时EMRFS，例如写入HDFS或使用 S3A 文件系统。

要确定您的应用程序是否在 Amazon EMR 5.14.0 或更高版本中使用直接写入，请启用 Spark INFO 日志记录。如果 Spark 驱动程序日志或 Spark 执行器容器日志中存在包含文本“Direct Write:ENABLED”的日志行，则您的 Spark 应用程序使用直接写入进行写入。

默认情况下，Amazon EMR 开启了推测执行功能。如果以下两个条件都为真，我们强烈建议您不要启用推测执行：

- 您正将数据写入 Amazon S3。
- 数据以 Apache Parquet 以外的格式写入，或者以 Apache Parquet 格式写入，而不是使用 EMRFS S3 优化的提交器。

如果您打开 Spark 推测执行并使用 EMRFS 直接写入方式将数据写入 Amazon S3，则可能会出现间歇性数据丢失的情况。当您使用 EMRFS S3 优化的提交器向 HDFS Parquet 写入数据或在 Parquet 中写入数据时，Amazon EMR 不使用直接写入，也不会出现此问题。

如果您需要使用从 Spark EMRFS 直接写入 Amazon S3 并使用推测执行的格式写入数据，我们建议您使用 `s3DistCP` 写入输出文件，HDFS 然后将输出文件传输到 Amazon S3。

## 主题

- [一致视图](#)
- [授权访问 Amazon S3 中的 EMRFS 数据](#)
- [管理默认 AWS Security Token Service 端点](#)
- [使用 EMRFS 属性指定 Amazon S3 加密](#)

## 一致视图

### Warning

2023年6月1日，EMRFS一致观点将终止对未来亚马逊EMR版本的标准支持。EMRFS一致的视图将继续适用于现有版本。

随着2020年12月1日发布的Amazon S3强read-after-write一致性，您不再需要对您的亚马逊EMR集群使用EMRFS一致视图(EMRFSCV)。EMRFSCV是一项可选功能，允许亚马逊EMR集群检查Amazon S3对象的列表和read-after-write一致性。当您创建集群并启用EMRFS CV时，Amazon EMR会创建一个Amazon DynamoDB数据库来存储用于跟踪S3对象列表read-after-write和一致性的对象元数据。现在，您可以关闭EMRFS CV并删除它使用的DynamoDB数据库，这样就不会产生额外费用。以下过程说明了如何检查、关闭CV功能以及删除该功能使用的DynamoDB数据库。

检查你是否在使用 C EMRFS V 功能

1. 导航到 Configuration (配置) 选项卡。如果您的集群具有以下配置，则它将使用 EMRFS CV。

```
Classification=emrfs-site,Property=fs.s3.consistent,Value=true
```

2. 或者，使用 AWS CLI 来描述您的集群 [describe-clusterAPI](#)。如果输出包含 `fs.s3.consistent: true`，则您的集群使用 EMRFS CV。

关闭您的 Amazon EMR 集群上的 EMRFS CV

要关闭 EMRFS CV 功能，请使用以下三个选项之一。在将这些选项应用到生产环境之前，应先在测试环境中测试它们。

1. 停止现有集群并启动一个没有 EMRFS CV 选项的新集群。
  - a. 在停止集群之前，请务必备份数据并通知用户。
  - b. 要停止集群，请按照 [终止集群](#) 中的说明操作。
  - c. 如果您使用 Amazon EMR 控制台创建新集群，请导航至高级选项。在编辑软件设置部分，取消选择开启 EMRFS CV 的选项。如果 EMRFS 一致视图复选框可用，请将其保持未选中状态。
  - d. 如果您 AWS CLI 使用创建新集群 [create-clusterAPI](#)，请不要使用开启 EMRFS CV 的 `--emrfs` 选项。



- e. 如果您使用SDK或 AWS CloudFormation 创建新集群，请不要使用配置[一致视图](#)中列出的任何配置。

## 2. 克隆集群并删除 EMRFS CV

- a. 在 Amazon EMR 控制台中，选择使用 EMRFS CV 的集群。
- b. 在 Cluster Details ( 集群详细信息 ) 页面顶部，选择 Clone ( 克隆 )。
- c. 选择 Previous ( 上一步 ) 并导航至 Step 1: Software and Steps ( 步骤 1：软件和步骤 )。
- d. 在“编辑软件设置”中，删除 EMRFS CV。在 Edit configuration ( 编辑配置 ) 椎间盘每个。删除 emrfs-site 分类中的以下配置。如果您JSON从 S3 存储桶加载，则必须修改您的 S3 对象。

```
[
  {"classification":
    "emrfs-site",
    "properties": {
      "fs.s3.consistent.retryPeriodSeconds": "10",
      "fs.s3.consistent": "true",
      "fs.s3.consistent.retryCount": "5",
      "fs.s3.consistent.metadata.tableName": "EmrFSMetadata"
    }
  }
]
```

## 3. 从使用实例组的集群中删除 EMRFS CV

- a. 使用以下命令检查单个EMR集群是否使用EMRFS与 CV 关联的 DynamoDB 表，或者多个集群是否共享该表。表名称在 fs.s3.consistent.metadata.tableName 中指定，如 [Configure consistent view](#) ( 配置一致视图 ) 中所述。EMRFS CV 使用的默认表名是 EmrFSMetadata。

```
aws emr describe-cluster --cluster-id j-XXXXX | grep
fs.s3.consistent.metadata.tableName
```

- b. 如果您的集群不与其他集群共享您的 DynamoDB 数据库，请使用以下命令重新配置集群并停用 CV。EMRFS有关更多信息，请参阅[重新配置正在运行的集群中的实例组](#)。

```
aws emr modify-instance-groups --cli-input-json file://disable-emrfs-1.json
```

此命令打开要修改的文件。请使用以下配置修改此文件。

```
{
  "ClusterId": "j-xxxx",
  "InstanceGroups": [
    {
      "InstanceGroupId": "ig-xxxx",
      "Configurations": [
        {
          "Classification": "emrfs-site",
          "Properties": {
            "fs.s3.consistent": "false"
          },
          "Configurations": []
        }
      ]
    }
  ]
}
```

- c. 如果您的集群与其他集群共享 DynamoDB 表，请在没有群集修改共享 S3 位置中的任何对象时EMRFS关闭所有集群上的 CV。

### 删除与 CV 关联的亚马逊 DynamoDB 资源 EMRFS

从 Amazon EMR 集群中移除 EMRFS CV 后，删除与 CV 关联的 DynamoDB 资源。EMRFS在此之前，您将继续产生与 CV 相关的 DynamoDB 费用。EMRFS

1. 检查您的 DynamoDB 表的 CloudWatch 指标，并确认该表未被任何集群使用。
2. 删除 DynamoDB 表。

```
aws dynamodb delete-table --table-name <your-table-name>
```

### 删除与 EMRFS CV 关联的 Amazon SQS 资源

1. 如果您将集群配置为向 Amazon 推送不一致通知SQS，则可以删除所有SQS队列。
2. 查找中指定的 Amazon SQS 队列名称  
称fs.s3.consistent.notification.SQS.queueName，如[配置一致视图](#)中所述。默认队列名称格式为 EMRFS-Inconsistency-*<j-cluster ID>*。

```
aws sqs list-queues | grep 'EMRFS-Inconsistency'
```

```
aws sqs delete-queue --queue-url <your-queue-url>
```

## 要停止使用 EMRFS CLI

- [EMRFSCLI](#)管理 EMRFS CV 生成的元数据。随着对 EMRFS CV的标准支持在Amazon的未来版本中接近尾声EMR，对简历的支持也EMRFSCLI将接近尾声。

## 主题

- [启用一致视图](#)
- [了解EMRFS一致视图如何跟踪 Amazon S3 中的对象](#)
- [重试逻辑](#)
- [EMRFS一致的视图元数据](#)
- [为 CloudWatch 和 Amazon 配置一致性通知 SQS](#)
- [配置一致视图](#)
- [EMRFSCLI命令参考](#)

## 启用一致视图

您可以启用 Amazon S3 服务器端加密或一致视图以EMRFS使用 AWS Management Console AWS CLI、或emrfs-site配置分类。

### 使用控制台配置一致视图

1. 导航到新的 Amazon EMR 控制台，然后从侧面导航栏中选择“切换到旧主机”。有关切换到旧控制台后预期情况的更多信息，请参阅 [Using the old console](#)。
2. 依次选择 Create cluster (创建集群)、Go to advanced options (转到高级选项)。
3. 选择 Step 1: Software and Steps (步骤 1: 软件和步骤) 和 Step 2: Hardware (步骤 2: 硬件) 的设置。
4. 对于“步骤 3：常规群集设置”，在“其他选项”下，选择EMRFS一致视图。
5. 对于EMRFS元数据存储，键入元数据存储的名称。默认值为 **EmrFSMetadata**。如果 EmrFSMetadata 表不存在，则会在 DynamoDB 中为您创建该表。

**Note**

集群终止后，亚马逊EMR不会自动从 DynamoDB 中删除EMRFS元数据。

- 对于 Number of retries (重试次数)，键入一个整数值。如果检测到不一致，则EMRFS尝试调用 Amazon S3 此次数。默认值为 **5**。
- 对于 Retry period (in seconds) (重试期间 (秒))，键入一个整数值。这是在两次重试尝试之间 EMRFS等待的时间。默认值为 **10**。

**Note**

后续重试会使用指数退避。

要启动启用一致视图的集群，请使用 AWS CLI

我们建议您安装当前版本的 AWS CLI。要下载最新版本，请参阅<https://aws.amazon.com/cli/>。

**Note**

为了便于读取，包含 Linux 行继续符 (\)。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (^)。

```
aws emr create-cluster --instance-type m5.xlarge --instance-count 3 --emrfs  
  Consistent=true \  
  --release-label emr-7.2.0 --ec2-attributes KeyName=myKey
```

要检查一致视图是否已启用，请使用 AWS Management Console

- 要在控制台中检查是否启用了一致视图，请导航到 Cluster List (集群列表)，然后选择您的集群名称以查看 Cluster Details (集群详细信息)。“EMRFS一致视图”字段的值为 Enabled或Disabled。

## 通过检查 `emrfs-site.xml` 文件查看是否启用了一致视图

- 您可以通过检查集群主节点上的 `emrfs-site.xml` 配置文件，来查看是否启用了一致性。如果 `fs.s3.consistent` 的布尔值设置为 `true`，则表示已为涉及 Amazon S3 的文件系统操作启用了一致视图。

## 了解EMRFS一致视图如何跟踪 Amazon S3 中的对象

EMRFS通过将有关这些对象的信息添加到EMRFS元数据中，创建 Amazon S3 中对象的一致视图。EMRFS在以下情况下，将这些列表添加到其元数据中：

- 在 Amazon EMR 工作EMRFS过程中写入的对象。
- 使用将对象与元数据同步或导入到EMRFS元数据。EMRFS CLI

读取的EMRFS对象不会自动添加到元数据中。EMRFS删除对象时，列表仍保留在元数据中并处于已删除状态，直到使用清除该EMRFSCLI列表为止。要了解更多信息CLI，请参阅[EMRFSCLI命令参考](#)。有关清除EMRFS元数据中商品信息的更多信息，请参阅[EMRFS一致的视图元数据](#)。

对于每个 Amazon S3 操作，都会EMRFS检查元数据，以获得有关一致视图中该组对象的信息。如果在其中一个操作中EMRFS发现 Amazon S3 不一致，它将根据`emrfs-site`配置属性中定义参数重试该操作。在EMRFS用尽重试次数后，它要么抛出，要`ConsistencyException`么记录异常并继续工作流程。有关重试逻辑的更多信息，请参阅[重试逻辑](#)。您可以在日志中找到 `ConsistencyExceptions`，例如：

- `listStatus`: 元数据项没有 Amazon S3 对象 `/S3_bucket/dir/object`
- `getFileStatus`: 密钥存在`dir/file`于元数据中，但不存在于 Amazon S3 中

如果您直接从视图EMRFS一致跟踪的 Amazon S3 中删除对象，则EMRFS会将该对象视为不一致，因为它仍按照 Amazon S3 的形式列在元数据中。如果您的元数据与 Amazon S3 中的对象EMRFS轨迹不同步，则可以使用的`sync`子命令重置元数据，使其反映 Amazon S3。EMRFS CLI要了解元数据与 Amazon S3 之间的差异，请使用 `diff`。最后，EMRFS只有元数据中引用的对象具有一致的视图；同一 Amazon S3 路径中可能还有其他未被跟踪的对象。EMRFS列出 Amazon S3 路径中的对象时，它会返回元数据中正在跟踪的对象和该 Amazon S3 路径中的对象的超集。

## 重试逻辑

EMRFS尝试验证特定重试次数的元数据中跟踪的对象的列表一致性。默认值为 5。如果超过重试次数，则发起任务会返回错误，除非 `fs.s3.consistent.throwExceptionOnInconsistency` 设置为 `false` (此时仅将跟踪的对象记录为不一致)。EMRFS默认情况下使用指数退避重试策略，但您也可以将其设置为固定策略。用户还可能希望在重试特定一段时间之后继续任务的其余操作，而不是引发异常。可通过将 `fs.s3.consistent.throwExceptionOnInconsistency` 设置为 `false`，将 `fs.s3.consistent.retryPolicyType` 设置为 `fixed`，将 `fs.s3.consistent.retryPeriodSeconds` 设置为所需的值，来实现此目的。以下示例创建一个启用了一致性的集群，将记录不一致并设置 10 秒的固定重试间隔：

Example 将重试期间设置为固定量

```
aws emr create-cluster --release-label emr-7.2.0 \  
--instance-type m5.xlarge --instance-count 1 \  
--emrfs Consistent=true,Args=[fs.s3.consistent.throwExceptionOnInconsistency=false,  
fs.s3.consistent.retryPolicyType=fixed,fs.s3.consistent.retryPeriodSeconds=10] --ec2-  
attributes KeyName=myKey
```

### Note

为了便于读取，包含 Linux 行继续符 (\)。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (^)。

有关更多信息，请参阅 [一致视图](#)。

## EMRFSIMDS获取区域呼叫的配置

EMRFS依靠IMDS (实例元数据服务) 来获取实例区域和 Amazon S3、DynamoDB 或终端节点。AWS KMS 但是，IMDS它可以处理的请求数量有限，超过该限制的请求将失败。此IMDS限制可能导致初始化EMRFS失败并导致查询或命令失败。您可以使用以下随机的指数回退重试机制和 `emrfs-site.xml` 中的回退区域配置属性来解决所有重试都失败的情况。

```
<property>  
  <name>fs.s3.region.retryCount</name>  
  <value>3</value>  
  <description>  
    Maximum retries that would be attempted to get AWS region.</description>  
</property>
```

```
</description>
</property>
<property>
  <name>fs.s3.region.retryPeriodSeconds</name>
  <value>3</value>
  <description>
    Base sleep time in second for each get-region retry.
  </description>
</property>
<property>
  <name>fs.s3.region.fallback</name>
  <value>us-east-1</value>
  <description>
    Fallback to this region after maximum retries for getting AWS region have been
    reached.
  </description>
</property>
```

## EMRFS一致的视图元数据

EMRFS一致视图使用 DynamoDB 表跟踪一致性，以跟踪 Amazon S3 中已同步或由其创建的对象。EMRFS元数据用于跟踪所有操作 (读取、写入、更新和复制)。其中不存储任何实际内容。此元数据用于验证从 Amazon S3 接收的对象或元数据是否与预期内容匹配。通过此确认EMRFS，可以检查EMRFS写入 Amazon S3 的新对象或与EMRFS之同步的对象的列表 read-after-write一致性和一致性。多个集群可共享相同的元数据。

### 如何向元数据添加条目

您可以使用 `sync` 或 `import` 子命令向元数据添加条目。`sync` 反映路径中 Amazon S3 对象的状态，而 `import` 用于向元数据添加新条目。有关更多信息，请参阅[EMRFSCLI命令参考](#)。

### 如何检查元数据与 Amazon S3 中的对象之间的差异

要检查元数据与 Amazon S3 之间的差异，请使用的 `diffEMRFSCLI` 子命令。有关更多信息，请参阅[EMRFSCLI命令参考](#)。

### 如何了解元数据操作是否受限制

EMRFS将元数据读取和写入操作的默认吞吐容量限制分别设置为 500 和 100 个单位。大量对象或存储桶可能会导致操作超过此容量，此时 DynamoDB 会对它们进行限制。例如，`ProvisionedThroughputExceededException` 如果您执行的操作超过了这些容量限制，则应

用程序可能会导致EMRFS抛出。限制后，该EMRFSCLI工具会尝试使用[指数退避](#)重试写入 DynamoDB 表，直到操作完成或达到将对象从 Amazon 写入 Amazon S3 的最大重试值。EMR

您可以配置自己的吞吐容量限制。但是，对于读取和写入操作，DynamoDB 有严格的分区限制，分别为每秒 3000 个读取容量单位 RCUs WCU 和 1000 个写入容量单位 ( )。为避免限制导致sync故障，我们建议您将读取操作的吞吐量限制在 3000 RCUs 以下，将写入操作的吞吐量限制在 1000 以下。WCUs有关设置自定义吞吐容量限制的说明，请参阅 [配置一致视图](#)。

您还可以在 DynamoDB 控制台中查看EMRFS元数据的亚马逊 CloudWatch 指标，您可以在其中查看受限制的读取和写入请求数量。如果受限制的请求数不为零值，则增加为读取或写入操作分配的吞吐容量可能会使应用程序受益。如果您发现操作长时间接近分配的最大读取或写入吞吐量容量，则这样做也可能会获得性能好处。

### 重要EMRFS操作的吞吐量特性

读取和写入操作的默认值分别为 400 和 100 个吞吐量容量单位。您可以通过以下性能特征了解特定操作所需的吞吐量。这些测试是使用单节点 m3.large 集群执行的。所有操作都是单线程执行。特定应用程序特征会对性能造成很大影响，可能需要通过实验来优化文件系统操作。

操作	平均值 read-per-second	平均值 write-per-second
create ( 对象 )	26.79	6.70
delete ( 对象 )	10.79	10.79
delete ( 包含 1000 个对象的目录 )	21.79	338.40
getFileStatus ( 对象 )	34.70	0
getFileStatus ( 目录 )	19.96	0
listStatus ( 包含 1 个对象的目录 )	43.31	0
listStatus ( 包含 10 个对象的目录 )	44.34	0



操作	平均值 read-per-second	平均值 write-per-second
listStatus ( 包含 100 个对象的目录 )	84.44	0
listStatus ( 包含 1,000 个对象的目录 )	308.81	0
listStatus ( 包含 10,000 个对象的目录 )	416.05	0
listStatus ( 包含 100,000 个对象的目录 )	823.56	0
listStatus ( 包含 100 万个对象的目录 )	882.36	0
mkdir ( 持续 120 秒 )	24.18	4.03
mkdir	12.59	0
rename ( 对象 )	19.53	4.88
rename ( 包含 1000 个对象的目录 )	23.22	339.34

### 提交从元数据存储中清除旧数据的步骤

用户可能希望在基于 DynamoDB 的元数据中删除特定条目。这样可以帮助降低与表关联的存储成本。用户可以使用子命令手动或以编程方式清除特定条目。EMRFS CLI delete但是，如果您从元数据中删除条目，则EMRFS不会再进行任何一致性检查。

在任务完成后以编程方式进行清除，方法是向您的集群提交最后一个步骤，该集群将在上执行命令。EMRFS CLI例如，键入以下命令可向集群提交删除两天之前的所有条目的步骤。

```
aws emr add-steps --cluster-id j-2AL4XXXXXX5T9 --steps Name="emrfsCLI",Jar="command-runner.jar",Args=["emrfs","delete","--time","2","--time-unit","days"]
{
  "StepIds": [
```

```
    "s-B12345678902"  
  ]  
}
```

使用返回的 `StepId` 值在日志中查看操作结果。

## 为 CloudWatch 和 Amazon 配置一致性通知 SQS

您可以在 Amazon S3 中启用 CloudWatch 指标和亚马逊SQS消息，EMRFS以解决 Amazon S3 最终一致性问题。

### CloudWatch

启用 CloudWatch 指标后，由于 Amazon S3 的最终一致性，每次FileSystemAPI调用失败时，都会推送名为“不一致”的指标。

查看 Amazon S3 最终一致性问题的 CloudWatch 指标

要在 CloudWatch 控制台中查看不一致性指标，请选择EMRFS指标，然后选择 JobFlowId/指标名称对。例如：j-162XXXXXXM2CU ListStatus、j-162XXXXXXM2CU GetFileStatus 等。

1. 打开 CloudWatch 控制台，网址为<https://console.aws.amazon.com/cloudwatch/>。
2. 在控制面板的“指标”部分中，选择EMRFS。
3. 在 Job Flow Metrics 窗格中，选择一个或多个 JobFlowId/指标名称对。将在下面的窗口中显示指标的图形表示。

### 亚马逊 SQS

启用亚马逊SQS通知后，将在初始化时EMRFS创建名称EMRFS-Inconsistency-`<jobFlowId>`为的亚马逊SQS队列。由于 Amazon S3 的最终一致性，当FileSystemAPI呼叫失败时，Amazon SQS 消息会被推送到队列中。该消息包含诸如 JobFlowId、API、不一致路径列表、堆栈跟踪等信息。可以使用 Amazon SQS 控制台或EMRFSread-sqs命令读取消息。

管理 Amazon S3 最终一致性问题的亚马逊SQS消息

使用 Amazon S3 的最终一致性问题可以读取 Amazon S3 的SQS消息EMRFSCLI。要从 A EMRFS mazon SQS 队列读取消息，请键入read-sqs命令并在主节点的本地文件系统上为生成的输出文件指定输出位置。

您也可以使用delete-sqs命令删除 A EMRFS mazon SQS 队列。

1. 要从 Amazon SQS 队列读取消息，请键入以下命令。Replace ( 替换 ) *queuename* 使用您配置并替换的 Amazon SQS 队列的名称 */path/filename* 包含输出文件的路径：

```
emrfs read-sqs --queue-name queuename --output-file /path/filename
```

例如，要读取和输出默认队列中的 Amazon SQS 消息，请键入：

```
emrfs read-sqs --queue-name EMRFS-Inconsistency-j-162XXXXXXM2CU --output-file /path/filename
```

#### Note

还可以分别使用 `-q` 和 `-o` 快捷方式代替 `--queue-name` 和 `--output-file`。

2. 要删除 Amazon SQS 队列，请键入以下命令：

```
emrfs delete-sqs --queue-name queuename
```

例如，要删除默认队列，请键入：

```
emrfs delete-sqs --queue-name EMRFS-Inconsistency-j-162XXXXXXM2CU
```

#### Note

还可以使用 `-q` 快捷方式代替 `--queue-name`。

## 配置一致视图

您可以为一致视图配置其它设置，方法是使用 `emrfs-site` 属性的配置属性来提供这些设置。例如，您可以选择不同的默认 DynamoDB 吞吐量，方法是为该选项提供以下参数，使用 `emrfs-site` 配置分类（仅限 EMR Amazon 发行版 4.x 及更高版本），或者使用引导操作在主节点上配置 `emrfs-site.xml` 文件：CLI--emrfs

Example 在集群启动时更改默认元数据读取和写入值

```
aws emr create-cluster --release-label emr-7.2.0 --instance-type m5.xlarge \  
--emrfs Consistent=true,Args=[fs.s3.consistent.metadata.read.capacity=600,\  

```

```
fs.s3.consistent.metadata.write.capacity=300] --ec2-attributes KeyName=myKey
```

或者，使用下面的配置文件并将其保存到本地或 Amazon S3 中：

```
[
  {
    "Classification": "emrfs-site",
    "Properties": {
      "fs.s3.consistent.metadata.read.capacity": "600",
      "fs.s3.consistent.metadata.write.capacity": "300"
    }
  }
]
```

按照下面的语法使用您创建的配置：

```
aws emr create-cluster --release-label emr-7.2.0 --applications Name=Hive \
--instance-type m5.xlarge --instance-count 2 --configurations file:///./myConfig.json
```

#### Note

为了便于读取，包含 Linux 行继续符 ( \ )。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 ( ^ )。

可以使用配置或 AWS CLI `--emrfs` 参数设置以下选项。有关这些实参的信息，请参阅 [AWS CLI 命令参考](#)。

一致视图的 `emrfs-site.xml` 属性

属性	默认值	描述
<code>fs.s3.consistent</code>	<b>false</b>	设置为 <b>true</b> ，此属性将配置为使用 DynamoDB EMRFS 来提供一致性。
<code>fs.s3.consistent.retryPolicyType</code>	<b>exponential</b>	此属性标识针对一致性问题进行重试时要使用的策略。选项

属性	默认值	描述
		包括：exponential、fixed 和 none。
fs.s3.consistent.retryPeriodSeconds	<b>1</b>	此属性设置两次一致性重试尝试之间等待的时间长度。
fs.s3.consistent.retryCount	<b>10</b>	此属性设置检测到不一致时的最大重试次数。
fs.s3.consistent.throwExceptionOnInconsistency	<b>true</b>	此属性确定是引发还是记录一致性异常。设置为 <b>true</b> 时，会引发 ConsistencyException。
fs.s3.consistent.metadata.autoCreate	<b>true</b>	设置为 <b>true</b> 时，此属性会启用元数据表的自动创建。
fs.s3.consistent.metadata.tagVerificationEnabled	<b>true</b>	在 Amazon EMR 5.29.0 中，此属性默认处于启用状态。启用后，EMRFS使用 S3 ETags 验证正在读取的对象是否为最新的可用版本。此功能对于 S3 上的文件被覆盖，同时保留相同名称的 read-after-update 用例非常有用。此ETag验证功能目前不适用于 S3 Select。
fs.s3.consistent.metadata.tableName	<b>EmrFSMetadata</b>	此属性指定 DynamoDB 中元数据表的名称。
fs.s3.consistent.metadata.readCapacity	<b>500</b>	此属性指定创建元数据表时要配置的 DynamoDB 读取容量。
fs.s3.consistent.metadata.writeCapacity	<b>100</b>	此属性指定创建元数据表时要配置的 DynamoDB 写入容量。

属性	默认值	描述
<code>fs.s3.consistent.fastList</code>	<b>true</b>	设置为 <b>true</b> 时，此属性会使用多个线程列出目录（需要时）。必须启用一致性才能使用此属性。
<code>fs.s3.consistent.fastList.prefetchMetadata</code>	<b>false</b>	设置为 <b>true</b> 时，此属性会为包含 20000 个以上的项目的目录启用元数据预取。
<code>fs.s3.consistent.notification.CloudWatch</code>	<b>false</b>	如果设置为 <b>true</b> ，则会为因 Amazon S3 最终一致性问题而失败的 FileSystem API调用启用 CloudWatch 指标。
<code>fs.s3.consistent.notification.SQS</code>	<b>false</b>	设置为 <b>true</b> 时，最终一致性通知将推送到 Amazon SQS 队列。
<code>fs.s3.consistent.notification.SQS.queueName</code>	<b>EMRFS-Inconsistency- &lt;jobFlowId&gt;</b>	更改此属性允许您为有关 Amazon S3 最终一致性问题的消息指定自己的 SQS 队列名称。
<code>fs.s3.consistent.notification.SQS.customMsg</code>	<b>none</b>	此属性允许您指定 SQS 消息中包含的有关 Amazon S3 最终一致性问题的自定义信息。如果没有为此属性指定值，则消息中的对应字段为空。
<code>fs.s3.consistent.dynamodb.endpoint</code>	<b>none</b>	此属性允许您为一致性视图元数据指定自定义 DynamoDB 终端节点。
<code>fs.s3.useRequesterPaysHeader</code>	<b>false</b>	当设置为 <b>true</b> 时，此属性允许在启用付款人选项请求的情况下向存储桶发出 Amazon S3 请求。

## EMRFSCLI命令参考

默认情况下，EMRFSCLI使用 Amazon EMR 发行版 3.2.1 或更高版本创建的所有集群主节点上都安装了。您可以使用来管理元数据EMRFSCLI以获得一致的视图。

### Note

只有 VT1 00 终端仿真支持该emrfs命令。但是，它可能适用于其它终端仿真器模式。

## emrfs 顶级命令

emrfs 顶级命令支持以下结构。

```
emrfs [describe-metadata | set-metadata-capacity | delete-metadata | create-metadata |
\
list-metadata-stores | diff | delete | sync | import ] [options] [arguments]
```

指定 [选项]，根据需要使用或不使用下表中描述的 [参数]。有关特定于子命令 (describe-metadata、set-metadata-capacity 等) 的 [选项]，请参阅下面的每个子命令。

### emrfs 的 [选项]

选项	描述	必填
-a <i>AWS_ACCESS_KEY_ID</i>   --access-key <i>AWS_ACCESS_KEY_ID</i>	用于向 Amazon S3 写入对象以及在 DynamoDB 中创建或访问元数据存储的访问密钥。AWS 默认情况下， <i>AWS_ACCESS_KEY_ID</i> 设置为用于创建集群的访问密钥。	否
-s <i>AWS_SECRET_ACCESS_KEY</i>   --secret-key <i>AWS_SECRET_ACCESS_KEY</i>	与您用于向 Amazon S3 写入对象以及在 DynamoDB 中创建或访问元数据存储的访问 AWS 密钥关联的密钥。默认情况下， <i>AWS_SECRET_ACCESS_KEY</i> 设置为与用于创建集群的访问密钥关联的密钥。	否
-v   --verbose	使输出为详细模式。	否

选项	描述	必填
<code>-h   --help</code>	通过用法语句显示 <code>emrfs</code> 命令的帮助消息。	否

## emrfs describe-metadata 子命令

### emrfs describe-metadata 的 [选项]

选项	描述	必填
<code>-m <i>METADATA_NAME</i>   --metadata-name <i>METADATA_NAME</i></code>	<i>METADATA_NAME</i> 是 DynamoDB 元数据表的名称。如果 <i>METADATA_NAME</i> 未提供参数，默认值为 <code>EmrFSMetadata</code> 。	否

### Example emrfs describe-metadata 示例

以下示例描述默认元数据表。

```
$ emrfs describe-metadata
EmrFSMetadata
  read-capacity: 400
  write-capacity: 100
  status: ACTIVE
  approximate-item-count (6 hour delay): 12
```

## emrfs set-metadata-capacity 子命令

### emrfs 的 [选项] set-metadata-capacity

选项	描述	必填
<code>-m <i>METADATA_NAME</i>   --metadata-name <i>METADATA_NAME</i></code>	<i>METADATA_NAME</i> 是 DynamoDB 元数据表的名称。如果 <i>METADATA_NAME</i> 未提供参数，默认值为 <code>EmrFSMetadata</code> 。	否
	对元数据表请求的读取吞吐容量。如果 <i>READ_CAPACITY</i> 未提供参数，默认值为 <code>400</code> 。	否



选项	描述	必填
<code>-r <i>READ_CAPACITY</i>   --read-capacity <i>READ_CAPACITY</i></code>		
<code>-w <i>WRITE_CAPACITY</i>   --write-capacity <i>WRITE_CAPACITY</i></code>	对元数据表请求的写入吞吐容量。如果 <i>WRITE_CAPACITY</i> 未提供参数，默认值为100。	否

### Example emrfs 示例 set-metadata-capacity

以下示例将名为 `600` 的元数据表的读取吞吐容量设置为 `150`，将写入容量设置为 `EmrMetadataAlt`。

```
$ emrfs set-metadata-capacity --metadata-name EmrMetadataAlt --read-capacity 600 --
write-capacity 150
  read-capacity: 400
  write-capacity: 100
  status: UPDATING
  approximate-item-count (6 hour delay): 0
```

### emrfs delete-metadata 子命令

emrfs delete-metadata 的 [选项]

选项	描述	必填
<code>-m <i>METADATA_NAME</i>   --metadata-name <i>METADATA_NAME</i></code>	<i>METADATA_NAME</i> 是 DynamoDB 元数据表的名称。如果 <i>METADATA_NAME</i> 未提供参数，默认值为 <code>EmrFSMetadata</code> 。	否

### Example emrfs delete-metadata 示例

以下示例删除默认元数据表。

```
$ emrfs delete-metadata
```

## emrfs create-metadata 子命令

emrfs create-metadata 的 [选项]

选项	描述	必填
-m <i>METADATA_NAME</i>   --metadata-name <i>METADATA_NAME</i>	<i>METADATA_NAME</i> 是 DynamoDB 元数据表的名称。如果 <i>METADATA_NAME</i> 未提供参数，默认值为 EmrFSMetadata。	否
-r <i>READ_CAPACITY</i>   --read-capacity <i>READ_CAPACITY</i>	对元数据表请求的读取吞吐容量。如果 <i>READ_CAPACITY</i> 未提供参数，默认值为400。	否
-w <i>WRITE_CAPACITY</i>   --write-capacity <i>WRITE_CAPACITY</i>	对元数据表请求的写入吞吐容量。如果 <i>WRITE_CAPACITY</i> 未提供参数，默认值为100。	否

### Example emrfs create-metadata 示例

以下示例创建一个名为 EmrFSMetadataAlt 的元数据表。

```
$ emrfs create-metadata -m EmrFSMetadataAlt
Creating metadata: EmrFSMetadataAlt
EmrFSMetadataAlt
  read-capacity: 400
  write-capacity: 100
  status: ACTIVE
  approximate-item-count (6 hour delay): 0
```

## emrfs list-metadata-stores 子命令

emrfs list-metadata-stores 子命令没有任何 [选项]。

### Example List-metadata-stores 示例

以下示例列出您的元数据表。

```
$ emrfs list-metadata-stores
EmrFSMetadata
```

## emrfs diff 子命令

### emrfs diff 的 [选项]

选项	描述	必填
<code>-m <i>METADATA_NAME</i></code>   <code>--metadata-name <i>METADATA_NAME</i></code>	<i>METADATA_NAME</i> 是 DynamoDB 元数据表的名称。如果 <i>METADATA_NAME</i> 未提供参数，默认值为 EmrFSMetadata。	否
<code><i>s3://s3Path</i></code>	与元数据表进行比较的 Amazon S3 存储桶的路径。存储桶以递归方式同步。	是

### Example emrfs diff 示例

以下示例将默认元数据表与 Amazon S3 存储桶进行比较。

```
$ emrfs diff s3://elasticmapreduce/samples/cloudfront
BOTH | MANIFEST ONLY | S3 ONLY
DIR elasticmapreduce/samples/cloudfront
DIR elasticmapreduce/samples/cloudfront/code/
DIR elasticmapreduce/samples/cloudfront/input/
DIR elasticmapreduce/samples/cloudfront/logprocessor.jar
DIR elasticmapreduce/samples/cloudfront/input/XABCD12345678.2009-05-05-14.WxYz1234
DIR elasticmapreduce/samples/cloudfront/input/XABCD12345678.2009-05-05-15.WxYz1234
DIR elasticmapreduce/samples/cloudfront/input/XABCD12345678.2009-05-05-16.WxYz1234
DIR elasticmapreduce/samples/cloudfront/input/XABCD12345678.2009-05-05-17.WxYz1234
DIR elasticmapreduce/samples/cloudfront/input/XABCD12345678.2009-05-05-18.WxYz1234
DIR elasticmapreduce/samples/cloudfront/input/XABCD12345678.2009-05-05-19.WxYz1234
DIR elasticmapreduce/samples/cloudfront/input/XABCD12345678.2009-05-05-20.WxYz1234
DIR elasticmapreduce/samples/cloudfront/code/cloudfront-loganalyzer.tgz
```

## emrfs delete 子命令

### emrfs delete 的 [选项]

选项	描述	必填
<code>-m <i>METADATA_NAME</i>   --metadata-name <i>METADATA_NAME</i></code>	<i>METADATA_NAME</i> 是 DynamoDB 元数据表的名称。如果 <i>METADATA_NAME</i> 未提供参数，默认值为 <code>EmrFSMetadata</code> 。	否
<code><i>s3://s3Path</i></code>	为实现一致视图而跟踪的 Amazon S3 存储桶的路径。存储桶以递归方式同步。	是
<code>-t <i>TIME</i>   --time <i>TIME</i></code>	过期时间 (使用时间单位参数进行解释)。所有早于 <i>TIME</i> 已删除指定存储桶的参数。	
<code>-u <i>UNIT</i>   --time-unit <i>UNIT</i></code>	用于解释时间参数的度量值 (纳秒、微秒、毫秒、秒、分钟、小时或天)。如果未指定参数，则默认值为 <code>days</code> 。	
<code>--read-consumption <i>READ_CONSUMPTION</i></code>	请求用于 delete 操作的可用读取吞吐量。如果 <i>READ_CONSUMPTION</i> 未指定参数，默认值为 <code>400</code> 。	否
<code>--write-consumption <i>WRITE_CONSUMPTION</i></code>	请求用于 delete 操作的可用写入吞吐量。如果 <i>WRITE_CONSUMPTION</i> 未指定参数，默认值为 <code>100</code> 。	否

### Example emrfs delete 示例

以下示例从一致视图的跟踪元数据中删除一个 Amazon S3 存储桶中的所有对象。

```
$ emrfs delete s3://elasticmapreduce/samples/cloudfront
entries deleted: 11
```

## emrfs import 子命令

for emrfs import 的 [选项]

选项	描述	必填
<code>-m <i>METADATA_NAME</i></code>   <code>--metadata-name <i>METADATA_NAME</i></code>	<i>METADATA_NAME</i> 是 DynamoDB 元数据表的名称。如果 <i>METADATA_NAME</i> 未提供参数，默认值为 <code>EmrFSMetadata</code> 。	否
<code><i>s3://s3Path</i></code>	为实现一致视图而跟踪的 Amazon S3 存储桶的路径。存储桶以递归方式同步。	是
<code>--read-consumption <i>READ_CONSUMPTION</i></code>	请求用于 delete 操作的可用读取吞吐量。如果 <i>READ_CONSUMPTION</i> 未指定参数，默认值为 400。	否
<code>--write-consumption <i>WRITE_CONSUMPTION</i></code>	请求用于 delete 操作的可用写入吞吐量。如果 <i>WRITE_CONSUMPTION</i> 未指定参数，默认值为 100。	否

### Example emrfs import 示例

以下示例随一致视图的跟踪元数据导入一个 Amazon S3 存储桶中的所有对象。忽略所有未知键。

```
$ emrfs import s3://elasticmapreduce/samples/cloudfront
```

## emrfs sync 子命令

emrfs sync 的 [选项]

选项	描述	必填
		否

选项	描述	必填
<code>-m <i>METADATA_NAME</i></code>   <code>--metadata-name <i>METADATA_NAME</i></code>	<i>METADATA_NAME</i> 是 DynamoDB 元数据表的名称。如果 <i>METADATA_NAME</i> 未提供参数，默认值为 <code>EmrFSMetadata</code> 。	
<code>s3://s3Path</code>	为实现一致视图而跟踪的 Amazon S3 存储桶的路径。存储桶以递归方式同步。	是
<code>--read-consumption <i>READ_CONSUMPTION</i></code>	请求用于 delete 操作的可用读取吞吐量。如果 <i>READ_CONSUMPTION</i> 未指定参数，默认值为 400。	否
<code>--write-consumption <i>WRITE_CONSUMPTION</i></code>	请求用于 delete 操作的可用写入吞吐量。如果 <i>WRITE_CONSUMPTION</i> 未指定参数，默认值为 100。	否

### Example emrfs sync 命令示例

以下示例随一致视图的跟踪元数据导入一个 Amazon S3 存储桶中的所有对象。删除所有未知键。

```
$ emrfs sync s3://elasticmapreduce/samples/cloudfront
Synching samples/cloudfront           0 added | 0 updated |
  0 removed | 0 unchanged
Synching samples/cloudfront/code/      1 added | 0 updated |
  0 removed | 0 unchanged
Synching samples/cloudfront/           2 added | 0 updated |
  0 removed | 0 unchanged
Synching samples/cloudfront/input/     9 added | 0 updated |
  0 removed | 0 unchanged
Done synching s3://elasticmapreduce/samples/cloudfront 9 added | 0 updated |
  1 removed | 0 unchanged
creating 3 folder key(s)
folders written: 3
```

## emrfs read-sqs 子命令

emrfs read-sqs 的 [选项]

选项	描述	必填
<code>-q <i>QUEUE_NAME</i>   --queue-name <i>QUEUE_NAME</i></code>	<i>QUEUE_NAME</i> 是在中配置的 Amazon SQS 队列的名称emrfs-site.xml 。默认值为 <b>EMRFS-Inc consistency-&lt;jobFlowId&gt;</b> 。	是
<code>-o <i>OUTPUT_FILE</i>   --output-file <i>OUTPUT_FILE</i></code>	<i>OUTPUT_FILE</i> 是主节点本地文件系统上输出文件的路径。从队列读取的消息会写入此文件。	是

## emrfs delete-sqs 子命令

emrfs delete-sqs 的 [选项]

选项	描述	必填
<code>-q <i>QUEUE_NAME</i>   --queue-name <i>QUEUE_NAME</i></code>	<i>QUEUE_NAME</i> 是在中配置的 Amazon SQS 队列的名称emrfs-site.xml 。默认值为 <b>EMRFS-Inc consistency-&lt;jobFlowId&gt;</b> 。	是

## 将EMRFSCLI命令作为步骤提交

以下示例说明如何在主节点上使用该emrfs实用程序，方法是利用 AWS CLI 或API和command-runner.jar将emrfs命令作为步骤运行。该示例使用 AWS SDK for Python (Boto3) 向集群添加一个步骤，该步骤将 Amazon S3 存储桶中的对象添加到默认EMRFS元数据表。

```
import boto3
from botocore.exceptions import ClientError

def add_emrfs_step(command, bucket_url, cluster_id, emr_client):
    """
    Add an EMRFS command as a job flow step to an existing cluster.
```

```
:param command: The EMRFS command to run.
:param bucket_url: The URL of a bucket that contains tracking metadata.
:param cluster_id: The ID of the cluster to update.
:param emr_client: The Boto3 Amazon EMR client object.
:return: The ID of the added job flow step. Status can be tracked by calling
        the emr_client.describe_step() function.
"""
job_flow_step = {
    "Name": "Example EMRFS Command Step",
    "ActionOnFailure": "CONTINUE",
    "HadoopJarStep": {
        "Jar": "command-runner.jar",
        "Args": ["/usr/bin/emrfs", command, bucket_url],
    },
}

try:
    response = emr_client.add_job_flow_steps(
        JobFlowId=cluster_id, Steps=[job_flow_step]
    )
    step_id = response["StepIds"][0]
    print(f"Added step {step_id} to cluster {cluster_id}.")
except ClientError:
    print(f"Couldn't add a step to cluster {cluster_id}.")
    raise
else:
    return step_id

def usage_demo():
    emr_client = boto3.client("emr")
    # Assumes the first waiting cluster has EMRFS enabled and has created metadata
    # with the default name of 'EmrFSMetadata'.
    cluster = emr_client.list_clusters(ClusterStates=["WAITING"])["Clusters"][0]
    add_emrfs_step(
        "sync", "s3://elasticmapreduce/samples/cloudfront", cluster["Id"], emr_client
    )

if __name__ == "__main__":
    usage_demo()
```



可以使用返回的 `step_id` 值检查日志以了解操作结果。

## 授权访问 Amazon S3 中的EMRFS数据

默认情况下，的EMR角色EC2决定访问 Amazon S3 中EMRFS数据的权限。无论通过哪个用户或群组发出请求，附加到此角色的IAM策略都适用EMRFS。默认值为 `EMR_EC2_DefaultRole`。有关更多信息，请参阅[集群EC2实例的服务角色 \( EC2实例配置文件 \)](#)。

从 Amazon EMR 发行版 5.10.0 开始，您可以使用安全配置为其指定IAM角色。EMRFS这允许您为拥有多个用户的集群自定义向 Amazon S3 发出的EMRFS请求的权限。您可以根据 Amazon S3 中的前缀为不同的用户和群组以及不同的 Amazon S3 存储桶位置指定不同的IAM角色。当向 Amazon S3 EMRFS 发出与用户、群组或您指定的位置相匹配的请求时，集群将使用您指定的相应角色而不是其中的EMR角色EC2。有关更多信息，请参阅[为向 Amazon S3 发出的EMRFS请求配置IAM角色](#)。

或者，如果您的 Amazon EMR 解决方案的要求超出了所EMRFS提供的IAM角色范围，则可以定义一个自定义凭证提供程序类别，该类允许您自定义对 Amazon S3 中EMRFS数据的访问权限。

### 为 Amazon S3 中的EMRFS数据创建自定义凭证提供程序

要创建自定义凭证提供程序，您需要实现[AWSCredentialsProvider](#)和 Hadoop [可配置](#)类。

有关此方法的详细说明，请参阅 AWS [大数据博客EMRFS中的安全分析来自其他 AWS 账户的数据](#)。博客文章包含一个教程，指导您完成从创建IAM角色到启动集群的整个过程end-to-end。其中还提供了实施自定义凭证提供程序类的 Java 代码示例。

基本步骤如下所示：

#### 指定自定义凭证提供程序

1. 创建编译为JAR文件的自定义凭证提供程序类。
2. 运行脚本作为引导操作，将自定义凭据提供程序JAR文件复制到集群主节点上的`/usr/share/aws/emr/emrfs/auxlib`位置。有关引导操作的更多信息，请参阅[\( 可选 \) 创建引导操作以安装其它软件](#)。
3. 自定义`emrfs-site`分类以指定您在JAR文件中实现的类。有关指定配置对象以自定义应用程序的更多信息，请参阅 Amazon EMR 发行指南中的[配置应用程序](#)。

以下示例演示了启动包含常见配置参数的 Hive 集群的 `create-cluster` 命令，并包括：

- 运行脚本 `copy_jar_file.sh` 的引导操作，该脚本已保存到 Amazon S3 中的 `mybucket`。

- 一种emrfs-site分类，它将JAR文件中定义的自定义凭证提供程序指定为 *MyCustomCredentialsProvider*

### Note

为了便于读取，包含 Linux 行继续符 (\)。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (^)。

```
aws emr create-cluster --applications Name=Hive \
--bootstrap-actions '[{"Path":"s3://mybucket/copy_jar_file.sh","Name":"Custom
action"}]' \
--ec2-attributes '{"KeyName":"MyKeyPair","InstanceProfile":"EMR_EC2_DefaultRole",\
"SubnetId":"subnet-xxxxxxx","EmrManagedSlaveSecurityGroup":"sg-xxxxxxx",\
"EmrManagedMasterSecurityGroup":"sg-xxxxxxx"}' \
--service-role EMR_DefaultRole_V2 --enable-debugging --release-label emr-7.2.0 \
--log-uri 's3n://my-emr-log-bucket/' --name 'test-awscredentialsprovider-emrfs' \
--instance-type=m5.xlarge --instance-count 3 \
--configurations '[{"Classification":"emrfs-site",\
"Properties":
{"fs.s3.customAWSCredentialsProvider":"MyAWSCredentialsProviderWithUri"},\
"Configurations":[]}]'
```

## 管理默认 AWS Security Token Service 端点

EMRFS使用 AWS Security Token Service (STS) 检索临时安全证书以访问您的 AWS 资源。早期的 Amazon EMR 发行版本将所有 AWS STS 请求发送到位于的单个全局终端节点 <https://sts.amazonaws.com>。亚马逊EMR发布的版本为 5.31.0 和 6.1.0 及更高版本，改为向区域 AWS STS 终端节点发出请求。这可以降低延迟并提高会话令牌的有效性。有关 AWS STS 终端节点的更多信息，请参阅《AWS Identity and Access Management 用户指南》中的在[AWS 区域中管理 AWS STS](#)。

当您使用亚马逊EMR发行版 5.31.0 和 6.1.0 及更高版本时，您可以覆盖默认终端节点。AWS STS 为此，您必须更改 emrfs-site 配置中的 fs.s3.sts.endpoint 属性。

以下 AWS CLI 示例将使用的默认 AWS STS 终端节点设置EMRFS为全局终端节点。

```
aws emr create-cluster --release-label <emr-5.33.0> --instance-type m5.xlarge \  
--emrfs Args=[fs.s3.sts.endpoint=https://sts.amazonaws.com]
```

### Note

为了便于读取，包含 Linux 行继续符 (\)。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (^)。

或者，您可以使用以下示例创建JSON配置文件，然后使用--configurations参数进行指定emr create-cluster。有关使用 --configurations，的更多信息，请参阅 [AWS CLI 命令参考](#)。

```
[  
  {  
    "classification": "emrfs-site",  
    "properties": {  
      "fs.s3.sts.endpoint": "https://sts.amazonaws.com"  
    }  
  }  
]
```

## 使用EMRFS属性指定 Amazon S3 加密

### Important


从 Amazon EMR 发行版 4.8.0 开始，您可以使用安全配置更轻松地应用加密设置，并提供更多选项。建议您使用安全配置。有关更多信息，请参阅[配置数据加密](#)。此部分中所述的控制台说明适用于 4.8.0 之前的版本。如果您使用在 AWS CLI 集群配置和后续版本的安全配置中配置 Amazon S3 加密，则安全配置将覆盖集群配置。

创建集群时，您可以使用控制台或通过或使用emrfs-site分类属性为 Amazon S3 中的EMRFS数据指定服务器端加密 (CSE) 或客户端加密 ()。SSE AWS CLI EMR SDK Amazon S3 SSE 和CSE是互斥的；您可以选择其中一个，但不能两者兼而有之。

有关 AWS CLI 说明，请参阅下面与您的加密类型对应的部分。

要指定EMRFS加密选项，请使用 AWS Management Console

1. 导航到新的 Amazon EMR 控制台，然后从侧面导航栏中选择“切换到旧主机”。有关切换到旧控制台后预期情况的更多信息，请参阅 [Using the old console](#)。
2. 依次选择 Create cluster (创建集群)、Go to advanced options (转到高级选项)。
3. 选择 Release (版本) 4.7.2 或更早版本。
4. 为 Software and Steps (软件和步骤) 选择适用于您的应用程序的其它选项，然后选择 Next (下一步)。
5. 在 Hardware (硬件) 和 General Cluster Settings (常规集群设置) 窗格中选择适用于您的应用程序的设置。
6. 在“安全”窗格的“身份验证和加密”下，选择要使用的 S3 加密 (使用EMRFS) 选项。

 Note

使用 Amazon EMR 发行版 4.4 或更早版本时，无法使用带有KMS密钥管理 (SSE-KMS) 的 S3 服务器端加密。

- 如果您选择使用AWS 密钥管理的选项，请选择密AWS KMS钥 ID。有关更多信息，请参阅 [AWS KMS keys 用于EMRFS加密](#)。
  - 如果您选择使用自定义材料提供程序进行 S3 客户端加密，请提供类名和JAR位置。有关更多信息，请参阅 [Amazon S3 客户端加密](#)。
7. 选择适用于您的应用程序的其它选项，然后选择 Create Cluster (创建集群)。

## AWS KMS keys 用于EMRFS加密

AWS KMS 加密密钥必须与您的 Amazon EMR 集群实例和与一起EMRFS使用的 Amazon S3 存储桶在同一区域创建。如果您指定的密钥与用于配置集群的账户不同，则必须使用其来指定密钥ARN。

Amazon EC2 实例配置文件的角色必须有权使用您指定的KMS密钥。Amazon 中实例配置文件的默认角色EMR是EMR\_EC2\_DefaultRole。如果您为实例配置文件使用不同的角色，或者您使用IAM角色来EMRFS请求到 Amazon S3，请确保将每个角色添加为适当的关键用户。这为角色提供了使用KMS密钥的权限。有关更多信息，请参阅AWS Key Management Service 开发人员指南中的[使用密钥策略](#)和[配置向 Amazon S3 EMRFS 请求的IAM角色](#)。

您可以使用将您的实例配置文件或EC2实例配置文件 AWS Management Console 添加到指定密钥的 KMS密钥用户列表中，也可以使用 AWS CLI 或附加 AWS SDK相应的密钥策略。

请注意，Amazon 仅EMR支持[对称KMS密钥](#)。您不能使用[非对称KMS密钥](#)对 Amazon EMR 集群中的静态数据进行加密。要帮助确定KMS密钥是对称密钥还是非对称密钥，请参阅[识别对称和非对称密钥](#)。KMS

以下步骤介绍如何使用添加默认 Amazon EMR 实例配置文件EMR\_EC2\_DefaultRole作为关键用户 AWS Management Console。它假设您已经创建了KMS密钥。要创建新KMS密钥，请参阅AWS Key Management Service 开发者指南中的[创建密钥](#)。

将 Amazon EMR 的EC2实例配置文件添加到加密密钥用户列表中

1. 登录 AWS Management Console 并打开 AWS Key Management Service (AWS KMS) 控制台，[网址为 https://console.aws.amazon.com/kms](https://console.aws.amazon.com/kms)。
2. 要更改 AWS 区域，请使用页面右上角的区域选择器。
3. 选择要修改的KMS密钥的别名。
4. 在密钥详细信息页面的 Key Users (密钥用户) 下，选择 Add (添加)。
5. 在 Add key users (添加密钥用户) 对话框中，选择适当的角色。默认角色的名称为 EMR\_EC2\_DefaultRole。
6. 选择添加。

## Amazon S3 服务器端加密

设置 Amazon S3 服务器端加密时，Amazon S3 在向磁盘写入数据时会在对象级别加密数据，并在访问数据时对数据进行解密。有关更多信息SSE，请参阅《Amazon 简单存储服务用户指南》中的[使用服务器端加密保护数据](#)。

在 Amazon SSE 中指定时，您可以在两种不同的密钥管理系统之间进行选择EMR：

- SSE-S3 — 亚马逊 S3 为您管理密钥。
- SSE-KMS — 您可以使用 AWS KMS key 来设置适用于 Amazon 的政策EMR。有关 Amazon 密钥要求的更多信息EMR，请参阅[AWS KMS keys 用于加密](#)。

SSE使用买家提供的密钥 (SSE-C) 不适用于亚马逊。EMR

要创建启用了 SSE-S3 的集群，请使用 AWS CLI

- 键入以下命令：

```
aws emr create-cluster --release-label emr-4.7.2 or earlier \
--instance-count 3 --instance-type m5.xlarge --emrfs Encryption=ServerSide
```

您也可以通过设置 `fs.s3` 来启用-S3。enableServerSide属性中的emrfs-site加密属性为 true。参见KMS下面的 SSE-示例，省略密钥 ID 的属性。

要创建KMS启用 SSE-的集群，请使用 AWS CLI

### Note

SSE-KMS 仅在 Amazon EMR 发行版 4.5.0 及更高版本中可用。

- 键入以下 AWS CLI 命令创建带有 SSE-的集群KMS，其中 *keyID* 例如 AWS KMS key，是 *a4567b8-9900-12ab-1234-123a45678901*:

```
aws emr create-cluster --release-label emr-4.7.2 or earlier --instance-count 3 \
--instance-type m5.xlarge --use-default-roles \
--emrfs Encryption=ServerSide,Args=[fs.s3.serverSideEncryption.kms.keyId=keyId]
```

--OR--

使用emrfs-site分类键入以下 AWS CLI 命令，并提供一个包含内容的配置JSON文件，如下例myConfig.json所示：

```
aws emr create-cluster --release-label emr-4.7.2 or earlier --instance-count 3
--instance-type m5.xlarge --applications Name=Hadoop --configurations file://
myConfig.json --use-default-roles
```

myConfig.json 的示例内容：

```
[
  {
    "Classification": "emrfs-site",
    "Properties": {
```

```

    "fs.s3.enableServerSideEncryption": "true",
    "fs.s3.serverSideEncryption.kms.keyId": "a4567b8-9900-12ab-1234-123a45678901"
  }
}
]

```

## SSE-S3 和 SSE-的 配置属性 KMS

可使用 `emrfs-site` 配置分类来配置这些属性。SSE-KMS 仅在 Amazon EMR 发行版 4.5.0 及更高版本中可用。

属性	默认值	描述
<code>fs.s3.enableServerSideEncryption</code>	<b>false</b>	设置为 <b>true</b> 时，使用服务器端加密对 Amazon S3 中存储的对象进行加密。如果未指定密钥，则使用 SSE-S3。
<code>fs.s3.serverSideEncryption.kms.keyId</code>	<b>n/a</b>	指定密 AWS KMS 键 ID 或 ARN。如果指定了密钥，KMS 则使用 SSE-。

## Amazon S3 客户端加密

使用 Amazon S3 客户端加密，Amazon S3 的加密和解密将在集群的 EMRFS 客户端中进行。在对象上载到 Amazon S3 之前对其进行加密，并在下载后对其进行解密。您指定的提供程序会提供客户端使用的加密密钥。客户端可以使用 AWS KMS (CSE-KMS) 提供的密钥或提供客户端根密钥的自定义 Java 类 (CSE-C)。CSE-KMS 和 CSE-C 之间的加密细节略有不同，具体取决于指定的提供程序以及正在解密或加密的对象的元数据。有关这些区别的更多信息，请参阅《Amazon Simple Storage Service 用户指南》中的[使用客户端加密保护数据](#)。

**Note**

Amazon S3 CSE 仅确保与 Amazon S3 交换EMRFS的数据经过加密；并非集群实例卷上的所有数据都经过加密。此外，由于 Hue 不使用 HueEMRFS，因此 Hue S3 文件浏览器写入到 Amazon S3 的对象不会被加密。

要指定 CSE-KMS 对于 Amazon S3 中的EMRFS数据，使用 AWS CLI

- 键入以下命令并替换 *MyKMSKeyID* 使用密钥 ID 或ARN要使用的KMS密钥：

```
aws emr create-cluster --release-label emr-4.7.2 or earlier
--emrfs Encryption=ClientSide,ProviderType=KMS,KMSKeyId=MyKMSKeyId
```

## 创建自定义密钥提供程序

根据您在创建自定义密钥提供程序时使用的加密类型，应用程序还必须实现不同的 EncryptionMaterialsProvider 接口。这两个接口均在 Java 版本 1.11.0 及更高版本中可用。AWS SDK

- 要实现亚马逊 S3 加密，请使用 [com.amazonaws.services.s3.model.EncryptionMaterialsProvider](#) 接口。
- 要实现本地磁盘加密，请使用 [com.amazonaws.services.elasticmapreduce.spi.security.EncryptionMaterialsProvider](#) 接口。

您可以使用任何策略为实施提供加密材料。例如，您可以选择提供静态加密材料或与更复杂的密钥管理系统集成。

如果您使用的是 Amazon S3 加密，则必须NoPadding对自定义加密材料使用加密算法 GCM/AES/。

如果您使用的是本地磁盘加密，则用于自定义加密材料的加密算法因EMR版本而异。对于亚马逊 EMR 7.0.0 及更低版本，您必须使用 AES//GCM。NoPadding对于亚马逊 EMR 7.1.0 及更高版本，您必须使用AES。

该 EncryptionMaterialsProvider 类通过加密上下文获取加密材料。Amazon EMR 会在运行时填充加密上下文信息，以帮助调用者确定要返回的正确加密材料。



## Example 示例：使用自定义密钥提供程序进行 Amazon S3 加密 EMRFS

当 Amazon 从 `EncryptionMaterialsProvider` 类中 EMR 提取加密材料以执行加密时，可以 EMRFS 选择在 `materialsDescription` 参数中填充两个字段：对象的 Amazon S3 URI 和集群的，该 `EncryptionMaterialsProvider` 类可以选择性地使用这两个字段来返回加密材料。 `JobFlowId`

例如，提供商可能会为不同的 Amazon S3 URI 前缀返回不同的密钥。它是对返回的加密材料的描述，这些材料最终与 Amazon S3 对象一起存储，而不是由提供商生成 EMRFS 并传递给提供商的 `materialsDescription` 值。解密 Amazon S3 对象时，加密材料描述会传递给该 `EncryptionMaterialsProvider` 类，这样它就可以有选择地返回匹配的密钥来解密该对象。

下面提供了一个 `EncryptionMaterialsProvider` 参考实现。另一个自定义提供程序可从中获得 GitHub。 [EMRFSRSAEncryptionMaterialsProvider](#)

```
import com.amazonaws.services.s3.model.EncryptionMaterials;
import com.amazonaws.services.s3.model.EncryptionMaterialsProvider;
import com.amazonaws.services.s3.model.KMSEncryptionMaterials;
import org.apache.hadoop.conf.Configurable;
import org.apache.hadoop.conf.Configuration;

import java.util.Map;

/**
 * Provides KMSEncryptionMaterials according to Configuration
 */
public class MyEncryptionMaterialsProviders implements EncryptionMaterialsProvider,
    Configurable{
    private Configuration conf;
    private String kmsKeyId;
    private EncryptionMaterials encryptionMaterials;

    private void init() {
        this.kmsKeyId = conf.get("my.kms.key.id");
        this.encryptionMaterials = new KMSEncryptionMaterials(kmsKeyId);
    }

    @Override
    public void setConf(Configuration conf) {
        this.conf = conf;
        init();
    }

    @Override
```

```
public Configuration getConf() {
    return this.conf;
}

@Override
public void refresh() {

}

@Override
public EncryptionMaterials getEncryptionMaterials(Map<String, String>
materialsDescription) {
    return this.encryptionMaterials;
}

@Override
public EncryptionMaterials getEncryptionMaterials() {
    return this.encryptionMaterials;
}
}
```

## 使用指定定制材料提供商 AWS CLI

要使用 AWS CLI，请将 Encryption、ProviderType、CustomProviderClass 和 CustomProviderLocation 参数传递给 emrfs 选项。

```
aws emr create-cluster --instance-type m5.xlarge --release-label emr-4.7.2 or earlier
--emrfs Encryption=ClientSide,ProviderType=Custom,CustomProviderLocation=s3://
mybucket/myfolder/provider.jar,CustomProviderClass=classname
```

设置 Encryption 为 ClientSide 启用客户端加密是您的 EncryptionMaterialsProvider 对象的名称，CustomProviderClass 也是 Amazon 从中 EMR 复制 CustomProviderClass 到群集中的每个节点并将其放置在类路径中的本地或 Amazon S3 位置。CustomProviderLocation

## 使用指定定制材料提供商 SDK

要使用 SDK，您可以将属性设置为 fs.s3.cse.encryptionMaterialsProvider.uri 将存储在 Amazon S3 中的自定义 EncryptionMaterialsProvider 类下载到集群中的每个节点。您可以在 emrfs-site.xml 文件中对其进行配置，同时配置自定义提供程序的 CSE 启用和正确位置。

例如，在 AWS SDK for Java 使用中 RunJobFlowRequest，您的代码可能如下所示：

```

<snip>
    Map<String,String> emrfsProperties = new HashMap<String,String>();
        emrfsProperties.put("fs.s3.cse.encryptionMaterialsProvider.uri","s3://mybucket/
MyCustomEncryptionMaterialsProvider.jar");
        emrfsProperties.put("fs.s3.cse.enabled","true");
        emrfsProperties.put("fs.s3.consistent","true");

    emrfsProperties.put("fs.s3.cse.encryptionMaterialsProvider","full.class.name.of.EncryptionMate

    Configuration myEmrfsConfig = new Configuration()
        .withClassification("emrfs-site")
        .withProperties(emrfsProperties);

    RunJobFlowRequest request = new RunJobFlowRequest()
        .withName("Custom EncryptionMaterialsProvider")
        .withReleaseLabel("emr-7.2.0")
        .withApplications(myApp)
        .withConfigurations(myEmrfsConfig)
        .withServiceRole("EMR_DefaultRole_V2")
        .withJobFlowRole("EMR_EC2_DefaultRole")
        .withLogUri("s3://myLogUri/")
        .withInstances(new JobFlowInstancesConfig()
            .withEc2KeyName("myEc2Key")
            .withInstanceCount(2)
            .withKeepJobFlowAliveWhenNoSteps(true)
            .withMasterInstanceType("m5.xlarge")
            .withSlaveInstanceType("m5.xlarge")
        );

    RunJobFlowResult result = emr.runJobFlow(request);
</snip>

```

## 带参数的自定义 EncryptionMaterialsProvider

您可能需要将参数直接传递给提供程序。要执行此操作，您可以将 `emrfs-site` 配置分类与定义为属性的自定义参数结合使用。下面显示了一个示例配置，该示例配置将另存为 `myConfig.json` 文件：

```

[
  {
    "Classification": "emrfs-site",
    "Properties": {
      "myProvider.arg1": "value1",
      "myProvider.arg2": "value2"
    }
  }
]

```

```

    }
  }
]

```

使用中的create-cluster命令 AWS CLI，您可以使用--configurations选项来指定文件，如下所示：

```

aws emr create-cluster --release-label emr-7.2.0 --instance-type m5.xlarge
--instance-count 2 --configurations file://myConfig.json --emrfs
Encryption=ClientSide,CustomProviderLocation=s3://mybucket/myfolder/
myprovider.jar,CustomProviderClass=classname

```

## 配置 EMRFS S3EC V2 支持

S3 Java SDK 版本 ( 1.11.837 及更高版本 ) 支持加密客户端版本 2 ( S3EC V2 )，并具有各种安全增强功能。有关更多信息，请参阅 S3 博客文章 [Updates to the Amazon S3 encryption client](#)。另请参阅《AWS SDK for Java 开发人员指南》中的 [Amazon S3 加密客户端迁移](#)。

加密客户端 V1 仍可在中使用，SDK以实现向后兼容。如果启用，默认情况下EMRFS将使用 S3EC V1 来加密和解密 S3 对象。CSE

在发行版本低于 emr-5.31.0 ( emr-5.30.1 及更早版本，emr-6.1.0 及更早版本，emr-6.1.0 及更早版本 ) 的EMR集群EMRFS上无法解密使用 S3EC V2 加密的 S3 对象。

Example 配置EMRFS为使用 S3EC V2

EMRFS要配置为使用 S3EC V2，请添加以下配置：

```

{
  "Classification": "emrfs-site",
  "Properties": {
    "fs.s3.cse.encryptionV2.enabled": "true"
  }
}

```

## Amazon S3 客户端加密的 `emrfs-site.xml` 属性

属性	默认值	描述
<code>fs.s3.cse.enabled</code>	<b>false</b>	

属性	默认值	描述
		设置为时 <b>true</b> ，存储在 Amazon S3 中的EMRFS对象将使用客户端加密进行加密。
<code>fs.s3.cse.encryptedV2.enabled</code>	<b>false</b>	设置为时 <b>true</b> ，EMRFS使用 S3 加密客户端版本 2 对 S3 上的对象进行加密和解密。适用于 5.31.0 及更高EMR版本。
<code>fs.s3.cse.encryptedMaterialsProvider.uri</code>	<b>N/A</b>	在使用自定义加密材料时适用。JAR带有的URI所在的 Amazon S3。EncryptionMaterialsProvider 当您提供此信息时URI，Amazon EMR 会自动将下载JAR到集群中的所有节点。
<code>fs.s3.cse.encryptedMaterialsProvider</code>	<b>N/A</b>	用于客户端加密的 EncryptionMaterialsProvider 类路径。使用 CSE-时KMS，请指定 <code>com.amazon.ws.emr.hadoop.fs.cse.KMSEncryptionMaterialsProvider</code> 。
<code>fs.s3.cse.materialsDescription.enabled</code>	<b>false</b>	如果设置为 <b>true</b> ，则使用对象 <code>materialsDescription</code> 的 Amazon S3 URI 填充加密对象和。JobFlowId在使用自定义加密材料时设置为 <b>true</b> 。

属性	默认值	描述
<code>fs.s3.cse.kms.keyId</code>	N/A	使用 CSE-时适用KMS。用于加密的KMS密钥的 KeyIdARN、或别名的值。
<code>fs.s3.cse.cryptoStorageMode</code>	<b>ObjectMet adata</b>	Amazon S3 存储模式。默认情况下，加密信息的描述存储在对象元数据中。也可以将描述存储在指令文件中。有效值为 <code>ObjectMetadata</code> 和 <code>InstructionFile</code> 。有关更多信息，请参阅使用 <a href="#">AWS SDK for Java</a> 和 <a href="#">Amazon S3 进行客户端数据加密</a> 。 。

## 亚马逊 CloudWatch 代理

亚马逊上的亚马逊 CloudWatch 代理EMR是一种可以监控EMR集群中亚马逊EC2实例的工具。您可以存储和查看通过 CloudWatch 代理收集的指标 CloudWatch。有关 CloudWatch 代理的更多信息，请参阅 [Amazon CloudWatch 用户指南](#)。

### Note

如果您使用其他 AWS 服务 方式发布、查询或查看 Amazon CloudWatch 代理指标，则会产生额外费用。有关更多定价信息，请参阅以下页面。

- [亚马逊 CloudWatch 定价](#)
- [适用于 Prometheus 的亚马逊托管服务定价](#)
- [亚马逊托管 Grafana 定价](#)

这些指标与 Amazon 在AWS/ElasticMapReduce命名空间下EMR提供的免费指标是分开的 CloudWatch。有关 CloudWatch 代理未发布的指标的更多信息，请参阅[使用监控 Amazon EMR 指标 CloudWatch](#)。

在 Amazon EMR 7.0 及更高版本中，您可以在集群上安装自定义版本的 Amazon CloudWatch 代理，以便从集EMR群中收集指标。代理从集群中的所有节点收集指标，在主节点上收集这些指标，然后将指标从该节点发布到云端。

您可以在控制台中创建新集群时安装代理，也可以在使用时安装代理create-clusterAPI。有关更多信息，请参阅 [创建使用 Amazon CloudWatch 代理的EMR集群](#)。默认情况下，运行 7.x 版系列的 Amazon EMR 集群以 60 秒的间隔向 CloudWatch 发布 [34 个系统级指标](#)，但您可以将代理配置为发布不同的指标。另一种选择是将指标发布到适用于 Prometheus 的亚马逊托管服务，您也可以选择发布哪些指标。对于不同的用例和设置，您可以将代理配置为在 CloudWatch 控制台、Amazon Managed Grafana 中或通过 for 或 Amazon Prometheus 托管APIs服务 CloudWatch 查看和查询指标。当您使用它们来存储和查询指标时，它们会 AWS 服务 产生费用。

Amazon 上的 CloudWatch 代理EMR可以在您的 Amazon EMR 集群上发布系统JMX指标以及以下服务的指标。

- Hadoop DataNode
- Hadoop NameNode

- 纱线 NodeManager
- 纱线 ResourceManager
- HBaseMaster — 仅限亚马逊 EMR 7.1+
- HBase RegionServer — 仅限亚马逊 EMR 7.1+
- HBase ThriftServer — 仅限亚马逊 EMR 7.1+

有关可用指标以及如何在亚马逊上配置 CloudWatch 代理的更多信息EMR，请参阅[为亚马逊配置 CloudWatch 代理EMR](#)。

下表列出了最新版本的 Amazon EMR 7.x 系列中 AmazonCloudWatchAgent 包含的版本以及亚马逊随之EMR安装的组件。 AmazonCloudWatchAgent

有关此版本 AmazonCloudWatchAgent 中安装的组件的版本，请参阅 [7.2.0 版组件版本](#)。

AmazonCloudWatchAgent emr-7.2.0 的版本信息

亚马逊EMR发行标签	AmazonCloudWatchAgent 版本	安装的组件 AmazonCloudWatchAgent
emr-7.2.0	AmazonCloudWatchAgent 1.300032.2	adot-java-agent, emrfs, emr-amazon-cloudwatch-agent, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server

## 主题

- [创建使用 Amazon CloudWatch 代理的EMR集群](#)
- [Amazon CloudWatch 代理的默认指标 EMR](#)



- [为 Amazon 配置 CloudWatch 代理 EMR](#)
- [注意事项和限制](#)
- [CloudWatch 代理发布历史记录](#)

## 创建使用 Amazon CloudWatch 代理的EMR集群

本节中的过程描述了使用和中的亚马逊 CloudWatch 代理在 Amaz EMR on 中创建集群 AWS Management Console 的步骤 AWS CLI。

### 主题

- [CloudWatch 代理所需的IAM权限](#)
- [所需的 CloudWatch 代理端点](#)
- [创建 EMR 集群](#)

## CloudWatch 代理所需的IAM权限

CloudWatch 代理需要亚马逊EC2实例配置文件中的 AWS Identity and Access Management (IAM) `cloudwatch:PutMetricData` 权限EMR。Amazon EMR 默认角色已经拥有此权限。您可以使用创建默认角色`aws emr create-default-roles`。AWS CLI 有关更多信息，请参阅 Amazon EMR 管理指南中的[集群EC2EC2实例服务角色 \( 实例配置文件 \)](#)。

以下示例IAM策略包括`cloudwatch:PutMetricData`权限：

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "cloudwatch:PutMetricData",
      "Resource": "*"
    }
  ]
}
```

## 所需的 CloudWatch 代理端点

要向私有子网中的EMR集群发布指标，请创建 CloudWatch 代理终端节点并与私有子网所在VPC的代理终端节点相关联。CloudWatch

有关每个 CloudWatch 终端节点的更多信息 AWS 区域，请参阅《AWS 通用参考指南》中的 [Amazon CloudWatch 终端节点和配额](#)。

## 创建 EMR 集群

设置了与 CloudWatch 代理一起使用的所需权限和终端节点后，使用 AWS Management Console 或创建安装了代理的新集群。AWS CLI

### Console

通过控制台使用 Amazon CloudWatch 代理创建集群

1. 导航至 Amazon EMR 控制台。
2. 选择创建集群。
3. 在“名称和应用程序”下，选择 7.0.0 或更高EMR版本的 Amazon 版本。
4. 在“应用程序捆绑包”下，选择要安装到集群中的一个或多个应用程序，并在您的选择中加入CloudWatch 代理。
5. 继续创建集群以满足您的应用场景需求。

### AWS CLI

在中 AWS CLI，您可以使用--applications参数将 Amazon CloudWatch 代理添加到集群create-cluster。

要使用 Amazon CloudWatch 代理创建集群，请访问 AWS CLI

- 创建集群时，使用类似于以下内容的命令来包含 Amazon CloudWatch 代理。*myKey*替换为您的 EC2 key pair 的名称。

```
aws emr create-cluster --name "Spark cluster with CloudWatch agent" \  
--release-label emr-7.0.0 \  
--applications Name=Spark Name=AmazonCloudWatchAgent \  
--ec2-attributes KeyName=myKey --instance-type m7g.2xlarge \  
--instance-count 3 --use-default-roles
```

有关如何将 Amazon EMR 与配合使用的更多详情 AWS CLI，请参阅[AWS CLI 命令参考](#)。

# Amazon CloudWatch 代理的默认指标 EMR

当您在亚马逊上安装 Amazon CloudWatch 代理时EMR，除非您对[代理进行不同的配置](#)，否则默认配置会发布集群中所有实例的以下系统指标。有关每个[指标的定义](#)，请参阅 [Amazon CloudWatch 用户指南](#)中的 [CloudWatch 代理收集](#)的指标。

## CPU

### CPU指标

- `cpu_usage_active`
- `cpu_usage_guest`
- `cpu_usage_guest_nice`
- `cpu_usage_idle`
- `cpu_usage_iowait`
- `cpu_usage_irq`
- `cpu_usage_nice`
- `cpu_usage_softirq`
- `cpu_usage_steal`
- `cpu_usage_system`
- `cpu_usage_user`

## Disk

### 磁盘指标

- `disk_free`
- `disk_total`
- `disk_used`
- `disk_used_percent`

## Memory

### 内存指标

- `mem_active`

- mem\_available
- mem\_available\_percent
- mem\_free
- mem\_inactive
- mem\_total
- mem\_used
- mem\_used\_percent
- mem\_buffered
- mem\_cached

## Network IO

### 网络 IO 指标

- net\_bytes\_recv
- net\_bytes\_sent
- net\_packets\_recv
- net\_packets\_sent

## Process

### 进程指标

- processes\_running
- processes\_total

## Swap

### 交换指标

- swap\_free
- swap\_used
- swap\_used\_percent

默认情况下，代理会将所有指标发布到 CloudWatch 自定义命名空间CWAgent下和架构cluster.id、instance.id、node.type、service.name。使用以下步骤查看 CloudWatch 代理为 Amazon 发布的这些默认指标EMR：

1. 导航至 Amazon CloudWatch 控制台。
2. 选择指标选项卡，然后选择所有指标。
3. 在自定义命名空间下，选择 **CWAgent**。然后，选择架构：**cluster.id**、**instance.id**、**node.type**、**service.name**。
4. 根据您的用例需要，继续从 CloudWatch 界面查询指标。

Amazon EMR 在每个指标上贴上了以下标签:["cluster.id", "instance.id", "node.type", "service.name"]。关于这些标签，请考虑以下几点：

- cluster.id 标签的值是生成该指标的 EMR 集群的 ID。cluster.id 标签的示例值为：j-123456789ABC。
- instance.id 标签的值是 EMR 集群中生成该指标的实例的 ID。instance.id 标签的示例值为 i-01bcf5f140f3355777。
- node.type 标签的值表示集群中生成该指标的实例的节点类型。该值可以是 mastercore、或 task。
- service.name 标签的值是生成该指标的服务的名称。默认系统指标的默认 service.name 值为 system。

#### Note

Amazon 7.0.0 EMR 版本使用的指标标签架构略有不同。cluster.id 是 jobflow.id，并且 node.type 不可用。instance.id 并且 service.name 可用。

## 为 Amazon 配置 CloudWatch 代理 EMR

亚马逊 EMR 7.0.0 及更高版本包括亚马逊 CloudWatch 代理。以下各节介绍如何在 EMR 集群上配置代理，如中所述 [创建使用 Amazon CloudWatch 代理的 EMR 集群](#)。

### 主题

- [为亚马逊 EMR 7.1.0 配置 CloudWatch 代理](#)

- [为亚马逊 EMR 7.0.0 配置 CloudWatch 代理](#)

## 为亚马逊 EMR 7.1.0 配置 CloudWatch 代理

从 Amazon EMR 7.1.0 开始，您可以使用亚马逊配置将亚马逊 CloudWatch 代理EMR配置API为使用其他系统指标、添加应用程序指标和更改指标目标。有关如何使用EMR配置API来配置集群应用程序的更多信息，请参阅[配置应用程序](#)。

### Note

7.1.0 仅支持重新配置类型。OVERWRITE有关重新配置类型的更多信息，请参阅[重新配置实例组时的注意事项](#)。

### 主题

- [配置架构](#)
- [系统指标配置示例](#)
- [应用程序指标配置示例](#)
- [适用于 Prometheus 的亚马逊托管服务示例](#)

## 配置架构

emr-metrics有以下分类：

- `emr-system-metrics`— 配置系统指标CPU，例如磁盘和内存。
- `emr-hadoop-hdfs-datanode-metrics`— 配置 Hadoop 指标 DataNode JMX
- `emr-hadoop-hdfs-namenode-metrics`— 配置 Hadoop 指标 NameNode JMX
- `emr-hadoop-yarn-nodemanager-metrics`— 配置 Yarn NodeManager JMX 指标
- `emr-hadoop-yarn-resourcemanager-metrics`— 配置 Yarn ResourceManager JMX 指标
- `emr-hbase-master-metrics`— 配置HBase主JMX指标
- `emr-hbase-region-server-metrics`— 配置HBase区域服务器JMX指标
- `emr-hbase-rest-server-metrics`— 配置HBaseREST服务器JMX指标
- `emr-hbase-thrift-server-metrics`— 配置 HBase Thrift Server 指标 JMX

下表描述了所有分类的可用属性和配置。

### emr-metrics 属性

属性	必需	描述	默认值	可能的值	注意
<code>metrics_destination</code>	可选	确定集群指标是发布到亚马逊 CloudWatch 还是发布到亚马逊 Prometheus 托管服务。	"CLOUDWATCH"	"CLOUDWATCH", "PROMETHEUS"	此属性不区分大小写。例如，“Cloudwatch”与“CLOUDWATCH”相同。
<code>prometheus_endpoint</code>	可选	如果设置 <code>metrics_destination</code> 为 "PROMETHEUS"，则此属性将 CloudWatch 代理配置为向提供的亚马逊托管服务 Prometheus 远程写入端点发送指标。	不适用	任何适用于 Prometheus 的有效亚马逊托管服务远程写入。URL 远程写入 URL 格式为  <pre>https:// aps- worksp aces. &lt;region&gt; s.com/ wor kspaces/ &lt;workspace_id&gt; / api/v1/r emote_wri te</pre>	如果设置为 "PROMETHEUS"， <code>metrics_destination</code> 则此字段为必填字段。如果您不提供密钥或值为空字符串，则配置将失败。

### emr-system-metrics 属性

属性	必需	描述	默认值	可能的值	注意
metrics_collection_interval	可选	从 CloudWatch 代理收集和发布指标的频率 (以秒为单位)。	“60”	指定秒数的字符串。只接受整数。	您可以使用各个指标组中的 metrics_collection_interval 属性来覆盖此属性。

## emr-system-metrics 配置

### cpu

属性	必需	描述	默认值	可能的值	注意
metrics	可选	代理要收集的CPU指标列表。	查看 <a href="#">Amazon CloudWatch 代理的默认指标 EMR</a>	以逗号分隔的带或不带cpu_前缀的有效CPU指标名称列表，例如usage_active 和。cpu_idle 有关有效指标，请参阅 <a href="#">CloudWatch 代理收集的指标</a> 。	指定空字符串意味着不发布任何CPU指标。
metrics_collection_interval	可选	代理收集和发布CPU指标的频率 (以秒为单位)。	全球的值 metrics_collection_interval。	指定秒数的字符串。只接受整数。	此值仅会覆盖CPU指标的全局 metrics_collection



属性	必需	描述	默认值	可能的值	注意
					<code>_interval</code> 属性。
<code>drop_original_metrics</code>	可选	不发布未汇总CPU指标的指标列表。	未发布任何未汇总的CPU指标。	以逗号分隔的CPU指标列表，这些指标也在 <code>metrics</code> 属性中指定。空字符串表示发布所有CPU指标。	CloudWatch 代理按集群 ID、实例 ID、节点类型和服务名称汇总所有指标。默认情况下，CloudWatch 代理不会为具有多个资源的指标发布每种资源的指标。
<code>resources</code>	可选	确定代理是否将发布每个内核的指标。	<code>"*"</code>	<code>"*"</code> 启用每内核指标。 <code>""</code> 禁用每内核指标。	CloudWatch 代理仅发布未被删除的CPU指标的每核指标。 <code>drop_original_metrics</code>

## disk

属性	必需	描述	默认值	可能的值	注意
<code>metrics</code>	可选	代理要收集的磁盘指标列表。	查看 <a href="#">Amazon CloudWatch 代理的默认指标 EMR</a>	以逗号分隔的有效磁盘指标名称列表，带或不	指定空字符串意味着不发布任何磁盘指标。

属性	必需	描述	默认值	可能的值	注意
				带disk_前缀，例如disk_total和used_percent有关有效指标，请参阅 <a href="#">CloudWatch代理收集的指标</a> 。	
metrics_collection_interval	可选	代理收集和发布磁盘指标的频率（以秒为单位）。	全球的值metrics_collection_interval。	指定秒数的字符串。只接受整数。	此值仅覆盖磁盘指标的全局metrics_collection_interval属性。
drop_original_metrics	可选	不发布未聚合指标的磁盘指标列表。	未发布未汇总的磁盘指标。	以逗号分隔的磁盘指标列表，这些指标也在metrics属性中指定。空字符串表示发布所有磁盘指标。	CloudWatch代理按集群ID、实例ID、节点类型和服务名称汇总所有指标。默认情况下，CloudWatch代理不会为具有多个资源的指标发布每种资源的指标。

属性	必需	描述	默认值	可能的值	注意
resources	可选	确定代理是否将发布 per-mount-point 指标。	"*"	"*" 表示所有挂载点，"" 表示没有挂载点或以逗号分隔的挂载点列表。例如，"/,/emr"。	CloudWatch 代理仅发布 per-mount-point 未放入的磁盘指标的指标 drop_original_metrics。

## diskio

属性	必需	描述	默认值	可能的值	注意
metrics	可选	代理要收集的磁盘 IO 指标列表。	查看 <a href="#">Amazon CloudWatch 代理的默认指标 EMR</a>	以逗号分隔的有效磁盘 IO 指标名称列表，带或不带 diskio_ 前缀，例如 diskio_reads 和。write 有关有效指标，请参阅 <a href="#">CloudWatch 代理收集的指标</a> 。	指定空字符串意味着不发布任何磁盘 IO 指标。
metrics_collection_interval	可选	代理收集和发布磁盘 IO 指标的频率（以秒为单位）。	全球的值 metrics_collection_interval。	指定秒数的字符串。只接受整数。	此值仅覆盖磁盘 IO 指标的全局 metrics_collection

属性	必需	描述	默认值	可能的值	注意
					<code>_interval</code> 属性。
<code>drop_original_metrics</code>	可选	不发布未聚合指标的磁盘 IO 指标列表。	未发布未聚合的磁盘 IO 指标。	以逗号分隔的磁盘 IO 指标列表，这些指标也在 <code>metrics</code> 属性中指定。空字符串表示发布所有磁盘 IO 指标。	CloudWatch 代理按集群 ID、实例 ID、节点类型和服务名称汇总所有指标。默认情况下，CloudWatch 代理不会为具有多个资源的指标发布每种资源的指标。
<code>resources</code>	可选	确定代理是否会发布每台设备的指标。	<code>"*"</code>	<code>"*"</code> 表示所有存储设备， <code>""</code> 表示没有存储设备或以逗号分隔的设备名称列表。例如， <code>"nvme0n1 nvme1n1"</code> 。	CloudWatch 代理仅发布未放入的磁盘 IO 指标的每台设备指标。 <code>drop_original_metrics</code>

## mem

属性	必需	描述	默认值	可能的值	注意
<code>metrics</code>	可选	代理要收集的内存指标列表。	查看 <a href="#">Amazon CloudWatch</a>	以逗号分隔的有效内存指标名称列	指定空字符串意味着不

属性	必需	描述	默认值	可能的值	注意
			<a href="#">代理的默认指标 EMR</a>	表，带或不带mem_前缀，例如mem_available 和。ava_percent 有关有效 <a href="#">指标</a> ，请参见 <a href="#">CloudWatch 代理收集的指标</a> 。	发布任何内存指标。
metrics_collection_interval	可选	代理收集和发布内存指标的频率（以秒为单位）。	全球的值metrics_collection_interval。	指定秒数的字符串。只接受整数。	此值仅覆盖内存指标的 <a href="#">全局metrics_collection_interval 属性</a> 。

## net

属性	必需	描述	默认值	可能的值	注意
metrics	可选	代理要收集的网络指标列表。	查看 <a href="#">Amazon CloudWatch 代理的默认指标 EMR</a>	以逗号分隔的有效网络指标名称列表，带或不带net_前缀，例如net_packets_sent 和。ecv 有关有效 <a href="#">指标</a>	指定空字符串意味着不发布任何网络指标。

属性	必需	描述	默认值	可能的值	注意
				标，请参阅 <a href="#">CloudWatch 代理收集的指标</a> 。	
<code>metrics_collection_interval</code>	可选	代理收集和发布网络指标的频率（以秒为单位）。	全球的值 <code>metrics_collection_interval</code> 。	指定秒数的字符串。只接受整数。	此值仅覆盖网络指标的 <code>metrics_collection_interval</code> 属性。
<code>drop_original_metrics</code>	可选	不发布未汇总指标的网络指标列表。	未发布未汇总的网络指标。	以逗号分隔的网络指标列表，这些指标也在 <code>metrics</code> 属性中指定。空字符串表示发布所有网络指标。	CloudWatch 代理按集群 ID、实例 ID、节点类型和服务名称汇总所有指标。默认情况下，CloudWatch 代理不会为具有多个资源的指标发布每种资源的指标。

属性	必需	描述	默认值	可能的值	注意
resources	可选	确定代理是否将发布每个接口的指标。	"*"	"*" 表示所有网络接口，"" 表示没有网络接口，或者是以逗号分隔的接口名称列表。例如，"eth0,eth1"。	CloudWatch 代理仅发布未放入的网络指标的每个接口指标。drop_original_metrics

## netstat

属性	必需	描述	默认值	可能的值	注意
metrics	可选	代理要收集的网络统计数据指标列表。	查看 <a href="#">Amazon CloudWatch 代理的默认指标 EMR</a>	以逗号分隔的有效内存指标名称列表，带或不带netstat_前缀，例如tcp_listen 和。netstat_socket 有关有效指标，请参阅 <a href="#">CloudWatch 代理收集的指标</a> 。	指定空字符串意味着不发布任何网络统计指标。
metrics_collection_interval	可选	代理收集和发布网络统计指标的频	全球的价值metrics_collection	指定秒数的字符串。只接受整数。	此值仅覆盖网络统计指标的全局metrics_c

属性	必需	描述	默认值	可能的值	注意
		率 ( 以秒为单位 )。	<code>_interval</code> 。		<code>collection_interval</code> 属性。

## processes

属性	必需	描述	默认值	可能的值	注意
<code>metrics</code>	可选	代理要收集的流程指标列表。	查看 <a href="#">Amazon CloudWatch 代理的默认指标 EMR</a>	以逗号分隔的有效内存指标名称列表，带或不带 <code>processes_</code> 前缀，例如 <code>processes_running</code> 和 <code>processes_running</code> 。有关有效指标，请参阅 <a href="#">CloudWatch 代理收集的指标</a> 。	指定空字符串意味着不发布任何流程指标。
<code>metrics_collection_interval</code>	可选	代理收集和发布系统进程指标的频率 ( 以秒为单位 )。	全球的值 <code>metrics_collection_interval</code> 。	指定秒数的字符串。只接受整数。	此值仅覆盖系统进程指标的全局 <code>metrics_collection_interval</code> 属性。



## swap

属性	必需	描述	默认值	可能的值	注意
metrics	可选	代理要收集的交换指标列表。	查看 <a href="#">Amazon CloudWatch 代理的默认指标 EMR</a>	以逗号分隔的有效内存指标名称列表，带或不带swap_前缀，例如swap_free和。used_percent 有关有效指标，请参阅 <a href="#">CloudWatch 代理收集的指标</a> 。	指定空字符串意味着不发布任何交换指标。
metrics_collection_interval	可选	代理收集和发布掉期指标的频率（以秒为单位）。	全球的值metrics_collection_interval。	指定秒数的字符串。只接受整数。	此值仅覆盖掉期指标的全局metrics_collection_interval属性。

## emr-hadoop-hdfs-datanode-指标属性

属性	必需	描述	默认值	可能的值	
<code>&lt;custom_bean_name&gt;</code>	可选	不适用	MBean该CloudWatch代理应从中收集指标，例如Hadoop:se	一个字符串，其中包含与相关联的以逗号分隔的指标列表。MBean例	

属性	必需	描述	默认值	可能的值
			rvice=DataNode,name=DataNodeActivity。 您可以在 Amazon 7.0 EMR 版本的 <a href="#">示例 JMX YAML 文件中找到示例 MBean</a> 名称及其对应的指标。	如 , BlocksCached, BlocksRead 。
otel.metric.export.interval	可选	收集 Hadoop 指标的频率 ( 以毫秒为单位 )。DataNode	“60000”	一个指定毫秒数的字符串。只接受整数。

### emr-hadoop-hdfs-namenode-指标属性

属性	必需	描述	默认值	可能的值
<i>&lt;custom_bean_name&gt;</i>	可选	不适用	MBean该 CloudWatch 代理应从中收集指标 , 例如Hadoop:service=NameNode,name=FSNamesystem 。 您可以在	一个字符串 , 其中包含与相关联的以逗号分隔的指标列表。MBean例如 , BlockCapacity,CapacityUsedGB 。

属性	必需	描述	默认值	可能的值
			Amazon 7.0 EMR 版本的 <a href="#">示例 JMX YAML 文件中找到示例 MBean 名称及其对应的指标</a> 。	
<code>otel.metric.export.interval</code>	可选	收集 Hadoop 指标的频率（以毫秒为单位）。NameNode	“60000”	一个指定毫秒数的字符串。只接受整数。

#### emr-hadoop-yarn-nodemanager-指标属性

属性	必需	描述	默认值	可能的值
<code>&lt;custom_bean_name&gt;</code>	可选	不适用	MBean 该 CloudWatch 代理应从中收集指标，例如 <code>Hadoop:service=NodeManager,name=NodeManagerMetrics</code> 。您可以在 Amazon 7.0 EMR 版本的 <a href="#">示例 JMX YAML 文</a>	一个字符串，其中包含与相关联的以逗号分隔的指标列表。MBean 例如， <code>MaxCapacity,AllocatedGB</code> 。

属性	必需	描述	默认值	可能的值
			<a href="#">件中找到示例 MBean 名称及其对应的指标</a> 。	
<code>otel.metric.export.interval</code>	可选	收集 Hadoop 指标的频率（以毫秒为单位）。YARN NodeManager	“60000”	一个指定毫秒数的字符串。只接受整数。

### emr-hadoop-yarn-resourcemanager-指标属性

属性	必需	描述	默认值	可能的值
<code>&lt;custom_bean_name&gt;</code>	可选	不适用	MBean 该 CloudWatch 代理应从中收集指标，例如 <code>Hadoop:service=ResourceManager,name=PartitionQueueMetrics</code> 。您可以在 Amazon 7.0 EMR 版本的 <a href="#">示例 JMX YAML 文件中找到示例 MBean 名称及</a>	一个字符串，其中包含与相关联的以逗号分隔的指标列表。MBean 例如， <code>MaxCapacity</code> ， <code>MaxCapacityVCores</code> 。

属性	必需	描述	默认值	可能的值
			其对应的指标。 。	
<code>otel.metric.export.interval</code>	可选	收集 Hadoop 指标的频率（以毫秒为单位）。YARN Resource Manager	“60000”	一个指定毫秒数的字符串。只接受整数。

### emr-hbase-master-metrics 属性

属性	必需	描述	默认值	可能的值
<code>&lt;custom_bean_name&gt;</code>	可选	不适用	MBean该 CloudWatch 代理应从中收集指标，例如Hadoop:service=HBase,name=Master,sub=AssignmentManager。 您可以在 Amazon 7.0 EMR 版本的 <a href="#">示例 JMX YAML 文件中找到示例 MBean 名称及其对应的指标</a> 。 。	一个字符串，其中包含与相关联的以逗号分隔的指标列表。MBean例如，AssignFailedCount, AssignSubmittedCount。

属性	必需	描述	默认值	可能的值
otel.metric.export.interval	可选	收集HBase主指标的频率（以毫秒为单位）。	“60000”	一个指定毫秒数的字符串。只接受整数。

### emr-hbase-region-server-指标属性

属性	必需	描述	默认值	可能的值
<i>&lt;custom_mbean_name&gt;</i>	可选	不适用	MBean该CloudWatch代理应从中收集指标，例如Hadoop:service=HBase,name=RegionServer,sub=IPC。您可以在Amazon 7.0 EMR版本的 <a href="#">示例JMX YAML文件中找到示例MBean名称及其对应的指标</a> 。	一个字符串，其中包含与相关联的以逗号分隔的指标列表。MBean例如，numActiveHandler,numActivePriorityHandler。
otel.metric.export.interval	可选	收集HBase区域服务器指标的频率（以毫秒为单位）。	“60000”	一个指定毫秒数的字符串。只接受整数。

## emr-hbase-rest-server-指标属性

属性	必需	描述	默认值	可能的值
<code>&lt;custom_bean_name&gt;</code>	可选	不适用	MBean该 CloudWatch 代理应从中收集指标，例如Hadoop:service=HBase,name=REST。您可以在 Amazon 7.0 EMR 版本的 <a href="#">示例 JMXAML文件中找到示例 MBean名称及其对应的指标</a> 。	一个字符串，其中包含与相关联的以逗号分隔的指标列表。MBean例如，successfulPut,successfulScanCount。
<code>otel.metric.export.interval</code>	可选	收集 HBase Rest Server 指标的频率（以毫秒为单位）。	“60000”	一个指定毫秒数的字符串。只接受整数。

## emr-hbase-thrift-server-指标属性

属性	必需	描述	默认值	可能的值
<code>&lt;custom_bean_name&gt;</code>	可选	不适用	MBean该 CloudWatch 代理应从中收集指标，例如Hadoop:se	一个字符串，其中包含与相关联的以逗号分隔的指标列表。MBean例

属性	必需	描述	默认值	可能的值
			rvice=HBase, name=Thrift, sub=ThriftOne。您可以在 Amazon 7.0 EMR 版本的 <a href="#">示例 JMX YAML 文件中找到示例 MBean 名称及其对应的指标</a> 。	如 , BatchGet_max, BatchGet_mean 。
otel.metric.export.interval	可选	收集 T HBase Thrift 服务器指标的频率 (以毫秒为单位)。	"60000"	一个指定毫秒数的字符串。只接受整数。

## 系统指标配置示例

以下示例演示如何将 CloudWatch 代理配置为停止导出所有系统指标。

```
[
  {
    "Classification": "emr-metrics",
    "Properties": {},
    "Configurations": [
      {
        "Classification": "emr-system-metrics",
        "Properties": {},
        "Configurations": []
      }
    ]
  }
]
```



```
]
```

以下示例将 CloudWatch 代理配置为导出默认系统指标。如果您已经至少重新配置过一次系统指标，那么这样做是将代理重置回仅导出默认系统指标的快速方法。此重置还会删除之前重新配置的所有应用程序指标。

```
[
  {
    "Classification": "emr-metrics",
    "Properties": {},
    "Configurations": []
  }
]
```

以下示例将集群配置为导出cpumem、和disk指标。

```
[
  {
    "Classification": "emr-metrics",
    "Properties": {},
    "Configurations": [
      {
        "Classification": "emr-system-metrics",
        "Properties": {
          "metrics_collection_interval": "20"
        },
        "Configurations": [
          {
            "Classification": "cpu",
            "Properties": {
              "metrics": "cpu_usage_guest,cpu_usage_idle",
              "metrics_collection_interval": "30",
              "drop_original_metrics": "cpu_usage_guest"
            }
          },
          {
            "Classification": "mem",
            "Properties": {
              "metrics": "mem_active"
            }
          },
          {
            "Classification": "disk",
```

```

        "Properties": {
            "metrics": "disk_used_percent",
            "resources": "/,/mnt",
            "drop_original_metrics": ""
        }
    ]
}
]

```

前面的示例配置具有以下属性：

- 每隔 30 秒，代理就会收集所有人的cpu\_guest指标CPUs。您可以在 CloudWatch 命名空间下找到聚合指标CWAgent > cluster.id, instance.id, node.type, service.name。
- 每隔 30 秒，代理就会收集所有人的cpu\_idle指标CPUs。您可以在 CloudWatch 命名空间下找到聚合指标CWAgent > cluster.id, instance.id, node.type, service.name。代理还会收集每个 CPU 的指标。你可以在同一个命名空间中找到它们。代理收集此指标是因为该drop\_original\_metrics属性不包含cpu\_idle，因此代理不会忽略该指标。
- 代理每 20 秒收集一次mem\_active指标。您可以在 CloudWatch 命名空间下找到聚合指标CWAgent > cluster.id, instance.id, node.type, service.name。
- 代理每 20 秒收集一次/和/mnt磁盘装载的disk\_used\_percent指标。您可以在 CloudWatch 命名空间下找到聚合指标CWAgent > cluster.id, instance.id, node.type, service.name。代理还会收集每次挂载的指标。你可以在同一个命名空间中找到它们。代理收集此指标是因为该drop\_original\_metrics属性不包含disk\_used\_percent，因此代理不会忽略该指标。

## 应用程序指标配置示例

以下示例将 CloudWatch 代理配置为停止导出 Hadoop Namenode 服务的指标。

```

[
  {
    "Classification": "emr-metrics",
    "Properties": {},
    "Configurations": [
      {
        "Classification": "emr-hadoop-hdfs-namenode-metrics",

```

```

    "Properties": {},
    "Configurations": []
  }
]
}
]

```

以下示例将集群配置为导出 Hadoop 应用程序指标。

```

[
  {
    "Classification": "emr-metrics",
    "Properties": {},
    "Configurations": [
      {
        "Classification": "emr-hadoop-hdfs-namenode-metrics",
        "Properties": {
          "Hadoop:service=NameNode,name=FSNamesystem": "BlockCapacity,CapacityUsedGB",
          "otel.metric.export.interval": "20000"
        },
        "Configurations": []
      },
      {
        "Classification": "emr-hadoop-hdfs-datanode-metrics",
        "Properties": {
          "Hadoop:service=DataNode,name=JvmMetrics": "MemNonHeapUsedM",
          "otel.metric.export.interval": "30000"
        },
        "Configurations": []
      },
      {
        "Classification": "emr-hadoop-yarn-resourcemanager-metrics",
        "Properties": {
          "Hadoop:service=ResourceManager,name=CapacitySchedulerMetrics":
"AllocateNumOps,NodeUpdateNumOps"
        },
        "Configurations": []
      }
    ]
  }
]

```

前面的示例具有以下属性：

- 代理每 20 秒就会从运行 Hadoop Namenode 服务的实例收集BlockCapacity和CapacityUsedGB指标。
- 代理每 30 秒从运行 Hadoop Datanode 服务的实例收集MemNonHeapUsedM指标。
- 代理每 30 秒就会从运行 Hadoop YARN ResourceManaager 的实例收集AllocateNumOps和NodeUpdateNumOps指标。

## 适用于 Prometheus 的亚马逊托管服务示例

以下示例演示如何配置 CloudWatch 代理以将指标导出到适用于 Prometheus 的亚马逊托管服务。

如果您当前正在将指标导出到适用于 Prometheus 的亚马逊托管服务，并且想要重新配置集群的指标并继续将指标导出到适用于 Prometheus 的亚马逊托管服务，则必须包括属性和和。metrics\_destination prometheus\_endpoint

```
[
  {
    "Classification": "emr-metrics",
    "Properties": {
      "metrics_destination": "prometheus",
      "prometheus_endpoint": "http://amp-workspace/api/v1/remote_write"
    },
    "Configurations": []
  }
]
```

要使用 CloudWatch 代理将指标导出到 CloudWatch，请使用以下示例。

```
[
  {
    "Classification": "emr-metrics",
    "Properties": {
      "metrics_destination": "cloudwatch"
    },
    "Configurations": []
  }
]
```

**Note**

CloudWatch 代理有一个 Prometheus 导出器，可以重命名某些属性。对于默认指标标签，亚马逊 Prometheus 托管服务使用下划线字符代替亚马逊使用的句点。CloudWatch 如果您使用 Amazon Managed Grafana 来可视化适用于 Prometheus 的亚马逊托管服务中的默认指标，则标签将显示为、和。cluster\_id instance\_id node\_type service\_name

## 为亚马逊 EMR 7.0.0 配置 CloudWatch 代理

您可以将 Amazon CloudWatch 代理配置为使用[默认 CloudWatch 代理配置](#)提供的系统指标之外的其他系统指标。7.0.0 的配置需要使用引导操作，我们在以下各部分中提供了这些操作的示例。在即将发布的版本中，亚马逊EMR将通过亚马逊提供其他配置选项EMRAPI。

### 主题

- [使用 Amazon EMR 7.0.0 配置其他系统指标](#)
- [使用 Amazon EMR 7.0.0 配置应用程序指标](#)
- [使用亚马逊 7.0.0 将适用于 Prometheus 的亚马逊托管服务配置为指标的云存储 EMR](#)

## 使用 Amazon EMR 7.0.0 配置其他系统指标

使用以下步骤将代理配置为使用 Amazon EMR 7.0.0 中的一组不同的系统指标：

1. 在您的 Amazon S3 账户中创建或选择一个存储桶，用于存储指定 CloudWatch 代理指标的配置文件。
2. 使用您指定的首选指标创建 emr-amazon-cloudwatch-agent.json 配置文件。为此，请使用[创建 CloudWatch 代理配置文件中介绍](#)的方法之一。有关 CloudWatch 代理配置文件结构的更多信息，请参阅 Amazon CloudWatch 用户指南中的[手动创建或编辑 CloudWatch 代理配置文件](#)。
3. 接下来，导航到aws-emr-utilities存储库 GitHub 并下载以下系统指标脚本：
  - [install\\_system\\_metrics\\_launcher.sh](#) – 下载然后在后台运行 install\_system\_metrics.sh 的脚本，以便节点可以完成引导。
  - [install\\_system\\_metrics.sh](#)— 一种脚本，它等待其运行的实例完成引导，然后下载并应用文件中的配置。JSON
4. 打开每个 SH 文件并将 *my-s3-bucket* 替换为步骤 1 中的存储桶的名称。
5. 将一个JSON和两个 SH 文件上传到您的 S3 存储桶。

- 现在，您可以导航到 Amazon EMR 控制台并使用 CloudWatch 代理创建新集群。EC2在左侧导航栏的“开”下EMR，选择“集群”，然后选择“创建集群”。
- 在“名称和应用程序”部分，选择 7.0.0 或更高EMR版本的 Amazon 版本。
- 在应用程序捆绑包下，选择要安装到集群中的一个或多个应用程序，并在您的选择中包含 Amazon A CloudWatch gent。
- 在引导操作部分中，选择添加。
  - 对于名称，请插入 **install\_system\_metrics\_launcher.sh**。
  - 对于脚本位置，请插入 **s3://my-s3-bucket/install\_system\_metrics\_launcher.sh**。将 **my-s3-bucket** 替换为 S3 存储桶的路径。
  - 将参数块留为空白。
- 选择 Add bootstrap action ( 添加引导操作 ) 。
- 继续创建集群以满足您的工作负载需求。

当您的集群启动时，CloudWatch 代理会将您在配置文件中指定的系统指标发布到 CloudWatch。

## 使用 Amazon EMR 7.0.0 配置应用程序指标

您可以将 Amazon CloudWatch 代理配置为HDFS YARN除了系统指标之外还发布应用程序指标。按照以下步骤将代理配置为发布应用程序指标：

- 在您的 Amazon S3 账户中创建或选择一个存储桶，用于存储指定 CloudWatch 代理指标的配置文件。
- 接下来，导航到aws-emr-utilities存储库 GitHub 并下载以下脚本：
  - [install\\_app\\_metrics\\_launcher.sh](#) – 下载然后在后台运行 `install_app_metrics.sh` 的脚本，以便节点可以完成引导。
  - [install\\_app\\_metrics.sh](#)— 一种脚本，它等待其运行的实例完成引导，然后下载并应用文件中的配置，这些YAML文件将在接下来的步骤中下载。
- 打开每个文件并将 **my-s3-bucket** 替换为步骤 1 中的存储桶的名称。
- 接下来，下载以下YAML映射文件。有关这些YAML文件结构的信息，请参阅 OpenTelemetry Instrumentation for Java GitHub repo [javaagent](#)中的。
  - [datanode-metrics.yaml](#) – Hadoop DataNode 指标的配置。
  - [namenode-metrics.yaml](#) – Hadoop NameNode 指标的配置。
  - [nodemanager-metrics.yaml](#) – Yarn NodeManager 指标的配置。

- [resourcemanager-metrics.yaml](#) – Yarn ResourceManager 指标的配置。
5. 将两个 SH 和四个YAML文件上传到您的 S3 存储桶。
  6. 现在，您可以导航到 Amazon EMR 控制台并使用 CloudWatch 代理创建新集群。EC2在左侧导航栏的“开”下EMR，选择“集群”，然后选择“创建集群”。
  7. 在“名称和应用程序”部分，选择 7.0.0 或更高EMR版本的 Amazon 版本。
  8. 在“应用程序捆绑包”下，选择要安装到集群的捆绑包或自定义应用程序组，并在您的选择中加入CloudWatch 代理。
  9. 在引导操作部分中，选择添加。
    - 对于名称，请插入 `install_app_metrics_launcher.sh`。
    - 对于脚本位置，请插入 `s3://my-s3-bucket/install_app_metrics_launcher.sh`。将 `my-s3-bucket` 替换为 S3 存储桶的路径。
    - 将参数块留为空白。
  10. 选择 Add bootstrap action ( 添加引导操作 ) 。
  11. 继续创建集群以满足您的工作负载需求。

当您的集群启动时，CloudWatch 代理会将您指定的应用程序指标以及系统指标发布到 CloudWatch。

使用亚马逊 7.0.0 将适用于 Prometheus 的亚马逊托管服务配置为指标的云存储 EMR

您可以将亚马逊 CloudWatch 代理配置为将指标发布到适用于 Prometheus 的亚马逊托管服务，而不是。CloudWatch

#### Note

您可以将指标从亚马逊 CloudWatch 代理发布到适用于 Prometheus 的亚马逊托管服务或 CloudWatch 亚马逊，但不能将指标发布到同一集群的两个服务。

要将代理配置为将指标发布到适用于 Prometheus 的亚马逊托管服务，您必须在亚马逊EC2实例配置 `aps:RemoteWrite` AWS Identity and Access Management 文件中添加 IAM () 权限。EMR以下示例策略包含所需的权限：

```
{
  "Version": "2012-10-17",
  "Statement": [
```

```

    {
      "Effect": "Allow",
      "Action": "aps:RemoteWrite",
      "Resource": "*"
    }
  ]
}

```

使用EMR集群上的 CloudWatch 代理将指标发布到适用于 Prometheus 的亚马逊托管服务

服务策略获得正确权限后，使用以下步骤启动一个集群，该集群使用 CloudWatch 代理向适用于 Prometheus 的亚马逊托管服务发布指标。

1. 使用 AWS Management Console 或 AWS CLI 创建适用于 Prometheus 的亚马逊托管服务工作区。有关更多信息，请参阅《Amazon Managed Service for Prometheus 用户指南》中的[创建工作区](#)。
2. 在您的 Amazon S3 账户中创建或选择一个存储桶，该存储桶用于存储将 Amazon Managed Service for Prometheus 指定为云存储的启动文件。
3. 接下来，导航到aws-emr-utilities存储库 GitHub 并下载以下脚本：
  - [add\\_prometheus\\_endpoint\\_launcher.sh](#) – 下载然后在后台运行 add\_prometheus\_endpoint.sh 的脚本，以便节点可以完成引导。
  - [add\\_prometheus\\_endpoint.sh](#)— 一种脚本，它等待其运行的实例完成引导，然后将 CloudWatch 代理配置为发布到你在启动集群时作为参数提供的亚马逊托管服务 Prometheus 终端节点。
4. 打开每个文件并将 *my-s3-bucket* 替换为步骤 2 中的存储桶的名称。
5. 使用创建带有add\_prometheus\_endpoint\_launcher.sh引导操作的EMR集群。AWS CLI 在以下命令中，将 *my-s3-bucket* 替换为包含引导操作的存储桶，并将 *managedpro-remote-write-workspace-url* 替换为 Amazon Managed Service for Prometheus 工作区的远程写入端点。请务必指定一个等于emr-7.0.0或更高的 Amazon EMR 发行标签。

```

aws emr create-cluster --name managedpro-cluster \
  --release-label emr-7.0.0 \
  --applications Name=Hadoop Name=AmazonCloudWatchAgent \
  --ec2-attributes KeyName=myKey --instance-type m7g.2xlarge \
  --instance-count 3 --use-default-roles
  --bootstrap-actions Name='Add Prometheus Endpoint',Path=s3://my-s3-bucket/
  add_prometheus_endpoint_launcher.sh,Args='managedpro-remote-write-workspace-url'

```



当您的集群启动时，CloudWatch 代理会将其收集的指标发布到适用于 Prometheus 的亚马逊托管服务。

将 Amazon Managed Service for Prometheus 用作 Amazon Managed Grafana 的数据来源

亚马逊EMR将集群指标发布到亚马逊 Prometheus 托管服务后，您可以使用以下步骤使用亚马逊托管 Grafana 对指标进行可视化：

1. 使用创建亚马逊托管 Grafana 工作空间和具有相应权限的用户。AWS Management Console 有关更多信息，请参阅《Amazon Managed Grafana 用户指南》中的[创建工作区](#)。
2. 将 Amazon Managed Service for Prometheus 工作区添加作为 Amazon Managed Grafana 的数据来源。有关更多信息，请参阅《Amazon Managed Grafana 用户指南》中的[使用 AWS 数据来源配置添加 Amazon Managed Service for Prometheus 作为数据来源](#)。

#### Note

CloudWatch 代理有一个 Prometheus 导出器，可以重命名某些属性。对于默认指标标签，亚马逊 Prometheus 托管服务使用下划线字符代替亚马逊使用的句点。CloudWatch 因此，如果您使用 Amazon Managed Grafana 来可视化 Amazon Managed Service for Prometheus 中的默认指标，则标签将显示为 `jobflow_id`、`instance_id` 和 `service_name`。

此外，CloudWatch 代理发布到亚马逊 Prometheus 托管服务的所有应用程序指标都使用该标签代替。job service\_name但是，系统指标继续使用 `service_name` 标签。

## 注意事项和限制

- 本机 Amazon CloudWatch 代理适用于您在亚马逊 7.0.0 及更高EMR版本中创建的集群。
- Amazon EMR 7.0.0 中的 CloudWatch 代理配置需要使用[引导](#)操作。在即将发布的版本中，亚马逊 EMR将通过亚马逊提供其他配置选项EMRAPI。
- 如果您已经在创建集群的区域中通过其他方法（例如 CloudWatch AWS Systems Manager 代理（代理））部署了SSM代理，则无法安装 Amazon EMR CloudWatch 代理。这样做将导致您的集群因错误终止。

## CloudWatch 代理发布历史记录

下表列出了 Amazon 的每个发行版本中 AmazonCloudWatchAgent 包含的版本EMR，以及随应用程序一起安装的组件。有关每个发行版本中的组件版本，请参阅 [亚马逊 EMR 7.x 发布版本](#)、[亚马逊 EMR 6.x 发布版本](#) 或 [亚马逊 EMR 5.x 发布版本](#) 中的发行版“组件版本”部分。

### AmazonCloudWatchAgent 版本信息

亚马逊EMR发布标签	AmazonCloudWatchAgent 版本	安装在一起的组件 AmazonCloudWatchAgent
emr-7.2.0	1.300032.2	adot-java-agent, emrfs, emr-amazon-cloudwatch-agent, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-7.1.0	1.300032.2	adot-java-agent, emrfs, emr-amazon-cloudwatch-agent, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

亚马逊EMR发布标签	AmazonCloudWatchAgent 版本	安装在一起的组件 AmazonCloudWatchAgent
emr-7.0.0	1.300031.1	adot-java-agent, emr-amazon-cloudwatch-agent

## Delta Lake

Delta Lake 是一种存储层框架，用于构建通常在 Amazon S3 上构建的数据湖仓一体架构。在亚马逊 EMR 版本 6.9.0 及更高版本中，您可以在带有 Delta Lake 表的 EMR 亚马逊集群上使用 [Apache Spark 3.x](#)。有关使用 Delta Lake 构建湖仓一体架构的更多信息，请参阅 <https://delta.io/>。

下表列出了最新版本的亚马逊 EMR 7.x 系列中包含的 Delta 版本，以及亚马逊与 Delta 一起 EMR 安装的组件。

有关此版本中与 Delta 一起安装的组件的版本，请参阅 [7.2.0 版组件版本](#)。

emr-7.2.0 的 Delta 版本信息

Amazon EMR 发布标签	Delta 版本	随 Delta 安装的组件
emr-7.2.0	Delta 3.1.0	Not available.

下表列出了最新版本的亚马逊 EMR 6.x 系列中包含的 Delta 版本，以及亚马逊与 Delta 一起 EMR 安装的组件。

有关此发行版中随 Delta 安装的组件版本，请参阅 [发行版 6.15.0 组件版本](#)。

emr-6.15.0 的 Delta 版本信息

Amazon EMR 发布标签	Delta 版本	随 Delta 安装的组件
emr-6.15.0	Delta 2.4.0	Not available.

## Delta Lake 简介

Delta Lake 是一个开源项目，可帮助实施通常构建在 Amazon S3 上的现代数据湖架构。Delta Lake 提供以下功能：

- Spark 上的原子、一致、隔离、持久 (ACID) 事务。在 Spark 作业期间，读者可以看到一致的表格视图。
- 可扩展的元数据处理，由 Spark 进行分布式处理。
- 使用相同的 Delta 表结合流处理和批处理使用案例。
- 强制执行自动架构以避免数据摄取期间出现错误记录。

- 使用数据版本控制进行时空旅行。
- 支持针对复杂用例的合并、更新和删除操作，例如更改数据捕获 (CDC)、串流 upsert 等。

## 使用安装有 Delta Lake 的集群

### 主题

- [将 Delta Lake 集群与 Flink 结合使用](#)
- [将 Delta Lake 集群与 Trino 结合使用](#)
- [将 Delta Lake 集群与 Spark 结合使用](#)
- [将 Delta Lake 集群与 Spark 和 AWS Glue 结合使用](#)

## 将 Delta Lake 集群与 Flink 结合使用

在亚马逊 6.11 及更高EMR版本中，您可以将 Delta Lake 与 Flink 集群配合使用。以下示例使用在 Amazon EMR Flink 集群上使用 Delta Lake。AWS CLI

### Note

当您将 Delta Lake 与 Flink 集群配合使用 DataStream API时，亚马逊EMR支持 Flink。

## 创建 Delta Lake 集群

1. 创建文件 `delta_configurations.json` 并输入以下内容：

```
[{"Classification":"delta-defaults",  
  "Properties":{"delta.enabled":"true"}}]
```

2. 使用以下配置创建集群。在该 URL 中，将 example Amazon S3 bucket path 和 subnet ID 替换为您自己的值。

```
aws emr create-cluster  
--release-label emr-6.11.0  
--applications Name=Flink  
--configurations file://delta_configurations.json  
--region us-east-1 --name My_Spark_Delta_Cluster  
--log-uri s3://DOC-EXAMPLE-BUCKET/
```

```
--instance-type m5.xlarge
--instance-count 3
--service-role EMR_DefaultRole_V2
--ec2-attributes
  InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-1234567890abcdef0
```

## 初始化 Flink yarn 会话

要初始化 Flink yarn 会话，请运行以下命令：

```
flink-yarn-session -d
```

## 使用 Delta Lake 创建 Flink 作业

以下示例展示如何使用 sbt 或 Maven 在 Delta Lake 中构建 Flink 作业。

### sbt

[sbt](#) 是 Scala 的构建工具，当您处理小型项目时，只需很少甚至不需要配置即可使用。

```
libraryDependencies += Seq(
  "io.delta" %% "delta-flink" % deltaConnectorsVersion % "provided",
  "io.delta" %% "delta-standalone" % deltaConnectorsVersion % "provided",
  "org.apache.flink" %% "flink-clients" % flinkVersion % "provided",
  "org.apache.flink" %% "flink-parquet" % flinkVersion % "provided",
  "org.apache.hadoop" % "hadoop-client" % hadoopVersion % "provided",
  "org.apache.flink" % "flink-table-common" % flinkVersion % "provided",
  "org.apache.flink" %% "flink-table-runtime" % flinkVersion % "provided")
```

### Maven

[Maven](#) 是 Apache Software Foundation 推出的开源构建自动化工具。使用 Maven，你可以在亚马逊上使用 Delta Lake 构建、发布和部署 Flink 作业。EMR

```
<project>
<properties>
  <scala.main.version>2.12</scala.main.version>
  <delta-connectors-version>0.6.0</delta-connectors-version>
  <flink-version>1.16.1</flink-version>
  <hadoop-version>3.1.0</hadoop-version>
</properties>
```

```
<dependencies>
  <dependency>
    <groupId>io.delta</groupId>
    <artifactId>delta-flink</artifactId>
    <version>${delta-connectors-version}</version>
    <scope>provided</scope>
  </dependency>
  <dependency>
    <groupId>io.delta</groupId>
    <artifactId>delta-standalone_${scala-main-version}</artifactId>
    <version>${delta-connectors-version}</version>
    <scope>provided</scope>
  </dependency>
  <dependency>
    <groupId>org.apache.flink</groupId>
    <artifactId>flink-clients</artifactId>
    <version>${flink-version}</version>
    <scope>provided</scope>
  </dependency>
  <dependency>
    <groupId>org.apache.flink</groupId>
    <artifactId>flink-parquet</artifactId>
    <version>${flink-version}</version>
    <scope>provided</scope>
  </dependency>
  <dependency>
    <groupId>org.apache.hadoop</groupId>
    <artifactId>hadoop-client</artifactId>
    <version>${hadoop-version}</version>
    <scope>provided</scope>
  </dependency>
  <dependency>
    <groupId>org.apache.flink</groupId>
    <artifactId>flink-table-common</artifactId>
    <version>${flink-version}</version>
    <scope>provided</scope>
  </dependency>
  <dependency>
    <groupId>org.apache.flink</groupId>
    <artifactId>flink-table-runtime</artifactId>
    <version>${flink-version}</version>
    <scope>provided</scope>
  </dependency>
</dependencies>
```

```
</dependencies>
```

## 使用 Flink 数据流写入增量表 API

使用以下示例创建要写入 DeltaSink 到带有 a 的表中 deltaTablePath:

```
public static DataStream<RowData> createDeltaSink(
    DataStream<RowData> stream,
    String deltaTablePath,
    RowType rowType) {
    Configuration configuration = new Configuration();
    DeltaSink<RowData> deltaSink = DeltaSink
        .forRowData(
            new org.apache.flink.core.fs.Path(deltaTablePath),
            configuration,
            rowType)
        .build();
    stream.sinkTo(deltaSink);
    return stream;
}
```

## 使用 Flink 数据流从增量表中读取 API

使用以下示例创建要从表中读 DeltaSource 取的有界值，其中带有 deltaTablePath:

```
public static DataStream<RowData> createBoundedDeltaSourceAllColumns(
    StreamExecutionEnvironment env,
    String deltaTablePath) {
    Configuration configuration = new Configuration();
    DeltaSource<RowData> deltaSource = DeltaSource
        .forBoundedRowData(
            new org.apache.flink.core.fs.Path(deltaTablePath),
            configuration)
        .build();

    return env.fromSource(deltaSource, WatermarkStrategy.noWatermarks(), "delta-
source");
}
```



## 使用对 Delta Lake 独立版的多集群支持创建接收器

使用以下示例创建 DeltaSink [支持deltaTablePath和多集群](#)的待写入表：

```
public DataStream<RowData> createDeltaSink(
    DataStream<RowData> stream,
    String deltaTablePath) {
    Configuration configuration = new Configuration();
    configuration.set("spark.delta.logStore.s3.impl",
        "io.delta.storage.S3DynamoDBLogStore");
    configuration.set("spark.io.delta.storage.S3DynamoDBLogStore.ddb.tableName",
        "delta_log");
    configuration.set("spark.io.delta.storage.S3DynamoDBLogStore.ddb.region", "us-
        east-1");

    DeltaSink<RowData> deltaSink = DeltaSink
        .forRowData(
            new Path(deltaTablePath),
            configuration,
            rowType)
        .build();
    stream.sinkTo(deltaSink);
    return stream;
}
```

## 运行 Flink 作业

使用下列命令以运行您的作业：

```
flink run FlinkJob.jar
```

## 将 Delta Lake 集群与 Trino 结合使用

在 Amazon 6.9.0 及更高EMR版本中，您可以将 Delta Lake 与 Trino 集群配合使用。

在本教程中，我们将使用在 Amazon EMR Trino 集群上使用 Delta Lake。AWS CLI

### 创建 Delta Lake 集群

1. 创建文件 `delta_configurations.json`，然后为您选择的目录设置值。例如，假设您想将 Hive 元存储作为目录使用，则您的文件应包含以下内容：

```
[{"Classification":"delta-defaults",
  "Properties":{"delta.enabled":"true"}},
 {"Classification":"trino-connector-delta",
  "Properties":{"hive.metastore.uri":"thrift://localhost:9083"}}]
```

如果您想使用 AWS Glue 目录作为商店，则您的文件应包含以下内容：

```
[{"Classification":"delta-defaults",
  "Properties":{"delta.enabled":"true"}},
 {"Classification":"trino-connector-delta",
  "Properties":{"hive.metastore":"glue"}}]
```

2. 使用以下配置创建集群，将 **example Amazon S3 bucket path** 和 **subnet ID** 替换为您自己的值。

```
aws emr create-cluster
  --release-label emr-6.9.0
  --applications Name=Trino
  --configurations file://delta_configurations.json
  --region us-east-1 --name My_Spark_Delta_Cluster
  --log-uri s3://DOC-EXAMPLE-BUCKET/
  --instance-type m5.xlarge
  --instance-count 2
  --service-role EMR_DefaultRole_V2
  --ec2-attributes
  InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-1234567890abcdef0
```

## 初始化 Delta Lake 的 Trino 会话

要初始化 Trino 会话，请运行以下命令。

```
trino-cli --catalog delta
```

## 写入 Delta Lake 表

使用以下 SQL 命令创建并写入您的表：

```
SHOW SCHEMAS;
```

```
CREATE TABLE default.delta_table (id int, data varchar, category varchar) WITH
( location = 's3://DOC-EXAMPLE-BUCKET/<prefix>');

INSERT INTO default.delta_table VALUES (1,'a','c1'), (2,'b','c2'), (3,'c','c3');
```

## 从 Delta Lake 表中读取

使用以下SQL命令从表中读取：

```
SELECT * from default.delta_table;
```

## 将 Delta Lake 集群与 Spark 结合使用

从 Amazon 6.9.0 EMR 版本开始，您无需执行引导操作即可将 Delta Lake 与 Spark 集群配合使用。对于 Amazon 6.8.0 及更低EMR版本，您可以使用引导操作来预安装必要的依赖项。

以下示例使用在 Amazon EMR Spark 集群上使用 Delta Lake。AWS CLI

要将 Amazon 上的 Delta Lake EMR 与配合使用 AWS Command Line Interface，请先创建一个集群。有关如何使用指定 Delta Lake 分类的信息 AWS Command Line Interface，请参阅[在创建集群 AWS Command Line Interface 时使用提供配置或在创建集群SDK时使用 Java 提供配置](#)。

1. 创建文件 `configurations.json` 并输入以下内容：

```
[{"Classification":"delta-defaults", "Properties":{"delta.enabled":"true"} ]]
```

2. 使用以下配置创建集群，将示例 Amazon S3 **bucket path** 和 **subnet ID** 替换为您自己的值。

```
aws emr create-cluster
  --release-label emr-6.9.0
  --applications Name=Spark
  --configurations file://delta_configurations.json
  --region us-east-1
  --name My_Spark_Delta_Cluster
  --log-uri s3://DOC-EXAMPLE-BUCKET/
  --instance-type m5.xlarge
  --instance-count 2
  --service-role EMR_DefaultRole_V2
```

```
--ec2-attributes  
InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-1234567890abcdef0
```

或者，您可以创建 Amazon EMR 集群和 Spark 应用程序，并将以下文件作为 Spark 任务中的JAR 依赖项：

```
/usr/share/aws/delta/lib/delta-core.jar,  
/usr/share/aws/delta/lib/delta-storage.jar,  
/usr/share/aws/delta/lib/delta-storage-s3-dynamodb.jar
```

#### Note

如果您使用的是亚马逊 6.9.0 或更高EMR版本，请/`usr/share/aws/delta/lib/delta-spark.jar`改用。`/usr/share/aws/delta/lib/delta-core.jar`

有关更多信息，请参阅[提交应用程序](#)。

要将 jar 依赖项包含在 Spark 任务中，您可以将以下配置属性添加到 Spark 应用程序中：

```
--conf "spark.jars=/usr/share/aws/delta/lib/delta-core.jar,  
/usr/share/aws/delta/lib/delta-storage.jar,  
/usr/share/aws/delta/lib/delta-storage-s3-dynamodb.jar"
```

有关 Spark 任务依赖项的更多信息，请参阅 [Dependency Management](#) ( 依赖项管理 )。

如果您使用的是 Amazon 6.9.0 或更高EMR版本，请改为添加/`usr/share/aws/delta/lib/delta-spark.jar`配置。

```
--conf "spark.jars=/usr/share/aws/delta/lib/delta-spark.jar,  
/usr/share/aws/delta/lib/delta-storage.jar,  
/usr/share/aws/delta/lib/delta-storage-s3-dynamodb.jar"
```

## 初始化 Delta Lake 的 Spark 会话

以下示例展示了如何启动交互式 Spark 外壳、使用 Spark 提交或使用亚马逊EMR笔记本在亚马逊上使用 Delta Lake EMR。

## spark-shell

1. 使用连接到主节点SSH。有关更多信息，请参阅 [《Amazon EMR 管理指南》SSH中的“使用连接到主节点”](#)。
2. 输入以下命令以启动 Spark shell。要使用 PySpark 外壳，请spark-shell替换为pyspark。

```
spark-shell \  
  --conf "spark.sql.extensions=io.delta.sql.DeltaSparkSessionExtension" \  
  --conf  
  "spark.sql.catalog.spark_catalog=org.apache.spark.sql.delta.catalog.DeltaCatalog"
```

如果您运行亚马逊EMR版本 6.15.0 或更高版本，则还必须使用以下配置来使用基于带有 Delta Lake 的 Lake Formation 的精细访问控制。

```
spark-shell \  
  --conf  
  spark.sql.extensions=io.delta.sql.DeltaSparkSessionExtension,com.amazonaws.emr.recordserver  
  \  
  --conf  
  spark.sql.catalog.spark_catalog=org.apache.spark.sql.delta.catalog.DeltaCatalog \  
  
  --conf spark.sql.catalog.spark_catalog.lf.managed=true
```

## spark-submit

1. 使用连接到主节点SSH。有关更多信息，请参阅 [《Amazon EMR 管理指南》SSH中的“使用连接到主节点”](#)。
2. 输入以下命令以启动 Delta Lake 的 Spark 会话。

```
spark-submit  
-conf "spark.sql.extensions=io.delta.sql.DeltaSparkSessionExtension"  
-conf  
  "spark.sql.catalog.spark_catalog=org.apache.spark.sql.delta.catalog.DeltaCatalog"
```

如果您运行亚马逊EMR版本 6.15.0 或更高版本，则还必须使用以下配置来使用基于带有 Delta Lake 的 Lake Formation 的精细访问控制。

```
spark-submit \  
  `
```

```
--conf
  spark.sql.extensions=io.delta.sql.DeltaSparkSessionExtension,com.amazonaws.emr.recordserver
--conf
  spark.sql.catalog.spark_catalog=org.apache.spark.sql.delta.catalog.DeltaCatalog \
--conf spark.sql.catalog.spark_catalog.lf.managed=true
```

## EMR Studio notebooks

要使用 Amazon EMR Studio 笔记本初始化 Spark 会话，请在亚马逊EMR笔记本中使用 `%configure` 神奇命令配置 Spark 会话，如下例所示。有关更多信息，请参阅《Amazon EMR 管理指南》中的[“使用EMR笔记本魔法”](#)。

```
%%configure -f
{
  "conf": {
    "spark.sql.extensions": "io.delta.sql.DeltaSparkSessionExtension",
    "spark.sql.catalog.spark_catalog":
"org.apache.spark.sql.delta.catalog.DeltaCatalog"
  }
}
```

如果您运行亚马逊EMR版本 6.15.0 或更高版本，则还必须使用以下配置来使用基于带有 Delta Lake 的 Lake Formation 的精细访问控制。

```
%%configure -f
{
  "conf": {
    "spark.sql.extensions":
"io.delta.sql.DeltaSparkSessionExtension,com.amazonaws.emr.recordserver.connector.spark.sql
    "spark.sql.catalog.spark_catalog":
"org.apache.spark.sql.delta.catalog.DeltaCatalog",
    "spark.sql.catalog.spark_catalog.lf.managed": "true"
  }
}
```

## 写入 Delta Lake 表

以下示例说明如何创建 DataFrame 并将其写为 Delta Lake 数据集。该示例说明了如何在使用默认 hadoop 用户连接到主节点时使用 SSH Spark shell 处理数据集。

**Note**

要将代码示例粘贴到 Spark Shell 中，请在提示符处键入 `:paste`，粘贴示例，然后按 `CTRL + D`。

## PySpark

Spark 包含基于 Python 的 Shell `pyspark`，您可以用它来设计以 Python 编写的 Spark 程序的原型。就像使用 `spark-shell` 一样，在主节点上调用 `pyspark`。

```
## Create a DataFrame
data = spark.createDataFrame([("100", "2015-01-01", "2015-01-01T13:51:39.340396Z"),
                              ("101", "2015-01-01", "2015-01-01T12:14:58.597216Z"),
                              ("102", "2015-01-01", "2015-01-01T13:51:40.417052Z"),
                              ("103", "2015-01-01", "2015-01-01T13:51:40.519832Z")],
                              ["id", "creation_date", "last_update_time"])

## Write a DataFrame as a Delta Lake dataset to the S3 location
spark.sql("""CREATE TABLE IF NOT EXISTS delta_table (id string, creation_date
string,
last_update_time string)
USING delta location
's3://DOC-EXAMPLE-BUCKET/example-prefix/db/delta_table'""");

data.writeTo("delta_table").append()
```

## Scala

```
import org.apache.spark.sql.SaveMode
import org.apache.spark.sql.functions._

// Create a DataFrame
val data = Seq(("100", "2015-01-01", "2015-01-01T13:51:39.340396Z"),
              ("101", "2015-01-01", "2015-01-01T12:14:58.597216Z"),
              ("102", "2015-01-01", "2015-01-01T13:51:40.417052Z"),
              ("103", "2015-01-01", "2015-01-01T13:51:40.519832Z")).toDF("id", "creation_date",
"last_update_time")

// Write a DataFrame as a Delta Lake dataset to the S3 location
spark.sql("""CREATE TABLE IF NOT EXISTS delta_table (id string,
creation_date string,
```

```
last_update_time string)
USING delta location
's3://DOC-EXAMPLE-BUCKET/example-prefix/db/delta_table'");

data.write.format("delta").mode("append").saveAsTable("delta_table")
```

## SQL

```
-- Create a Delta Lake table with the S3 location
CREATE TABLE delta_table(id string,
creation_date string,
last_update_time string)
USING delta LOCATION
's3://DOC-EXAMPLE-BUCKET/example-prefix/db/delta_table';

-- insert data into the table
INSERT INTO delta_table VALUES ("100", "2015-01-01",
"2015-01-01T13:51:39.340396Z"),
("101", "2015-01-01", "2015-01-01T12:14:58.597216Z"),
("102", "2015-01-01", "2015-01-01T13:51:40.417052Z"),
("103", "2015-01-01", "2015-01-01T13:51:40.519832Z");
```

## 从 Delta Lake 表中读取

### PySpark

```
ddf = spark.table("delta_table")
ddf.show()
```

### Scala

```
val ddf = spark.table("delta_table")
ddf.show()
```

### SQL

```
SELECT * FROM delta_table;
```



## 将 Delta Lake 集群与 Spark 和 AWS Glue 结合使用

要使用 Glue AWS Catalog 作为 Delta Lake 表格的元数据仓，请按照以下步骤创建一个集群。有关使用指定 Delta Lake 分类的信息 AWS Command Line Interface，请参阅[在创建集群 AWS Command Line Interface 时使用提供配置或在创建集群SDK时使用 Java 提供配置](#)。

### 创建 Delta Lake 集群

1. 创建文件 `configurations.json` 并输入以下内容：

```
[{"Classification":"delta-defaults",
  "Properties":{"delta.enabled":"true"}},
{"Classification":"spark-hive-site",
  "Properties":
  {"hive.metastore.client.factory.class":"com.amazonaws.glue.catalog.metastore.AWSGlueDataCat
```

2. 使用以下配置创建集群，将 **example Amazon S3 bucket path** 和 **subnet ID** 替换为您自己的值。

```
aws emr create-cluster
  --release-label emr-6.9.0
  --applications Name=Spark
  --configurations file://delta_configurations.json
  --region us-east-1
  --name My_Spark_Delta_Cluster
  --log-uri s3://DOC-EXAMPLE-BUCKET/
  --instance-type m5.xlarge
  --instance-count 2
  --service-role EMR_DefaultRole_V2
  --ec2-attributes
  InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-1234567890abcdef0
```

## 注意事项和限制

- 亚马逊 6.9.0 及更高EMR版本支持 Delta Lake。你可以在带增量表的亚马逊EMR集群上使用 [Apache Spark 3.x](#)。
- 我们建议您对 S3 位置路径使用 s3 URI 方案，而不是 s3a，以获得最佳性能、安全性和可靠性。有关更多信息，请参阅[使用存储和文件系统](#)。

- 在 Amazon EMR 7.0 中 Delta Universal Format (UniForm)，不支持转换为冰山语句。
- 在 Amazon EMR 6.9 和 6.10 中，当您将三角湖表数据存储在 Amazon S3 中时，列数据将在列重命名操作 NULL 后变为列数据。此问题已在 Amazon EMR 6.11 中得到解决。有关此实验性列重命名操作的更多信息，请参阅《Delta Lake User Guide》中的 [Column rename operation](#)。
- 如果您在 Apache Spark 之外的 AWS Glue 数据目录中创建数据库，则该数据库可能有一个空 LOCATION 字段。由于 Spark 不允许使用空位置属性创建数据库，因此如果您在 Amazon EMR 中使用 Spark 在 Glue 数据库中创建 Delta 表，并且该数据库的 LOCATION 属性为空，则会出现以下错误：

```
IllegalArgumentException: Can not create a Path from an empty string
```

要解决此问题，请在数据目录中创建数据库，并且 LOCATION 字段使用有效的非空路径。有关实现此解决方案的步骤，请参阅《Amazon Athena 用户指南》中的 [创建表时出现非法参数异常](#)。

## Delta 发行版历史记录

下表列出了 Amazon 的每个发行版本中包含的 Delta 版本 EMR，以及随应用程序一起安装的组件。有关每个发行版本中的组件版本，请参阅 [亚马逊 EMR 7.x 发布版本](#)、[亚马逊 EMR 6.x 发布版本](#) 或 [亚马逊 EMR 5.x 发布版本](#) 中的发行版“组件版本”部分。

### Delta 版本信息

Amazon EMR 发布标签	Delta 版本	随 Delta 安装的组件
emr-7.2.0	3.1.0	Not available.
emr-7.1.0	3.0.0	Not available.
emr-7.0.0	3.0.0	Not available.
emr-6.15.0	2.4.0	Not available.
emr-6.14.0	2.4.0	Not available.
emr-6.13.0	2.4.0	Not available.
emr-6.12.0	2.4.0	Not available.
emr-6.11.1	2.2.0	Not available.

Amazon EMR 发布标签	Delta 版本	随 Delta 安装的组件
emr-6.11.0	2.2.0	Not available.
emr-6.10.1	2.2.0	Not available.
emr-6.10.0	2.2.0	Not available.
emr-6.9.1	2.1.0	Not available.
emr-6.9.0	2.1.0	Not available.

# Apache Flink

[Apache Flink](#) 是一个流式处理数据流引擎，您可以使用此引擎在高吞吐量数据源上轻松运行实时流处理。Flink 支持事件的事件时间语义、exactly-once 语义、背压控制，并针对编写流式和批处理应用程序APIs进行了优化。out-of-order

此外，Flink 具有适用于第三方数据源的连接器，例如以下内容：

- [Amazon Kinesis Data Streams](#)
- [Apache Kafka](#)
- [Flink Elasticsearch Connector](#)
- [推特直播 API](#)
- [Cassandra](#)

Amazon EMR 支持 Flink 作为YARN应用程序，因此您可以管理集群中的资源以及其他应用程序。Flink-on-YARN 允许您提交临时的 Flink 作业，或者您可以创建一个长时间运行的集群，该集群接受多个任务并根据总体预留量分配资源。YARN

Flink 包含在亚马逊EMR发行版 5.1.0 及更高版本中。

## Note

亚马逊EMR发行版 5.2.1 中增加了对该FlinkKinesisConsumer类的支持。

下表列出了最新版本的亚马逊 EMR 7.x 系列中包含的 Flink 版本，以及亚马逊与 Flink 一起EMR安装的组件。

有关此版本中与 Flink 一起安装的组件的版本，请参阅 [7.2.0 版本的组件](#)版本。

emr-7.2.0 的 Flink 版本信息

亚马逊EMR发行标签	Flink 版本	随 Flink 安装的组件
emr-7.2.0	Flink 1.18.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano

亚马逊EMR发行标签	Flink 版本	随 Flink 安装的组件
		de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta

下表列出了最新版本的亚马逊 EMR 6.x 系列中包含的 Flink 版本，以及亚马逊与 Flink 一起EMR安装的组件。

有关此发行版中随 Flink 安装的组件版本，请参阅[发行版 6.15.0 组件版本](#)。

emr-6.15.0 的 Flink 版本信息

亚马逊EMR发行标签	Flink 版本	随 Flink 安装的组件
emr-6.15.0	Flink 1.17.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta-standalone-connectors

下表列出了最新版本的亚马逊 EMR 5.x 系列中包含的 Flink 版本，以及亚马逊与 Flink 一起EMR安装的组件。

有关此版本中与 Flink 一起安装的组件的版本，请参阅[5.36.2 版本](#)的组件版本。

## emr-5.36.2 的 Flink 版本信息

亚马逊EMR发行标签	Flink 版本	随 Flink 安装的组件
emr-5.36.2	Flink 1.14.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config

### 主题

- [使用 Flink 创建集群](#)
- [在亚马逊中配置 Flink EMR](#)
- [在 Amazon EMR 中使用 Flink 作业](#)
- [使用 Scala Shell](#)
- [查找 Flink Web 界面](#)
- [Flink Autoscaler](#)
- [优化任务恢复和扩展操作的作业重启时间](#)
- [在 Amazon EMR 中通过 Zeppelin 使用 Flink 作业](#)
- [Flink 发布历史记录](#)

## 使用 Flink 创建集群

您可以使用 AWS Management Console AWS CLI、或启动集群 AWS SDK。

使用控制台启动安装了 Flink 的集群

1. 打开亚马逊EMR控制台，[网址为 https://console.aws.amazon.com/emr](https://console.aws.amazon.com/emr)。
2. 依次选择 Create cluster (创建集群)、Go to advanced options (转到高级选项)。

3. 对于软件配置，请选择EMR版本 `emr-5.1.0` 或更高版本。
4. 选择 Flink 作为应用程序（与要安装的任何其它应用程序一起）。
5. 根据需要选择其它选项，然后选择 Create cluster (创建集群)。

要使用 Flink 启动集群，请从 AWS CLI

- 使用下面的命令创建集群：

```
aws emr create-cluster --release-label emr-7.2.0 \  
--applications Name=Flink \  
--configurations file:///./configurations.json \  
--region us-east-1 \  
--log-uri s3://myLogUri \  
--instance-type m5.xlarge \  
--instance-count 2 \  
--service-role EMR_DefaultRole_V2 \  
--ec2-attributes KeyName=MyKeyName,InstanceProfile=EMR_EC2_DefaultRole \  
--steps Type=CUSTOM_JAR,Jar=command-runner.jar,Name=Flink_Long_Running_Session,\  
Args=flink-yarn-session,-d
```

#### Note

为了便于读取，包含 Linux 行继续符 (\)。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (^)。

## 在亚马逊中配置 Flink EMR

### 使用 Hive 元存储和 Glue 目录配置 Flink

亚马逊EMR发布的 6.9.0 及更高版本支持 Hive Metastore 和 Glue Catalog，并 AWS 通过 Apache Flink 连接到 Hive。本部分概括介绍了使用 Flink 配置 [AWS Glue 目录](#) 和 [Hive 元存储](#) 所需的步骤。

#### 主题

- [使用 Hive 元存储](#)
- [使用 AWS Glue 数据目录](#)

## 使用 Hive 元存储

1. 创建 6.9.0 或更高版本的EMR集群以及至少两个应用程序：Hive 和 Flink。
2. 使用[脚本运行程序](#)将以下脚本作为步骤函数执行：

hive-metastore-setup.sh

```
sudo cp /usr/lib/hive/lib/antlr-runtime-3.5.2.jar /usr/lib/flink/lib
sudo cp /usr/lib/hive/lib/hive-exec-3.1.3*.jar /lib/flink/lib
sudo cp /usr/lib/hive/lib/libfb303-0.9.3.jar /lib/flink/lib
sudo cp /usr/lib/flink/opt/flink-connector-hive_2.12-1.15.2.jar /lib/flink/lib
sudo chmod 755 /usr/lib/flink/lib/antlr-runtime-3.5.2.jar
sudo chmod 755 /usr/lib/flink/lib/hive-exec-3.1.3*.jar
sudo chmod 755 /usr/lib/flink/lib/libfb303-0.9.3.jar
sudo chmod 755 /usr/lib/flink/lib/flink-connector-hive_2.12-1.15.2.jar
```

**Add step**
✕

**Step type** Custom JAR

**Name\***

**JAR location\***  JAR location maybe a path into S3 or a fully qualified java class in the classpath.

**Arguments**  These are passed to the main function in the JAR. If the JAR does not specify a main class in its manifest file you can specify another class name as the first argument.

**Action on failure**  What happens if the step fails

Cancel Save

## 使用 AWS Glue 数据目录

1. 创建 6.9.0 或更高版本的EMR集群以及至少两个应用程序：Hive 和 Flink。
2. 在 AWS Glue 数据目录设置中选择用于 Hive 表元数据，以在集群中启用数据目录。
3. 使用[脚本运行器](#)将以下脚本作为步骤函数执行：[在 Amazon EMR 集群上运行命令和脚本](#)：

glue-catalog-setup.sh



```

sudo cp /usr/lib/hive/auxlib/aws-glue-datacatalog-hive3-client.jar /usr/lib/flink/lib
sudo cp /usr/lib/hive/lib/antlr-runtime-3.5.2.jar /usr/lib/flink/lib
sudo cp /usr/lib/hive/lib/hive-exec-3.1.3*.jar /usr/lib/flink/lib
sudo cp /usr/lib/hive/lib/libfb303-0.9.3.jar /usr/lib/flink/lib
sudo cp /usr/lib/flink/opt/flink-connector-hive_2.12-1.15.2.jar /usr/lib/flink/lib
sudo chmod 755 /usr/lib/flink/lib/aws-glue-datacatalog-hive3-client.jar
sudo chmod 755 /usr/lib/flink/lib/antlr-runtime-3.5.2.jar
sudo chmod 755 /usr/lib/flink/lib/hive-exec-3.1.3*.jar
sudo chmod 755 /usr/lib/flink/lib/libfb303-0.9.3.jar
sudo chmod 755 /usr/lib/flink/lib/flink-connector-hive_2.12-1.15.2.jar

```

**Add step** ✕

**Step type** Custom JAR

**Name\***

**JAR location\***  JAR location maybe a path into S3 or a fully qualified java class in the classpath.

**Arguments**  These are passed to the main function in the JAR. If the JAR does not specify a main class in its manifest file you can specify another class name as the first argument.

**Action on failure**  What happens if the step fails

Cancel
Save

## 使用配置文件配置 Flink

您可以使用 Amazon EMR 配置API通过配置文件配置 Flink。可在中配置的文件API有：

- flink-conf.yaml
- log4j.properties
- flink-log4j-session
- log4j-cli.properties

Flink 的主配置文件的名称为 flink-conf.yaml。

## 从 AWS CLI 配置用于 Flink 的任务槽的数目

1. 创建文件 `configurations.json` 并输入以下内容：

```
[
  {
    "Classification": "flink-conf",
    "Properties": {
      "taskmanager.numberOfTaskSlots": "2"
    }
  }
]
```

2. 接下来，使用以下配置创建集群：

```
aws emr create-cluster --release-label emr-7.2.0 \
--applications Name=Flink \
--configurations file://./configurations.json \
--region us-east-1 \
--log-uri s3://myLogUri \
--instance-type m5.xlarge \
--instance-count 2 \
--service-role EMR_DefaultRole_V2 \
--ec2-attributes KeyName=YourKeyName,InstanceProfile=EMR_EC2_DefaultRole
```

### Note

您也可以使用 Flink API 更改某些配置。有关更多信息，请参阅 Flink 文档中的[概念](#)。在 Amazon 5.21.0 及更高 EMR 版本中，您可以覆盖集群配置，并为正在运行的集群中的每个实例组指定其他配置分类。为此，您可以使用 Amazon EMR 控制台、AWS Command Line Interface (AWS CLI) 或 AWS SDK。有关更多信息，请参阅[为运行的集群中的实例组提供配置](#)。

## 并行选项

作为应用程序所有者，您最了解应将哪些资源分配给 Flink 中的任务。对于本文档中的示例，请使用与您用于应用程序的任务实例相同的任务数量。通常，我们建议对初始并行级别执行此操作，但您也可以使用任务槽来增加并行粒度，它一般不应超过每实例[虚拟内核](#)数量。有关 Flink 架构的更多信息，请参阅 Flink 文档中的 [Concepts](#)。

## 在具有多个主EMR节点的集群上配置 Flink

在具有多个主节点的 Amazon EMR 集群的主节点故障转移过程中，Flink 的 `of` 仍然可用。JobManager 从 Amazon EMR 5.28.0 开始，还会自动启用 JobManager 高可用性。无需手动配置。

在 Amazon 5.27.0 或更早EMR版本中，JobManager 这是单点故障。当 JobManager 失败时，它会丢失所有作业状态，并且不会恢复正在运行的作业。您可以通过配置应用程序尝试次数、检查点并启用 Flink 的状态存储来实现 JobManager ZooKeeper 高可用性，如以下示例所示：

```
[
  {
    "Classification": "yarn-site",
    "Properties": {
      "yarn.resourcemanager.am.max-attempts": "10"
    }
  },
  {
    "Classification": "flink-conf",
    "Properties": {
      "yarn.application-attempts": "10",
      "high-availability": "zookeeper",
      "high-availability.zookeeper.quorum": "%{hiera('hadoop:zk')}",
      "high-availability.storageDir": "hdfs:///user/flink/recovery",
      "high-availability.zookeeper.path.root": "/flink"
    }
  }
]
```

您必须同时配置 Flink 的最大应用程序主尝试次数YARN和应用程序尝试次数。有关更多信息，请参阅[YARN集群高可用性配置](#)。您可能还需要配置 Flink 检查点以使重新启动的作业从之前完成的检查点 JobManager 恢复。有关更多信息，请参阅[开展 Flink 检查点检验](#)。

### 配置内存进程大小

对于使用 Flink 1.11.x 的 Amazon EMR 版本，您必须在中同时配置 JobManager (`jobmanager.memory.process.size`) 和 TaskManager (`taskmanager.memory.process.size`) 的内存进程总大小。`flink-conf.yaml`您可以通过配置集群来设置这些值，也可以通过SSH手动取消对这些字段的注释。APIFlink 提供以下默认值。

- `jobmanager.memory.process.size` : 1600m
- `taskmanager.memory.process.size` : 1728m

要排除JVM元空间和开销，请使用 Flink 内存总大小 (`taskmanager.memory.flink.size`) 代替。`taskmanager.memory.process.size`和`taskmanager.memory.process.size`的默认值为 1280m。不建议同时设置 `taskmanager.memory.process.size` 和 `taskmanager.memory.process.size`。

所有使用 Flink 1.12.0 及更高EMR版本的亚马逊版本都将 Flink 的开源集合中列出的默认值作为亚马逊上的默认值EMR，因此您无需自己进行配置。

## 配置日志输出文件大小

Flink 应用程序容器创建并写入三种类型的日志文件：`.out` 文件、`.log` 文件和 `.err` 文件。仅限将 `.err` 文件压缩并从文件系统中删除，而将 `.log` 和 `.out` 日志文件保留在文件系统中。为确保这些输出文件保持可管理以及集群保持稳定，您可以在 `log4j.properties` 设置文件的上限数量并限制其大小。

### 亚马逊EMR版本 5.30.0 及更高版本

从 Amazon EMR 5.30.0 开始，Flink 使用配置分类名称的 `log4j2` 日志框架。以下示例配置演示了 `log4j2 flink-log4j` 格式。

```
[
  {
    "Classification": "flink-log4j",
    "Properties": {
      "appender.main.name": "MainAppender",
      "appender.main.type": "RollingFile",
      "appender.main.append" : "false",
      "appender.main.fileName" : "${sys:log.file}",
      "appender.main.filePattern" : "${sys:log.file}.%i",
      "appender.main.layout.type" : "PatternLayout",
      "appender.main.layout.pattern" : "%d{yyyy-MM-dd HH:mm:ss,SSS} %-5p %-60c %x - %m
%n",
      "appender.main.policies.type" : "Policies",
      "appender.main.policies.size.type" : "SizeBasedTriggeringPolicy",
      "appender.main.policies.size.size" : "100MB",
      "appender.main.strategy.type" : "DefaultRolloverStrategy",
      "appender.main.strategy.max" : "10"
    },
  }
]
```

### 亚马逊 5.29.0 及更早EMR版本

在亚马逊 5.29.0 及更早EMR版本中，Flink 使用 log4j 日志框架。下面的示例配置演示了 log4j 格式。

```
[
  {
    "Classification": "flink-log4j",
    "Properties": {
      "log4j.appender.file": "org.apache.log4j.RollingFileAppender",
      "log4j.appender.file.append": "true",
      # keep up to 4 files and each file size is limited to 100MB
      "log4j.appender.file.MaxFileSize": "100MB",
      "log4j.appender.file.MaxBackupIndex": 4,
      "log4j.appender.file.layout": "org.apache.log4j.PatternLayout",
      "log4j.appender.file.layout.ConversionPattern": "%d{yyyy-MM-dd HH:mm:ss,SSS} %-5p
%-60c %x - %m%n"
    },
  }
]
```

## 将 Flink 配置为使用 Java 11 运行

亚马逊EMR版本 6.12.0 及更高版本为 Flink 提供了 Java 11 运行时支持。以下各节介绍如何配置集群以为 Flink 提供 Java 11 运行时系统支持。

### 主题

- [在创建集群时配置 Flink for Java 11](#)
- [在正在运行的集群上配置 Flink for Java 11](#)
- [在正在运行的集群上确认 Flink 的 Java 运行时系统](#)

## 在创建集群时配置 Flink for Java 11

使用以下步骤创建带有 Flink 和 Java 11 运行时的EMR集群。添加 Java 11 运行时系统支持所在的配置文件是 `flink-conf.yaml`。

### Console

在控制台中使用 Flink 和 Java 11 运行时创建集群

1. 登录并打开亚马逊EMR控制台 AWS Management Console，[网址为 https://console.aws.amazon.com/emr](https://console.aws.amazon.com/emr)。

2. 在导航窗格EC2中选择“EMR on”下的“集群”，然后选择“创建集群”。
3. 选择亚马逊 6.12.0 或更高EMR版本，然后选择安装 Flink 应用程序。选择要在集群上安装的任何其他应用程序。
4. 继续设置您的集群。在可选的软件设置部分，使用默认的输入配置选项，并输入以下配置：

```
[
  {
    "Classification": "flink-conf",
    "Properties": {
      "containerized.taskmanager.env.JAVA_HOME":"/usr/lib/jvm/jre-11",
      "containerized.master.env.JAVA_HOME":"/usr/lib/jvm/jre-11",
      "env.java.home":"/usr/lib/jvm/jre-11"
    }
  }
]
```

5. 继续设置并启动您的集群。

## AWS CLI

### 要使用 Flink 和 Java 11 运行时创建集群 CLI

1. 创建一个将 Flink 配置为使用 Java 11 的配置文件 `configurations.json`。

```
[
  {
    "Classification": "flink-conf",
    "Properties": {
      "containerized.taskmanager.env.JAVA_HOME":"/usr/lib/jvm/jre-11",
      "containerized.master.env.JAVA_HOME":"/usr/lib/jvm/jre-11",
      "env.java.home":"/usr/lib/jvm/jre-11"
    }
  }
]
```

2. 从中 AWS CLI，使用 Amazon 6.12.0 或更高EMR版本创建新EMR集群，然后安装 Flink 应用程序，如以下示例所示：

```
aws emr create-cluster --release-label emr-6.12.0 \  
--applications Name=Flink \  
--configurations file://./configurations.json \  

```

```
--region us-east-1 \
--log-uri s3://myLogUri \
--instance-type m5.xlarge \
--instance-count 2 \
--service-role EMR_DefaultRole_V2 \
--ec2-attributes KeyName=YourKeyName,InstanceProfile=EMR_EC2_DefaultRole
```

## 在正在运行的集群上配置 Flink for Java 11

使用以下步骤使用 Flink 和 Java 11 运行时更新正在运行的EMR集群。添加 Java 11 运行时系统支持所在的配置文件是 `flink-conf.yaml`。

### Console

在控制台中使用 Flink 和 Java 11 运行时更新正在运行的集群

1. 登录并打开亚马逊EMR控制台 AWS Management Console，[网址为 https://console.aws.amazon.com/emr](https://console.aws.amazon.com/emr)。
2. 在导航窗格EC2中选择“EMRon”下的 Clusters，然后选择要更新的集群。

#### Note

集群必须使用亚马逊 6.12.0 或更高EMR版本才能支持 Java 11。

3. 选择配置选项卡。
4. 在实例组配置部分，选择要更新的正在运行的实例组，然后从列表操作菜单中选择重新配置。
5. 使用编辑属性选项重新配置实例组，如下所示。在每个配置之后选择添加新配置。

分类	属性	值
flink-conf	containerized.task manager.env.JAVA_H OME	/usr/lib/jvm/jre-1 1
flink-conf	containerized.mast er.env.JAVA_HOME	/usr/lib/jvm/jre-1 1

分类	属性	值
flink-conf	env.java.home	/usr/lib/jvm/jre-11

6. 选择保存更改以添加配置。

## AWS CLI

要更新正在运行的集群以使用来自 Flink 和 Java 11 运行时 CLI

使用 `modify-instance-groups` 命令为运行的集群中的一个实例组指定新配置。

1. 首先，创建一个将 Flink 配置为使用 Java 11 的配置文件 `configurations.json`。在以下示例中，替换 `ig-1xxxxxxx9` 使用您要重新配置的实例组的 ID。将文件保存在您将要运行 `modify-instance-groups` 命令的同一目录中。

```
[
  {
    "InstanceGroupId": "ig-1xxxxxxx9",
    "Configurations": [
      {
        "Classification": "flink-conf",
        "Properties": {
          "containerized.taskmanager.env.JAVA_HOME": "/usr/lib/jvm/jre-11",
          "containerized.master.env.JAVA_HOME": "/usr/lib/jvm/jre-11",
          "env.java.home": "/usr/lib/jvm/jre-11"
        },
        "Configurations": []
      }
    ]
  }
]
```

2. 从中 AWS CLI，运行以下命令。替换您要重新配置的实例组的 ID：

```
aws emr modify-instance-groups --cluster-id j-2AL4XXXXXX5T9 \
--instance-groups file://configurations.json
```



## 在正在运行的集群上确认 Flink 的 Java 运行时系统

要确定正在运行的集群的 Java 运行时，请使用登录到主节点，SSH如[使用连接主节点](#)中所述SSH。然后运行以下命令：

```
ps -ef | grep flink
```

包含 `-ef` 选项的 `ps` 命令列出了系统上所有正在运行的进程。您可以使用 `grep` 过滤该输出，以查找提及 `flink` 字符串的内容。查看 Java 运行时环境 (JRE) 值的输出，`jre-XX`。在以下输出中，`jre-11` 表示在运行时系统为 Flink 选择了 Java 11。

```
flink    19130    1  0 09:17 ?          00:00:15 /usr/lib/jvm/jre-11/bin/
java -Djava.io.tmpdir=/mnt/tmp -Dlog.file=/usr/lib/flink/log/flink-flink-
historyserver-0-ip-172-31-32-127.log -Dlog4j.configuration=file:/usr/lib/flink/conf/
log4j.properties -Dlog4j.configurationFile=file:/usr/lib/flink/conf/log4j.properties
-Dlogback.configurationFile=file:/usr/lib/flink/conf/logback.xml -classpath /usr/lib/
flink/lib/flink-cep-1.17.0.jar:/usr/lib/flink/lib/flink-connector-files-1.17.0.jar:/
usr/lib/flink/lib/flink-csv-1.17.0.jar:/usr/lib/flink/lib/flink-json-1.17.0.jar:/usr/
lib/flink/lib/flink-scala_2.12-1.17.0.jar:/usr/lib/flink/lib/flink-table-api-java-
uber-1.17.0.jar:/usr/lib/flink/lib/flink-table-api-scala-bridge_2.12-1.17.0.
```

或者，[使用命令 `flink-yarn-session -d` 登录主节点SSH并启动 Flink YARN 会话](#)。输出显示了 Flink 的 Java 虚拟机 (JVM)，如以下示例 `java-11-amazon-corretto` 所示：

```
2023-05-29 10:38:14,129 INFO  org.apache.flink.configuration.GlobalConfiguration
    [] - Loading configuration property: containerized.master.env.JAVA_HOME, /usr/lib/
jvm/java-11-amazon-corretto.x86_64
```

## 在 Amazon EMR 中使用 Flink 作业

在 Amazon EMR 上与 Flink 进行交互的方式有多种：通过控制台、跟踪界面上 Resource Manager 的 Flink 界面和命令行。您可通过以上任一方式将 JAR 文件提交到 Flink 应用程序。提交 JAR 文件后，它就会变成由 Flink Job Manager 管理的作业。Job Manager 位于托管 Flink 会话 Application Master 守护程序的 YARN 节点上。

可以将 Flink 应用程序作为长时间运行集群或临时集群上的 YARN 作业。在长时间运行集群上，您可以将多个 Flink 作业提交给 Amazon EMR 上运行的一个 Flink 集群。如果您在临时集群上运行 Flink 作业，则 Amazon EMR 集群仅在其运行 Flink 应用程序的时间内存在，因此您只需为使用的资源和时间

付费。您可以使用 Amazon EMR AddSteps API 操作，并通过 AWS CLI `add-steps` 或 `create-cluster` 命令来提交 Flink 作业，作为 `RunJobFlow` 操作的步骤参数。

## 启动 Flink YARN 应用程序，作为长时间运行集群上的步骤

要启动 Flink 应用程序，使多个客户端能够通过 YARN API 操作向其提交工作，需要您创建集群或将 Flink 应用程序添加到现有集群中。有关如何创建新集群的说明，请参阅[使用 Flink 创建集群](#)。要在现有集群上启动 YARN 会话，可从控制台、AWS CLI 或 Java SDK 使用以下步骤。

### Note

向 Amazon EMR 5.5.0 版本中添加了 `flink-yarn-session` 命令作为 `yarn-session.sh` 脚本的包装程序以简化执行。如果您使用 Amazon EMR 的更早版本，请将 `bash -c "/usr/lib/flink/bin/yarn-session.sh -d"` 在控制台中替换为 Arguments (参数) 或在 AWS CLI 命令中替换为 Args。

使用控制台在现有集群上提交 Flink 作业

使用 `flink-yarn-session` 命令在现有集群中提交 Flink 会话。

1. [打开亚马逊 EMR 控制台](https://console.aws.amazon.com/emr)，网址为 <https://console.aws.amazon.com/emr>。
2. 在集群列表中，选择先前已启动的集群。
3. 在集群详细信息页面上，选择 Steps (步骤)，再选择 Add Step (添加步骤)。
4. 使用随后提供的指南输入参数，然后选择添加。

参数	描述
Step type (步骤类型)	自定义 JAR
名称	可帮助您标识步骤的名称。例如， <code>&lt; example-flink-step-name &gt;</code> 。
Jar location (Jar 位置)	<code>command-runner.jar</code>
Arguments (参数)	带适合您的应用的参数的 <code>flink-yarn-session</code> 命令。例如， <code>flink-yarn-session -d</code> 在您的 YARN 集群中以分离状

参数	描述
	态启动 Flink 会话 () -d。有关参数详细信息，请参阅新版 Flink 文档中的 <a href="#">YARN 设置</a> 。

使用 AWS CLI 在现有集群上提交 Flink 作业

- 使用 `add-steps` 命令将 Flink 任务添加到长时间运行的集群。以下示例命令指定 `Args="flink-yarn-session", "-d"` 在分离状态下 (-d) 在 YARN 集群中启动 Flink 会话。有关参数详细信息，请参阅新版 Flink 文档中的 [YARN 设置](#)。

```
aws emr add-steps --cluster-id <j-XXXXXXXX> --steps Type=CUSTOM_JAR,Name=<example-flink-step-name>,Jar=command-runner.jar,Args="flink-yarn-session","-d"
```

## 将工作提交到长时间运行集群上的现有 Flink 应用程序

如果您在长时间运行的集群上已有 Flink 应用程序，则可以指定集群的 Flink 应用程序 ID，以便向其提交工作。要获取应用程序 ID，请在 `yarn application -list` AWS CLI 或通过 [YarnClient](#) API 操作运行：

```
$ yarn application -list
16/09/07 19:32:13 INFO client.RMProxy: Connecting to ResourceManager at
ip-10-181-83-19.ec2.internal/10.181.83.19:8032
Total number of applications (application-types: [] and states: [SUBMITTED, ACCEPTED,
RUNNING]):1
Application-Id      Application-Name      Application-Type      User      Queue      State
Final-State        Progress             Tracking-URL
application_1473169569237_0002  Flink session with 14 TaskManagers (detached)
  Apache Flink      hadoop              default              RUNNING      UNDEFINED
100% http://ip-10-136-154-194.ec2.internal:33089
```

此 Flink 会话的应用程序 ID 为 `application_1473169569237_0002`，支持您使用 AWS CLI 或 SDK 来将作业提交到应用程序。

### Example SDK for Java

```
List<StepConfig> stepConfigs = new ArrayList<StepConfig>();

HadoopJarStepConfig flinkWordCountConf = new HadoopJarStepConfig()
```

```

        .withJar("command-runner.jar")
        .withArgs("flink", "run", "-m", "yarn-cluster", "-yid",
"application_1473169569237_0002", "-yn", "2", "/usr/lib/flink/examples/streaming/
WordCount.jar",
        "--input", "s3://myBucket/pg11.txt", "--output", "s3://myBucket/alice2/");

StepConfig flinkRunWordCount = new StepConfig()
    .withName("Flink add a wordcount step")
    .withActionOnFailure("CONTINUE")
    .withHadoopJarStep(flinkWordCountConf);

stepConfigs.add(flinkRunWordCount);

AddJobFlowStepsResult res = emr.addJobFlowSteps(new AddJobFlowStepsRequest()
    .withJobFlowId("myClusterId")
    .withSteps(stepConfigs));

```

## Example AWS CLI

```

aws emr add-steps --cluster-id <j-XXXXXXXX> \
--steps Type=CUSTOM_JAR,Name=Flink_Submit_To_Long_Running,Jar=command-runner.jar,\
Args="flink","run","-m","yarn-cluster","-yid","application_1473169569237_0002",\
"/usr/lib/flink/examples/streaming/WordCount.jar",\
"--input","s3://myBucket/pg11.txt","--output","s3://myBucket/alice2/" \
--region <region-code>

```

## 提交临时 Flink 作业

以下示例启动一个临时集群，它运行 Flink 作业并在完成时将其终止。

### Example SDK for Java

```

import java.util.ArrayList;
import java.util.List;
import com.amazonaws.AmazonClientException;
import com.amazonaws.auth.AWSCredentials;
import com.amazonaws.auth.AWSStaticCredentialsProvider;
import com.amazonaws.auth.profile.ProfileCredentialsProvider;
import com.amazonaws.services.elasticmapreduce.AmazonElasticMapReduce;
import com.amazonaws.services.elasticmapreduce.AmazonElasticMapReduceClientBuilder;
import com.amazonaws.services.elasticmapreduce.model.*;

public class Main_test {

```

```
public static void main(String[] args) {
    AWSCredentials credentials_profile = null;
    try {
        credentials_profile = new ProfileCredentialsProvider("default").getCredentials();
    } catch (Exception e) {
        throw new AmazonClientException(
            "Cannot load credentials from .aws/credentials file. " +
            "Make sure that the credentials file exists and the profile name is specified
within it.",
            e);
    }

    AmazonElasticMapReduce emr = AmazonElasticMapReduceClientBuilder.standard()
        .withCredentials(new AWSStaticCredentialsProvider(credentials_profile))
        .withRegion(Regions.US_WEST_1)
        .build();

    List<StepConfig> stepConfigs = new ArrayList<StepConfig>();
    HadoopJarStepConfig flinkWordCountConf = new HadoopJarStepConfig()
        .withJar("command-runner.jar")
        .withArgs("bash", "-c", "flink", "run", "-m", "yarn-cluster", "-yn", "2",
            "/usr/lib/flink/examples/streaming/WordCount.jar", "--input", "s3://path/to/
input-file.txt", "--output",
            "s3://path/to/output/");

    StepConfig flinkRunWordCountStep = new StepConfig()
        .withName("Flink add a wordcount step and terminate")
        .withActionOnFailure("CONTINUE")
        .withHadoopJarStep(flinkWordCountConf);

    stepConfigs.add(flinkRunWordCountStep);

    Application flink = new Application().withName("Flink");

    RunJobFlowRequest request = new RunJobFlowRequest()
        .withName("flink-transient")
        .withReleaseLabel("emr-5.20.0")
        .withApplications(flink)
        .withServiceRole("EMR_DefaultRole")
        .withJobFlowRole("EMR_EC2_DefaultRole")
        .withLogUri("s3://path/to/my/logfiles")
        .withInstances(new JobFlowInstancesConfig()
            .withEc2KeyName("myEc2Key"))
```

```

        .withEc2SubnetId("subnet-12ab3c45")
        .withInstanceCount(3)
        .withKeepJobFlowAliveWhenNoSteps(false)
        .withMasterInstanceType("m4.large")
        .withSlaveInstanceType("m4.large"))
        .withSteps(stepConfigs);

    RunJobFlowResult result = emr.runJobFlow(request);
    System.out.println("The cluster ID is " + result.toString());
}
}

```

## Example AWS CLI

使用 `create-cluster` 子命令创建一个临时集群，该集群在 Flink 作业完成时终止：

```

aws emr create-cluster --release-label emr-5.2.1 \
--name "Flink_Transient" \
--applications Name=Flink \
--configurations file:///./configurations.json \
--region us-east-1 \
--log-uri s3://myLogUri \
--auto-terminate
--instance-type m5.xlarge \
--instance-count 2 \
--service-role EMR_DefaultRole_V2 \
--ec2-attributes KeyName=<YourKeyName>,InstanceProfile=EMR_EC2_DefaultRole \
--steps Type=CUSTOM_JAR,Jar=command-runner.jar,Name=Flink_Long_Running_Session,\
Args="bash","-c","\fink run -m yarn-cluster /usr/lib/flink/examples/streaming/
WordCount.jar
--input s3://myBucket/pg11.txt --output s3://myBucket/alice/"

```

## 使用 Scala Shell

适用于 EMR 集群的 Flink Scala Shell 仅配置为启动新的 YARN 会话。您可以通过以下过程使用 Scala Shell。

在主节点上使用 Flink Scala Shell

1. 使用 SSH 登录主节点，如 [Connect to the primary node with SSH](#) 中所述。

## 2. 键入以下命令启动 Shell :

在 Amazon EMR 5.5.0 及更高版本中，您可以借助一个 TaskManager 来使用以下命令启动 Scala Shell Yarn 集群。

```
% flink-scala-shell yarn 1
```

在 Amazon EMR 的更早版本中，使用：

```
% /usr/lib/flink/bin/start-scala-shell.sh yarn 1
```

这将启动 Flink Scala Shell，以便您能以交互方式使用 Flink。与使用其它接口和选项一样，您可以基于要从 Shell 运行的任务数来缩放示例中使用的 `-n` 选项值。

有关更多信息，请参阅官方 Apache Flink 文档中的 [Scala REPL](#)。

## 查找 Flink Web 界面

属于 Flink 应用程序的 Application Master 托管 Flink Web 界面。这是将 JAR 作为作业提交或查看其他作业当前状态的另一种方式。只要有 Flink 会话在运行，Flink Web 界面就处于活动状态。如果您有长时间运行的 YARN 作业并且已处于活动状态，则可以按照 Amazon EMR Management Guide 中的 [Connect to the primary node with SSH](#) 主题中的说明，来连接到 YARN ResourceManager。例如，如果您已设置 SSH 隧道并且已在浏览器中激活代理，则可在 EMR 集群详细信息页面中的 Connections (连接) 下选择 ResourceManager 连接。



The screenshot shows a cluster named "Development Cluster" with a status of "Waiting" and a message "Cluster ready after last step completed." Below this, under the "Connections:" section, there is a red arrow pointing to a link labeled "Resource Manager ... (View All)".

在找到 ResourceManager 后，选择正在托管 Flink 会话的 YARN 应用程序。选择 Tracking UI (跟踪 UI) 列下的链接。

# All Applications

Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved	Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealed Nodes
2	2 GB	11.25 GB	0 B	2	8	0	1	0	0	0	0

Scheduling Resource Type	Minimum Allocation	Maximum Allocation
MEMORY	<memory:32, vCores:1>	<memory:11520, vCores:8>

Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Progress	Tracking UI
Flink session with 1 TaskManagers (detached)	Apache Flink	default	Mon Oct 10 14:42:47 -0700 2016	N/A	RUNNING	UNDEFINED		<a href="#">ApplicationMaster</a>

在 Flink Web 界面中，您可以查看配置，将自己的自定义 JAR 作为作业提交或监视正在进行的作业。

The screenshot shows the Apache Flink Dashboard with the following components:

- Overview (Version: 1.1.1):**
  - Task Managers: 1
  - Task Slots: 1
  - Available Task Slots: 1
- Total Jobs:**
  - Running: 0
  - Finished: 0
  - Canceled: 0
  - Failed: 0
- Running Jobs:** (Table with columns: Start Time, End Time, Duration, Job Name, Job ID, Tasks, Status)
- Completed Jobs:** (Table with columns: Start Time, End Time, Duration, Job Name, Job ID, Tasks, Status)

## Flink Autoscaler

### 概述

Amazon EMR 发行版 6.15.0 及更高版本支持 Flink Autoscaler。作业自动扩缩器功能从正在运行的 Flink 流式传输作业中收集指标，并自动扩展单个作业顶点。这可以降低反向压力并满足您设定的利用率目标。



有关更多信息，请参阅 Apache Flink Kubernetes Operator 文档中的 [Autoscaler](#) 部分。

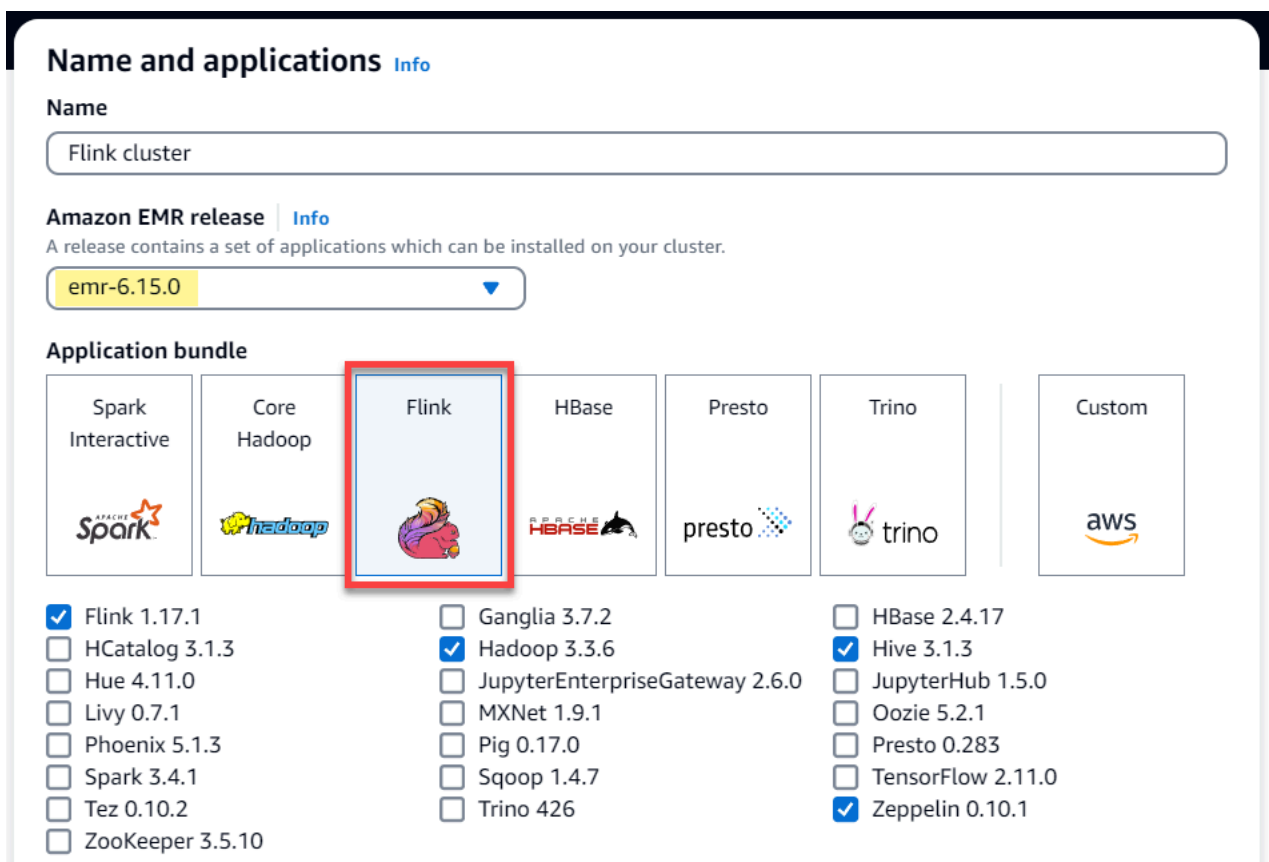
## 注意事项

- Amazon EMR 6.15.0 及更高版本支持 Flink Autoscaler。
- Flink Autoscaler 仅支持流式传输作业。
- 仅支持自适应计划程序。不支持默认计划程序。
- 我们建议您启用集群扩展以允许动态资源预调配。首选 Amazon EMR 托管式自动扩缩功能，因为指标评估每 5-10 秒进行一次。在此间隔内，您的集群可以更轻松地适应所需集群资源的变化。

## 启用 Autoscaler

创建 Amazon EMR on EC2 集群时，使用以下步骤启用 Flink Autoscaler。

1. 在 Amazon EMR 控制台中，创建新的 EMR 集群：
  - a. 选择 Amazon EMR 发行版 `emr-6.15.0` 或更高版本。选择 Flink 应用程序捆绑包，然后选择您可能想要包含在集群中的任何其他应用程序。



**Name and applications** [Info](#)

Name

Flink cluster

**Amazon EMR release** [Info](#)

A release contains a set of applications which can be installed on your cluster.

emr-6.15.0

**Application bundle**

Spark Interactive | Core Hadoop | **Flink** | HBase | Presto | Trino | Custom

Flink 1.17.1 |  HCatalog 3.1.3 |  Hue 4.11.0 |  Livy 0.7.1 |  Phoenix 5.1.3 |  Spark 3.4.1 |  Tez 0.10.2 |  ZooKeeper 3.5.10

Ganglia 3.7.2 |  Hadoop 3.3.6 |  JupyterEnterpriseGateway 2.6.0 |  MXNet 1.9.1 |  Pig 0.17.0 |  Sqoop 1.4.7 |  Trino 426

HBase 2.4.17 |  Hive 3.1.3 |  JupyterHub 1.5.0 |  Oozie 5.2.1 |  Presto 0.283 |  TensorFlow 2.11.0 |  Zeppelin 0.10.1

- b. 在集群扩展和预调配选项下，选择使用 EMR 托管扩展。

### Cluster scaling and provisioning Info

Set up scaling and provisioning configurations for the core and task node groups for your cluster.

Choose an option

**Set cluster size manually**  
 Use this option if you know your workload patterns in advance.

**Use EMR-managed scaling**  
 Monitor key workload metrics so that EMR can optimize the cluster size and resource utilization.

**Use custom automatic scaling**  
 To programmatically scale core and task nodes, create custom automatic scaling policies.

2. 在软件设置部分，输入以下配置以启用 Flink Autoscaler。对于测试场景，请将决策间隔、指标窗口间隔和稳定间隔设置为较低的值，以便作业立即做出扩展决策，从而更便于验证。

```
[
  {
    "Classification": "flink-conf",
    "Properties": {
      "job.autoscaler.enabled": "true",
      "jobmanager.scheduler": "adaptive",
      "job.autoscaler.stabilization.interval": "60s",
      "job.autoscaler.metrics.window": "60s",
      "job.autoscaler.decision.interval": "10s",
      "job.autoscaler.debug.logs.interval": "60s"
    }
  }
]
```

3. 选择或配置您喜欢的任何其他设置，然后创建支持 Flink Autoscaler 的集群。

## Autoscaler 配置

本部分涵盖了您可以根据自己的特定需求更改的大多数配置。

### Note

对于基于时间的配置（如 `time`、`interval` 和 `window` 设置），如果未指定单位，则默认单位为毫秒。因此，不带后缀的值 30 等于 30 毫秒。对于其他时间单位，对于秒，包含适当的后缀 `s`，对于分钟，包含 `m`，或者对于小时，包含 `h`。

## 主题

- [Autoscaler 循环配置](#)
- [指标聚合和历史记录配置](#)
- [作业顶点级别配置](#)
- [积压作业处理配置](#)
- [扩展操作配置](#)

## Autoscaler 循环配置

Autoscaler 每隔几个可配置的时间间隔获取一次作业顶点级别指标，将其转换为扩展操作项，估计新的作业顶点并行度，并将其推荐给作业计划程序。只有在作业重启时间和集群稳定间隔之后才会收集指标。

配置密钥	默认值	描述	示例值
<code>job.autoscaler.enabled</code>	<code>false</code>	在 Flink 集群上启用自动扩缩。	<code>true, false</code>
<code>job.autoscaler.decision.interval</code>	<code>60s</code>	Autoscaler 决策时间间隔。	<code>30</code> (默认单位为毫秒)、 <code>5m</code> 、 <code>1h</code>
<code>job.autoscaler.restart.time</code>	<code>3m</code>	在操作员能够根据历史记录可靠地确定重启时间之前，将使用预期重启时间。	<code>30</code> (默认单位为毫秒)、 <code>5m</code> 、 <code>1h</code>
<code>job.autoscaler.stabilization.interval</code>	<code>300s</code>	不会执行新扩展的稳定期。	<code>30</code> (默认单位为毫秒)、 <code>5m</code> 、 <code>1h</code>
<code>job.autoscaler.debug.logs.interval</code>	<code>300s</code>	Autoscaler 调试日志间隔。	<code>30</code> (默认单位为毫秒)、 <code>5m</code> 、 <code>1h</code>

## 指标聚合和历史记录配置

Autoscaler 获取指标，在基于时间的滑动窗口内聚合这些指标，然后对这些指标进行评估以做出扩展决策。利用每个作业顶点的扩展决策历史记录来估计新的并行度。它们既有基于年龄的到期时间，也有历史记录大小（至少 1）。

配置密钥	默认值	描述	示例值
<code>job.autoscaler.metrics.window</code>	600s	Scaling metrics aggregation window size.	30（默认单位为毫秒）、5m、1h
<code>job.autoscaler.history.max.count</code>	3	每个顶点要保留的过去的扩展决策的最大数量。	1 到 Integer.MAX_VALUE
<code>job.autoscaler.history.max.age</code>	24h	每个顶点要保留的过去的扩展决策的最小数量。	30（默认单位为毫秒）、5m、1h

## 作业顶点级别配置

每个作业顶点的并行度根据目标利用率进行修改，并受最小-最大并行度限制的局限。不建议将目标利用率设置为接近 100%（即，值 1），利用率界限可以作为缓冲区来处理中间负载波动。

配置密钥	默认值	描述	示例值
<code>job.autoscaler.target.utilization</code>	0.7	目标顶点利用率。	0 - 1
<code>job.autoscaler.target.utilization.boundary</code>	0.4	目标顶点利用率界限。如果当前处理速率在 $[\text{target\_rate} / (\text{target\_utilization} - \text{boundary})]$ 内并且为 $(\text{target\_rate} / (\text{target\_utilization}$	0 - 1

配置密钥	默认值	描述	示例值
		+ boundary)] , 则不会执行扩展	
job.autoscaler.vertex.min-parallelism	1	Autoscaler 可以使用的最小并行度。	0 - 200
job.autoscaler.vertex.max-parallelism	200	Autoscaler 可以使用的最大并行度。请注意, 如果该限值高于 Flink 配置中配置的最大并行度或直接在每个 Operator 上设定的最大并行度, 则该限值会被忽略。	0 - 200

## 积压作业处理配置

作业顶点需要额外的资源来处理在扩展操作期间积累的待处理事件或积压作业。这也称为 catch-up 持续时间。如果处理积压作业的时间超过配置的 lag -threshold 值, 则作业顶点目标利用率将提高到最大级别。这有助于防止在积压作业处理过程中进行不必要的扩展操作。

配置密钥	默认值	描述	示例值
job.autoscaler.backlog-processing.lag-threshold	5m	延迟阈值可防止不必要的扩展, 同时移除导致延迟的待处理消息。	30 ( 默认单位为毫秒 )、5m、1h
job.autoscaler.catch-up.duration	15m	扩展操作后完全处理所有积压作业的目标持续时间。设置为 0 可禁用基于积压作业的扩展。	30 ( 默认单位为毫秒 )、5m、1h

## 扩展操作配置

在宽限期内, Autoscaler 不会在纵向扩展操作后立即执行缩减操作。这样可以防止由于临时负载波动而导致的不必要的纵向扩展-缩减-扩展-缩减的操作循环。

我们可以使用缩减操作比率来逐步降低并行度并释放资源，以应对临时的负载峰值。它还有助于防止在大规模缩减操作后进行不必要的轻微纵向扩展操作。

我们可以根据过去的作业顶点扩展决策历史记录来检测无效的纵向扩展操作，以防止进一步的并行度变化。

配置密钥	默认值	描述	示例值
<code>job.autoscaler.scale-up.grace-period</code>	1h	纵向扩展顶点后不允许缩减顶点的持续时间。	30 (默认单位为毫秒)、5m、1h
<code>job.autoscaler.scale-down.max-factor</code>	0.6	最大缩减系数。值为 1 表示对缩减没有限制；0.6 意味着只能使用原始并行度的 60% 来缩减作业。	0 - 1
<code>job.autoscaler.scale-up.max-factor</code>	100000.	最大纵向扩展比。值为 2.0 意味着只能使用当前并行度的 200% 来纵向扩展作业。	0 - Integer.MAX_VALUE
<code>job.autoscaler.scaling.effectiveness.detection.enabled</code>	false	是否启用对无效扩展操作的检测，并允许 Autoscaler 阻止进一步的纵向扩展。	true, false

## 优化任务恢复和扩展操作的作业重启时间

当任务失败或发生扩展操作时，Flink 会尝试从上一次完成的检查点重新执行任务。重启过程可能需要一分钟或更长时间才能执行，具体取决于检查点状态的大小以及并行任务的数量。重启期间，可以累积作业的积压任务。但是，Flink 可以通过一些方法来优化执行图的恢复和重启速度，从而提高作业稳定性。

本页介绍了 Amazon EMR Flink 可以在任务恢复或扩展操作期间缩短作业的重启时间的一些方法。

### 主题

- [任务的本地恢复](#)
- [基于日志的通用增量检查点](#)

- [精细恢复](#)
- [自适应计划程序中的组合重启机制](#)

## 任务的本地恢复

### Note

Amazon EMR 6.0.0 及更高版本都支持任务的本地恢复。

使用 Flink 检查点，每个任务都会生成其状态的快照，Flink 会将该快照写入分布式存储（如 Amazon S3）。在恢复的情况下，任务会从分布式存储中恢复其状态。分布式存储提供容错能力，并且可以在重新扩展期间重新分配状态，因为它可供所有节点访问。

但远程分布式存储也有一个缺点：所有任务都必须通过网络从远程位置读取其状态。在任务恢复或扩展操作期间，这可能会导致大规模状态的恢复时间很长。

通过任务本地恢复可以解决恢复时间长这一问题。任务将其在检查点上的状态写入任务本地的辅助存储（例如本地磁盘）。它们还将状态存储在主存储中，或者存储在 Amazon S3 中（在本例中）。恢复期间，计划程序将任务计划在任务之前运行所在的同一个任务管理器上，这样它们就可以从本地状态存储中恢复，而不是从远程状态存储中读取。有关更多信息，请参阅 Apache Flink 文档中的[任务本地恢复](#)。

我们对示例作业进行的基准测试表明，启用任务本地恢复后，恢复时间已从几分钟缩短到几秒。

要启用任务本地恢复，请在 `flink-conf.yaml` 文件中设置以下配置。指定检查点间隔值，以毫秒为单位。

```
state.backend.local-recovery: true
state.backend: hasmap or rocksdb
state.checkpoints.dir: s3://storage-location-bucket-path/checkpoint
execution.checkpointing.interval: 15000
```

## 基于日志的通用增量检查点

### Note

Amazon EMR 6.10.0 及更高版本支持基于日志的通用增量检查点。

Flink 1.16 中添加了基于日志的通用增量检查点功能，以提高检查点的速度。较快的检查点间隔通常会导致恢复工作减少，因为恢复后需要重新处理的事件较少。有关更多信息，请参阅 Apache Flink 博客上的[使用基于日志的通用增量检查点提高检查点的速度和稳定性](#)。

对于示例作业，我们的基准测试表明，使用基于日志的通用增量检查点时，检查点时间从几分钟缩短到几秒。

要启用基于日志的通用增量检查点，请在 `flink-conf.yaml` 文件中设置以下配置。指定检查点间隔值，以毫秒为单位。

```
state.backend.changelog.enabled: true
state.backend.changelog.storage: filesystem
dstl.dfs.base-path: s3://bucket-path/changelog
state.backend.local-recovery: true
state.backend: rocksdb
state.checkpoints.dir: s3://bucket-path/checkpoint
execution.checkpointing.interval: 15000
```

## 精细恢复

### Note

Amazon EMR 6.0.0 及更高版本提供默认计划程序的精细恢复支持。Amazon EMR 6.15.0 及更高版本提供自适应计划程序的精细恢复支持。

当任务在执行过程中失败时，Flink 会重置整个执行图，并从上次完成的检查点触发完整的重新执行。这比仅重新执行失败的任务更昂贵。精细恢复仅重新启动失败的任务与管道连接的组件。在以下示例中，作业图有 5 个顶点 (A 到 E)。顶点之间的所有连接都使用逐点分布进行管道化处理，作业的 `parallelism.default` 设置为 2。

```
A # B # C # D # E
```

在本示例中，总共有 10 个任务在运行。第一个管道 (a1 到 e1) 在 TaskManager (TM1) 上运行，第二个管道 (a2 到 e2) 在另一个 TaskManager (TM2) 上运行。

```
a1 # b1 # c1 # d1 # e1
a2 # b2 # c2 # d2 # e2
```



有两个管道连接的组件：a1 # e1 和 a2 # e2。如果 TM1 或 TM2 其中一个失败，则故障仅影响 TaskManager 正在其中运行的管道中的 5 个任务。重启策略仅会启动受影响的管道化组件。

精细恢复仅适用于完全并行的 Flink 作业。keyBy() 或 redistribute() 操作不支持。有关更多信息，请参阅 Flink 改进提案 Jira 项目中的 [FLIP-1：从任务失败中进行精细恢复](#)。

要启用精细恢复，请在 flink-conf.yaml 文件中设置以下配置。

```
jobmanager.execution.failover-strategy: region
restart-strategy: exponential-delay or fixed-delay
```

## 自适应计划程序中的组合重启机制

### Note

Amazon EMR 6.15.0 及更高版本支持自适应计划程序中的组合重启机制。

自适应计划程序可以根据可用插槽调整作业的并行度。如果没有足够的插槽来适应配置的作业并行度，它将自动降低并行度。如果有新的插槽可用，则任务将再次纵向扩展到配置的作业并行度。当没有足够的可用资源时，自适应计划程序将避免作业停机。这是 Flink Autoscaler 支持的计划程序。出于这些原因，我们建议在 Amazon EMR Flink 中使用自适应计划程序。但是，自适应计划程序可能会在短时间内进行多次重启，每添加一个新资源就会重启一次。这可能导致作业性能下降。

在 Amazon EMR 6.15.0 及更高版本中，Flink 在自适应计划程序中具有组合重启机制，可在添加第一个资源时打开一个重启窗口，然后等到配置的默认 1 分钟窗口间隔时。当有足够的资源可用来运行具有配置并行性的作业时，或者当间隔超时时，它会执行一次重启。

对于示例作业，我们的基准测试表明，当您使用自适应计划程序和 Flink Autoscaler 时，此功能处理的记录比默认行为多 10%。

要启用组合重启机制，请在 flink-conf.yaml 文件中设置以下配置。

```
jobmanager.adaptive-scheduler.combined-restart.enabled: true
jobmanager.adaptive-scheduler.combined-restart.window-interval: 1m
```

# 在 Amazon EMR 中通过 Zeppelin 使用 Flink 作业

## 简介

Amazon EMR 发布了 6.10.0 及更高版本，支持与 Apache Flink 的 [Apache Zeppelin](#) 集成。您可以通过 Zeppelin 笔记本以交互方式提交 Flink 作业。使用 Flink 解释器，您可以执行 Flink 查询、定义 Flink 流媒体和批处理作业，以及在 Zeppelin 笔记本中可视化输出。Flink 解释器基于 Flink REST API 构建。这使您可以从 Zeppelin 环境中访问和操作 Flink 作业，以执行实时数据处理和分析。

Flink 解释器中有四个子解释器。它们的用途不同，但都在 JVM 中，与 Flink 共享相同的预配置入口点（`ExecutionEnvironment`、`StreamExecutionEnvironment`、`BatchTableEnvironment`、`StreamTableEnvironment`）。解释器如下：

- `%flink` – 创建 `ExecutionEnvironment`、`StreamExecutionEnvironment`、`BatchTableEnvironment`、`StreamTableEnvironment` 并提供 Scala 环境
- `%flink.pyflink` – 提供一个 Python 环境
- `%flink.ssql` – 提供流式 SQL 环境
- `%flink.bsql` – 提供批处理 SQL 环境

## 先决条件

- 使用 Amazon EMR 6.10.0 及更高版本创建的集群支持 Zeppelin 与 Flink 集成。
- 要根据这些步骤的要求查看 EMR 集群上托管的 Web 界面，必须配置 SSH 隧道以允许入站访问。有关更多信息，请参阅 [Configure proxy settings to view websites hosted on the primary node](#)。

## 在 EMR 集群上配置 Zeppelin-Flink

使用以下步骤将 Apache Zeppelin 上的 Apache Flink 配置为在 EMR 集群上运行：


1. 从 Amazon EMR 控制台创建新集群。为 Amazon EMR 版本选择 `emr-6.10.0` 或更高版本。然后，选择使用“自定义”选项自定义您的应用程序捆绑包。在您的捆绑包中至少包含 Flink、Hadoop 和 Zeppelin。


Amazon EMR release [Info](#)


A release contains a set of applications which can be installed on your cluster.


emr-6.10.0 ▼


Application bundle


Spark  


Core  
Hadoop  


HBase  


Presto  


Trino  


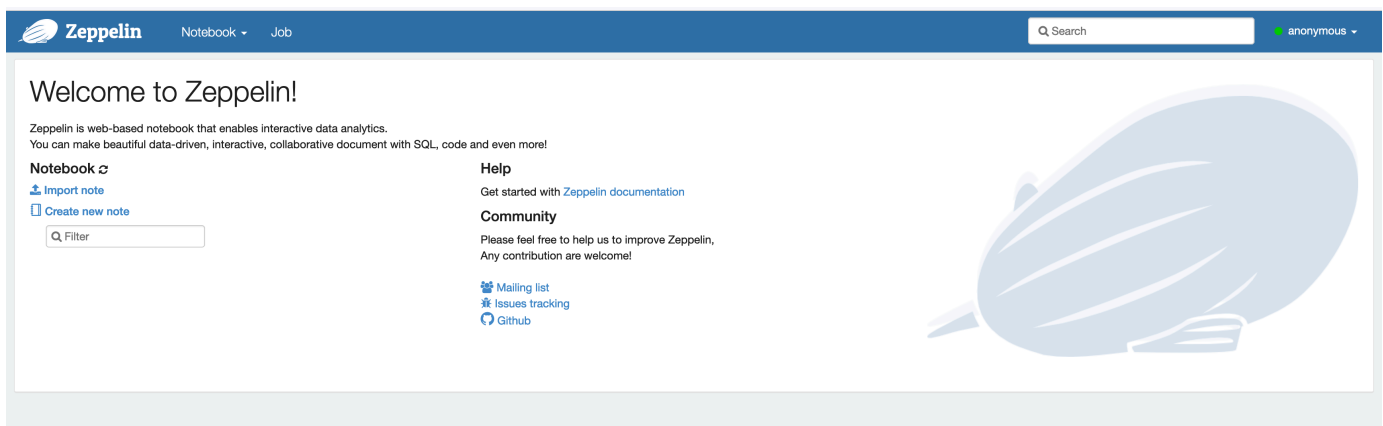
Custom  


▼ Customize your application bundle

Applications included in bundle

<input checked="" type="checkbox"/> Flink 1.16.0	<input type="checkbox"/> Ganglia 3.7.2
<input type="checkbox"/> HBase 2.4.15	<input type="checkbox"/> HCatalog 3.1.3
<input checked="" type="checkbox"/> Hadoop 3.3.3	<input type="checkbox"/> Hive 3.1.3
<input type="checkbox"/> Hue 4.10.0	<input type="checkbox"/> JupyterEnterpriseGateway 2.6.0
<input type="checkbox"/> JupyterHub 1.5.0	<input type="checkbox"/> Livy 0.7.1
<input type="checkbox"/> MXNet 1.9.1	<input type="checkbox"/> Oozie 5.2.1
<input type="checkbox"/> Phoenix 5.1.2	<input type="checkbox"/> Pig 0.17.0
<input type="checkbox"/> Presto 0.278	<input type="checkbox"/> Spark 3.3.1
<input type="checkbox"/> Sqoop 1.4.7	<input type="checkbox"/> TensorFlow 2.11.0
<input type="checkbox"/> Tez 0.10.2	<input type="checkbox"/> Trino 403
<input checked="" type="checkbox"/> Zeppelin 0.10.1	<input type="checkbox"/> ZooKeeper 3.5.10

- 使用您首选的设置创建集群的其余部分。
- 一旦集群开始运行，在控制台中选择集群以查看其详细信息并打开“应用程序”选项卡。从“应用程序”用户界面部分选择“Zeppelin”，以打开 Zeppelin 网页界面。请确保您已设置了对 Zeppelin Web 界面的访问，包含连接到主节点的 SSH 隧道和代理连接，如 [先决条件](#) 中所述。



Zeppelin Notebook - Job

Welcome to Zeppelin!

Zeppelin is web-based notebook that enables interactive data analytics. You can make beautiful data-driven, interactive, collaborative document with SQL, code and even more!

**Notebook** ↕

- Import note
- Create new note

Q Filter

**Help**

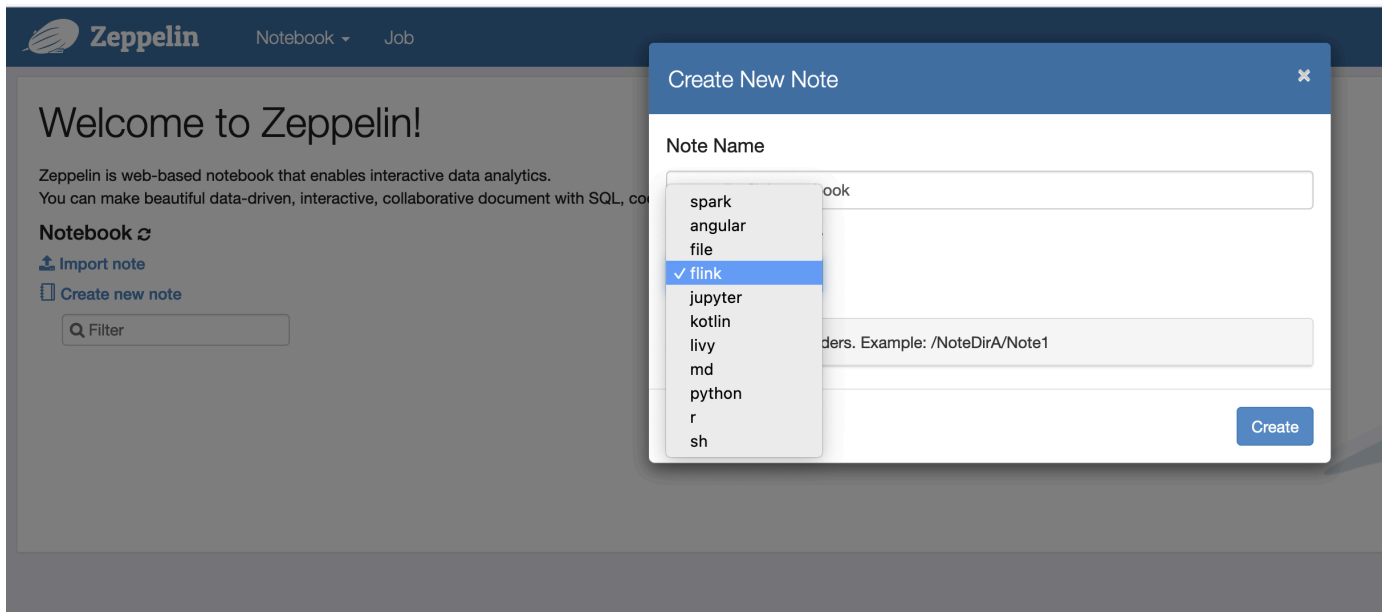
- Get started with [Zeppelin documentation](#)

**Community**

Please feel free to help us to improve Zeppelin. Any contribution are welcome!

- Mailing list
- Issues tracking
- GitHub

- 现在，您可以使用 Flink 作为默认解释器在 Zeppelin 笔记本中创建新笔记。



5. 请参阅以下代码示例，这些示例演示了如何从 Zeppelin 笔记本运行 Flink 作业。

## 在 EMR 集群上使用 Zeppelin-Flink 运行 Flink 作业

- 示例 1，Flink Scala

### a) Batch Ex WordCount ample (SCALA)

```
%flink

val data = benv.fromElements("hello world", "hello flink", "hello hadoop")
data.flatMap(line => line.split("\\s"))
    .map(w => (w, 1))
    .groupBy(0)
    .sum(1)
    .print()
```

### b) 直播 WordCount 示例 (SCALA)

```
%flink

val data = senv.fromElements("hello world", "hello flink", "hello hadoop")
data.flatMap(line => line.split("\\s"))
    .map(w => (w, 1))
    .keyBy(0)
```

```
.sum(1)
.print

senv.execute()
```

**Batch WordCount** FLINK JOB FINISHED

```
%flink
val data = benv.fromElements("hello world", "hello flink", "hello hadoop")
data.flatMap(line => line.split("\\s"))
  .map(w => (w, 1))
  .groupBy(0)
  .sum(1)
  .print()

data: org.apache.flink.api.scala.DataSet[String] = org.apache.flink.api.scala.DataSet@22fe7dd5
(flink,1)
(hadoop,1)
(hello,3)
(world,1)

Took 56 sec. Last updated by anonymous at May 04 2023, 2:19:24 PM. (outdated)
```

**Streaming WordCount** FLINK JOB FINISHED

```
%flink
val data = senv.fromElements("hello world", "hello flink", "hello hadoop")
data.flatMap(line => line.split("\\s"))
  .map(w => (w, 1))
  .keyBy(0)
  .sum(1)
  .print

senv.execute()

data: org.apache.flink.streaming.api.scala.DataStream[String] = org.apache.flink.streaming.api.scala.DataStream@
202340f2
warning: there was one deprecation warning; for details, enable `:setting -deprecation` or `:replay -deprecation`
res2: org.apache.flink.streaming.api.datastream.DataStreamSink[(String, Int)] = org.apache.flink.streaming.api.d
atastream.DataStreamSink@7a0e7b2f
res3: org.apache.flink.api.common.JobExecutionResult =
Program execution finished
Job with JobID a78d1a9b33c0fb71b8d53f00da51030 has finished.
Job Runtime: 8485 ms

Took 12 sec. Last updated by anonymous at May 04 2023, 2:21:05 PM. (outdated)
```

- 示例 2，Flink 流式传输 SQL

```
%flink.ssql
SET 'sql-client.execution.result-mode' = 'tableau';
SET 'table.dml-sync' = 'true';
SET 'execution.runtime-mode' = 'streaming';

create table dummy_table (
  id int,
  data string
) with (
  'connector' = 'filesystem',
  'path' = 's3://s3-bucket/dummy_table',
  'format' = 'csv'
);

INSERT INTO dummy_table SELECT * FROM (VALUES (1, 'Hello World'), (2, 'Hi'), (2,
'Hi'), (3, 'Hello'), (3, 'World'), (4, 'ADD'), (5, 'LINE'));

SELECT * FROM dummy_table;
```

```
%flink.ssql
SET 'sql-client.execution.result-mode' = 'tableau';
SET 'table.dml-sync' = 'true';
SET 'execution.runtime-mode' = 'streaming';

create table t1 (
  id int,
  data string
) with (
  'connector' = 'filesystem',
  'path' = 's3://dbsamrat-emr-dev/glue-catalog-test/dbsamrat/t1/',
  'format' = 'csv'
);

INSERT INTO t1 SELECT * FROM (VALUES (1, 'Hello World'), (2, 'Hi'), (2, 'Hi'), (3, 'Hello'), (3, 'World'), (4, 'ADD'), (5, 'LINE'));
SELECT * FROM t1;
```

id	data
3	Hello
3	World
3	Hello
3	World
4	ADD
4	ADD
5	LINE
5	LINE

Took 20 sec. Last updated by anonymous at May 04 2023, 3:07:03 PM. (outdated)

- 示例 3，Pyflink。请注意，您必须将自己的名为的示例文本文件上传word.txt到 S3 存储桶。

```
%flink.pyflink

import argparse
import logging
import sys

from pyflink.common import Row
from pyflink.table import (EnvironmentSettings, TableEnvironment, TableDescriptor,
                           Schema,
                           DataTypes, FormatDescriptor)
from pyflink.table.expressions import lit, col
from pyflink.table.udf import udtf

def word_count(input_path, output_path):
    t_env = TableEnvironment.create(EnvironmentSettings.in_streaming_mode())
    # write all the data to one file
    t_env.get_config().set("parallelism.default", "1")

    # define the source
    if input_path is not None:
        t_env.create_temporary_table(
            'source',
            TableDescriptor.for_connector('filesystem')
                .schema(Schema.new_builder()
```

```

        .column('word', DataTypes.STRING())
        .build()
        .option('path', input_path)
        .format('csv')
        .build()
    tab = t_env.from_path('source')
else:
    print("Executing word_count example with default input data set.")
    print("Use --input to specify file input.")
    tab = t_env.from_elements(map(lambda i: (i,), word_count_data),
                              DataTypes.ROW([DataTypes.FIELD('line',
DataTypes.STRING())]))

# define the sink
if output_path is not None:
    t_env.create_temporary_table(
        'sink',
        TableDescriptor.for_connector('filesystem')
            .schema(Schema.new_builder()
                    .column('word', DataTypes.STRING())
                    .column('count', DataTypes.BIGINT())
                    .build())
            .option('path', output_path)
            .format(FormatDescriptor.for_format('canal-json')
                    .build())
            .build())
else:
    print("Printing result to stdout. Use --output to specify output path.")
    t_env.create_temporary_table(
        'sink',
        TableDescriptor.for_connector('print')
            .schema(Schema.new_builder()
                    .column('word', DataTypes.STRING())
                    .column('count', DataTypes.BIGINT())
                    .build())
            .build())

@udtf(result_types=[DataTypes.STRING()])
def split(line: Row):
    for s in line[0].split():
        yield Row(s)

# compute word count
tab.flat_map(split).alias('word') \

```

```
.group_by(col('word')) \  
.select(col('word'), lit(1).count) \  
.execute_insert('sink') \  
.wait()
```

```
logging.basicConfig(stream=sys.stdout, level=logging.INFO, format="%(message)s")
```

```
word_count("s3://s3_bucket/word.txt", "s3://s3_bucket/demo_output.txt")
```

1. 在 Zeppelin 用户界面中选择 FLINK 作业即可访问和查看 Flink Web 用户界面。



**Batch WordCount**

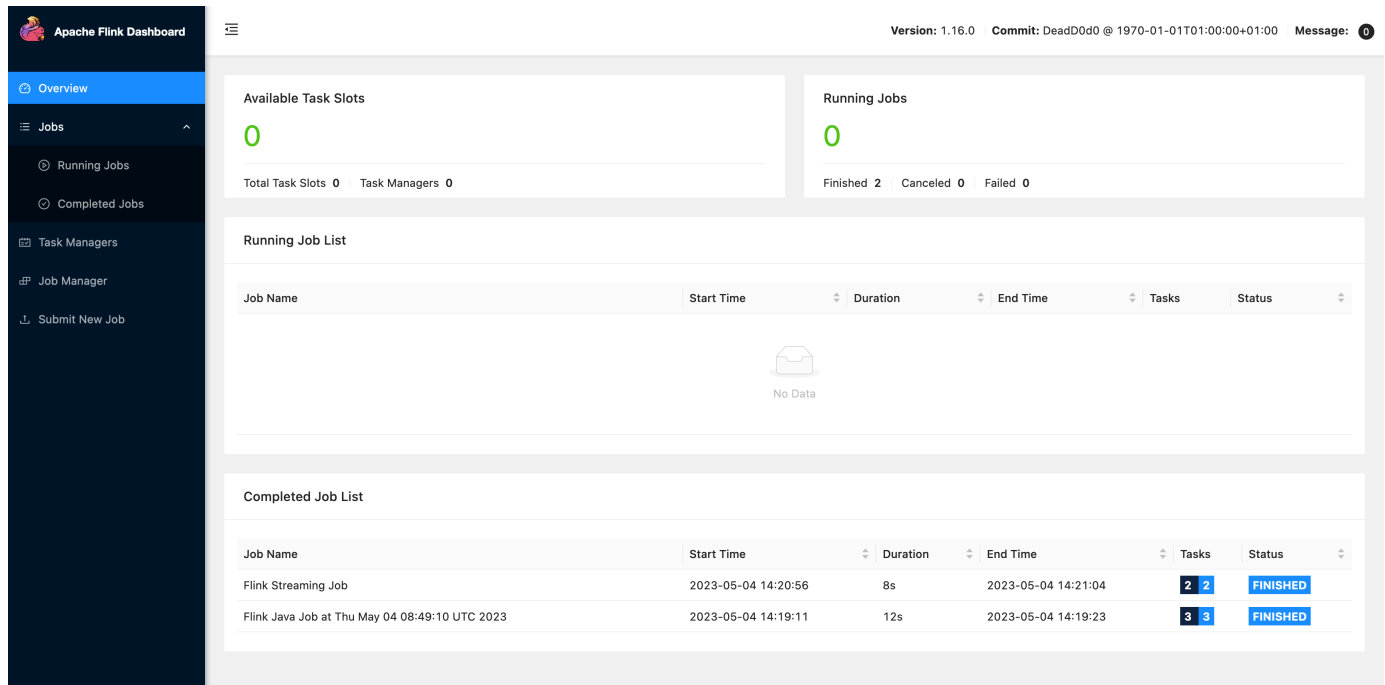
```
%flink  
  
val data = benv.fromElements("hello world", "hello flink", "hello hadoop")  
data.flatMap(line => line.split("\\s"))  
  .map(w => (w, 1))  
  .groupBy(0)  
  .sum(1)  
  .print()
```

**data:** `org.apache.flink.api.scala.DataSet[String]` = org.apache.flink.api.scala.DataSet@22fe7dd5  
(flink,1)  
(hadoop,1)  
(hello,3)  
(world,1)

Took 56 sec. Last updated by anonymous at May 04 2023, 2:19:24 PM. (outdated)

2. 在浏览器的另一个选项卡中选择 FLINK 作业，会路由到 Flink Web 控制台。





## Flink 发布历史记录

下表列出了 Amazon 的每个发行版本中包含的 Flink 版本EMR，以及随应用程序一起安装的组件。有关每个发行版本中的组件版本，请参阅 [亚马逊 EMR 7.x 发布版本](#)、[亚马逊 EMR 6.x 发布版本](#) 或 [亚马逊 EMR 5.x 发布版本](#) 中的发行版“组件版本”部分。

### Flink 版本信息

Amazon EMR 发布标签	Flink 版本	随 Flink 安装的组件
emr-7.2.0	1.18.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta

Amazon EMR 发布标签	Flink 版本	随 Flink 安装的组件
emr-5.36.2	1.14.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config
emr-7.1.0	1.18.1	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta
emr-7.0.0	1.18.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta-standalone-connectors

Amazon EMR 发布标签	Flink 版本	随 Flink 安装的组件
emr-6.15.0	1.17.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta-standalone-connectors
emr-6.14.0	1.17.1	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta-standalone-connectors
emr-6.13.0	1.17.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta-standalone-connectors

Amazon EMR 发布标签	Flink 版本	随 Flink 安装的组件
emr-6.12.0	1.17.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta-standalone-connectors
emr-6.11.1	1.16.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta-standalone-connectors
emr-6.11.0	1.16.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta-standalone-connectors

Amazon EMR 发布标签	Flink 版本	随 Flink 安装的组件
emr-6.10.1	1.16.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi
emr-6.10.0	1.16.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi
emr-6.9.1	1.15.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi

Amazon EMR 发布标签	Flink 版本	随 Flink 安装的组件
emr-6.9.0	1.15.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi
emr-6.8.1	1.15.1	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi
emr-6.8.0	1.15.1	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi

Amazon EMR 发布标签	Flink 版本	随 Flink 安装的组件
emr-6.7.0	1.14.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi
emr-5.36.1	1.14.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config
emr-5.36.0	1.14.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config

Amazon EMR 发布标签	Flink 版本	随 Flink 安装的组件
emr-6.6.0	1.14.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi
emr-5.35.0	1.14.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config
emr-6.5.0	1.14.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi



Amazon EMR 发布标签	Flink 版本	随 Flink 安装的组件
emr-6.4.0	1.13.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi
emr-6.3.1	1.12.1	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config
emr-6.3.0	1.12.1	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config

Amazon EMR 发布标签	Flink 版本	随 Flink 安装的组件
emr-6.2.1	1.11.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config
emr-6.2.0	1.11.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config
emr-6.1.1	1.11.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client

Amazon EMR 发布标签	Flink 版本	随 Flink 安装的组件
emr-6.1.0	1.11.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.34.0	1.13.1	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config
emr-5.33.1	1.12.1	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config

Amazon EMR 发布标签	Flink 版本	随 Flink 安装的组件
emr-5.33.0	1.12.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config
emr-5.32.1	1.11.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config
emr-5.32.0	1.11.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config

Amazon EMR 发布标签	Flink 版本	随 Flink 安装的组件
emr-5.31.1	1.11.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config
emr-5.31.0	1.11.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client, flink-jobmanager-config
emr-5.30.2	1.10.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client

Amazon EMR 发布标签	Flink 版本	随 Flink 安装的组件
emr-5.30.1	1.10.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.30.0	1.10.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.29.0	1.9.1	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client

Amazon EMR 发布标签	Flink 版本	随 Flink 安装的组件
emr-5.28.1	1.9.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.28.0	1.9.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.27.1	1.8.1	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client

Amazon EMR 发布标签	Flink 版本	随 Flink 安装的组件
emr-5.27.0	1.8.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.26.0	1.8.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.25.0	1.8.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client



Amazon EMR 发布标签	Flink 版本	随 Flink 安装的组件
emr-5.24.1	1.8.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.24.0	1.8.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.23.1	1.7.1	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client

Amazon EMR 发布标签	Flink 版本	随 Flink 安装的组件
emr-5.23.0	1.7.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.22.0	1.7.1	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.21.2	1.7.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client

Amazon EMR 发布标签	Flink 版本	随 Flink 安装的组件
emr-5.21.1	1.7.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.21.0	1.7.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.20.1	1.6.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client

Amazon EMR 发布标签	Flink 版本	随 Flink 安装的组件
emr-5.20.0	1.6.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.19.1	1.6.1	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.19.0	1.6.1	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client

Amazon EMR 发布标签	Flink 版本	随 Flink 安装的组件
emr-5.18.1	1.6.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.18.0	1.6.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.17.2	1.5.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client

Amazon EMR 发布标签	Flink 版本	随 Flink 安装的组件
emr-5.17.1	1.5.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.17.0	1.5.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.16.1	1.5.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client

Amazon EMR 发布标签	Flink 版本	随 Flink 安装的组件
emr-5.16.0	1.5.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.15.1	1.4.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.15.0	1.4.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client

Amazon EMR 发布标签	Flink 版本	随 Flink 安装的组件
emr-5.14.2	1.4.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.14.1	1.4.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.14.0	1.4.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client



Amazon EMR 发布标签	Flink 版本	随 Flink 安装的组件
emr-5.13.1	1.4.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.13.0	1.4.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.12.3	1.4.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client

Amazon EMR 发布标签	Flink 版本	随 Flink 安装的组件
emr-5.12.2	1.4.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.12.1	1.4.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.12.0	1.4.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client

Amazon EMR 发布标签	Flink 版本	随 Flink 安装的组件
emr-5.11.4	1.3.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.11.3	1.3.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.11.2	1.3.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client

Amazon EMR 发布标签	Flink 版本	随 Flink 安装的组件
emr-5.11.1	1.3.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.11.0	1.3.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.10.1	1.3.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client

Amazon EMR 发布标签	Flink 版本	随 Flink 安装的组件
emr-5.10.0	1.3.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.9.1	1.3.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.9.0	1.3.2	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client

Amazon EMR 发布标签	Flink 版本	随 Flink 安装的组件
emr-5.8.3	1.3.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.8.2	1.3.1	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.8.1	1.3.1	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client

Amazon EMR 发布标签	Flink 版本	随 Flink 安装的组件
emr-5.8.0	1.3.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.7.1	1.3.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.7.0	1.3.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, hadoop-ya rn-timeline-server, flink-client

Amazon EMR 发布标签	Flink 版本	随 Flink 安装的组件
emr-5.6.1	1.2.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.6.0	1.2.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.5.4	1.2.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, flink-client



Amazon EMR 发布标签	Flink 版本	随 Flink 安装的组件
emr-5.5.3	1.2.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, flink-client
emr-5.5.2	1.2.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, flink-client
emr-5.5.1	1.2.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, flink-client
emr-5.5.0	1.2.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, flink-client

Amazon EMR 发布标签	Flink 版本	随 Flink 安装的组件
emr-5.4.1	1.2.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, flink-client
emr-5.4.0	1.2.0	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, flink-client
emr-5.3.2	flink-client	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, flink-client
emr-5.3.1	flink-client	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, flink-client

Amazon EMR 发布标签	Flink 版本	随 Flink 安装的组件
emr-5.3.0	flink-client	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, flink-client
emr-5.2.3	1.1.3	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, flink-client
emr-5.2.2	1.1.3	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, flink-client
emr-5.2.1	1.1.3	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, flink-client

Amazon EMR 发布标签	Flink 版本	随 Flink 安装的组件
emr-5.2.0	1.1.3	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, flink-client
emr-5.1.1	1.1.3	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, flink-client
emr-5.1.0	1.1.3	emrfs, hadoop-client, hadoop- mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop- kms-server, hadoop-yarn- nodemanager, hadoop-yarn- resourcemanager, flink-client

## 按版本划分的 Flink 发行说明

有关完整的发行说明，请参阅以下部分。

### 亚马逊 EMR 7.2.0-Flink 发行说明

类型	描述
改进	Support 支持通过配置为每个 Flink 作业向 Kubernetes 服务添加自定义标签。kubernetes.service.labels

# Ganglia

## Note

使用 Ganglia 的最后一个 Amazon EMR 版本为 Amazon EMR 6.15.0。为了监控您的集群，6.15.0 以上的版本中包含了 [Amazon CloudWatch](#) 代理。

Ganglia 开源项目是一个可扩展的分布式系统，旨在监控集群和网格，同时尽量减少对其性能的影响。当您在集群上启用 Ganglia 时，您可以生成报告并查看整个集群的性能，还可以检查单个节点实例的性能。还配置 Ganglia 以提取和可视化 Hadoop 和 Spark 指标。有关 Ganglia 开源项目的更多信息，请转到 <http://ganglia.info/>。

当您在浏览器中查看 Ganglia Web UI 时，可以看到集群的性能概览，通过图形详细介绍了负载、内存使用率、CPU 使用率和集群的网络流量。位于集群统计数据下方的是集群中每个单独服务器的图形。

下表列出了 Amazon EMR 6.x 系列的最新发行版附带的 Ganglia 的版本，以及 Amazon EMR 随 Ganglia 一起安装的组件。

有关此发行版中随 Ganglia 安装的组件版本，请参阅[发行版 6.15.0 组件版本](#)。

emr-6.15.0 的 Ganglia 版本信息

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-6.15.0	Ganglia 3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver

下表列出了 Amazon EMR 5.x 系列的最新发行版附带的 Ganglia 的版本，以及 Amazon EMR 随 Ganglia 一起安装的组件。

有关此版本中与 Ganglia 一起安装的组件的版本，请参阅 [5.36.2 版组件版本](#)。

### emr-5.36.2 的 Ganglia 版本信息

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-5.36.2	Ganglia 3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

### 主题

- [使用 Ganglia 创建集群](#)
- [查看 Ganglia 指标](#)
- [Ganglia 中的 Hadoop 和 Spark 指标](#)
- [Ganglia 发行版历史记录](#)

## 使用 Ganglia 创建集群

### Note

使用 Ganglia 的最后一个 Amazon EMR 版本为 Amazon EMR 6.15.0。为了监控您的集群，6.15.0 以上的版本中包含了 [Amazon CloudWatch](#) 代理。

### 使用控制台通过 Ganglia 创建集群


1. 导航到 Amazon EMR 新控制台，然后从侧面导航栏中选择切换到旧控制台。有关切换到旧控制台后预期情况的更多信息，请参阅 [Using the old console](#)。
2. 选择 Create cluster (创建集群)。

3. 在 Software configuration (软件配置) 中，选择 All Applications (所有应用程序)、Core Hadoop (核心 Hadoop) 或者 Spark。
4. 根据需要进行配置创建集群。

要将 Ganglia 添加到集群中，请使用 AWS CLI

在中 AWS CLI，您可以通过使用 `create-cluster--applications` 参数将 Ganglia 添加到集群中。如果使用 `--applications` 参数仅指定 Ganglia，则 Ganglia 是唯一安装的应用程序。


- 键入以下命令以在创建集群时添加 Ganglia，将 *myKey* 替换为您的 EC2 密钥对的名称。

 Note

为了便于读取，包含 Linux 行继续符 (\)。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (^)。

```
aws emr create-cluster --name "Spark cluster with Ganglia" --release-label  
emr-6.15.0; \  
--applications Name=Spark Name=Ganglia \  
--ec2-attributes KeyName=myKey --instance-type m5.xlarge \  
--instance-count 3 --use-default-roles
```

如果不使用 `--instance-groups` 参数指定实例计数，则将启动单个主节点，其余实例将作为核心节点启动。所有节点都使用该命令中指定的实例类型。

 Note

如果您之前未创建默认 EMR 服务角色和 EC2 实例配置文件，请先键入 `aws emr create-default-roles` 创建它们，然后再键入 `create-cluster` 子命令。

有关在中使用 Amazon EMR 命令的更多信息 AWS CLI，请参阅 <https://docs.aws.amazon.com/cli/latest/reference/emr>



## 查看 Ganglia 指标

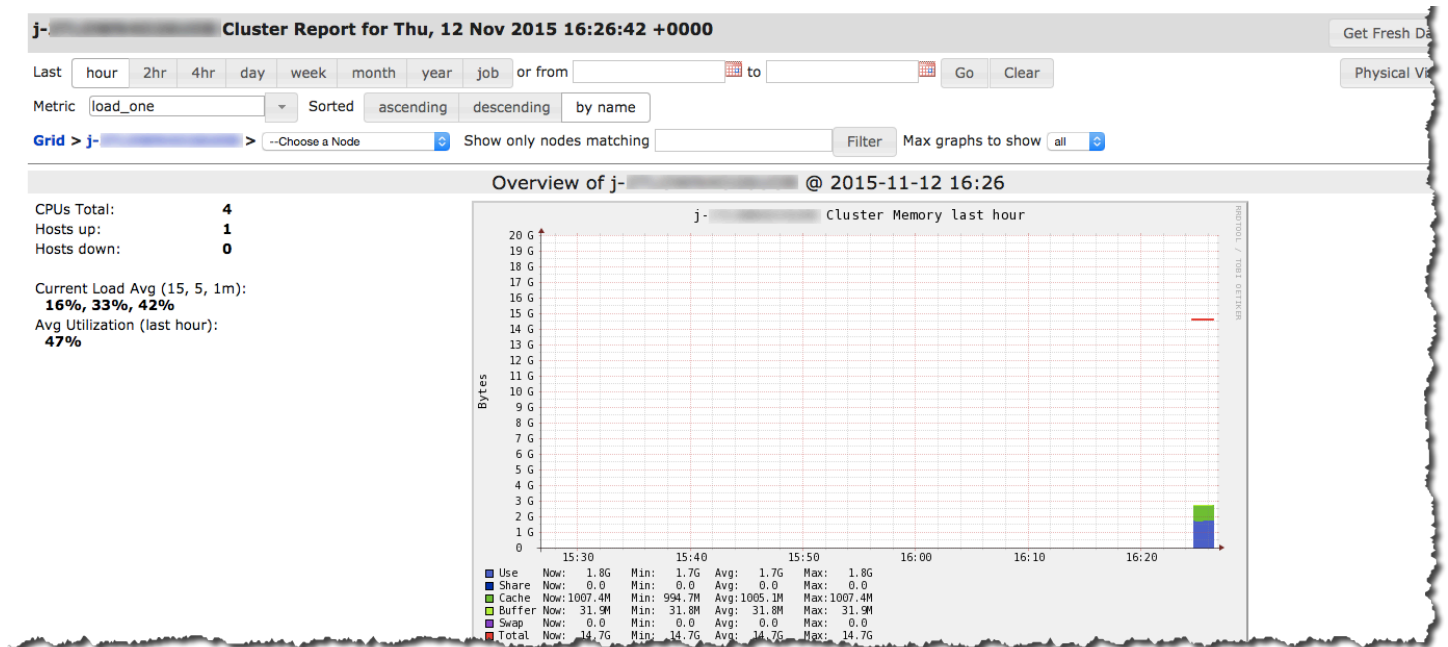
### Note

使用 Ganglia 的最后一个 Amazon EMR 版本为 Amazon EMR 6.15.0。为了监控您的集群，6.15.0 以上的版本中包含了 [Amazon CloudWatch](#) 代理。

Ganglia 提供基于 Web 的用户界面，您可以使用该界面查看 Ganglia 收集的指标。当您在 Amazon EMR 上运行 Ganglia 时，Web 界面会在主节点上运行，并可以使用端口转发（也称为创建 SSH 隧道）进行查看。有关在 Amazon EMR 上查看 Web 界面的更多信息，请参阅《Amazon EMR 管理指南》中的[查看 Amazon EMR 集群上托管的 Web 界面](#)。

### 查看 Ganglia Web 界面

1. 使用 SSH 隧道进入主节点并创建安全连接。有关如何创建到主节点的 SSH 隧道的信息，请参阅《Amazon EMR 管理指南》中的[选项 2，第 1 部分：使用动态端口转发设置到主节点的 SSH 隧道](#)。
2. 安装带有代理工具（例如适用于 Firefox 的 FoxyProxy 插件）的 Web 浏览器，为 \*ec2\*.amazonaws.com\* 类型的域名创建 SOCKS 代理。有关更多信息，请参阅《Amazon EMR 管理指南》中的[选项 2，第 2 部分：配置代理设置以查看主节点上托管的网站](#)。
3. 设置代理并打开 SSH 连接后，你可以打开带有 `http: master-public-dns-name//ganglia/` 的浏览器窗口来查看 Ganglia 用户界面，其中 *master-public-dns-name* 是 EMR 集群中主服务器的公有 DNS 地址。



## Ganglia 中的 Hadoop 和 Spark 指标

### Note

使用 Ganglia 的最后一个 Amazon EMR 版本为 Amazon EMR 6.15.0。为了监控您的集群，6.15.0 以上的版本中包含了 [Amazon CloudWatch](#) 代理。

Ganglia 报告每个实例的 Hadoop 指标。各种类型的指标按类别开头：分布式文件系统 (dfs.\*)、Java 虚拟机 (jvm.\*)、(mapred.\*) 和远程过程调用 MapReduce (rpc.\*)。

基于 YARN 的 Ganglia 指标（如 Spark 和 Hadoop）对于 EMR 发行版 4.4.0 和 4.5.0 不可用。利用更高版本来使用这些指标。

Ganglia 中的 Spark 指标通常具有 YARN 应用程序 ID 和 Spark DAGScheduler 的前缀。前缀遵循以下形式：

- DAGScheduler.\*
- application\_XXXXXXXXXX\_XXXX.driver.\*
- application\_XXXXXXXXXX\_XXXX.executor.\*

## Ganglia 发行版历史记录

### Note

最新发布的包含 Ganglia 的 Amazon EMR 是 Amazon EMR 6.15.0。为了监控您的集群，6.15.0 以上的版本中包含了 A [ma CloudWatch](#) zon 代理。

下表列出了 Amazon EMR 每个发行版中包含的 Ganglia 版本，以及随应用程序一起安装的组件。有关每个发行版本中的组件版本，请参阅 [亚马逊 EMR 7.x 发布版本](#)、[亚马逊 EMR 6.x 发布版本](#) 或 [亚马逊 EMR 5.x 发布版本](#) 中的发行版“组件版本”部分。

### Ganglia 版本信息

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-5.36.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-6.15.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-6.14.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.13.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.12.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-6.11.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.11.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.10.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-6.10.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.9.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.9.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-6.8.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-6.8.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-6.7.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-5.36.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.36.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.6.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver



Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-5.35.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.5.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.4.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-6.3.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.3.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.2.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-6.2.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.1.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.1.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-6.0.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.0.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.34.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-5.33.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.33.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.32.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-5.32.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.31.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.31.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-5.30.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.30.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.30.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-5.29.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.28.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.28.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver



Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-5.27.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.27.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.26.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-5.25.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.24.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.24.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-5.23.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.23.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.22.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-5.21.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.21.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.21.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-5.20.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.20.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.19.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-5.19.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.18.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.18.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-5.17.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.17.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.17.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-5.16.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.16.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.15.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver



Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-5.15.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.14.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.14.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-5.14.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.13.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.13.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-5.12.3	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.12.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.12.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-5.12.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.11.4	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.11.3	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-5.11.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.11.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.11.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-5.10.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.10.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.9.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-5.9.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.8.3	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.8.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-5.8.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.8.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.7.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver



Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-5.7.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.6.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.6.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-5.5.4	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-5.5.3	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-5.5.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-5.5.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-5.5.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-5.4.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-5.4.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-5.3.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-5.3.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-5.3.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-5.2.3	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-5.2.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-5.2.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-5.2.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-5.1.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-5.1.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-5.0.3	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-5.0.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-5.0.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-5.0.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-4.9.6	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver



Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-4.9.5	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-4.9.4	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-4.9.3	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-4.9.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-4.9.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-4.8.5	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-4.8.4	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.8.3	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.8.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-4.8.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-4.8.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-4.7.4	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-4.7.3	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-4.7.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-4.7.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-4.7.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.6.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.6.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver

Amazon EMR 发行版标签	Ganglia 版本	随 Ganglia 安装的组件
emr-4.5.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.4.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.3.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.2.0	3.6.0	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver

# Apache Hadoop

[Apache Hadoop](#) 是一种开源 Java 软件框架，支持跨越实例集群处理大量数据。它可以在单个实例或数千个实例上运行。Hadoop 使用各种处理模型（例如 MapReduce 和 Tez）将处理分布到多个实例，还使用 HDFS 为的分布式文件系统跨多个实例存储数据。Hadoop 监控集群中实例的运行状况，并可从一个或多个节点的故障中恢复。通过这种方式，Hadoop 可增加处理和存储容量以及高可用性。有关更多信息，请参阅 [Hadoop 文档](#)。

下表列出了最新版本的亚马逊 EMR 7.x 系列中包含的 Hadoop 版本，以及亚马逊与 Hadoop 一起EMR 安装的组件。

有关此版本中与 Hadoop 一起安装的组件的版本，请参阅 [7.2.0 版本](#) 组件版本。

emr-7.2.0 的 Hadoop 版本信息

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-7.2.0	Hadoop 3.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server

下表列出了最新版本的亚马逊 EMR 6.x 系列中包含的 Hadoop 版本，以及亚马逊与 Hadoop 一起EMR 安装的组件。

有关此发行版中随 Hadoop 安装的组件版本，请参阅[发行版 6.15.0 组件版本](#)。



## emr-6.15.0 的 Hadoop 版本信息

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-6.15.0	Hadoop 3.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

下表列出了最新版本的亚马逊 EMR 5.x 系列中包含的 Hadoop 版本，以及亚马逊与 Hadoop 一起EMR安装的组件。

有关此版本中与 Hadoop 一起安装的组件的版本，请参阅[发行版 5.36.2](#) 组件版本。

## emr-5.36.2 的 Hadoop 版本信息

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-5.36.2	Hadoop 2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

从 Amazon EMR 5.18.0 开始，您可以使用 Amazon 构建 EMR 存储库针对特定亚马逊版本中提供的库和依赖项的确切版本构建任务代码。EMR 有关更多信息，请参阅 [使用 Amazon EMR 项目存储库检查依赖项](#)。

## 主题

- [配置 Hadoop](#)
- [Amazon HDFS 上有透明加密 EMR](#)
- [创建或运行 Hadoop 应用程序](#)
- [读取已恢复的对象](#)
- [为容器开启非均匀内存访问感知 YARN](#)
- [Hadoop 版本历史记录](#)

## 配置 Hadoop

以下各节提供了 Hadoop 守护程序、任务和的默认配置设置。HDFS

## 主题

- [任务配置](#)
- [Hadoop 守护进程配置设置](#)
- [HDFS配置](#)

## 任务配置

您可以设置配置变量来调整 MapReduce 作业的性能。本节提供了一些重要设置的默认值。默认值因集群中使用的节点的 EC2 实例类型而异。HBase 在使用 Amazon EMR 发行版 4.6.0 及更高版本时可用。安装时 HBase 使用不同的默认值。这些值将与初始默认值一起提供。

Hadoop 2 使用两个参数，即 `mapreduce.map.java.opts` 和 `mapreduce.reduce.java.opts`，分别为 `map` 和 `re` JVMs `duce` 配置内存。它们替代之前的 Hadoop 版本中的单个 `mapreduce.map.java.opts` 配置选项。

同样，在 Hadoop 2.7.2 及更高版本中，`mapred.job.jvm.num.tasks` 替换 `mapred.job.reuse.jvm.num.tasks`。无论 EC2 实例类型如何，Amazon 都会将此值 EMR 设置为 20。您可以使用 `mapred-site` 配置分类覆盖此设置。将值设置为 -1 表示 JVM 可以在单个作业中重复使用无限数量的任务，而值为 1 表示每个任务都会生成一个新 JVM 任务。

例如，要将 `mapred.job.jvm.num.tasks` 的值设置为 `-1`，您可以创建一个包含以下内容的文件：

```
[
  {
    "Classification": "mapred-site",
    "Properties": {
      "mapred.job.jvm.num.tasks": "-1"
    }
  }
]
```

当您使用中的 `create-cluster` 命令或 `modify-instance-groups` 命令时 AWS CLI，可以引用 JSON 配置文件。在以下示例中，该配置文件将另存为 `myConfig.json` 并存储在 Amazon S3 中。

#### Note

为了便于读取，包含 Linux 行继续符 (`\`)。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (^)。

```
aws emr create-cluster --release-label emr-7.2.0 --instance-type m5.xlarge \
--instance-count 3 --applications Name=Hadoop --configurations https://
s3.amazonaws.com/mybucket/myfolder/myConfig.json \
--use-default-roles
```

您可以使用相同的方式使用 `mapred-site` 配置分类更改下面列出的默认值，并使用单个 JSON 文件设置多个值和多个配置分类。有关更多信息，请参阅 [配置应用程序](#)。

在 Amazon 5.21.0 及更高 EMR 版本中，您可以覆盖集群配置，并为正在运行的集群中的每个实例组指定其他配置分类。您可以使用 Amazon EMR 控制台、AWS Command Line Interface (AWS CLI) 或 AWS SDK。有关更多信息，请参阅 [为运行的集群中的实例组提供配置](#)。

## 任务配置设置的默认值

### 实例类型

- [c1 实例](#)
- [c3 实例](#)

- [c4 实例](#)
- [c5 实例](#)
- [c5a 实例](#)
- [c5ad 实例](#)
- [c5d 实例](#)
- [c5n 实例](#)
- [c6a 实例](#)
- [c6g 实例](#)
- [c6gd 实例](#)
- [c6gn 实例](#)
- [c6i 实例](#)
- [c6id 实例](#)
- [c6in 实例](#)
- [c7a 实例](#)
- [c7g 实例](#)
- [c7gd 实例](#)
- [c7gn 实例](#)
- [c7i 实例](#)
- [d2 实例](#)
- [d3 实例](#)
- [d3en 实例](#)
- [g3 实例](#)
- [g3s 实例](#)
- [g4dn 实例](#)
- [g5 实例](#)
- [h1 实例](#)
- [i2 实例](#)
- [i3 实例](#)

- [i3en 实例](#)
- [i4g 实例](#)
- [i4i 实例](#)
- [im4gn 实例](#)
- [is4gen 实例](#)
- [m1 实例](#)
- [m2 实例](#)
- [m3 实例](#)
- [m4 实例](#)
- [m5 实例](#)
- [m5a 实例](#)
- [m5ad 实例](#)
- [m5d 实例](#)
- [m5dn 实例](#)
- [m5n 实例](#)
- [m5zn 实例](#)
- [m6a 实例](#)
- [m6g 实例](#)
- [m6gd 实例](#)
- [m6i 实例](#)
- [m6id 实例](#)
- [m6idn 实例](#)
- [m6in 实例](#)
- [m7a 实例](#)
- [m7g 实例](#)
- [m7gd 实例](#)
- [m7i 实例](#)
- [m7i-flex 实例](#)
- [p2 实例](#)

- [p3 实例](#)
- [p5 实例](#)
- [r3 实例](#)
- [r4 实例](#)
- [r5 实例](#)
- [r5a 实例](#)
- [r5ad 实例](#)
- [r5b 实例](#)
- [r5d 实例](#)
- [r5dn 实例](#)
- [r5n 实例](#)
- [r6a 实例](#)
- [r6g 实例](#)
- [r6gd 实例](#)
- [r6i 实例](#)
- [r6id 实例](#)
- [r6idn 实例](#)
- [r6in 实例](#)
- [r7a 实例](#)
- [r7g 实例](#)
- [r7gd 实例](#)
- [r7i 实例](#)
- [r7iz 实例](#)
- [x1 实例](#)
- [x1e 实例](#)
- [x2gd 实例](#)
- [x2idn 实例](#)
- [x2iedn 实例](#)
- [z1d 实例](#)

## c1 实例

## c1.medium

配置选项	默认值	HBase安装完毕后
mapreduce.map.java.opts	-Xmx288m	-Xmx288m
mapreduce.java.opts	-Xmx288m	-Xmx288m
mapreduce.map.memory.mb	512	512
mapreduce.reduce.memory.mb	512	512
yarn.app.mapreduce.am.resource.mb	512	512
yarn.scheduler.minimum-allocation-mb	256	256
yarn.scheduler.maximum-allocation-mb	512	512
yarn.nodemanager.resource.memory-mb	1024	512

## c1.xlarge

配置选项	默认值	HBase安装完毕后
mapreduce.map.java.opts	-Xmx864m	-Xmx864m
mapreduce.java.opts	-Xmx1536m	-Xmx1536m
mapreduce.map.memory.mb	1024	1024
mapreduce.reduce.memory.mb	2048	2048
yarn.app.mapreduce.am.resource.mb	2048	2048

配置选项	默认值	HBase安装完毕后
yarn.scheduler.minimum-allocation-mb	256	256
yarn.scheduler.maximum-allocation-mb	2048	2560
yarn.nodemanager.resource.memory-mb	5120	2560

## c3 实例

### c3.xlarge

配置选项	默认值	HBase安装完毕后
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816



## c3.2xlarge

配置选项	默认值	HBase安装完毕后
mapreduce.map.java.opts	-Xmx1152m	-Xmx1152m
mapreduce.java.opts	-Xmx2304m	-Xmx2304m
mapreduce.map.memory.mb	1440	1440
mapreduce.reduce.memory.mb	2880	2 880
yarn.app.mapreduce.am.resource.mb	2 880	2 880
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11520	5760
yarn.nodemanager.resource.memory-mb	11520	5760

## c3.4xlarge

配置选项	默认值	HBase安装完毕后
mapreduce.map.java.opts	-Xmx1152m	-Xmx1152m
mapreduce.java.opts	-Xmx2304m	-Xmx2304m
mapreduce.map.memory.mb	1440	1440
mapreduce.reduce.memory.mb	2880	2 880
yarn.app.mapreduce.am.resource.mb	2 880	2 880

配置选项	默认值	HBase安装完毕后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23040	11520
yarn.nodemanager.resource.memory-mb	23040	11520

## c3.8xlarge

配置选项	默认值	HBase安装完毕后
mapreduce.map.java.opts	-Xmx1331m	-Xmx1331m
mapreduce.java.opts	-Xmx2662m	-Xmx2662m
mapreduce.map.memory.mb	1664	1664
mapreduce.reduce.memory.mb	3328	3328
yarn.app.mapreduce.am.resource.mb	3328	3328
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	53248	26624
yarn.nodemanager.resource.memory-mb	53248	26624

## c4 实例

## c4.large

配置选项	默认值	HBase安装完毕后
mapreduce.map.java.opts	-Xmx717m	-Xmx717m
mapreduce.java.opts	-Xmx1434m	-Xmx1434m
mapreduce.map.memory.mb	896	896
mapreduce.reduce.memory.mb	1792	1792
yarn.app.mapreduce.am.resource.mb	1792	1792
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	1792	896
yarn.nodemanager.resource.memory-mb	1792	896

## c4.xlarge

配置选项	默认值	HBase安装完毕后
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816

配置选项	默认值	HBase安装完毕后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

## c4.2xlarge

配置选项	默认值	HBase安装完毕后
mapreduce.map.java.opts	-Xmx1152m	-Xmx1152m
mapreduce.java.opts	-Xmx2304m	-Xmx2304m
mapreduce.map.memory.mb	1440	1440
mapreduce.reduce.memory.mb	2880	2 880
yarn.app.mapreduce.am.resource.mb	2 880	2 880
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11520	5760
yarn.nodemanager.resource.memory-mb	11520	5760

## c4.4xlarge

配置选项	默认值	HBase安装完毕后
mapreduce.map.java.opts	-Xmx1152m	-Xmx1152m
mapreduce.java.opts	-Xmx2304m	-Xmx2304m
mapreduce.map.memory.mb	1440	1440
mapreduce.reduce.memory.mb	2880	2 880
yarn.app.mapreduce.am.resource.mb	2 880	2 880
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23040	11520
yarn.nodemanager.resource.memory-mb	23040	11520

## c4.8xlarge

配置选项	默认值	HBase安装完毕后
mapreduce.map.java.opts	-Xmx1183m	-Xmx1183m
mapreduce.java.opts	-Xmx2366m	-Xmx2366m
mapreduce.map.memory.mb	1479	1479
mapreduce.reduce.memory.mb	2958	2958
yarn.app.mapreduce.am.resource.mb	2958	2958

配置选项	默认值	HBase安装完毕后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	53248	26624
yarn.nodemanager.resource.memory-mb	53248	26624

## c5 实例

## c5.xlarge

配置选项	默认值	HBase安装完毕后
mapreduce.map.java.opts	-Xmx1229m	-Xmx1229m
mapreduce.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.map.memory.mb	1536	1536
mapreduce.reduce.memory.mb	3072	3072
yarn.app.mapreduce.am.resource.mb	3072	3072
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	6144	3072
yarn.nodemanager.resource.memory-mb	6144	3072

## c5.2xlarge

配置选项	默认值	HBase安装完毕后
mapreduce.map.java.opts	-Xmx1229m	-Xmx1229m
mapreduce.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.map.memory.mb	1536	1536
mapreduce.reduce.memory.mb	3072	3072
yarn.app.mapreduce.am.resource.mb	3072	3072
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	12288	6144
yarn.nodemanager.resource.memory-mb	12288	6144

## c5.4xlarge

配置选项	默认值	HBase安装完毕后
mapreduce.map.java.opts	-Xmx1229m	-Xmx1229m
mapreduce.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.map.memory.mb	1536	1536
mapreduce.reduce.memory.mb	3072	3072
yarn.app.mapreduce.am.resource.mb	3072	3072

配置选项	默认值	HBase安装完毕后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24 576	12288
yarn.nodemanager.resource.memory-mb	24 576	12288

## c5.9xlarge

配置选项	默认值	HBase安装完毕后
mapreduce.map.java.opts	-Xmx1456m	-Xmx1456m
mapreduce.java.opts	-Xmx2912m	-Xmx2912m
mapreduce.map.memory.mb	1820	1820
mapreduce.reduce.memory.mb	3640	3640
yarn.app.mapreduce.am.resource.mb	3640	3640
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	65536	32768
yarn.nodemanager.resource.memory-mb	65536	32768



## c5.12xlarge

配置选项	默认值	HBase安装完毕后
mapreduce.map.java.opts	-Xmx1502m	-Xmx1502m
mapreduce.java.opts	-Xmx3004m	-Xmx3004m
mapreduce.map.memory.mb	1877	1877
mapreduce.reduce.memory.mb	3754	3754
yarn.app.mapreduce.am.resource.mb	3754	3754
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	90112	30048
yarn.nodemanager.resource.memory-mb	90112	30048

## c5.18xlarge

配置选项	默认值	HBase安装完毕后
mapreduce.map.java.opts	-Xmx1547m	-Xmx1547m
mapreduce.java.opts	-Xmx3094m	-Xmx3094m
mapreduce.map.memory.mb	1934	1934
mapreduce.reduce.memory.mb	3868	3868
yarn.app.mapreduce.am.resource.mb	3868	3868

配置选项	默认值	HBase安装完毕后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	139264	30960
yarn.nodemanager.resource.memory-mb	139264	30960

## c5.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1570m	-Xmx1570m
mapreduce.java.opts	-Xmx3140m	-Xmx3140m
mapreduce.map.memory.mb	1963	1963
mapreduce.reduce.memory.mb	3926	3926
yarn.app.mapreduce.am.resource.mb	3926	3926
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	188416	31376
yarn.nodemanager.resource.memory-mb	188416	31376

## c5a 实例

## c5a.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

## c5a.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## c5a.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## c5a.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1357m	-Xmx1357m
mapreduce.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.map.memory.mb	1696	1696
mapreduce.reduce.memory.mb	3392	3392
yarn.app.mapreduce.am.resource.mb	3392	3392
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## c5a.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1502m	-Xmx1502m
mapreduce.java.opts	-Xmx3004m	-Xmx3004m
mapreduce.map.memory.mb	1877	1877
mapreduce.reduce.memory.mb	3754	3754
yarn.app.mapreduce.am.resource.mb	3754	3754

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	90112	30048
yarn.nodemanager.resource.memory-mb	90112	30048

## c5a.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## c5a.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1494m	-Xmx1494m
mapreduce.java.opts	-Xmx2988m	-Xmx2988m
mapreduce.map.memory.mb	1867	1867
mapreduce.reduce.memory.mb	3734	3734
yarn.app.mapreduce.am.resource.mb	3734	3734
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29840
yarn.nodemanager.resource.memory-mb	179200	29840

## c5ad 实例

## c5ad.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

## c5ad.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856



## c5ad.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## c5ad.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1357m	-Xmx1357m
mapreduce.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.map.memory.mb	1696	1696
mapreduce.reduce.memory.mb	3392	3392
yarn.app.mapreduce.am.resource.mb	3392	3392

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## c5ad.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1425m	-Xmx1425m
mapreduce.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.map.memory.mb	1781	1781
mapreduce.reduce.memory.mb	3562	3562
yarn.app.mapreduce.am.resource.mb	3562	3562
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32074
yarn.nodemanager.resource.memory-mb	85504	32074

## c5ad.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## c5ad.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1494m	-Xmx1494m
mapreduce.java.opts	-Xmx2988m	-Xmx2988m
mapreduce.map.memory.mb	1867	1867
mapreduce.reduce.memory.mb	3734	3734
yarn.app.mapreduce.am.resource.mb	3734	3734

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29840
yarn.nodemanager.resource.memory-mb	179200	29840

## c5d 实例

## c5d.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1229m	-Xmx1229m
mapreduce.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.map.memory.mb	1536	1536
mapreduce.reduce.memory.mb	3072	3072
yarn.app.mapreduce.am.resource.mb	3072	3072
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	6144	3072
yarn.nodemanager.resource.memory-mb	6144	3072

## c5d.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1229m	-Xmx1229m
mapreduce.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.map.memory.mb	1536	1536
mapreduce.reduce.memory.mb	3072	3072
yarn.app.mapreduce.am.resource.mb	3072	3072
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	12288	6144
yarn.nodemanager.resource.memory-mb	12288	6144

## c5d.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1229m	-Xmx1229m
mapreduce.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.map.memory.mb	1536	1536
mapreduce.reduce.memory.mb	3072	3072
yarn.app.mapreduce.am.resource.mb	3072	3072

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24 576	12288
yarn.nodemanager.resource.memory-mb	24 576	12288

## c5d.9xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1456m	-Xmx1456m
mapreduce.java.opts	-Xmx2912m	-Xmx2912m
mapreduce.map.memory.mb	1820	1820
mapreduce.reduce.memory.mb	3640	3640
yarn.app.mapreduce.am.resource.mb	3640	3640
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	65536	32768
yarn.nodemanager.resource.memory-mb	65536	32768

## c5d.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1502m	-Xmx1502m
mapreduce.java.opts	-Xmx3004m	-Xmx3004m
mapreduce.map.memory.mb	1877	1877
mapreduce.reduce.memory.mb	3754	3754
yarn.app.mapreduce.am.resource.mb	3754	3754
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	90112	30048
yarn.nodemanager.resource.memory-mb	90112	30048

## c5d.18xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1547m	-Xmx1547m
mapreduce.java.opts	-Xmx3094m	-Xmx3094m
mapreduce.map.memory.mb	1934	1934
mapreduce.reduce.memory.mb	3868	3868
yarn.app.mapreduce.am.resource.mb	3868	3868

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	139264	30960
yarn.nodemanager.resource.memory-mb	139264	30960

## c5d.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1570m	-Xmx1570m
mapreduce.java.opts	-Xmx3140m	-Xmx3140m
mapreduce.map.memory.mb	1963	1963
mapreduce.reduce.memory.mb	3926	3926
yarn.app.mapreduce.am.resource.mb	3926	3926
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	188416	31376
yarn.nodemanager.resource.memory-mb	188416	31376



## c5n 实例

## c5n.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1613m	-Xmx1613m
mapreduce.java.opts	-Xmx3226m	-Xmx3226m
mapreduce.map.memory.mb	2016	2016
mapreduce.reduce.memory.mb	4032	4032
yarn.app.mapreduce.am.resource.mb	4032	4032
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	8064	4032
yarn.nodemanager.resource.memory-mb	8064	4032

## c5n.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1613m	-Xmx1613m
mapreduce.java.opts	-Xmx3226m	-Xmx3226m
mapreduce.map.memory.mb	2016	2016
mapreduce.reduce.memory.mb	4032	4032
yarn.app.mapreduce.am.resource.mb	4032	4032

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	16128	8064
yarn.nodemanager.resource.memory-mb	16128	8064

## c5n.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1741m	-Xmx1741m
mapreduce.java.opts	-Xmx3482m	-Xmx3482m
mapreduce.map.memory.mb	2176	2176
mapreduce.reduce.memory.mb	4352	4352
yarn.app.mapreduce.am.resource.mb	4352	4352
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	34816	17408
yarn.nodemanager.resource.memory-mb	34816	17408

## c5n.9xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2002m	-Xmx2002m
mapreduce.java.opts	-Xmx4004m	-Xmx4004m
mapreduce.map.memory.mb	2503	2503
mapreduce.reduce.memory.mb	5006	5006
yarn.app.mapreduce.am.resource.mb	5006	5006
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	90112	30040
yarn.nodemanager.resource.memory-mb	90112	30040

## c5n.18xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2094m	-Xmx2094m
mapreduce.java.opts	-Xmx4188m	-Xmx4188m
mapreduce.map.memory.mb	2617	2617
mapreduce.reduce.memory.mb	5234	5234
yarn.app.mapreduce.am.resource.mb	5234	5234

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	188416	31396
yarn.nodemanager.resource.memory-mb	188416	31396

## c6a 实例

## c6a.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

## c6a.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## c6a.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## c6a.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1357m	-Xmx1357m
mapreduce.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.map.memory.mb	1696	1696
mapreduce.reduce.memory.mb	3392	3392
yarn.app.mapreduce.am.resource.mb	3392	3392
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## c6a.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1425m	-Xmx1425m
mapreduce.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.map.memory.mb	1781	1781
mapreduce.reduce.memory.mb	3562	3562
yarn.app.mapreduce.am.resource.mb	3562	3562
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32074
yarn.nodemanager.resource.memory-mb	85504	32074

## c6a.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## c6a.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1494m	-Xmx1494m
mapreduce.java.opts	-Xmx2988m	-Xmx2988m
mapreduce.map.memory.mb	1867	1867
mapreduce.reduce.memory.mb	3734	3734
yarn.app.mapreduce.am.resource.mb	3734	3734
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29840
yarn.nodemanager.resource.memory-mb	179200	29840



## c6a.32xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1510m	-Xmx1510m
mapreduce.java.opts	-Xmx3020m	-Xmx3020m
mapreduce.map.memory.mb	1888	1888
mapreduce.reduce.memory.mb	3776	3776
yarn.app.mapreduce.am.resource.mb	3776	3776
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## c6a.48xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1527m	-Xmx1527m
mapreduce.java.opts	-Xmx3054m	-Xmx3054m
mapreduce.map.memory.mb	1909	1909
mapreduce.reduce.memory.mb	3818	3818
yarn.app.mapreduce.am.resource.mb	3818	3818

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30608
yarn.nodemanager.resource.memory-mb	366592	30608

## c6g 实例

## c6g.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

## c6g.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## c6g.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## c6g.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1357m	-Xmx1357m
mapreduce.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.map.memory.mb	1696	1696
mapreduce.reduce.memory.mb	3392	3392
yarn.app.mapreduce.am.resource.mb	3392	3392
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## c6g.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1425m	-Xmx1425m
mapreduce.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.map.memory.mb	1781	1781
mapreduce.reduce.memory.mb	3562	3562
yarn.app.mapreduce.am.resource.mb	3562	3562
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32074
yarn.nodemanager.resource.memory-mb	85504	32074

## c6g.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## c6gd 实例

## c6gd.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

## c6gd.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## c6gd.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## c6gd.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1357m	-Xmx1357m
mapreduce.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.map.memory.mb	1696	1696
mapreduce.reduce.memory.mb	3392	3392
yarn.app.mapreduce.am.resource.mb	3392	3392
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136



## c6gd.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1425m	-Xmx1425m
mapreduce.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.map.memory.mb	1781	1781
mapreduce.reduce.memory.mb	3562	3562
yarn.app.mapreduce.am.resource.mb	3562	3562
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32074
yarn.nodemanager.resource.memory-mb	85504	32074

## c6gd.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## c6gn 实例

## c6gn.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

## c6gn.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## c6gn.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## c6gn.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1357m	-Xmx1357m
mapreduce.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.map.memory.mb	1696	1696
mapreduce.reduce.memory.mb	3392	3392
yarn.app.mapreduce.am.resource.mb	3392	3392
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## c6gn.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1425m	-Xmx1425m
mapreduce.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.map.memory.mb	1781	1781
mapreduce.reduce.memory.mb	3562	3562
yarn.app.mapreduce.am.resource.mb	3562	3562
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32074
yarn.nodemanager.resource.memory-mb	85504	32074

## c6gn.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## c6i 实例

## c6i.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

## c6i.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## c6i.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## c6i.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1357m	-Xmx1357m
mapreduce.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.map.memory.mb	1696	1696
mapreduce.reduce.memory.mb	3392	3392
yarn.app.mapreduce.am.resource.mb	3392	3392
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136



## c6i.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1425m	-Xmx1425m
mapreduce.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.map.memory.mb	1781	1781
mapreduce.reduce.memory.mb	3562	3562
yarn.app.mapreduce.am.resource.mb	3562	3562
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32074
yarn.nodemanager.resource.memory-mb	85504	32074

## c6i.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## c6i.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1494m	-Xmx1494m
mapreduce.java.opts	-Xmx2988m	-Xmx2988m
mapreduce.map.memory.mb	1867	1867
mapreduce.reduce.memory.mb	3734	3734
yarn.app.mapreduce.am.resource.mb	3734	3734
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29840
yarn.nodemanager.resource.memory-mb	179200	29840

## c6i.32xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1510m	-Xmx1510m
mapreduce.java.opts	-Xmx3020m	-Xmx3020m
mapreduce.map.memory.mb	1888	1888
mapreduce.reduce.memory.mb	3776	3776
yarn.app.mapreduce.am.resource.mb	3776	3776
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## c6id 实例

## c6id.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

## c6id.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## c6id.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## c6id.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1357m	-Xmx1357m
mapreduce.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.map.memory.mb	1696	1696
mapreduce.reduce.memory.mb	3392	3392
yarn.app.mapreduce.am.resource.mb	3392	3392

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## c6id.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1425m	-Xmx1425m
mapreduce.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.map.memory.mb	1781	1781
mapreduce.reduce.memory.mb	3562	3562
yarn.app.mapreduce.am.resource.mb	3562	3562
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32074
yarn.nodemanager.resource.memory-mb	85504	32074

## c6id.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## c6id.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1494m	-Xmx1494m
mapreduce.java.opts	-Xmx2988m	-Xmx2988m
mapreduce.map.memory.mb	1867	1867
mapreduce.reduce.memory.mb	3734	3734
yarn.app.mapreduce.am.resource.mb	3734	3734

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29840
yarn.nodemanager.resource.memory-mb	179200	29840

## c6id.32xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1510m	-Xmx1510m
mapreduce.java.opts	-Xmx3020m	-Xmx3020m
mapreduce.map.memory.mb	1888	1888
mapreduce.reduce.memory.mb	3776	3776
yarn.app.mapreduce.am.resource.mb	3776	3776
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208



## c6in 实例

## c6in.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

## c6in.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## c6in.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## c6in.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1357m	-Xmx1357m
mapreduce.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.map.memory.mb	1696	1696
mapreduce.reduce.memory.mb	3392	3392
yarn.app.mapreduce.am.resource.mb	3392	3392
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## c6in.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1425m	-Xmx1425m
mapreduce.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.map.memory.mb	1781	1781
mapreduce.reduce.memory.mb	3562	3562
yarn.app.mapreduce.am.resource.mb	3562	3562

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32074
yarn.nodemanager.resource.memory-mb	85504	32074

## c6in.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## c6in.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1494m	-Xmx1494m
mapreduce.java.opts	-Xmx2988m	-Xmx2988m
mapreduce.map.memory.mb	1867	1867
mapreduce.reduce.memory.mb	3734	3734
yarn.app.mapreduce.am.resource.mb	3734	3734
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29840
yarn.nodemanager.resource.memory-mb	179200	29840

## c6in.32xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1510m	-Xmx1510m
mapreduce.java.opts	-Xmx3020m	-Xmx3020m
mapreduce.map.memory.mb	1888	1888
mapreduce.reduce.memory.mb	3776	3776
yarn.app.mapreduce.am.resource.mb	3776	3776

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## c7a 实例

## c7a.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

## c7a.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## c7a.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## c7a.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1357m	-Xmx1357m
mapreduce.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.map.memory.mb	1696	1696
mapreduce.reduce.memory.mb	3392	3392
yarn.app.mapreduce.am.resource.mb	3392	3392
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136



## c7a.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1425m	-Xmx1425m
mapreduce.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.map.memory.mb	1781	1781
mapreduce.reduce.memory.mb	3562	3562
yarn.app.mapreduce.am.resource.mb	3562	3562
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32074
yarn.nodemanager.resource.memory-mb	85504	32074

## c7a.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## c7a.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1494m	-Xmx1494m
mapreduce.java.opts	-Xmx2988m	-Xmx2988m
mapreduce.map.memory.mb	1867	1867
mapreduce.reduce.memory.mb	3734	3734
yarn.app.mapreduce.am.resource.mb	3734	3734
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29840
yarn.nodemanager.resource.memory-mb	179200	29840

## c7a.32xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1510m	-Xmx1510m
mapreduce.java.opts	-Xmx3020m	-Xmx3020m
mapreduce.map.memory.mb	1888	1888
mapreduce.reduce.memory.mb	3776	3776
yarn.app.mapreduce.am.resource.mb	3776	3776
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## c7a.48xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1527m	-Xmx1527m
mapreduce.java.opts	-Xmx3054m	-Xmx3054m
mapreduce.map.memory.mb	1909	1909
mapreduce.reduce.memory.mb	3818	3818
yarn.app.mapreduce.am.resource.mb	3818	3818

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30608
yarn.nodemanager.resource.memory-mb	366592	30608

## c7g 实例

## c7g.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

## c7g.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## c7g.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## c7g.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1357m	-Xmx1357m
mapreduce.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.map.memory.mb	1696	1696
mapreduce.reduce.memory.mb	3392	3392
yarn.app.mapreduce.am.resource.mb	3392	3392
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## c7g.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1425m	-Xmx1425m
mapreduce.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.map.memory.mb	1781	1781
mapreduce.reduce.memory.mb	3562	3562
yarn.app.mapreduce.am.resource.mb	3562	3562
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32074
yarn.nodemanager.resource.memory-mb	85504	32074

## c7g.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## c7gd 实例

## c7gd.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816



## c7gd.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## c7gd.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## c7gd.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1357m	-Xmx1357m
mapreduce.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.map.memory.mb	1696	1696
mapreduce.reduce.memory.mb	3392	3392
yarn.app.mapreduce.am.resource.mb	3392	3392
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## c7gd.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1425m	-Xmx1425m
mapreduce.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.map.memory.mb	1781	1781
mapreduce.reduce.memory.mb	3562	3562
yarn.app.mapreduce.am.resource.mb	3562	3562
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32074
yarn.nodemanager.resource.memory-mb	85504	32074

## c7gd.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## c7gn 实例

## c7gn.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

## c7gn.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## c7gn.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## c7gn.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1357m	-Xmx1357m
mapreduce.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.map.memory.mb	1696	1696
mapreduce.reduce.memory.mb	3392	3392
yarn.app.mapreduce.am.resource.mb	3392	3392
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## c7gn.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1425m	-Xmx1425m
mapreduce.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.map.memory.mb	1781	1781
mapreduce.reduce.memory.mb	3562	3562
yarn.app.mapreduce.am.resource.mb	3562	3562
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32074
yarn.nodemanager.resource.memory-mb	85504	32074

## c7gn.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## c7i 实例

## c7i.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1408	1408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816



## c7i.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## c7i.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## c7i.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1357m	-Xmx1357m
mapreduce.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.map.memory.mb	1696	1696
mapreduce.reduce.memory.mb	3392	3392
yarn.app.mapreduce.am.resource.mb	3392	3392
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## c7i.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1425m	-Xmx1425m
mapreduce.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.map.memory.mb	1781	1781
mapreduce.reduce.memory.mb	3562	3562
yarn.app.mapreduce.am.resource.mb	3562	3562
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32074
yarn.nodemanager.resource.memory-mb	85504	32074

## c7i.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## c7i.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1494m	-Xmx1494m
mapreduce.java.opts	-Xmx2988m	-Xmx2988m
mapreduce.map.memory.mb	1867	1867
mapreduce.reduce.memory.mb	3734	3734
yarn.app.mapreduce.am.resource.mb	3734	3734
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29840
yarn.nodemanager.resource.memory-mb	179200	29840

## c7i.48xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1527m	-Xmx1527m
mapreduce.java.opts	-Xmx3054m	-Xmx3054m
mapreduce.map.memory.mb	1909	1909
mapreduce.reduce.memory.mb	3818	3818
yarn.app.mapreduce.am.resource.mb	3818	3818
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30608
yarn.nodemanager.resource.memory-mb	366592	30608

## d2 实例

## d2.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## d2.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## d2.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## d2.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2417m	-Xmx2417m
mapreduce.java.opts	-Xmx4834m	-Xmx4834m
mapreduce.map.memory.mb	3021	3021
mapreduce.reduce.memory.mb	6042	6042
yarn.app.mapreduce.am.resource.mb	6042	6042

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30194
yarn.nodemanager.resource.memory-mb	241664	30194

### d3 实例

#### d3.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712



## d3.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## d3.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## d3.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## d3en 实例

## d3en.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## d3en.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## d3en.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## d3en.6xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.java.opts	-Xmx5700m	-Xmx5700m
mapreduce.map.memory.mb	3563	3563
mapreduce.reduce.memory.mb	7126	7126
yarn.app.mapreduce.am.resource.mb	7126	7126
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	28496
yarn.nodemanager.resource.memory-mb	85504	28496

## d3en.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## d3en.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29880
yarn.nodemanager.resource.memory-mb	179200	29880

## g3 实例

## g3.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## g3.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## g3.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720



## g3s 实例

## g3s.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## g4dn 实例

## g4dn.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.java.opts	-Xmx4916m	-Xmx4916m
mapreduce.map.memory.mb	3072	3072
mapreduce.reduce.memory.mb	6144	6144

配置选项	默认值	HBase安装后
yarn.app.mapreduce.am.resource.mb	6144	6144
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	12288	6144
yarn.nodemanager.resource.memory-mb	12288	6144

## g4dn.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.java.opts	-Xmx4916m	-Xmx4916m
mapreduce.map.memory.mb	3072	3072
mapreduce.reduce.memory.mb	6144	6144
yarn.app.mapreduce.am.resource.mb	6144	6144
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24 576	12288
yarn.nodemanager.resource.memory-mb	24 576	12288

## g4dn.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2867m	-Xmx2867m
mapreduce.java.opts	-Xmx5734m	-Xmx5734m
mapreduce.map.memory.mb	3584	3584
mapreduce.reduce.memory.mb	7168	7168
yarn.app.mapreduce.am.resource.mb	7168	7168
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	57344	28672
yarn.nodemanager.resource.memory-mb	57344	28672

## g4dn.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3072m	-Xmx3072m
mapreduce.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.map.memory.mb	3840	3840
mapreduce.reduce.memory.mb	7680	7680
yarn.app.mapreduce.am.resource.mb	7680	7680

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	122880	30720
yarn.nodemanager.resource.memory-mb	122880	30720

## g4dn.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3140m	-Xmx3140m
mapreduce.java.opts	-Xmx6280m	-Xmx6280m
mapreduce.map.memory.mb	3925	3925
mapreduce.reduce.memory.mb	7850	7850
yarn.app.mapreduce.am.resource.mb	7850	7850
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	188416	31416
yarn.nodemanager.resource.memory-mb	188416	31416

## g4dn.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3174m	-Xmx3174m
mapreduce.java.opts	-Xmx6348m	-Xmx6348m
mapreduce.map.memory.mb	3968	3968
mapreduce.reduce.memory.mb	7936	7936
yarn.app.mapreduce.am.resource.mb	7936	7936
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	253952	31744
yarn.nodemanager.resource.memory-mb	253952	31744

## g5 实例

## g5.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## g5.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## g5.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## g5.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## g5.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29880
yarn.nodemanager.resource.memory-mb	179200	29880



## g5.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## g5.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3055m	-Xmx3055m
mapreduce.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.map.memory.mb	3819	3819
mapreduce.reduce.memory.mb	7638	7638
yarn.app.mapreduce.am.resource.mb	7638	7638

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30520
yarn.nodemanager.resource.memory-mb	366592	30520

## g5.48xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3089m	-Xmx3089m
mapreduce.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.map.memory.mb	3861	3861
mapreduce.reduce.memory.mb	7722	7722
yarn.app.mapreduce.am.resource.mb	7722	7722
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30952
yarn.nodemanager.resource.memory-mb	741376	30952

## h1 实例

## h1.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.java.opts	-Xmx4916m	-Xmx4916m
mapreduce.map.memory.mb	3072	3072
mapreduce.reduce.memory.mb	6144	6144
yarn.app.mapreduce.am.resource.mb	6144	6144
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24 576	12288
yarn.nodemanager.resource.memory-mb	24 576	12288

## h1.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2867m	-Xmx2867m
mapreduce.java.opts	-Xmx5734m	-Xmx5734m
mapreduce.map.memory.mb	3584	3584
mapreduce.reduce.memory.mb	7168	7168
yarn.app.mapreduce.am.resource.mb	7168	7168

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	57344	28672
yarn.nodemanager.resource.memory-mb	57344	28672

## h1.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3072m	-Xmx3072m
mapreduce.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.map.memory.mb	3840	3840
mapreduce.reduce.memory.mb	7680	7680
yarn.app.mapreduce.am.resource.mb	7680	7680
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	122880	30720
yarn.nodemanager.resource.memory-mb	122880	30720

## h1.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3174m	-Xmx3174m
mapreduce.java.opts	-Xmx6348m	-Xmx6348m
mapreduce.map.memory.mb	3968	3968
mapreduce.reduce.memory.mb	7936	7936
yarn.app.mapreduce.am.resource.mb	7936	7936
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	253952	31744
yarn.nodemanager.resource.memory-mb	253952	31744

## i2 实例

## i2.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## i2.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## i2.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## i2.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

### i3 实例

#### i3.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712



## i3.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## i3.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## i3.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## i3.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

## i3en 实例

## i3en.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx4915m	-Xmx4915m
mapreduce.java.opts	-Xmx9830m	-Xmx9830m
mapreduce.map.memory.mb	6144	6144
mapreduce.reduce.memory.mb	12288	12288
yarn.app.mapreduce.am.resource.mb	12288	12288

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24 576	12288
yarn.nodemanager.resource.memory-mb	24 576	12288

## i3en.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5734m	-Xmx5734m
mapreduce.java.opts	-Xmx11468m	-Xmx11468m
mapreduce.map.memory.mb	7168	7168
mapreduce.reduce.memory.mb	14336	14336
yarn.app.mapreduce.am.resource.mb	14336	14336
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	57344	28672
yarn.nodemanager.resource.memory-mb	57344	28672

## i3en.3xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6007m	-Xmx6007m
mapreduce.java.opts	-Xmx12014m	-Xmx12014m
mapreduce.map.memory.mb	7509	7509
mapreduce.reduce.memory.mb	15018	15018
yarn.app.mapreduce.am.resource.mb	15018	15018
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	90112	30040
yarn.nodemanager.resource.memory-mb	90112	30040

## i3en.6xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6281m	-Xmx6281m
mapreduce.java.opts	-Xmx12562m	-Xmx12562m
mapreduce.map.memory.mb	7851	7851
mapreduce.reduce.memory.mb	15702	15702
yarn.app.mapreduce.am.resource.mb	15702	15702

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	188416	31396
yarn.nodemanager.resource.memory-mb	188416	31396

## i3en.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6417m	-Xmx6417m
mapreduce.java.opts	-Xmx12834m	-Xmx12834m
mapreduce.map.memory.mb	8021	8021
mapreduce.reduce.memory.mb	16042	16042
yarn.app.mapreduce.am.resource.mb	16042	16042
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	385024	32100
yarn.nodemanager.resource.memory-mb	385024	32100

## i3en.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6486m	-Xmx6486m
mapreduce.java.opts	-Xmx12972m	-Xmx12972m
mapreduce.map.memory.mb	8107	8107
mapreduce.reduce.memory.mb	16214	16214
yarn.app.mapreduce.am.resource.mb	16214	16214
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	778240	32396
yarn.nodemanager.resource.memory-mb	778240	32396

## i4g 实例

## i4g.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## i4g.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136



## i4g.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## i4g.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## i4g.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

## i4i 实例

## i4i.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resou rce.mb	11712	11712
yarn.scheduler.minimum-allo cation-mb	32	32
yarn.scheduler.maximum-allo cation-mb	23424	11712
yarn.nodemanager.resource.m emory-mb	23424	11712

## i4i.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resou rce.mb	13568	13568

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## i4i.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## i4i.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## i4i.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

## i4i.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

## i4i.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.java.opts	-Xmx12356m	-Xmx12356m
mapreduce.map.memory.mb	7723	7723
mapreduce.reduce.memory.mb	15446	15446
yarn.app.mapreduce.am.resource.mb	15446	15446
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30860
yarn.nodemanager.resource.memory-mb	741376	30860

## i4i.32xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6195m	-Xmx6195m
mapreduce.java.opts	-Xmx12390m	-Xmx12390m
mapreduce.map.memory.mb	7744	7744
mapreduce.reduce.memory.mb	15488	15488
yarn.app.mapreduce.am.resource.mb	15488	15488

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	991232	30976
yarn.nodemanager.resource.memory-mb	991232	30976

## im4gn 实例

## im4gn.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856



## im4gn.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## im4gn.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## im4gn.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## im4gn.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## is4gen 实例

## is4gen.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3514m	-Xmx3514m
mapreduce.java.opts	-Xmx7028m	-Xmx7028m
mapreduce.map.memory.mb	4393	4393
mapreduce.reduce.memory.mb	8786	8786
yarn.app.mapreduce.am.resource.mb	8786	8786

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	17572.12	8786.06
yarn.nodemanager.resource.memory-mb	17572.12	8786.06

## is4gen.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3866m	-Xmx3866m
mapreduce.java.opts	-Xmx7732m	-Xmx7732m
mapreduce.map.memory.mb	4832	4832
mapreduce.reduce.memory.mb	9664	9664
yarn.app.mapreduce.am.resource.mb	9664	9664
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	38656	19328
yarn.nodemanager.resource.memory-mb	38656	19328

## is4gen.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx4275m	-Xmx4275m
mapreduce.java.opts	-Xmx8550m	-Xmx8550m
mapreduce.map.memory.mb	5344	5344
mapreduce.reduce.memory.mb	10688	10688
yarn.app.mapreduce.am.resource.mb	10688	10688
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32064
yarn.nodemanager.resource.memory-mb	85504	32064

## is4gen.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx4480m	-Xmx4480m
mapreduce.java.opts	-Xmx8960m	-Xmx8960m
mapreduce.map.memory.mb	5600	5600
mapreduce.reduce.memory.mb	11200	11200
yarn.app.mapreduce.am.resource.mb	11200	11200

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	22400
yarn.nodemanager.resource.memory-mb	179200	22400

## m1 实例

## m1.small

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx288m	-Xmx288m
mapreduce.java.opts	-Xmx288m	-Xmx288m
mapreduce.map.memory.mb	512	512
mapreduce.reduce.memory.mb	512	512
yarn.app.mapreduce.am.resource.mb	512	512
yarn.scheduler.minimum-allocation-mb	256	256
yarn.scheduler.maximum-allocation-mb	512	512
yarn.nodemanager.resource.memory-mb	1024	512

## m1.medium

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx512m	-Xmx512m
mapreduce.java.opts	-Xmx768m	-Xmx768m
mapreduce.map.memory.mb	768	768
mapreduce.reduce.memory.mb	1024	1024
yarn.app.mapreduce.am.resource.mb	1024	1024
yarn.scheduler.minimum-allocation-mb	256	256
yarn.scheduler.maximum-allocation-mb	2048	1024
yarn.nodemanager.resource.memory-mb	2048	1024

## m1.large

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx512m	-Xmx512m
mapreduce.java.opts	-Xmx1024m	-Xmx1024m
mapreduce.map.memory.mb	768	768
mapreduce.reduce.memory.mb	1536	1536
yarn.app.mapreduce.am.resource.mb	1536	1536

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	256	256
yarn.scheduler.maximum-allocation-mb	3072	2560
yarn.nodemanager.resource.memory-mb	5120	2560

## m1.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx512m	-Xmx512m
mapreduce.java.opts	-Xmx1536m	-Xmx1536m
mapreduce.map.memory.mb	768	768
mapreduce.reduce.memory.mb	2048	2048
yarn.app.mapreduce.am.resource.mb	2048	2048
yarn.scheduler.minimum-allocation-mb	256	256
yarn.scheduler.maximum-allocation-mb	8192	6144
yarn.nodemanager.resource.memory-mb	12288	6144



## m2 实例

## m2.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx864m	-Xmx864m
mapreduce.java.opts	-Xmx1536m	-Xmx1536m
mapreduce.map.memory.mb	1024	1024
mapreduce.reduce.memory.mb	2048	2048
yarn.app.mapreduce.am.resource.mb	2048	2048
yarn.scheduler.minimum-allocation-mb	256	256
yarn.scheduler.maximum-allocation-mb	7168	7168
yarn.nodemanager.resource.memory-mb	14336	7168

## m2.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1280m	-Xmx1280m
mapreduce.java.opts	-Xmx2304m	-Xmx2304m
mapreduce.map.memory.mb	1536	1536
mapreduce.reduce.memory.mb	2560	2560
yarn.app.mapreduce.am.resource.mb	2560	2560

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	256	256
yarn.scheduler.maximum-allocation-mb	8192	15360
yarn.nodemanager.resource.memory-mb	30720	15360

## m2.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1280m	-Xmx1280m
mapreduce.java.opts	-Xmx2304m	-Xmx2304m
mapreduce.map.memory.mb	1536	1536
mapreduce.reduce.memory.mb	2560	2560
yarn.app.mapreduce.am.resource.mb	2560	2560
yarn.scheduler.minimum-allocation-mb	256	256
yarn.scheduler.maximum-allocation-mb	8192	30720
yarn.nodemanager.resource.memory-mb	61440	30720

## m3 实例

## m3.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1152m	-Xmx1152m
mapreduce.java.opts	-Xmx2304m	-Xmx2304m
mapreduce.map.memory.mb	1440	1440
mapreduce.reduce.memory.mb	2880	2 880
yarn.app.mapreduce.am.resou rce.mb	2 880	2 880
yarn.scheduler.minimum-alloc ation-mb	32	32
yarn.scheduler.maximum-alloc ation-mb	11520	5760
yarn.nodemanager.resource.m emory-mb	11520	5760

## m3.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1152m	-Xmx1152m
mapreduce.java.opts	-Xmx2304m	-Xmx2304m
mapreduce.map.memory.mb	1440	1440
mapreduce.reduce.memory.mb	2880	2 880
yarn.app.mapreduce.am.resou rce.mb	2 880	2 880

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23040	11520
yarn.nodemanager.resource.memory-mb	23040	11520

## m4 实例

## m4.large

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1229m	-Xmx1229m
mapreduce.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.map.memory.mb	1536	1536
mapreduce.reduce.memory.mb	3072	3072
yarn.app.mapreduce.am.resource.mb	3072	3072
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	6144	3072
yarn.nodemanager.resource.memory-mb	6144	3072

## m4.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1229m	-Xmx1229m
mapreduce.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.map.memory.mb	1536	1536
mapreduce.reduce.memory.mb	3072	3072
yarn.app.mapreduce.am.resource.mb	3072	3072
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	12288	6144
yarn.nodemanager.resource.memory-mb	12288	6144

## m4.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1229m	-Xmx1229m
mapreduce.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.map.memory.mb	1536	1536
mapreduce.reduce.memory.mb	3072	3072
yarn.app.mapreduce.am.resource.mb	3072	3072

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24 576	12288
yarn.nodemanager.resource.memory-mb	24 576	12288

## m4.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1434m	-Xmx1434m
mapreduce.java.opts	-Xmx2868m	-Xmx2868m
mapreduce.map.memory.mb	1792	1792
mapreduce.reduce.memory.mb	3584	3584
yarn.app.mapreduce.am.resource.mb	3584	3584
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	57344	28672
yarn.nodemanager.resource.memory-mb	57344	28672

## m4.10xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1557m	-Xmx1557m
mapreduce.java.opts	-Xmx3114m	-Xmx3114m
mapreduce.map.memory.mb	1946	1946
mapreduce.reduce.memory.mb	3892	3892
yarn.app.mapreduce.am.resource.mb	3892	3892
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	155648	311104
yarn.nodemanager.resource.memory-mb	155648	311104

## m4.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx1587m	-Xmx1587m
mapreduce.java.opts	-Xmx3174m	-Xmx3174m
mapreduce.map.memory.mb	1984	1984
mapreduce.reduce.memory.mb	3968	3968
yarn.app.mapreduce.am.resource.mb	3968	3968

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	253952	31744
yarn.nodemanager.resource.memory-mb	253952	31744

## m5 实例

## m5.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.java.opts	-Xmx4916m	-Xmx4916m
mapreduce.map.memory.mb	3072	3072
mapreduce.reduce.memory.mb	6144	6144
yarn.app.mapreduce.am.resource.mb	6144	6144
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	12288	6144
yarn.nodemanager.resource.memory-mb	12288	6144



## m5.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.java.opts	-Xmx4916m	-Xmx4916m
mapreduce.map.memory.mb	3072	3072
mapreduce.reduce.memory.mb	6144	6144
yarn.app.mapreduce.am.resource.mb	6144	6144
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24 576	12288
yarn.nodemanager.resource.memory-mb	24 576	12288

## m5.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2867m	-Xmx2867m
mapreduce.java.opts	-Xmx5734m	-Xmx5734m
mapreduce.map.memory.mb	3584	3584
mapreduce.reduce.memory.mb	7168	7168
yarn.app.mapreduce.am.resource.mb	7168	7168

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	57344	28672
yarn.nodemanager.resource.memory-mb	57344	28672

## m5.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3072m	-Xmx3072m
mapreduce.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.map.memory.mb	3840	3840
mapreduce.reduce.memory.mb	7680	7680
yarn.app.mapreduce.am.resource.mb	7680	7680
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	122880	30720
yarn.nodemanager.resource.memory-mb	122880	30720

## m5.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3140m	-Xmx3140m
mapreduce.java.opts	-Xmx6280m	-Xmx6280m
mapreduce.map.memory.mb	3925	3925
mapreduce.reduce.memory.mb	7850	7850
yarn.app.mapreduce.am.resource.mb	7850	7850
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	188416	31416
yarn.nodemanager.resource.memory-mb	188416	31416

## m5.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3174m	-Xmx3174m
mapreduce.java.opts	-Xmx6348m	-Xmx6348m
mapreduce.map.memory.mb	3968	3968
mapreduce.reduce.memory.mb	7936	7936
yarn.app.mapreduce.am.resource.mb	7936	7936

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	253952	31744
yarn.nodemanager.resource.memory-mb	253952	31744

## m5.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3209m	-Xmx3209m
mapreduce.java.opts	-Xmx6418m	-Xmx6418m
mapreduce.map.memory.mb	4011	4011
mapreduce.reduce.memory.mb	8022	8022
yarn.app.mapreduce.am.resource.mb	8022	8022
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	385024	32056
yarn.nodemanager.resource.memory-mb	385024	32056

## m5a 实例

## m5a.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.java.opts	-Xmx4916m	-Xmx4916m
mapreduce.map.memory.mb	3072	3072
mapreduce.reduce.memory.mb	6144	6144
yarn.app.mapreduce.am.resou rce.mb	6144	6144
yarn.scheduler.minimum-allo cation-mb	32	32
yarn.scheduler.maximum-allo cation-mb	12288	6144
yarn.nodemanager.resource.m emory-mb	12288	6144

## m5a.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.java.opts	-Xmx4916m	-Xmx4916m
mapreduce.map.memory.mb	3072	3072
mapreduce.reduce.memory.mb	6144	6144
yarn.app.mapreduce.am.resou rce.mb	6144	6144

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24 576	12288
yarn.nodemanager.resource.memory-mb	24 576	12288

## m5a.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2867m	-Xmx2867m
mapreduce.java.opts	-Xmx5734m	-Xmx5734m
mapreduce.map.memory.mb	3584	3584
mapreduce.reduce.memory.mb	7168	7168
yarn.app.mapreduce.am.resource.mb	7168	7168
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	57344	28672
yarn.nodemanager.resource.memory-mb	57344	28672

## m5a.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3072m	-Xmx3072m
mapreduce.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.map.memory.mb	3840	3840
mapreduce.reduce.memory.mb	7680	7680
yarn.app.mapreduce.am.resource.mb	7680	7680
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	122880	30720
yarn.nodemanager.resource.memory-mb	122880	30720

## m5a.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3140m	-Xmx3140m
mapreduce.java.opts	-Xmx6280m	-Xmx6280m
mapreduce.map.memory.mb	3925	3925
mapreduce.reduce.memory.mb	7850	7850
yarn.app.mapreduce.am.resource.mb	7850	7850

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	188416	31416
yarn.nodemanager.resource.memory-mb	188416	31416

## m5a.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3174m	-Xmx3174m
mapreduce.java.opts	-Xmx6348m	-Xmx6348m
mapreduce.map.memory.mb	3968	3968
mapreduce.reduce.memory.mb	7936	7936
yarn.app.mapreduce.am.resource.mb	7936	7936
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	253952	31744
yarn.nodemanager.resource.memory-mb	253952	31744



## m5a.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3209m	-Xmx3209m
mapreduce.java.opts	-Xmx6418m	-Xmx6418m
mapreduce.map.memory.mb	4011	4011
mapreduce.reduce.memory.mb	8022	8022
yarn.app.mapreduce.am.resource.mb	8022	8022
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	385024	32056
yarn.nodemanager.resource.memory-mb	385024	32056

## m5ad 实例

## m5ad.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## m5ad.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## m5ad.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## m5ad.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## m5ad.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29880
yarn.nodemanager.resource.memory-mb	179200	29880

## m5ad.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## m5ad.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3055m	-Xmx3055m
mapreduce.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.map.memory.mb	3819	3819
mapreduce.reduce.memory.mb	7638	7638
yarn.app.mapreduce.am.resource.mb	7638	7638

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30520
yarn.nodemanager.resource.memory-mb	366592	30520

## m5d 实例

## m5d.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.java.opts	-Xmx4916m	-Xmx4916m
mapreduce.map.memory.mb	3072	3072
mapreduce.reduce.memory.mb	6144	6144
yarn.app.mapreduce.am.resource.mb	6144	6144
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	12288	6144
yarn.nodemanager.resource.memory-mb	12288	6144

## m5d.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.java.opts	-Xmx4916m	-Xmx4916m
mapreduce.map.memory.mb	3072	3072
mapreduce.reduce.memory.mb	6144	6144
yarn.app.mapreduce.am.resource.mb	6144	6144
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24 576	12288
yarn.nodemanager.resource.memory-mb	24 576	12288

## m5d.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2867m	-Xmx2867m
mapreduce.java.opts	-Xmx5734m	-Xmx5734m
mapreduce.map.memory.mb	3584	3584
mapreduce.reduce.memory.mb	7168	7168
yarn.app.mapreduce.am.resource.mb	7168	7168

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	57344	28672
yarn.nodemanager.resource.memory-mb	57344	28672

## m5d.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3072m	-Xmx3072m
mapreduce.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.map.memory.mb	3840	3840
mapreduce.reduce.memory.mb	7680	7680
yarn.app.mapreduce.am.resource.mb	7680	7680
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	122880	30720
yarn.nodemanager.resource.memory-mb	122880	30720



## m5d.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3140m	-Xmx3140m
mapreduce.java.opts	-Xmx6280m	-Xmx6280m
mapreduce.map.memory.mb	3925	3925
mapreduce.reduce.memory.mb	7850	7850
yarn.app.mapreduce.am.resource.mb	7850	7850
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	188416	31416
yarn.nodemanager.resource.memory-mb	188416	31416

## m5d.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3174m	-Xmx3174m
mapreduce.java.opts	-Xmx6348m	-Xmx6348m
mapreduce.map.memory.mb	3968	3968
mapreduce.reduce.memory.mb	7936	7936
yarn.app.mapreduce.am.resource.mb	7936	7936

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	253952	31744
yarn.nodemanager.resource-memory-mb	253952	31744

## m5d.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3209m	-Xmx3209m
mapreduce.java.opts	-Xmx6418m	-Xmx6418m
mapreduce.map.memory.mb	4011	4011
mapreduce.reduce.memory.mb	8022	8022
yarn.app.mapreduce.am.resource.mb	8022	8022
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	385024	32056
yarn.nodemanager.resource-memory-mb	385024	32056

## m5dn 实例

## m5dn.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## m5dn.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## m5dn.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## m5dn.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## m5dn.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29880
yarn.nodemanager.resource.memory-mb	179200	29880

## m5dn.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## m5dn.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3055m	-Xmx3055m
mapreduce.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.map.memory.mb	3819	3819
mapreduce.reduce.memory.mb	7638	7638
yarn.app.mapreduce.am.resource.mb	7638	7638
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30520
yarn.nodemanager.resource.memory-mb	366592	30520

## m5n 实例

## m5n.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## m5n.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712



## m5n.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## m5n.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## m5n.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29880
yarn.nodemanager.resource.memory-mb	179200	29880

## m5n.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## m5n.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3055m	-Xmx3055m
mapreduce.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.map.memory.mb	3819	3819
mapreduce.reduce.memory.mb	7638	7638
yarn.app.mapreduce.am.resource.mb	7638	7638

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30520
yarn.nodemanager.resource.memory-mb	366592	30520

m5zn 实例

m5zn.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2304m	-Xmx2304m
mapreduce.java.opts	-Xmx4608m	-Xmx4608m
mapreduce.map.memory.mb	2880	2 880
mapreduce.reduce.memory.mb	5760	5760
yarn.app.mapreduce.am.resource.mb	5760	5760
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11520	5760
yarn.nodemanager.resource.memory-mb	11520	5760

## m5zn.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## m5zn.3xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2577m	-Xmx2577m
mapreduce.java.opts	-Xmx5154m	-Xmx5154m
mapreduce.map.memory.mb	3221	3221
mapreduce.reduce.memory.mb	6442	6442
yarn.app.mapreduce.am.resource.mb	6442	6442

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	38656	19328
yarn.nodemanager.resource.memory-mb	38656	19328

## m5zn.6xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.java.opts	-Xmx5700m	-Xmx5700m
mapreduce.map.memory.mb	3563	3563
mapreduce.reduce.memory.mb	7126	7126
yarn.app.mapreduce.am.resource.mb	7126	7126
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	28496
yarn.nodemanager.resource.memory-mb	85504	28496

## m5zn.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29880
yarn.nodemanager.resource.memory-mb	179200	29880

## m6a 实例

## m6a.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## m6a.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712



## m6a.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## m6a.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## m6a.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29880
yarn.nodemanager.resource.memory-mb	179200	29880

## m6a.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## m6a.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3055m	-Xmx3055m
mapreduce.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.map.memory.mb	3819	3819
mapreduce.reduce.memory.mb	7638	7638
yarn.app.mapreduce.am.resource.mb	7638	7638

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30520
yarn.nodemanager.resource.memory-mb	366592	30520

## m6a.32xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3072m	-Xmx3072m
mapreduce.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.map.memory.mb	3840	3840
mapreduce.reduce.memory.mb	7680	7680
yarn.app.mapreduce.am.resource.mb	7680	7680
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

## m6a.48xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3089m	-Xmx3089m
mapreduce.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.map.memory.mb	3861	3861
mapreduce.reduce.memory.mb	7722	7722
yarn.app.mapreduce.am.resource.mb	7722	7722
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30952
yarn.nodemanager.resource.memory-mb	741376	30952

## m6g 实例

## m6g.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## m6g.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## m6g.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## m6g.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## m6g.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	181248	30208
yarn.nodemanager.resource.memory-mb	181248	30208



## m6g.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## m6gd 实例

## m6gd.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## m6gd.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## m6gd.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## m6gd.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## m6gd.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	181248	30208
yarn.nodemanager.resource.memory-mb	181248	30208

## m6gd.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## m6i 实例

## m6i.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## m6i.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## m6i.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## m6i.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## m6i.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	181248	30208
yarn.nodemanager.resource.memory-mb	181248	30208



## m6i.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## m6i.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3055m	-Xmx3055m
mapreduce.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.map.memory.mb	3819	3819
mapreduce.reduce.memory.mb	7638	7638
yarn.app.mapreduce.am.resource.mb	7638	7638

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30520
yarn.nodemanager.resource.memory-mb	366592	30520

## m6i.32xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3072m	-Xmx3072m
mapreduce.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.map.memory.mb	3840	3840
mapreduce.reduce.memory.mb	7680	7680
yarn.app.mapreduce.am.resource.mb	7680	7680
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

## m6id 实例

## m6id.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## m6id.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## m6id.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## m6id.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## m6id.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29880
yarn.nodemanager.resource.memory-mb	179200	29880

## m6id.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## m6id.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3055m	-Xmx3055m
mapreduce.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.map.memory.mb	3819	3819
mapreduce.reduce.memory.mb	7638	7638
yarn.app.mapreduce.am.resource.mb	7638	7638
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30520
yarn.nodemanager.resource.memory-mb	366592	30520

## m6id.32xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3072m	-Xmx3072m
mapreduce.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.map.memory.mb	3840	3840
mapreduce.reduce.memory.mb	7680	7680
yarn.app.mapreduce.am.resource.mb	7680	7680

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

## m6idn 实例

## m6idn.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856



## m6idn.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## m6idn.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## m6idn.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## m6idn.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29880
yarn.nodemanager.resource.memory-mb	179200	29880

## m6idn.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## m6idn.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3055m	-Xmx3055m
mapreduce.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.map.memory.mb	3819	3819
mapreduce.reduce.memory.mb	7638	7638
yarn.app.mapreduce.am.resource.mb	7638	7638
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30520
yarn.nodemanager.resource.memory-mb	366592	30520

## m6idn.32xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3072m	-Xmx3072m
mapreduce.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.map.memory.mb	3840	3840
mapreduce.reduce.memory.mb	7680	7680
yarn.app.mapreduce.am.resource.mb	7680	7680
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

## m6in 实例

## m6in.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## m6in.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## m6in.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## m6in.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## m6in.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	181248	30208
yarn.nodemanager.resource.memory-mb	181248	30208



## m6in.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## m6in.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3055m	-Xmx3055m
mapreduce.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.map.memory.mb	3819	3819
mapreduce.reduce.memory.mb	7638	7638
yarn.app.mapreduce.am.resource.mb	7638	7638

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30520
yarn.nodemanager.resource.memory-mb	366592	30520

## m6in.32xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3072m	-Xmx3072m
mapreduce.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.map.memory.mb	3840	3840
mapreduce.reduce.memory.mb	7680	7680
yarn.app.mapreduce.am.resource.mb	7680	7680
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

## m7a 实例

## m7a.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resou rce.mb	5856	5856
yarn.scheduler.minimum-allo cation-mb	32	32
yarn.scheduler.maximum-allo cation-mb	11712	5856
yarn.nodemanager.resource.m emory-mb	11712	5856

## m7a.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resou rce.mb	5856	5856

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## m7a.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## m7a.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## m7a.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29880
yarn.nodemanager.resource.memory-mb	179200	29880

## m7a.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## m7a.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3055m	-Xmx3055m
mapreduce.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.map.memory.mb	3819	3819
mapreduce.reduce.memory.mb	7638	7638
yarn.app.mapreduce.am.resource.mb	7638	7638
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30520
yarn.nodemanager.resource.memory-mb	366592	30520

## m7a.32xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3072m	-Xmx3072m
mapreduce.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.map.memory.mb	3840	3840
mapreduce.reduce.memory.mb	7680	7680
yarn.app.mapreduce.am.resource.mb	7680	7680

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

## m7a.48xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3089m	-Xmx3089m
mapreduce.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.map.memory.mb	3861	3861
mapreduce.reduce.memory.mb	7722	7722
yarn.app.mapreduce.am.resource.mb	7722	7722
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30952
yarn.nodemanager.resource.memory-mb	741376	30952



## m7g 实例

## m7g.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## m7g.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## m7g.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## m7g.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## m7g.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29880
yarn.nodemanager.resource.memory-mb	179200	29880

## m7g.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## m7gd 实例

## m7gd.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## m7gd.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## m7gd.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## m7gd.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## m7gd.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29880
yarn.nodemanager.resource.memory-mb	179200	29880

## m7gd.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208



## m7i 实例

## m7i.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resou rce.mb	5856	5856
yarn.scheduler.minimum-allo cation-mb	32	32
yarn.scheduler.maximum-allo cation-mb	11712	5856
yarn.nodemanager.resource.m emory-mb	11712	5856

## m7i.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resou rce.mb	5856	5856

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## m7i.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## m7i.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## m7i.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	181248	30208
yarn.nodemanager.resource.memory-mb	181248	30208

## m7i.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## m7i.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3055m	-Xmx3055m
mapreduce.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.map.memory.mb	3819	3819
mapreduce.reduce.memory.mb	7638	7638
yarn.app.mapreduce.am.resource.mb	7638	7638
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30520
yarn.nodemanager.resource.memory-mb	366592	30520

## m7i.48xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3089m	-Xmx3089m
mapreduce.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.map.memory.mb	3861	3861
mapreduce.reduce.memory.mb	7722	7722
yarn.app.mapreduce.am.resource.mb	7722	7722

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30952
yarn.nodemanager.resource.memory-mb	741376	30952

### m7i-flex 实例

#### m7i-flex.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

## m7i-flex.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## m7i-flex.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## m7i-flex.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184



## p2 实例

## p2.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.java.opts	-Xmx21708m	-Xmx21708m
mapreduce.map.memory.mb	13568	13568
mapreduce.reduce.memory.mb	27136	27136
yarn.app.mapreduce.am.resource.mb	27136	27136
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## p2.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.java.opts	-Xmx24576m	-Xmx24576m
mapreduce.map.memory.mb	15360	15360
mapreduce.reduce.memory.mb	30720	30720
yarn.app.mapreduce.am.resource.mb	30720	30720

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

## p2.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx9267m	-Xmx9267m
mapreduce.java.opts	-Xmx18534m	-Xmx18534m
mapreduce.map.memory.mb	11584	11584
mapreduce.reduce.memory.mb	23168	23168
yarn.app.mapreduce.am.resource.mb	23168	23168
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	23168
yarn.nodemanager.resource.memory-mb	741376	23168

## p3 实例

## p3.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## p3.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## p3.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

## p5 实例

## p5.48xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx8294m	-Xmx8294m
mapreduce.java.opts	-Xmx16588m	-Xmx16588m
mapreduce.map.memory.mb	10368	10368
mapreduce.reduce.memory.mb	20736	20736
yarn.app.mapreduce.am.resource.mb	20736	20736
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	1990656	20736
yarn.nodemanager.resource.memory-mb	1990656	20736

## r3 实例

## r3.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856

配置选项	默认值	HBase安装后
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## r3.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## r3.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## r3.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## r4 实例

## r4.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712



## r4.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## r4.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## r4.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## r4.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

## r5 实例

## r5.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx4915m	-Xmx4915m
mapreduce.java.opts	-Xmx9830m	-Xmx9830m
mapreduce.map.memory.mb	6144	6144
mapreduce.reduce.memory.mb	12288	12288
yarn.app.mapreduce.am.resource.mb	12288	12288

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24 576	12288
yarn.nodemanager.resource.memory-mb	24 576	12288

## r5.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5734m	-Xmx5734m
mapreduce.java.opts	-Xmx11468m	-Xmx11468m
mapreduce.map.memory.mb	7168	7168
mapreduce.reduce.memory.mb	14336	14336
yarn.app.mapreduce.am.resource.mb	14336	14336
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	57344	28672
yarn.nodemanager.resource.memory-mb	57344	28672

## r5.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	122880	30720
yarn.nodemanager.resource.memory-mb	122880	30720

## r5.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6349m	-Xmx6349m
mapreduce.java.opts	-Xmx12698m	-Xmx12698m
mapreduce.map.memory.mb	7936	7936
mapreduce.reduce.memory.mb	15872	15872
yarn.app.mapreduce.am.resource.mb	15872	15872

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	253952	31744
yarn.nodemanager.resource.memory-mb	253952	31744

## r5.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6417m	-Xmx6417m
mapreduce.java.opts	-Xmx12834m	-Xmx12834m
mapreduce.map.memory.mb	8021	8021
mapreduce.reduce.memory.mb	16042	16042
yarn.app.mapreduce.am.resource.mb	16042	16042
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	385024	32100
yarn.nodemanager.resource.memory-mb	385024	32100

## r5.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6451m	-Xmx6451m
mapreduce.java.opts	-Xmx12902m	-Xmx12902m
mapreduce.map.memory.mb	8064	8064
mapreduce.reduce.memory.mb	16128	16128
yarn.app.mapreduce.am.resource.mb	16128	16128
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	516096	32256
yarn.nodemanager.resource.memory-mb	516096	32256

## r5.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6486m	-Xmx6486m
mapreduce.java.opts	-Xmx12972m	-Xmx12972m
mapreduce.map.memory.mb	8107	8107
mapreduce.reduce.memory.mb	16214	16214
yarn.app.mapreduce.am.resource.mb	16214	16214

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	778240	32396
yarn.nodemanager.resource.memory-mb	778240	32396

## r5a 实例

## r5a.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx4915m	-Xmx4915m
mapreduce.java.opts	-Xmx9830m	-Xmx9830m
mapreduce.map.memory.mb	6144	6144
mapreduce.reduce.memory.mb	12288	12288
yarn.app.mapreduce.am.resource.mb	12288	12288
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24 576	12288
yarn.nodemanager.resource.memory-mb	24 576	12288



## r5a.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5734m	-Xmx5734m
mapreduce.java.opts	-Xmx11468m	-Xmx11468m
mapreduce.map.memory.mb	7168	7168
mapreduce.reduce.memory.mb	14336	14336
yarn.app.mapreduce.am.resource.mb	14336	14336
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	57344	28672
yarn.nodemanager.resource.memory-mb	57344	28672

## r5a.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	122880	30720
yarn.nodemanager.resource.memory-mb	122880	30720

## r5a.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6349m	-Xmx6349m
mapreduce.java.opts	-Xmx12698m	-Xmx12698m
mapreduce.map.memory.mb	7936	7936
mapreduce.reduce.memory.mb	15872	15872
yarn.app.mapreduce.am.resource.mb	15872	15872
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	253952	31744
yarn.nodemanager.resource.memory-mb	253952	31744

## r5a.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6417m	-Xmx6417m
mapreduce.java.opts	-Xmx12834m	-Xmx12834m
mapreduce.map.memory.mb	8021	8021
mapreduce.reduce.memory.mb	16042	16042
yarn.app.mapreduce.am.resource.mb	16042	16042
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	385024	32100
yarn.nodemanager.resource.memory-mb	385024	32100

## r5a.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6451m	-Xmx6451m
mapreduce.java.opts	-Xmx12902m	-Xmx12902m
mapreduce.map.memory.mb	8064	8064
mapreduce.reduce.memory.mb	16128	16128
yarn.app.mapreduce.am.resource.mb	16128	16128

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	516096	32256
yarn.nodemanager.resource.memory-mb	516096	32256

## r5a.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6486m	-Xmx6486m
mapreduce.java.opts	-Xmx12972m	-Xmx12972m
mapreduce.map.memory.mb	8107	8107
mapreduce.reduce.memory.mb	16214	16214
yarn.app.mapreduce.am.resource.mb	16214	16214
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	778240	32396
yarn.nodemanager.resource.memory-mb	778240	32396

## r5ad 实例

## r5ad.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## r5ad.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## r5ad.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## r5ad.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## r5ad.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

## r5ad.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6246m	-Xmx6246m
mapreduce.java.opts	-Xmx12492m	-Xmx12492m
mapreduce.map.memory.mb	7808	7808
mapreduce.reduce.memory.mb	15616	15616
yarn.app.mapreduce.am.resource.mb	15616	15616
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	499712	31232
yarn.nodemanager.resource.memory-mb	499712	31232



## r5ad.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.java.opts	-Xmx12356m	-Xmx12356m
mapreduce.map.memory.mb	7723	7723
mapreduce.reduce.memory.mb	15446	15446
yarn.app.mapreduce.am.resource.mb	15446	15446
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30860
yarn.nodemanager.resource.memory-mb	741376	30860

## r5b 实例

## r5b.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## r5b.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## r5b.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## r5b.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## r5b.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

## r5b.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

## r5b.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.java.opts	-Xmx12356m	-Xmx12356m
mapreduce.map.memory.mb	7723	7723
mapreduce.reduce.memory.mb	15446	15446
yarn.app.mapreduce.am.resource.mb	15446	15446

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30860
yarn.nodemanager.resource.memory-mb	741376	30860

## r5d 实例

## r5d.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx4915m	-Xmx4915m
mapreduce.java.opts	-Xmx9830m	-Xmx9830m
mapreduce.map.memory.mb	6144	6144
mapreduce.reduce.memory.mb	12288	12288
yarn.app.mapreduce.am.resource.mb	12288	12288
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24 576	12288
yarn.nodemanager.resource.memory-mb	24 576	12288

## r5d.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5734m	-Xmx5734m
mapreduce.java.opts	-Xmx11468m	-Xmx11468m
mapreduce.map.memory.mb	7168	7168
mapreduce.reduce.memory.mb	14336	14336
yarn.app.mapreduce.am.resource.mb	14336	14336
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	57344	28672
yarn.nodemanager.resource.memory-mb	57344	28672

## r5d.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	122880	30720
yarn.nodemanager.resource.memory-mb	122880	30720

## r5d.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6349m	-Xmx6349m
mapreduce.java.opts	-Xmx12698m	-Xmx12698m
mapreduce.map.memory.mb	7936	7936
mapreduce.reduce.memory.mb	15872	15872
yarn.app.mapreduce.am.resource.mb	15872	15872
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	253952	31744
yarn.nodemanager.resource.memory-mb	253952	31744



## r5d.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6417m	-Xmx6417m
mapreduce.java.opts	-Xmx12834m	-Xmx12834m
mapreduce.map.memory.mb	8021	8021
mapreduce.reduce.memory.mb	16042	16042
yarn.app.mapreduce.am.resource.mb	16042	16042
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	385024	32100
yarn.nodemanager.resource.memory-mb	385024	32100

## r5d.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6451m	-Xmx6451m
mapreduce.java.opts	-Xmx12902m	-Xmx12902m
mapreduce.map.memory.mb	8064	8064
mapreduce.reduce.memory.mb	16128	16128
yarn.app.mapreduce.am.resource.mb	16128	16128

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	516096	32256
yarn.nodemanager.resource.memory-mb	516096	32256

## r5d.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6486m	-Xmx6486m
mapreduce.java.opts	-Xmx12972m	-Xmx12972m
mapreduce.map.memory.mb	8107	8107
mapreduce.reduce.memory.mb	16214	16214
yarn.app.mapreduce.am.resource.mb	16214	16214
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	778240	32396
yarn.nodemanager.resource.memory-mb	778240	32396

## r5dn 实例

## r5dn.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## r5dn.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## r5dn.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## r5dn.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## r5dn.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

## r5dn.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

## r5dn.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.java.opts	-Xmx12356m	-Xmx12356m
mapreduce.map.memory.mb	7723	7723
mapreduce.reduce.memory.mb	15446	15446
yarn.app.mapreduce.am.resource.mb	15446	15446
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30860
yarn.nodemanager.resource.memory-mb	741376	30860

## r5n 实例

## r5n.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## r5n.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136



## r5n.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## r5n.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## r5n.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

## r5n.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

## r5n.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.java.opts	-Xmx12356m	-Xmx12356m
mapreduce.map.memory.mb	7723	7723
mapreduce.reduce.memory.mb	15446	15446
yarn.app.mapreduce.am.resource.mb	15446	15446

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30860
yarn.nodemanager.resource.memory-mb	741376	30860

## r6a 实例

## r6a.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## r6g.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## r6a.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## r6a.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## r6a.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

## r6a.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

## r6a.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.java.opts	-Xmx12356m	-Xmx12356m
mapreduce.map.memory.mb	7723	7723
mapreduce.reduce.memory.mb	15446	15446
yarn.app.mapreduce.am.resource.mb	15446	15446
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30860
yarn.nodemanager.resource.memory-mb	741376	30860



## r6a.32xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6195m	-Xmx6195m
mapreduce.java.opts	-Xmx12390m	-Xmx12390m
mapreduce.map.memory.mb	7744	7744
mapreduce.reduce.memory.mb	15488	15488
yarn.app.mapreduce.am.resource.mb	15488	15488
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	991232	30976
yarn.nodemanager.resource.memory-mb	991232	30976

## r6a.48xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6212m	-Xmx6212m
mapreduce.java.opts	-Xmx12424m	-Xmx12424m
mapreduce.map.memory.mb	7765	7765
mapreduce.reduce.memory.mb	15530	15530
yarn.app.mapreduce.am.resource.mb	15530	15530

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	1490944	31124
yarn.nodemanager.resource.memory-mb	1490944	31124

r6g 实例

r6g.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## r6g.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## r6g.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource-memory-mb	116736	29184

## r6g.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource-memory-mb	241664	30208

## r6g.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

## r6g.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

## r6gd 实例

## r6gd.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## r6gd.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## r6gd.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## r6gd.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208



## r6gd.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

## r6gd.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

r6i 实例

r6i.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## r6i.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## r6i.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## r6i.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## r6i.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

## r6i.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

## r6i.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.java.opts	-Xmx12356m	-Xmx12356m
mapreduce.map.memory.mb	7723	7723
mapreduce.reduce.memory.mb	15446	15446
yarn.app.mapreduce.am.resource.mb	15446	15446
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30860
yarn.nodemanager.resource.memory-mb	741376	30860

## r6i.32xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6029m	-Xmx6029m
mapreduce.java.opts	-Xmx12058m	-Xmx12058m
mapreduce.map.memory.mb	7536	7536
mapreduce.reduce.memory.mb	15072	15072
yarn.app.mapreduce.am.resource.mb	15072	15072
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	964608	30144
yarn.nodemanager.resource.memory-mb	964608	30144

## r6id 实例

## r6id.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## r6gd.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136



## r6id.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## r6id.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## r6id.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

## r6id.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

## r6id.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.java.opts	-Xmx12356m	-Xmx12356m
mapreduce.map.memory.mb	7723	7723
mapreduce.reduce.memory.mb	15446	15446
yarn.app.mapreduce.am.resource.mb	15446	15446

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30860
yarn.nodemanager.resource.memory-mb	741376	30860

## r6id.32xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6195m	-Xmx6195m
mapreduce.java.opts	-Xmx12390m	-Xmx12390m
mapreduce.map.memory.mb	7744	7744
mapreduce.reduce.memory.mb	15488	15488
yarn.app.mapreduce.am.resource.mb	15488	15488
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	991232	30976
yarn.nodemanager.resource.memory-mb	991232	30976

## r6idn 实例

## r6idn.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resou rce.mb	11712	11712
yarn.scheduler.minimum-allo cation-mb	32	32
yarn.scheduler.maximum-allo cation-mb	23424	11712
yarn.nodemanager.resource.m emory-mb	23424	11712

## r6idn.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resou rce.mb	13568	13568

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## r6idn.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## r6idn.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## r6idn.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

## r6idn.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720



## r6idn.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.java.opts	-Xmx12356m	-Xmx12356m
mapreduce.map.memory.mb	7723	7723
mapreduce.reduce.memory.mb	15446	15446
yarn.app.mapreduce.am.resource.mb	15446	15446
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30860
yarn.nodemanager.resource.memory-mb	741376	30860

## r6idn.32xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6195m	-Xmx6195m
mapreduce.java.opts	-Xmx12390m	-Xmx12390m
mapreduce.map.memory.mb	7744	7744
mapreduce.reduce.memory.mb	15488	15488
yarn.app.mapreduce.am.resource.mb	15488	15488

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	991232	30976
yarn.nodemanager.resource.memory-mb	991232	30976

r6in 实例

r6in.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## r6in.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## r6in.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## r6in.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## r6in.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

## r6in.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

## r6in.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.java.opts	-Xmx12356m	-Xmx12356m
mapreduce.map.memory.mb	7723	7723
mapreduce.reduce.memory.mb	15446	15446
yarn.app.mapreduce.am.resource.mb	15446	15446
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30860
yarn.nodemanager.resource.memory-mb	741376	30860

## r6in.32xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6195m	-Xmx6195m
mapreduce.java.opts	-Xmx12390m	-Xmx12390m
mapreduce.map.memory.mb	7744	7744
mapreduce.reduce.memory.mb	15488	15488
yarn.app.mapreduce.am.resource.mb	15488	15488
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	991232	30976
yarn.nodemanager.resource.memory-mb	991232	30976

## r7a 实例

## r7a.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## r7a.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136



## r7a.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## r7a.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## r7a.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

## r7a.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

## r7a.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.java.opts	-Xmx12356m	-Xmx12356m
mapreduce.map.memory.mb	7723	7723
mapreduce.reduce.memory.mb	15446	15446
yarn.app.mapreduce.am.resource.mb	15446	15446

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30860
yarn.nodemanager.resource-memory-mb	741376	30860

## r7a.32xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6195m	-Xmx6195m
mapreduce.java.opts	-Xmx12390m	-Xmx12390m
mapreduce.map.memory.mb	7744	7744
mapreduce.reduce.memory.mb	15488	15488
yarn.app.mapreduce.am.resource.mb	15488	15488
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	991232	30976
yarn.nodemanager.resource-memory-mb	991232	30976

## r7a.48xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6212m	-Xmx6212m
mapreduce.java.opts	-Xmx12424m	-Xmx12424m
mapreduce.map.memory.mb	7765	7765
mapreduce.reduce.memory.mb	15530	15530
yarn.app.mapreduce.am.resource.mb	15530	15530
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	1490944	31124
yarn.nodemanager.resource.memory-mb	1490944	31124

## r7g 实例

## r7g.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## r7g.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## r7g.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## r7g.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## r7g.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564



## r7g.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

## r7gd 实例

## r7gd.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## r7gd.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## r7gd.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## r7gd.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## r7gd.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

## r7gd.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

## r7i 实例

## r7i.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## r7i.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## r7i.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## r7i.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## r7i.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564



## r7i.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

## r7i.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.java.opts	-Xmx12356m	-Xmx12356m
mapreduce.map.memory.mb	7723	7723
mapreduce.reduce.memory.mb	15446	15446
yarn.app.mapreduce.am.resource.mb	15446	15446

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30860
yarn.nodemanager.resource.memory-mb	741376	30860

## r7i.48xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6212m	-Xmx6212m
mapreduce.java.opts	-Xmx12424m	-Xmx12424m
mapreduce.map.memory.mb	7765	7765
mapreduce.reduce.memory.mb	15530	15530
yarn.app.mapreduce.am.resource.mb	15530	15530
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	1490944	31124
yarn.nodemanager.resource.memory-mb	1490944	31124

## r7iz 实例

## r7iz.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

## r7iz.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

## r7iz.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## r7iz.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## r7iz.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

## r7iz.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

## r7iz.32xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx6029m	-Xmx6029m
mapreduce.java.opts	-Xmx12058m	-Xmx12058m
mapreduce.map.memory.mb	7536	7536
mapreduce.reduce.memory.mb	15072	15072
yarn.app.mapreduce.am.resource.mb	15072	15072
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	964608	30144
yarn.nodemanager.resource.memory-mb	964608	30144

## x1 实例

## x1.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx12058m	-Xmx12058m
mapreduce.java.opts	-Xmx24116m	-Xmx24116m
mapreduce.map.memory.mb	15072	15072
mapreduce.reduce.memory.mb	30144	30144
yarn.app.mapreduce.am.resource.mb	30144	30144

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	964608	30144
yarn.nodemanager.resource.memory-mb	964608	30144

## x1.32xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx12109m	-Xmx12109m
mapreduce.java.opts	-Xmx24218m	-Xmx24218m
mapreduce.map.memory.mb	15136	15136
mapreduce.reduce.memory.mb	30272	30272
yarn.app.mapreduce.am.resource.mb	30272	30272
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	1937408	30272
yarn.nodemanager.resource.memory-mb	1937408	30272



## x1e 实例

## x1e.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx22682m	-Xmx22682m
mapreduce.java.opts	-Xmx45364m	-Xmx45364m
mapreduce.map.memory.mb	28352	28352
mapreduce.reduce.memory.mb	56704	56704
yarn.app.mapreduce.am.resource.mb	56704	56704
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	113408	0
yarn.nodemanager.resource.memory-mb	113408	0

## x1e.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx23501m	-Xmx23501m
mapreduce.java.opts	-Xmx47002m	-Xmx47002m
mapreduce.map.memory.mb	29376	29376
mapreduce.reduce.memory.mb	58752	58752
yarn.app.mapreduce.am.resource.mb	58752	58752

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	235008	0
yarn.nodemanager.resource.memory-mb	235008	0

## x1e.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx23910m	-Xmx23910m
mapreduce.java.opts	-Xmx47820m	-Xmx47820m
mapreduce.map.memory.mb	29888	29888
mapreduce.reduce.memory.mb	59776	59776
yarn.app.mapreduce.am.resource.mb	59776	59776
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	478208	0
yarn.nodemanager.resource.memory-mb	478208	0

## x1e.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx24115m	-Xmx24115m
mapreduce.java.opts	-Xmx48230m	-Xmx48230m
mapreduce.map.memory.mb	30144	30144
mapreduce.reduce.memory.mb	60288	60288
yarn.app.mapreduce.am.resource.mb	60288	60288
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	964608	0
yarn.nodemanager.resource.memory-mb	964608	0

## x1e.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx24218m	-Xmx24218m
mapreduce.java.opts	-Xmx48436m	-Xmx48436m
mapreduce.map.memory.mb	30272	30272
mapreduce.reduce.memory.mb	60544	60544
yarn.app.mapreduce.am.resource.mb	60544	60544

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	1937408	0
yarn.nodemanager.resource.memory-mb	1937408	0

## x1e.32xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx24269m	-Xmx24269m
mapreduce.java.opts	-Xmx48538m	-Xmx48538m
mapreduce.map.memory.mb	30336	30336
mapreduce.reduce.memory.mb	60672	60672
yarn.app.mapreduce.am.resource.mb	60672	60672
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	3883008	0
yarn.nodemanager.resource.memory-mb	3883008	0

## x2gd 实例

## x2gd.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.java.opts	-Xmx21708m	-Xmx21708m
mapreduce.map.memory.mb	13568	13568
mapreduce.reduce.memory.mb	27136	27136
yarn.app.mapreduce.am.resou rce.mb	27136	27136
yarn.scheduler.minimum-allo cation-mb	32	32
yarn.scheduler.maximum-allo cation-mb	54272	27136
yarn.nodemanager.resource.m emory-mb	54272	27136

## x2gd.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.java.opts	-Xmx23348m	-Xmx23348m
mapreduce.map.memory.mb	14592	14592
mapreduce.reduce.memory.mb	29184	29184
yarn.app.mapreduce.am.resou rce.mb	29184	29184

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

## x2gd.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx12083m	-Xmx12083m
mapreduce.java.opts	-Xmx24166m	-Xmx24166m
mapreduce.map.memory.mb	15104	15104
mapreduce.reduce.memory.mb	30208	30208
yarn.app.mapreduce.am.resource.mb	30208	30208
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

## x2gd.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.java.opts	-Xmx24576m	-Xmx24576m
mapreduce.map.memory.mb	15360	15360
mapreduce.reduce.memory.mb	30720	30720
yarn.app.mapreduce.am.resource.mb	30720	30720
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

## x2gd.12xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx12356m	-Xmx12356m
mapreduce.java.opts	-Xmx24712m	-Xmx24712m
mapreduce.map.memory.mb	15445	15445
mapreduce.reduce.memory.mb	30890	30890
yarn.app.mapreduce.am.resource.mb	30890	30890

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30906
yarn.nodemanager.resource.memory-mb	741376	30906

## x2gd.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx12390m	-Xmx12390m
mapreduce.java.opts	-Xmx24780m	-Xmx24780m
mapreduce.map.memory.mb	15488	15488
mapreduce.reduce.memory.mb	30976	30976
yarn.app.mapreduce.am.resource.mb	30976	30976
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	991232	30976
yarn.nodemanager.resource.memory-mb	991232	30976



## x2idn 实例

## x2idn.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx12390m	-Xmx12390m
mapreduce.java.opts	-Xmx24780m	-Xmx24780m
mapreduce.map.memory.mb	15488	15488
mapreduce.reduce.memory.mb	30976	30976
yarn.app.mapreduce.am.resource.mb	30976	30976
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	991232	30976
yarn.nodemanager.resource.memory-mb	991232	30976

## x2idn.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx12425m	-Xmx12425m
mapreduce.java.opts	-Xmx24850m	-Xmx24850m
mapreduce.map.memory.mb	15531	15531
mapreduce.reduce.memory.mb	31062	31062
yarn.app.mapreduce.am.resource.mb	31062	31062

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	1490944	31030
yarn.nodemanager.resource.memory-mb	1490944	31030

## x2idn.32xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx12442m	-Xmx12442m
mapreduce.java.opts	-Xmx24884m	-Xmx24884m
mapreduce.map.memory.mb	15552	15552
mapreduce.reduce.memory.mb	311104	311104
yarn.app.mapreduce.am.resource.mb	311104	311104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	1990656	311104
yarn.nodemanager.resource.memory-mb	1990656	311104

## x2iedn 实例

## x2iedn.xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx23347m	-Xmx23347m
mapreduce.java.opts	-Xmx46694m	-Xmx46694m
mapreduce.map.memory.mb	29184	29184
mapreduce.reduce.memory.mb	58368	58368
yarn.app.mapreduce.am.resource.mb	58368	58368
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	0
yarn.nodemanager.resource.memory-mb	116736	0

## x2iedn.2xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx24166m	-Xmx24166m
mapreduce.java.opts	-Xmx48332m	-Xmx48332m
mapreduce.map.memory.mb	30208	30208
mapreduce.reduce.memory.mb	60416	60416
yarn.app.mapreduce.am.resource.mb	60416	60416

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	0
yarn.nodemanager.resource.memory-mb	241664	0

## x2iedn.4xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx24576m	-Xmx24576m
mapreduce.java.opts	-Xmx49152m	-Xmx49152m
mapreduce.map.memory.mb	30720	30720
mapreduce.reduce.memory.mb	61440	61440
yarn.app.mapreduce.am.resource.mb	61440	61440
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	0
yarn.nodemanager.resource.memory-mb	491520	0

## x2iedn.8xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx24781m	-Xmx24781m
mapreduce.java.opts	-Xmx49562m	-Xmx49562m
mapreduce.map.memory.mb	30976	30976
mapreduce.reduce.memory.mb	61952	61952
yarn.app.mapreduce.am.resource.mb	61952	61952
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	991232	0
yarn.nodemanager.resource.memory-mb	991232	0

## x2iedn.16xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx24883m	-Xmx24883m
mapreduce.java.opts	-Xmx49766m	-Xmx49766m
mapreduce.map.memory.mb	311104	311104
mapreduce.reduce.memory.mb	62208	62208
yarn.app.mapreduce.am.resource.mb	62208	62208

配置选项	默认值	HBase安装后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	1990656	0
yarn.nodemanager.resource.memory-mb	1990656	0

## x2iedn.24xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx24918m	-Xmx24918m
mapreduce.java.opts	-Xmx49836m	-Xmx49836m
mapreduce.map.memory.mb	31147	31147
mapreduce.reduce.memory.mb	62294	62294
yarn.app.mapreduce.am.resource.mb	62294	62294
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	2990080	-32
yarn.nodemanager.resource.memory-mb	2990080	-32

## x2iedn.32xlarge

配置选项	默认值	HBase安装后
mapreduce.map.java.opts	-Xmx24934m	-Xmx24934m
mapreduce.java.opts	-Xmx49868m	-Xmx49868m
mapreduce.map.memory.mb	31168	31168
mapreduce.reduce.memory.mb	62336	62336
yarn.app.mapreduce.am.resource.mb	62336	62336
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	3989504	0
yarn.nodemanager.resource.memory-mb	3989504	0

## z1d 实例

## z1d.xlarge

配置选项	默认值	HBase安装完毕后
mapreduce.map.java.opts	-Xmx4915m	-Xmx4915m
mapreduce.java.opts	-Xmx9830m	-Xmx9830m
mapreduce.map.memory.mb	6144	6144
mapreduce.reduce.memory.mb	12288	12288
yarn.app.mapreduce.am.resource.mb	12288	12288

配置选项	默认值	HBase安装完毕后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24 576	12288
yarn.nodemanager.resource.memory-mb	24 576	12288

## z1d.2xlarge

配置选项	默认值	HBase安装完毕后
mapreduce.map.java.opts	-Xmx5734m	-Xmx5734m
mapreduce.java.opts	-Xmx11468m	-Xmx11468m
mapreduce.map.memory.mb	7168	7168
mapreduce.reduce.memory.mb	14336	14336
yarn.app.mapreduce.am.resource.mb	14336	14336
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	57344	28672
yarn.nodemanager.resource.memory-mb	57344	28672



## z1d.3xlarge

配置选项	默认值	HBase安装完毕后
mapreduce.map.java.opts	-Xmx6007m	-Xmx6007m
mapreduce.java.opts	-Xmx12014m	-Xmx12014m
mapreduce.map.memory.mb	7509	7509
mapreduce.reduce.memory.mb	15018	15018
yarn.app.mapreduce.am.resource.mb	15018	15018
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	90112	30040
yarn.nodemanager.resource.memory-mb	90112	30040

## z1d.6xlarge

配置选项	默认值	HBase安装完毕后
mapreduce.map.java.opts	-Xmx6281m	-Xmx6281m
mapreduce.java.opts	-Xmx12562m	-Xmx12562m
mapreduce.map.memory.mb	7851	7851
mapreduce.reduce.memory.mb	15702	15702
yarn.app.mapreduce.am.resource.mb	15702	15702

配置选项	默认值	HBase安装完毕后
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	188416	31396
yarn.nodemanager.resource.memory-mb	188416	31396

## z1d.12xlarge

配置选项	默认值	HBase安装完毕后
mapreduce.map.java.opts	-Xmx6417m	-Xmx6417m
mapreduce.java.opts	-Xmx12834m	-Xmx12834m
mapreduce.map.memory.mb	8021	8021
mapreduce.reduce.memory.mb	16042	16042
yarn.app.mapreduce.am.resource.mb	16042	16042
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	385024	32100
yarn.nodemanager.resource.memory-mb	385024	32100

## Hadoop 守护进程配置设置

Hadoop 守护进程设置因集群节点使用的 EC2 实例类型而异。以下表格列出了每种 EC2 实例类型的默认配置设置。

要自定义这些设置，请使用 `hadoop-env` 配置分类。有关更多信息，请参阅[配置应用程序](#)。

### 实例类型

- [c1 实例](#)
- [c3 实例](#)
- [c4 实例](#)
- [c5 实例](#)
- [c5a 实例](#)
- [c5ad 实例](#)
- [c5d 实例](#)
- [c5n 实例](#)
- [c6a 实例](#)
- [c6g 实例](#)
- [c6gd 实例](#)
- [c6gn 实例](#)
- [c6i 实例](#)
- [c6id 实例](#)
- [c6in 实例](#)
- [c7a 实例](#)
- [c7g 实例](#)
- [c7gd 实例](#)
- [c7gn 实例](#)
- [c7i 实例](#)
- [d2 实例](#)
- [d3 实例](#)
- [d3en 实例](#)

- [g3 实例](#)
- [g3s 实例](#)
- [g4dn 实例](#)
- [g5 实例](#)
- [h1 实例](#)
- [i2 实例](#)
- [i3 实例](#)
- [i3en 实例](#)
- [i4g 实例](#)
- [i4i 实例](#)
- [im4gn 实例](#)
- [is4gen 实例](#)
- [m1 实例](#)
- [m2 实例](#)
- [m3 实例](#)
- [m4 实例](#)
- [m5 实例](#)
- [m5a 实例](#)
- [m5ad 实例](#)
- [m5d 实例](#)
- [m5dn 实例](#)
- [m5n 实例](#)
- [m5zn 实例](#)
- [m6a 实例](#)
- [m6g 实例](#)
- [m6gd 实例](#)
- [m6i 实例](#)
- [m6id 实例](#)
- [m6idn 实例](#)

- [m6in 实例](#)
- [m7a 实例](#)
- [m7g 实例](#)
- [m7gd 实例](#)
- [m7i 实例](#)
- [m7i-flex 实例](#)
- [p2 实例](#)
- [p3 实例](#)
- [p5 实例](#)
- [r3 实例](#)
- [r4 实例](#)
- [r5 实例](#)
- [r5a 实例](#)
- [r5ad 实例](#)
- [r5b 实例](#)
- [r5d 实例](#)
- [r5dn 实例](#)
- [r5n 实例](#)
- [r6a 实例](#)
- [r6g 实例](#)
- [r6gd 实例](#)
- [r6i 实例](#)
- [r6id 实例](#)
- [r6idn 实例](#)
- [r6in 实例](#)
- [r7a 实例](#)
- [r7g 实例](#)
- [r7gd 实例](#)
- [r7i 实例](#)

- [r7iz 实例](#)
- [x1 实例](#)
- [x1e 实例](#)
- [x2gd 实例](#)
- [x2idn 实例](#)
- [x2iedn 实例](#)
- [z1d 实例](#)

## c1 实例

### c1.medium

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	192
YARN_PROXYSERVER_HEAPSIZE	96
YARN_NODEMANAGER_HEAPSIZE	128
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	128
HADOOP_NAMENODE_HEAPSIZE	192
HADOOP_DATANODE_HEAPSIZE	96

### c1.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	768
YARN_PROXYSERVER_HEAPSIZE	384
YARN_NODEMANAGER_HEAPSIZE	512

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	512
HADOOP_NAMENODE_HEAPSIZE	768
HADOOP_DATANODE_HEAPSIZE	384

## c3 实例

### c3.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

### c3.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2396
YARN_PROXYSERVER_HEAPSIZE	2396
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2396
HADOOP_NAMENODE_HEAPSIZE	1740
HADOOP_DATANODE_HEAPSIZE	757

### c3.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2703
YARN_PROXYSERVER_HEAPSIZE	2703
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2703
HADOOP_NAMENODE_HEAPSIZE	3276
HADOOP_DATANODE_HEAPSIZE	1064

### c3.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3317
YARN_PROXYSERVER_HEAPSIZE	3317
YARN_NODEMANAGER_HEAPSIZE	2048



参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3317
HADOOP_NAMENODE_HEAPSIZE	6348
HADOOP_DATANODE_HEAPSIZE	1679

## c4 实例

### c4.large

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	1152
YARN_PROXYSERVER_HEAPSIZE	1152
YARN_NODEMANAGER_HEAPSIZE	1152
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	1152
HADOOP_NAMENODE_HEAPSIZE	576
HADOOP_DATANODE_HEAPSIZE	384

### c4.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

## c4.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2396
YARN_PROXYSERVER_HEAPSIZE	2396
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2396
HADOOP_NAMENODE_HEAPSIZE	1740
HADOOP_DATANODE_HEAPSIZE	757

## c4.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2703
YARN_PROXYSERVER_HEAPSIZE	2703
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2703
HADOOP_NAMENODE_HEAPSIZE	3276
HADOOP_DATANODE_HEAPSIZE	1064

### c4.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3317
YARN_PROXYSERVER_HEAPSIZE	3317
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3317
HADOOP_NAMENODE_HEAPSIZE	6348
HADOOP_DATANODE_HEAPSIZE	1679

### c5 实例

#### c5.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2252
YARN_PROXYSERVER_HEAPSIZE	2252
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2252
HADOOP_NAMENODE_HEAPSIZE	1024
HADOOP_DATANODE_HEAPSIZE	614

## c5.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2416
YARN_PROXYSERVER_HEAPSIZE	2416
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2416
HADOOP_NAMENODE_HEAPSIZE	1843
HADOOP_DATANODE_HEAPSIZE	778

## c5.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

## c5.9xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3563
YARN_PROXYSERVER_HEAPSIZE	3563
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3563
HADOOP_NAMENODE_HEAPSIZE	7577
HADOOP_DATANODE_HEAPSIZE	1925

## c5.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4055
YARN_PROXYSERVER_HEAPSIZE	4055
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4055
HADOOP_NAMENODE_HEAPSIZE	10035
HADOOP_DATANODE_HEAPSIZE	2416

### c5.18xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	5038
YARN_PROXYSERVER_HEAPSIZE	5038
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	5038
HADOOP_NAMENODE_HEAPSIZE	14950
HADOOP_DATANODE_HEAPSIZE	3399

### c5.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	6021
YARN_PROXYSERVER_HEAPSIZE	6021
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	6021
HADOOP_NAMENODE_HEAPSIZE	19865
HADOOP_DATANODE_HEAPSIZE	4096

## c5a 实例

### c5a.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

### c5a.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

## c5a.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## c5a.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048



参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## c5a.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4055
YARN_PROXYSERVER_HEAPSIZE	4055
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4055
HADOOP_NAMENODE_HEAPSIZE	10035
HADOOP_DATANODE_HEAPSIZE	2416

## c5a.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

### c5a.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

### c5ad 实例

#### c5ad.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

## c5ad.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

## c5ad.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## c5ad.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## c5ad.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

## c5ad.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## c5ad.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

## c5d 实例

### c5d.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2252
YARN_PROXYSERVER_HEAPSIZE	2252
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2252
HADOOP_NAMENODE_HEAPSIZE	1024
HADOOP_DATANODE_HEAPSIZE	614

### c5d.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2416
YARN_PROXYSERVER_HEAPSIZE	2416
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2416
HADOOP_NAMENODE_HEAPSIZE	1843
HADOOP_DATANODE_HEAPSIZE	778

## c5d.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

## c5d.9xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3563
YARN_PROXYSERVER_HEAPSIZE	3563
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3563
HADOOP_NAMENODE_HEAPSIZE	7577
HADOOP_DATANODE_HEAPSIZE	1925

## c5d.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4055
YARN_PROXYSERVER_HEAPSIZE	4055
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4055
HADOOP_NAMENODE_HEAPSIZE	10035
HADOOP_DATANODE_HEAPSIZE	2416

## c5d.18xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	5038
YARN_PROXYSERVER_HEAPSIZE	5038
YARN_NODEMANAGER_HEAPSIZE	2048



参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	5038
HADOOP_NAMENODE_HEAPSIZE	14950
HADOOP_DATANODE_HEAPSIZE	3399

### c5d.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	6021
YARN_PROXYSERVER_HEAPSIZE	6021
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	6021
HADOOP_NAMENODE_HEAPSIZE	19865
HADOOP_DATANODE_HEAPSIZE	4096

### c5n 实例

#### c5n.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2304
YARN_PROXYSERVER_HEAPSIZE	2304
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2304
HADOOP_NAMENODE_HEAPSIZE	1280
HADOOP_DATANODE_HEAPSIZE	665

## c5n.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2519
YARN_PROXYSERVER_HEAPSIZE	2519
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2519
HADOOP_NAMENODE_HEAPSIZE	2355
HADOOP_DATANODE_HEAPSIZE	880

## c5n.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2949
YARN_PROXYSERVER_HEAPSIZE	2949
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2949
HADOOP_NAMENODE_HEAPSIZE	4505
HADOOP_DATANODE_HEAPSIZE	1310

## c5n.9xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4055
YARN_PROXYSERVER_HEAPSIZE	4055
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4055
HADOOP_NAMENODE_HEAPSIZE	10035
HADOOP_DATANODE_HEAPSIZE	2416

## c5n.18xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	6021
YARN_PROXYSERVER_HEAPSIZE	6021
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	6021
HADOOP_NAMENODE_HEAPSIZE	19865
HADOOP_DATANODE_HEAPSIZE	4096

## c6a 实例

### c6a.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

### c6a.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

## c6a.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## c6a.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## c6a.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

## c6a.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## c6a.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

## c6a.32xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

### c6a.48xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

### c6g 实例

#### c6g.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048



参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

## c6g.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

## c6g.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## c6g.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## c6g.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

### c6g.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

### c6gd 实例

#### c6gd.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

## c6gd.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

## c6gd.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## c6gd.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## c6gd.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

### c6gd.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

### c6gn 实例

#### c6gn.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

## c6gn.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

## c6gn.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## c6gn.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## c6gn.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048



参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

### c6gn.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

### c6i 实例

#### c6i.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

## c6i.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

## c6i.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## c6i.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## c6i.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

## c6i.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## c6i.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

### c6i.32xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

### c6id 实例

#### c6id.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

## c6id.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

## c6id.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## c6id.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## c6id.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

## c6id.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## c6id.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048



参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

### c6id.32xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

### c6in 实例

#### c6in.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

## c6in.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

## c6in.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## c6in.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## c6in.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

## c6in.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## c6in.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

### c6in.32xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

### c7a 实例

#### c7a.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

## c7a.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

## c7a.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

### c7a.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

### c7a.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

### c7a.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

### c7a.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048



参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

## c7a.32xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## c7a.48xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## c7g 实例

### c7g.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

### c7g.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

## c7g.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## c7g.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## c7g.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

## c7g.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## c7gd 实例

### c7gd.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

### c7gd.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

## c7gd.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## c7gd.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## c7gd.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

## c7gd.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## c7gn 实例

### c7gn.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

### c7gn.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048



参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

## c7gn.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## c7gn.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## c7gn.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

## c7gn.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## c7i 实例

### c7i.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

### c7i.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

### c7i.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

### c7i.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## c7i.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

## c7i.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

### c7i.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

### c7i.48xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## d2 实例

### d2.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

### d2.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## d2.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## d2.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048



参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## d3 实例

### d3.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

### d3.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

### d3.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

### d3.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## d3en 实例

### d3en.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

### d3en.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## d3en.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## d3en.6xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

## d3en.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## d3en.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

## g3 实例

### g3.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

### g3.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

### g3.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

### g3s 实例

#### g3s.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## g4dn 实例

### g4dn.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2416
YARN_PROXYSERVER_HEAPSIZE	2416
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2416
HADOOP_NAMENODE_HEAPSIZE	1843
HADOOP_DATANODE_HEAPSIZE	778

### g4dn.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048



参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

## g4dn.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

## g4dn.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4710
YARN_PROXYSERVER_HEAPSIZE	4710
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4710
HADOOP_NAMENODE_HEAPSIZE	13312
HADOOP_DATANODE_HEAPSIZE	3072

## g4dn.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	6021
YARN_PROXYSERVER_HEAPSIZE	6021
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	6021
HADOOP_NAMENODE_HEAPSIZE	19865
HADOOP_DATANODE_HEAPSIZE	4096

## g4dn.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7331
YARN_PROXYSERVER_HEAPSIZE	7331
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7331
HADOOP_NAMENODE_HEAPSIZE	26419
HADOOP_DATANODE_HEAPSIZE	4096

## g5 实例

### g5.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

### g5.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## g5.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## g5.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## g5.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

## g5.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## g5.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## g5.48xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

## h1 实例

### h1.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

### h1.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

## h1.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4710
YARN_PROXYSERVER_HEAPSIZE	4710
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4710
HADOOP_NAMENODE_HEAPSIZE	13312
HADOOP_DATANODE_HEAPSIZE	3072

## h1.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7331
YARN_PROXYSERVER_HEAPSIZE	7331
YARN_NODEMANAGER_HEAPSIZE	2048



参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7331
HADOOP_NAMENODE_HEAPSIZE	26419
HADOOP_DATANODE_HEAPSIZE	4096

## i2 实例

### i2.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

### i2.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

### i2.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

### i2.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## i3 实例

### i3.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

### i3.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

### i3.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

### i3.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

### i3.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

### i3en 实例

#### i3en.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

## i3en.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

## i3en.3xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4055
YARN_PROXYSERVER_HEAPSIZE	4055
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4055
HADOOP_NAMENODE_HEAPSIZE	10035
HADOOP_DATANODE_HEAPSIZE	2416

## i3en.6xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	6021
YARN_PROXYSERVER_HEAPSIZE	6021
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	6021
HADOOP_NAMENODE_HEAPSIZE	19865
HADOOP_DATANODE_HEAPSIZE	4096

## i3en.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9953
YARN_PROXYSERVER_HEAPSIZE	9953
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9953
HADOOP_NAMENODE_HEAPSIZE	39526
HADOOP_DATANODE_HEAPSIZE	4096

### i3en.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	17817
YARN_PROXYSERVER_HEAPSIZE	17817
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	17817
HADOOP_NAMENODE_HEAPSIZE	78848
HADOOP_DATANODE_HEAPSIZE	4096

### i4g 实例

#### i4g.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048



参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## i4g.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## i4g.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## i4g.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## i4g.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## i4i 实例

### i4i.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

### i4i.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

#### i4i.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

#### i4i.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## i4i.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## i4i.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## i4i.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

## i4i.32xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	22077
YARN_PROXYSERVER_HEAPSIZE	22077
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	22077
HADOOP_NAMENODE_HEAPSIZE	100147
HADOOP_DATANODE_HEAPSIZE	4096

## im4gn 实例

### im4gn.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

### im4gn.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## im4gn.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## im4gn.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048



参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

### im4gn.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

### is4gen 实例

#### is4gen.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2557
YARN_PROXYSERVER_HEAPSIZE	2557
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2557
HADOOP_NAMENODE_HEAPSIZE	2547
HADOOP_DATANODE_HEAPSIZE	919

## is4gen.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3025
YARN_PROXYSERVER_HEAPSIZE	3025
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3025
HADOOP_NAMENODE_HEAPSIZE	4889
HADOOP_DATANODE_HEAPSIZE	1387

## is4gen.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

### is4gen.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

### m1 实例

#### m1.small

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	256
YARN_PROXYSERVER_HEAPSIZE	96
YARN_NODEMANAGER_HEAPSIZE	192

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	128
HADOOP_NAMENODE_HEAPSIZE	192
HADOOP_DATANODE_HEAPSIZE	96

## m1.medium

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	384
YARN_PROXYSERVER_HEAPSIZE	192
YARN_NODEMANAGER_HEAPSIZE	256
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	256
HADOOP_NAMENODE_HEAPSIZE	384
HADOOP_DATANODE_HEAPSIZE	192

## m1.large

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	768
YARN_PROXYSERVER_HEAPSIZE	384
YARN_NODEMANAGER_HEAPSIZE	512

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	512
HADOOP_NAMENODE_HEAPSIZE	768
HADOOP_DATANODE_HEAPSIZE	384

## m1.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	1024
YARN_PROXYSERVER_HEAPSIZE	512
YARN_NODEMANAGER_HEAPSIZE	768
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	1024
HADOOP_NAMENODE_HEAPSIZE	2304
HADOOP_DATANODE_HEAPSIZE	384

## m2 实例

### m2.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	1536
YARN_PROXYSERVER_HEAPSIZE	1024
YARN_NODEMANAGER_HEAPSIZE	1024

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	1024
HADOOP_NAMENODE_HEAPSIZE	3072
HADOOP_DATANODE_HEAPSIZE	384

### m2.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	1536
YARN_PROXYSERVER_HEAPSIZE	1024
YARN_NODEMANAGER_HEAPSIZE	1024
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	1536
HADOOP_NAMENODE_HEAPSIZE	6144
HADOOP_DATANODE_HEAPSIZE	384

### m2.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2048
YARN_PROXYSERVER_HEAPSIZE	1024
YARN_NODEMANAGER_HEAPSIZE	1536

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	1536
HADOOP_NAMENODE_HEAPSIZE	12288
HADOOP_DATANODE_HEAPSIZE	384

## m3 实例

### m3.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2396
YARN_PROXYSERVER_HEAPSIZE	2396
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2396
HADOOP_NAMENODE_HEAPSIZE	1740
HADOOP_DATANODE_HEAPSIZE	757

### m3.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2703
YARN_PROXYSERVER_HEAPSIZE	2703
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2703
HADOOP_NAMENODE_HEAPSIZE	3276
HADOOP_DATANODE_HEAPSIZE	1064

## m4 实例

### m4.large

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2252
YARN_PROXYSERVER_HEAPSIZE	2252
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2252
HADOOP_NAMENODE_HEAPSIZE	1024
HADOOP_DATANODE_HEAPSIZE	614

### m4.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2416
YARN_PROXYSERVER_HEAPSIZE	2416
YARN_NODEMANAGER_HEAPSIZE	2048



参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2416
HADOOP_NAMENODE_HEAPSIZE	1843
HADOOP_DATANODE_HEAPSIZE	778

#### m4.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

#### m4.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

#### m4.10xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	5365
YARN_PROXYSERVER_HEAPSIZE	5365
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	5365
HADOOP_NAMENODE_HEAPSIZE	16588
HADOOP_DATANODE_HEAPSIZE	3727

#### m4.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7331
YARN_PROXYSERVER_HEAPSIZE	7331
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7331
HADOOP_NAMENODE_HEAPSIZE	26419
HADOOP_DATANODE_HEAPSIZE	4096

## m5 实例

### m5.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2416
YARN_PROXYSERVER_HEAPSIZE	2416
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2416
HADOOP_NAMENODE_HEAPSIZE	1843
HADOOP_DATANODE_HEAPSIZE	778

### m5.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

### m5.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

### m5.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4710
YARN_PROXYSERVER_HEAPSIZE	4710
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4710
HADOOP_NAMENODE_HEAPSIZE	13312
HADOOP_DATANODE_HEAPSIZE	3072

## m5.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	6021
YARN_PROXYSERVER_HEAPSIZE	6021
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	6021
HADOOP_NAMENODE_HEAPSIZE	19865
HADOOP_DATANODE_HEAPSIZE	4096

## m5.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7331
YARN_PROXYSERVER_HEAPSIZE	7331
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7331
HADOOP_NAMENODE_HEAPSIZE	26419
HADOOP_DATANODE_HEAPSIZE	4096

### m5.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9953
YARN_PROXYSERVER_HEAPSIZE	9953
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9953
HADOOP_NAMENODE_HEAPSIZE	39526
HADOOP_DATANODE_HEAPSIZE	4096

### m5a 实例

#### m5a.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2416
YARN_PROXYSERVER_HEAPSIZE	2416
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2416
HADOOP_NAMENODE_HEAPSIZE	1843
HADOOP_DATANODE_HEAPSIZE	778

### m5a.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

### m5a.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

## m5a.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4710
YARN_PROXYSERVER_HEAPSIZE	4710
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4710
HADOOP_NAMENODE_HEAPSIZE	13312
HADOOP_DATANODE_HEAPSIZE	3072

## m5a.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	6021
YARN_PROXYSERVER_HEAPSIZE	6021
YARN_NODEMANAGER_HEAPSIZE	2048



参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	6021
HADOOP_NAMENODE_HEAPSIZE	19865
HADOOP_DATANODE_HEAPSIZE	4096

## m5a.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7331
YARN_PROXYSERVER_HEAPSIZE	7331
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7331
HADOOP_NAMENODE_HEAPSIZE	26419
HADOOP_DATANODE_HEAPSIZE	4096

## m5a.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9953
YARN_PROXYSERVER_HEAPSIZE	9953
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9953
HADOOP_NAMENODE_HEAPSIZE	39526
HADOOP_DATANODE_HEAPSIZE	4096

## m5ad 实例

### m5ad.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

### m5ad.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## m5ad.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## m5ad.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## m5ad.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

## m5ad.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

### m5ad.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

### m5d 实例

#### m5d.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2416
YARN_PROXYSERVER_HEAPSIZE	2416
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2416
HADOOP_NAMENODE_HEAPSIZE	1843
HADOOP_DATANODE_HEAPSIZE	778

## m5d.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

## m5d.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

## m5d.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4710
YARN_PROXYSERVER_HEAPSIZE	4710
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4710
HADOOP_NAMENODE_HEAPSIZE	13312
HADOOP_DATANODE_HEAPSIZE	3072

## m5d.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	6021
YARN_PROXYSERVER_HEAPSIZE	6021
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	6021
HADOOP_NAMENODE_HEAPSIZE	19865
HADOOP_DATANODE_HEAPSIZE	4096

## m5d.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7331
YARN_PROXYSERVER_HEAPSIZE	7331
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7331
HADOOP_NAMENODE_HEAPSIZE	26419
HADOOP_DATANODE_HEAPSIZE	4096

## m5d.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9953
YARN_PROXYSERVER_HEAPSIZE	9953
YARN_NODEMANAGER_HEAPSIZE	2048



参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9953
HADOOP_NAMENODE_HEAPSIZE	39526
HADOOP_DATANODE_HEAPSIZE	4096

## m5dn 实例

### m5dn.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

### m5dn.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## m5dn.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## m5dn.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## m5dn.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

## m5dn.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

### m5dn.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

### m5n 实例

#### m5n.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

## m5n.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## m5n.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## m5n.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## m5n.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

## m5n.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## m5n.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## m5zn 实例

### m5zn.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2396
YARN_PROXYSERVER_HEAPSIZE	2396
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2396
HADOOP_NAMENODE_HEAPSIZE	1740
HADOOP_DATANODE_HEAPSIZE	757

### m5zn.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048



参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## m5zn.3xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3025
YARN_PROXYSERVER_HEAPSIZE	3025
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3025
HADOOP_NAMENODE_HEAPSIZE	4889
HADOOP_DATANODE_HEAPSIZE	1387

## m5zn.6xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

### m5zn.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

### m6a 实例

#### m6a.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

### m6a.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

### m6a.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## m6a.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## m6a.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

## m6a.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## m6a.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## m6a.32xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## m6a.48xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

## m6g 实例

### m6g.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

### m6g.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## m6g.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## m6g.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048



参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## m6g.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	5877
YARN_PROXYSERVER_HEAPSIZE	5877
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	5877
HADOOP_NAMENODE_HEAPSIZE	19148
HADOOP_DATANODE_HEAPSIZE	4096

## m6g.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## m6gd 实例

### m6gd.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

### m6gd.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## m6gd.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## m6gd.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## m6gd.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	5877
YARN_PROXYSERVER_HEAPSIZE	5877
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	5877
HADOOP_NAMENODE_HEAPSIZE	19148
HADOOP_DATANODE_HEAPSIZE	4096

## m6gd.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## m6i 实例

### m6i.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

### m6i.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## m6i.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## m6i.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## m6i.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	5877
YARN_PROXYSERVER_HEAPSIZE	5877
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	5877
HADOOP_NAMENODE_HEAPSIZE	19148
HADOOP_DATANODE_HEAPSIZE	4096

## m6i.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## m6i.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## m6i.32xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048



参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## m6id 实例

### m6id.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

### m6id.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## m6id.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## m6id.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## m6id.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

## m6id.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## m6id.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## m6id.32xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## m6idn 实例

### m6idn.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

### m6idn.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## m6idn.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## m6idn.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## m6idn.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

## m6idn.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## m6idn.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## m6idn.32xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048



参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## m6in 实例

### m6in.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

### m6in.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## m6in.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## m6in.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## m6in.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	5877
YARN_PROXYSERVER_HEAPSIZE	5877
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	5877
HADOOP_NAMENODE_HEAPSIZE	19148
HADOOP_DATANODE_HEAPSIZE	4096

## m6in.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## m6in.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## m6in.32xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## m7a 实例

### m7a.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

### m7a.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

### m7a.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

### m7a.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## m7a.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

## m7a.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## m7a.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## m7a.32xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048



参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

### m7a.48xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

### m7g 实例

#### m7g.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

## m7g.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## m7g.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## m7g.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## m7g.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

### m7g.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

### m7gd 实例

#### m7gd.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

## m7gd.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## m7gd.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## m7gd.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## m7gd.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

### m7gd.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

### m7i 实例

#### m7i.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

## m7i.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## m7i.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048



参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## m7i.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## m7i.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	5877
YARN_PROXYSERVER_HEAPSIZE	5877
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	5877
HADOOP_NAMENODE_HEAPSIZE	19148
HADOOP_DATANODE_HEAPSIZE	4096

## m7i.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## m7i.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

### m7i.48xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

### m7i-flex 实例

#### m7i-flex.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

## m7i-flex.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## m7i-flex.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

### m7i-flex.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

### p2 实例

#### p2.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## p2.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## p2.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

## p3 实例

### p3.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

### p3.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

### p3.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

### p5 实例

#### p5.48xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	42065
YARN_PROXYSERVER_HEAPSIZE	42065
YARN_NODEMANAGER_HEAPSIZE	2048



参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	42065
HADOOP_NAMENODE_HEAPSIZE	200089
HADOOP_DATANODE_HEAPSIZE	4096

## r3 实例

### r3.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

### r3.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

### r3.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

### r3.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## r4 实例

### r4.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

### r4.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

#### r4.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

#### r4.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

#### r4.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

#### r5 实例

##### r5.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

## r5.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

## r5.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4710
YARN_PROXYSERVER_HEAPSIZE	4710
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4710
HADOOP_NAMENODE_HEAPSIZE	13312
HADOOP_DATANODE_HEAPSIZE	3072

## r5.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7331
YARN_PROXYSERVER_HEAPSIZE	7331
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7331
HADOOP_NAMENODE_HEAPSIZE	26419
HADOOP_DATANODE_HEAPSIZE	4096

## r5.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9953
YARN_PROXYSERVER_HEAPSIZE	9953
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9953
HADOOP_NAMENODE_HEAPSIZE	39526
HADOOP_DATANODE_HEAPSIZE	4096

## r5.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	12574
YARN_PROXYSERVER_HEAPSIZE	12574
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	12574
HADOOP_NAMENODE_HEAPSIZE	52633
HADOOP_DATANODE_HEAPSIZE	4096

## r5.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	17817
YARN_PROXYSERVER_HEAPSIZE	17817
YARN_NODEMANAGER_HEAPSIZE	2048



参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	17817
HADOOP_NAMENODE_HEAPSIZE	78848
HADOOP_DATANODE_HEAPSIZE	4096

## r5a 实例

### r5a.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

### r5a.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

## r5a.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4710
YARN_PROXYSERVER_HEAPSIZE	4710
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4710
HADOOP_NAMENODE_HEAPSIZE	13312
HADOOP_DATANODE_HEAPSIZE	3072

## r5a.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7331
YARN_PROXYSERVER_HEAPSIZE	7331
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7331
HADOOP_NAMENODE_HEAPSIZE	26419
HADOOP_DATANODE_HEAPSIZE	4096

## r5a.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9953
YARN_PROXYSERVER_HEAPSIZE	9953
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9953
HADOOP_NAMENODE_HEAPSIZE	39526
HADOOP_DATANODE_HEAPSIZE	4096

## r5a.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	12574
YARN_PROXYSERVER_HEAPSIZE	12574
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	12574
HADOOP_NAMENODE_HEAPSIZE	52633
HADOOP_DATANODE_HEAPSIZE	4096

### r5a.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	17817
YARN_PROXYSERVER_HEAPSIZE	17817
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	17817
HADOOP_NAMENODE_HEAPSIZE	78848
HADOOP_DATANODE_HEAPSIZE	4096

### r5ad 实例

#### r5ad.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## r5ad.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## r5ad.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## r5ad.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## r5ad.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## r5ad.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	12247
YARN_PROXYSERVER_HEAPSIZE	12247
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	12247
HADOOP_NAMENODE_HEAPSIZE	50995
HADOOP_DATANODE_HEAPSIZE	4096

## r5ad.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

## r5b 实例

### r5b.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

### r5b.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048



参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## r5b.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## r5b.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## r5b.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## r5b.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

### r5b.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

### r5d 实例

#### r5d.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

## r5d.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

## r5d.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4710
YARN_PROXYSERVER_HEAPSIZE	4710
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4710
HADOOP_NAMENODE_HEAPSIZE	13312
HADOOP_DATANODE_HEAPSIZE	3072

## r5d.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7331
YARN_PROXYSERVER_HEAPSIZE	7331
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7331
HADOOP_NAMENODE_HEAPSIZE	26419
HADOOP_DATANODE_HEAPSIZE	4096

## r5d.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9953
YARN_PROXYSERVER_HEAPSIZE	9953
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9953
HADOOP_NAMENODE_HEAPSIZE	39526
HADOOP_DATANODE_HEAPSIZE	4096

## r5d.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	12574
YARN_PROXYSERVER_HEAPSIZE	12574
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	12574
HADOOP_NAMENODE_HEAPSIZE	52633
HADOOP_DATANODE_HEAPSIZE	4096

## r5d.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	17817
YARN_PROXYSERVER_HEAPSIZE	17817
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	17817
HADOOP_NAMENODE_HEAPSIZE	78848
HADOOP_DATANODE_HEAPSIZE	4096

## r5dn 实例

### r5dn.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

### r5dn.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## r5dn.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## r5dn.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048



参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## r5dn.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## r5dn.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

### r5dn.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

### r5n 实例

#### r5n.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## r5n.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## r5n.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## r5n.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## r5n.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## r5n.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## r5n.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

## r6a 实例

### r6a.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

### r6g.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## r6a.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## r6a.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## r6a.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## r6a.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048



参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## r6a.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

## r6a.32xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	22077
YARN_PROXYSERVER_HEAPSIZE	22077
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	22077
HADOOP_NAMENODE_HEAPSIZE	100147
HADOOP_DATANODE_HEAPSIZE	4096

### r6a.48xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	32071
YARN_PROXYSERVER_HEAPSIZE	32071
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	32071
HADOOP_NAMENODE_HEAPSIZE	150118
HADOOP_DATANODE_HEAPSIZE	4096

### r6g 实例

#### r6g.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## r6g.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## r6g.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## r6g.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## r6g.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

### r6g.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

### r6gd 实例

#### r6gd.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## r6gd.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## r6gd.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## r6gd.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## r6gd.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

### r6gd.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

### r6i 实例

#### r6i.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048



参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## r6i.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## r6i.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## r6i.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## r6i.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## r6i.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## r6i.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

### r6i.32xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	21544
YARN_PROXYSERVER_HEAPSIZE	21544
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	21544
HADOOP_NAMENODE_HEAPSIZE	97484
HADOOP_DATANODE_HEAPSIZE	4096

### r6id 实例

#### r6id.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## r6gd.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## r6id.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## r6id.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## r6id.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## r6id.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## r6id.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

### r6id.32xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	22077
YARN_PROXYSERVER_HEAPSIZE	22077
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	22077
HADOOP_NAMENODE_HEAPSIZE	100147
HADOOP_DATANODE_HEAPSIZE	4096

### r6idn 实例

#### r6idn.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048



参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## r6idn.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## r6idn.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## r6idn.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## r6idn.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## r6idn.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## r6idn.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

### r6idn.32xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	22077
YARN_PROXYSERVER_HEAPSIZE	22077
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	22077
HADOOP_NAMENODE_HEAPSIZE	100147
HADOOP_DATANODE_HEAPSIZE	4096

### r6in 实例

#### r6in.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## r6in.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## r6in.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## r6in.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## r6in.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## r6in.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## r6in.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

### r6in.32xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	22077
YARN_PROXYSERVER_HEAPSIZE	22077
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	22077
HADOOP_NAMENODE_HEAPSIZE	100147
HADOOP_DATANODE_HEAPSIZE	4096

### r7a 实例

#### r7a.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048



参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

## r7a.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## r7a.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## r7a.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## r7a.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## r7a.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## r7a.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

## r7a.32xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	22077
YARN_PROXYSERVER_HEAPSIZE	22077
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	22077
HADOOP_NAMENODE_HEAPSIZE	100147
HADOOP_DATANODE_HEAPSIZE	4096

## r7a.48xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	32071
YARN_PROXYSERVER_HEAPSIZE	32071
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	32071
HADOOP_NAMENODE_HEAPSIZE	150118
HADOOP_DATANODE_HEAPSIZE	4096

## r7g 实例

### r7g.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

### r7g.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## r7g.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## r7g.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## r7g.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## r7g.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## r7gd 实例

### r7gd.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

### r7gd.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048



参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## r7gd.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## r7gd.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## r7gd.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## r7gd.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## r7i 实例

### r7i.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

### r7i.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## r7i.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## r7i.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## r7i.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## r7i.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## r7i.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

## r7i.48xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	32071
YARN_PROXYSERVER_HEAPSIZE	32071
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	32071
HADOOP_NAMENODE_HEAPSIZE	150118
HADOOP_DATANODE_HEAPSIZE	4096

## r7iz 实例

### r7iz.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

### r7iz.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

## r7iz.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

## r7iz.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048



参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## r7iz.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

## r7iz.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

### r7iz.32xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	21544
YARN_PROXYSERVER_HEAPSIZE	21544
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	21544
HADOOP_NAMENODE_HEAPSIZE	97484
HADOOP_DATANODE_HEAPSIZE	4096

### x1 实例

#### x1.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	21544
YARN_PROXYSERVER_HEAPSIZE	21544
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	21544
HADOOP_NAMENODE_HEAPSIZE	97484
HADOOP_DATANODE_HEAPSIZE	4096

### x1.32xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	41000
YARN_PROXYSERVER_HEAPSIZE	41000
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	41000
HADOOP_NAMENODE_HEAPSIZE	194764
HADOOP_DATANODE_HEAPSIZE	4096

### x1e 实例

#### x1e.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4520
YARN_PROXYSERVER_HEAPSIZE	4520
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4520
HADOOP_NAMENODE_HEAPSIZE	12364
HADOOP_DATANODE_HEAPSIZE	2882

## x1e.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	6952
YARN_PROXYSERVER_HEAPSIZE	6952
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	6952
HADOOP_NAMENODE_HEAPSIZE	24524
HADOOP_DATANODE_HEAPSIZE	4096

## x1e.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	11816
YARN_PROXYSERVER_HEAPSIZE	11816
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	11816
HADOOP_NAMENODE_HEAPSIZE	48844
HADOOP_DATANODE_HEAPSIZE	4096

## x1e.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	21544
YARN_PROXYSERVER_HEAPSIZE	21544
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	21544
HADOOP_NAMENODE_HEAPSIZE	97484
HADOOP_DATANODE_HEAPSIZE	4096

## x1e.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	41000
YARN_PROXYSERVER_HEAPSIZE	41000
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	41000
HADOOP_NAMENODE_HEAPSIZE	194764
HADOOP_DATANODE_HEAPSIZE	4096

### x1e.32xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	79912
YARN_PROXYSERVER_HEAPSIZE	79912
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	79912
HADOOP_NAMENODE_HEAPSIZE	389324
HADOOP_DATANODE_HEAPSIZE	4096

### x2gd 实例

#### x2gd.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

### x2gd.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

### x2gd.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## x2gd.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## x2gd.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048



参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

### x2gd.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	22077
YARN_PROXYSERVER_HEAPSIZE	22077
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	22077
HADOOP_NAMENODE_HEAPSIZE	100147
HADOOP_DATANODE_HEAPSIZE	4096

### x2idn 实例

#### x2idn.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	22077
YARN_PROXYSERVER_HEAPSIZE	22077
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	22077
HADOOP_NAMENODE_HEAPSIZE	100147
HADOOP_DATANODE_HEAPSIZE	4096

## x2idn.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	32071
YARN_PROXYSERVER_HEAPSIZE	32071
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	32071
HADOOP_NAMENODE_HEAPSIZE	150118
HADOOP_DATANODE_HEAPSIZE	4096

## x2idn.32xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	42065
YARN_PROXYSERVER_HEAPSIZE	42065
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	42065
HADOOP_NAMENODE_HEAPSIZE	200089
HADOOP_DATANODE_HEAPSIZE	4096

## x2iedn 实例

### x2iedn.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

### x2iedn.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

## x2iedn.4xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

## x2iedn.8xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	22077
YARN_PROXYSERVER_HEAPSIZE	22077
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	22077
HADOOP_NAMENODE_HEAPSIZE	100147
HADOOP_DATANODE_HEAPSIZE	4096

## x2iedn.16xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	42065
YARN_PROXYSERVER_HEAPSIZE	42065
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	42065
HADOOP_NAMENODE_HEAPSIZE	200089
HADOOP_DATANODE_HEAPSIZE	4096

## x2iedn.24xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	62054
YARN_PROXYSERVER_HEAPSIZE	62054
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	62054
HADOOP_NAMENODE_HEAPSIZE	300032
HADOOP_DATANODE_HEAPSIZE	4096

### x2iedn.32xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	82042
YARN_PROXYSERVER_HEAPSIZE	82042
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	82042
HADOOP_NAMENODE_HEAPSIZE	399974
HADOOP_DATANODE_HEAPSIZE	4096

### z1d 实例

#### z1d.xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

## z1d.2xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

## z1d.3xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	4055
YARN_PROXYSERVER_HEAPSIZE	4055
YARN_NODEMANAGER_HEAPSIZE	2048

参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	4055
HADOOP_NAMENODE_HEAPSIZE	10035
HADOOP_DATANODE_HEAPSIZE	2416

## z1d.6xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	6021
YARN_PROXYSERVER_HEAPSIZE	6021
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	6021
HADOOP_NAMENODE_HEAPSIZE	19865
HADOOP_DATANODE_HEAPSIZE	4096

## z1d.12xlarge

参数	值
YARN_RESOURCEMANAGER_HEAPSIZE	9953
YARN_PROXYSERVER_HEAPSIZE	9953
YARN_NODEMANAGER_HEAPSIZE	2048



参数	值
HADOOP_JOB_HISTORY_SERVER_HEAPSIZE	9953
HADOOP_NAMENODE_HEAPSIZE	39526
HADOOP_DATANODE_HEAPSIZE	4096

## HDFS配置

下表描述了 Hadoop 分布式文件系统 (HDFS) 的默认参数及其设置。您可以使用 `hdfs-site` 配置分类更改这些值。有关更多信息，请参阅 [配置应用程序](#)。

### Warning

- 如果单个节点出现故障，在节点少于四个的群集上设置 `dfs.replication` 为 1 可能会导致 HDFS 数据丢失。如果您的群集有 HDFS 存储，我们建议您为群集配置至少四个核心节点，用于生产工作负载，以避免数据丢失。
- Amazon EMR 不允许群集在下方扩展核心节点 `dfs.replication`。例如，如果是 `dfs.replication = 2`，则最小核心节点数为 2。
- 当您使用托管式自动扩缩功能、自动扩缩功能或选择手动调整群集大小时，建议您将 `dfs.replication` 设置为 2 或更高。

参数	定义	默认值
<code>dfs.block.size</code>	方 HDFS 块的大小。对存储在中的数据进行操作时 HDFS，分割大小通常为 HDFS 块的大小。数字越大，提供的任务粒度越小，但群集 NameNode 受到的压力也越小。	134217728 (128 MB)
<code>dfs.replication</code>	要持久性存储的每个数据块的副本数量。Amazon EMR 根据为群集配置的核心节点数量来设置此值。调整该值以满足您的需求。要覆盖默认值，请使用 <code>hdfs-site</code> 分类。	1 适用于预置少于四个核心节点的群集

参数	定义	默认值
		2 适用于预置少于 10 个核心节点的集群
		3 适用于所有其他集群

## Amazon HDFS 上有透明加密 EMR

透明加密是通过使用加密区域来实现的，HDFS加密区域是您定义的HDFS路径。每个加密区域都有自己的密钥（存储在使用 `hdfs-site` 配置分类指定的密钥服务器中）。

从 Amazon EMR 发行版 4.8.0 开始，您可以使用亚马逊EMR安全配置更轻松地为集群配置数据加密设置。安全配置提供的设置可以保护亚马逊弹性区块存储 (AmazonEBS) 存储卷中的传输数据和静态 EMRFS数据以及 Amazon S3 中的数据。有关更多信息，请参阅 Amazon EMR 管理指南中的[加密传输和静态数据](#)。

亚马逊KMS默认EMR使用 Hadoop；但是，您可以使用另一个来实现KMS该 KeyProvider API操作。HDFS加密区域中的每个文件都有自己唯一的数据加密密钥，该密钥由加密区域密钥加密。HDFS由于加密和解密活动仅在客户端进行，因此当数据写入加密区域时，会对数据进行加密 end-to-end（静态和传输中）。

您无法在加密区域之间移动文件，也无法将文件从加密区域移至未加密路径。

NameNode 和HDFS客户端通过该操作与 HadoopKMS（或KMS您配置的替代方案）进行交互。KeyProvider APIKMS负责将加密密钥存储在备用密钥库中。此外，Amazon 还EMR包含JCE无限强度政策，因此您可以创建所需长度的密钥。

有关更多信息，请参阅 Hadoop HDFS 文档中的[透明加密](#)。

### Note

在亚马逊KMS中EMR，Hadoop KMS 默认HTTPS不启用 over。有关如何启KMS用的更多信息 HTTPS，请参阅 [Hadoop 文档KMS](#)。

## 配置HDFS透明加密

您可以EMR通过创建密钥和添加加密区域在 Amazon 中配置透明加密。有几种方式可以实现：

- 在创建集群时使用 Amazon EMR 配置API操作
- 在 `command-runner.jar` 中使用 Hadoop JAR 步骤
- 登录到 Hadoop 集群的主节点并使用 `hadoop key` 和 `hdfs crypto` 命令行客户端
- 使用 Hadoop 和 REST APIs KMS HDFS

有关的更多信息 REST APIs，请参阅 Hadoop KMS 和的相应文档。HDFS

要在创建集群时使用创建加密区域及其密钥 CLI

配置API操作中的`hdfs-encryption-zones`分类允许您在创建集群时指定密钥名称和加密区域。Amazon 在您的集群KMS上的 Hadoop 中EMR创建此密钥并配置加密区域。

- 使用以下命令创建集群。

```
aws emr create-cluster --release-label emr-7.2.0 --instance-type m5.xlarge --
instance-count 2 \
--applications Name=App1 Name=App2 --configurations https://s3.amazonaws.com/
mybucket/myfolder/myConfig.json
```

### Note

为了便于读取，包含 Linux 行继续符 ( \ )。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 ( ^ )。

myConfig.json:

```
[
  {
    "Classification": "hdfs-encryption-zones",
    "Properties": {
      "/myHDFSPath1": "path1_key",
      "/myHDFSPath2": "path2_key"
    }
  }
]
```

```
] ]
```

## 在主节点上手动创建加密区域及其密钥

1. 使用高于 4.1.0 的 Amazon EMR 版本启动您的集群。
2. 使用 Connect 连接到集群的主节点SSH。
3. 在 Hadoop KMS 中创建密钥。

```
$ hadoop key create path2_key
path2_key has been successfully created with options Options{cipher='AES/CTR/
NoPadding', bitLength=256, description='null', attributes=null}.
KMSClientProvider[http://ip-x-x-x-x.ec2.internal:16000/kms/v1/] has been updated.
```

### Important

Hadoop KMS 要求您的密钥名称必须为小写。如果您使用的密钥包含大写字符，则您的集群将在启动过程中失败。

4. 在中创建加密区域路径HDFS。

```
$ hadoop fs -mkdir /myHDFSPath2
```

5. 使用您创建的密钥将HDFS路径设置为加密区域。

```
$ hdfs crypto -createZone -keyName path2_key -path /myHDFSPath2
Added encryption zone /myHDFSPath2
```

## 要使用手动创建加密区域及其密钥 AWS CLI

- 使用以下命令添加手动创建KMS密钥和加密区域的步骤。

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF --steps Type=CUSTOM_JAR,Name="Create
  First Hadoop KMS Key",Jar="command-runner.jar",ActionOnFailure=CONTINUE,Args=[/
bin/bash,-c,"\"hadoop key create path1_key\""] \
Type=CUSTOM_JAR,Name="Create First Hadoop HDFS Path",Jar="command-
runner.jar",ActionOnFailure=CONTINUE,Args=[/bin/bash,-c,"\"hadoop fs -mkdir /
myHDFSPath1\""] \
```

```
Type=CUSTOM_JAR,Name="Create First Encryption Zone",Jar="command-  
runner.jar",ActionOnFailure=CONTINUE,Args=[/bin/bash,-c,"\"hdfs crypto -createZone  
-keyName path1_key -path /myHDFSPath1\""] \  
Type=CUSTOM_JAR,Name="Create Second Hadoop KMS Key",Jar="command-  
runner.jar",ActionOnFailure=CONTINUE,Args=[/bin/bash,-c,"\"hadoop key create  
path2_key\""] \  
Type=CUSTOM_JAR,Name="Create Second Hadoop HDFS Path",Jar="command-  
runner.jar",ActionOnFailure=CONTINUE,Args=[/bin/bash,-c,"\"hadoop fs -mkdir /  
myHDFSPath2\""] \  
Type=CUSTOM_JAR,Name="Create Second Encryption Zone",Jar="command-  
runner.jar",ActionOnFailure=CONTINUE,Args=[/bin/bash,-c,"\"hdfs crypto -createZone  
-keyName path2_key -path /myHDFSPath2\""]
```

### Note

为了便于读取，包含 Linux 行继续符 (\)。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (^)。

## HDFS透明加密的注意事项

最佳实践是为可能写入文件的每个应用程序创建一个加密区域。此外，您可以使用配置中的hdfs-encryption-zones 分类HDFS对所有内容进行加密，API并将根路径 (/) 指定为加密区域。

## Hadoop 密钥管理服务器

[Hadoop KMS](#) 是一款密钥管理服务器，它能够为 Hadoop 集群实现加密服务，并且可以充当 Hadoop 集群的密钥供应商。[Amazon HDFS 上有透明加密 EMR](#)当您在启动集群时选择 Hadoop 应用程序时，会默认安装并启用KMS亚马逊EMR中的 Hadoop。EMR除了临时缓存之外，Hadoop 本身KMS不存储密钥。Hadoop KMS 充当密钥提供商和客户受托人之间对后备密钥库的代理，它不是密钥库。为 Hadoop 创建的默认密钥库KMS是 Java 加密扩展 (JCE)。KeyStore JCEKS还包括JCE无限强度政策，因此您可以创建所需长度的密钥。Hadoop KMS 还支持一系列ACLs控制密钥访问权限和密钥操作的方式，独立于其他客户端应用程序，例如。HDFSAmazon 中的默认密钥长度EMR为 256 位。

要配置 HadoopKMS，请使用 hadoop-kms-site 分类来更改设置。要进行配置ACLs，您可以使用分类 kms-acls。

有关更多信息，请参阅 [Hadoop 文档KMS](#)。Hadoop 用KMS于 Hadoop 透明加密HDFS。要了解有关 HDFS透明加密的更多信息，请参阅 Apache Hadoop 文档中的[HDFS透明加密](#)主题。

**Note**

在亚马逊KMS中EMR，Hadoop KMS 默认HTTPS不启用 over。要了解如何启用 KMSHTTPS，请参阅 [Hadoop 文档KMS](#)。

**Important**

Hadoop KMS 要求您的密钥名称必须为小写。如果您使用的密钥包含大写字符，则您的集群将在启动过程中失败。

## 在亚马逊中配置 Hadoop KMS EMR

如果使用亚马逊EMR发布的4.6.0或更高版本，则为9700，kms-http-port为9701。kms-admin-port

您可以在创建集群KMS时使用API亚马逊EMR版本的配置来配置 Hadoop。以下是 Hadoop KMS 可用的配置对象分类：

### Hadoop 配置分类 KMS

分类	文件名
hadoop-kms-site	kms-site.xml
hadoop-kms-acls	kms-acls.xml
hadoop-kms-env	kms-env.sh
hadoop-kms-log4j	kms-log4j.properties

### 要使用 Hadoop 来设置 Hadoop KMS ACLs CLI

- 使用以下命令ACLs使用 Hadoop KMS 创建集群：

```
aws emr create-cluster --release-label emr-7.2.0 --instance-type m5.xlarge --
instance-count 2 \
```

```
--applications Name=App1 Name=App2 --configurations https://s3.amazonaws.com/mybucket/myfolder/myConfig.json
```

### Note

为了便于读取，包含 Linux 行继续符 (\)。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (^)。

myConfig.json:

```
[
  {
    "Classification": "hadoop-kms-acls",
    "Properties": {
      "hadoop.kms.blacklist.CREATE": "hdfs,foo,myBannedUser",
      "hadoop.kms.acl.ROLLOVER": "myAllowedUser"
    }
  }
]
```

要禁用 Hadoop KMS 缓存，请使用 CLI

- 使用以下命令创建 Hadoop KMS `hadoop.kms.cache.enable` 设置为 `false` 的集群：

```
aws emr create-cluster --release-label emr-7.2.0 --instance-type m5.xlarge --instance-count 2 \
--applications Name=App1 Name=App2 --configurations https://s3.amazonaws.com/mybucket/myfolder/myConfig.json
```

### Note

为了便于读取，包含 Linux 行继续符 (\)。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (^)。

myConfig.json:

```
[
  {
    "Classification": "hadoop-kms-site",
    "Properties": {
      "hadoop.kms.cache.enable": "false"
    }
  }
]
```

要在 `kms-env.sh` 脚本中设置环境变量，请使用 CLI

- 通过 `kms-env.sh` 配置更改 `hadoop-kms-env` 中的设置。使用以下命令KMS使用 Hadoop 创建集群：

```
aws emr create-cluster --release-label emr-7.2.0 --instance-type m5.xlarge --
instance-count 2 \
--applications Name=App1 Name=App2 --configurations https://s3.amazonaws.com/
mybucket/myfolder/myConfig.json
```

#### Note

为了便于读取，包含 Linux 行继续符 (\)。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (^)。

myConfig.json:

```
[
  {
    "Classification": "hadoop-kms-env",
    "Properties": {
    },
    "Configurations": [
      {
        "Classification": "export",
        "Properties": {
          "JAVA_LIBRARY_PATH": "/path/to/files",
          "KMS_SSL_KEYSTORE_FILE": "/non/Default/Path/.keystore",
          "KMS_SSL_KEYSTORE_PASS": "myPass"
        }
      }
    ]
  }
]
```



```

    },
    "Configurations": [
    ]
  }
]
}
]

```

[有关配置 Hadoop 的信息KMS](#)，请参阅 [Hadoop 文档](#)。KMS

## HDFS在具有多个主节点的EMR集群上进行透明加密

[Apache Ranger](#) KMS 用于具有多个主节点的亚马逊EMR集群中，用于在中进行透明加密。HDFS

RDS对于具有多个主节点的亚马逊EMR集群，Apache Ranger 将其根密钥和加密区域 (EZ) 密钥KMS 存储在您的亚马逊中。要在具有多个主节点的 Amazon EMR 集群HDFS上启用透明加密，您必须提供以下配置。

- Amazon RDS 或者你自己的 SQL My Server 连接URL来存储 Ranger KMS 根密钥和 EZ 密钥
- “我的” 的用户名和密码 SQL
- Ranger KMS 根密钥的密码
- 用于SSL连接到“我的SQL服务器”的证书颁发机构 (CA) PEM 文件

您可以使用 `ranger-kms-dbks-site` 分类和 `ranger-kms-db-ca` 分类提供这些配置，如以下示例所示。

```

[
  {
    "Classification": "ranger-kms-dbks-site",
    "Properties": {
      "ranger.ks.jpa.jdbc.url": "jdbc:log4jdbc:mysql://mysql-host-url.xx-
xxx-1.xxx.amazonaws.com:3306/rangerkms",
      "ranger.ks.jpa.jdbc.user": "mysql-user-name",
      "ranger.ks.jpa.jdbc.password": "mysql-password",
      "ranger.db.encrypt.key.password": "password-for-encrypting-a-master-key"
    }
  },
  {
    "Classification": "ranger-kms-db-ca",

```

```

    "Properties": {
      "ranger.kms.trust.ca.file.s3.url": "s3://rds-downloads/rds-ca-2019-root.pem"
    }
  }
]

```

以下是 Apache Ranger KMS er 的配置对象分类。

## Hadoop 配置分类 KMS

分类	描述
ranger-kms-dbks-site	更改 Ranger KMS 的 dbks-site.xml 文件中的值。
ranger-kms-site	更改 Ran ranger-kms-site KMS ger .xml 文件中的值。
ranger-kms-env	在游侠KMS环境中更改值。
ranger-kms-log4j	更改 Ranger 的 kms-log4j.properties 文件中的值。KMS
ranger-kms-db-ca	更改 S3 上我与 Ranger SQL SSL KMS 连接的 CA 文件值。

## 注意事项

- 强烈建议您加密您的 Amazon RDS 实例以提高安全性。有关更多信息，请参阅[加密 Amazon RDS 资源概述](#)。
- 强烈建议您为每个具有多个主节点的 Amazon EMR 集群使用单独的“我的SQL数据库”，以实现高安全性。
- 要在具有多个主节点的 Amazon EMR 集群HDFS上配置透明加密，您必须在创建集群时指定hdfs-encryption-zones分类。否则，KMS将无法配置或启动 Ranger。具有多个主节点的 Ama EMR zon 集群不支持在正在运行的集群上重新KMS配置hdfs-encryption-zones分类或任何 Hadoop 配置分类。

# 创建或运行 Hadoop 应用程序

## 主题

- [使用 Amazon EMR 构建二进制文件](#)
- [通过流式处理来处理数据](#)
- [使用自定义处理数据 JAR](#)

## 使用 Amazon EMR 构建二进制文件

您可以使用 Amazon EMR 作为构建环境，以编译用于您的集群的程序。在 Amazon EMR 中使用的程序必须在运行 Linux 的系统上进行编译且 Linux 版本与 Amazon EMR 所用的相同。对于 32 位版本，您应在 32 位机器上或在打开 32 位交叉编译选项的情况下进行编译。对于 64 位版本，您需要在 64 位机器上编译或打开 64 位交叉编译选项。有关 EC2 实例版本的更多信息，请参阅《Amazon EMR 管理指南》中的[计划和配置 EC2 实例](#)。支持的编程语言包括 C++、Python 和 C#。

下表概览了使用 Amazon EMR 构建和测试您的应用程序所涉及的步骤。

### 构建模块的过程

- 1 连接到集群的主节点。
- 2 将源文件复制到主节点。
- 3 使用任何必要的优化方法构建二进制文件。
- 4 将二进制文件从主节点复制到 Amazon S3。

每个步骤的详细信息请参阅下面的部分。

### 连接到集群的主节点

- 按照《Amazon EMR 管理指南》中的[使用 SSH 连接到主节点](#)中的说明操作。

### 将源文件复制到主节点。

1. 将源文件置于 Amazon S3 存储桶中。要了解如何创建存储桶以及如何将数据移至 Amazon S3，请参阅 [《Amazon Simple Storage Service 用户指南》](#)。

2. 通过输入类似于以下内容的命令，为您的源文件在 Hadoop 集群上创建文件夹：

```
mkdir SourceFiles
```

3. 通过键入类似以下内容的命令，将您的源文件从 Amazon S3 复制到主节点：

```
hadoop fs -get s3://mybucket/SourceFiles SourceFiles
```

## 使用任何必要的优化构建二进制文件

构建二进制文件的方式取决于多种因素。请按照具体构建工具的说明，设置和配置您的环境。您可以使用 Hadoop 系统规范命令获取集群信息，以确定如何安装您的构建环境。

## 识别系统规范

- 使用以下命令验证用于构建二进制文件的架构。
  - a. 要查看 Debian 版本，请输入以下命令：

```
master$ cat /etc/issue
```

该输出值看上去类似于以下内容。

```
Debian GNU/Linux 5.0
```

- b. 要查看公有 DNS 名称和处理器大小，请输入以下命令：

```
master$ uname -a
```

该输出值看上去类似于以下内容。

```
Linux domU-12-31-39-17-29-39.compute-1.internal 2.6.21.7-2.fc8xen #1 SMP Fri  
Feb 15 12:34:28 EST 2008 x86_64 GNU/Linux
```

- c. 要查看处理器速度，请输入以下命令：

```
master$ cat /proc/cpuinfo
```

该输出值看上去类似于以下内容。

```
processor : 0
vendor_id : GenuineIntel
model name : Intel(R) Xeon(R) CPU E5430 @ 2.66GHz
flags : fpu tsc msr pae mce cx8 apic mca cmov pat pse36 clflush dts acpi mmx
fxsr sse sse2 ss ht tm syscall nx lm constant_tsc pni monitor ds_cpl vmx est
tm2 ssse3 cx16 xtpr cda lahf_lm
...
```

构建了二进制文件后，您就可以将文件复制到 Amazon S3。

将二进制文件从主节点复制到 Amazon S3

- 键入以下命令以将二进制文件复制到 Amazon S3 存储桶：

```
hadoop fs -put BinaryFiles s3://mybucket/BinaryDestination
```

## 通过流式处理来处理数据

Hadoop 流媒体是 Hadoop 附带的一个实用程序，它使您能够使用 Java 以外的语言开发可 MapReduce 执行文件。流媒体以 JAR 文件形式实现，因此您可以像标准 JAR 文件一样从 Amazon EMR API 或命令行运行它。

本节介绍如何在 Amazon 上使用直播功能 EMR。

### Note

Apache Hadoop Streaming 是一种独立工具。因此，这里并不介绍其所有函数和参数。有关 Hadoop 直播的更多信息，请访问 <http://hadoop.apache.org/docs/stable/hadoop-streaming/HadoopStreaming.html>。

## 使用 Hadoop Streaming 实用工具

此部分介绍如何使用 Hadoop 的 Streaming 实用工具。

### Hadoop 进程

- 1 以您所选择的编程语言编写映射器和 Reducer 可执行文件。

按照 Hadoop 的文档中的指示编写流式处理可执行文件。该等程序应从标准输入读取其输入内容，并通过标准输出来输出数据。默认情况下，输入/输出的每一行都代表一条记录，并且每一行中的第一个制表符都用作密钥与值之间的分隔符。

- 2 在本地测试您的可执行文件，并将它们上传到 Amazon S3。
- 3 使用亚马逊EMR命令行界面或亚马逊EMR控制台来运行您的应用程序。

每个映射器脚本都会以单独进程的形式在集群中启动。每个 Reducer 可执行文件都会通过任务流程将映射器可执行文件的输出转到数据输出中。

大多数 Streaming 应用程序都需要 `input`、`output`、`mapper` 和 `reducer` 参数。下表描述了上述参数和其它可选参数。

参数	描述	必填
<code>-input</code>	<p>输入数据在 Amazon S3 上的位置。</p> <p>类型：字符串</p> <p>默认值：无</p> <p>约束：URI。如果没有指定协议，那么它就可以使用集群的默认文件系统。</p>	是
<code>-output</code>	<p>亚马逊在 Amazon S3 EMR 上上传处理过的数据的位置。</p> <p>类型：字符串</p> <p>默认值：无</p> <p>约束：URI</p> <p>默认：如果未指定位置，Amazon 会将数据EMR上传到指定的 <code>input</code> 位置。</p>	是
<code>-mapper</code>	<p>映射器可执行文件的名称。</p>	是

参数	描述	必填
	类型：字符串  默认值：无	
-reducer	Reducer 可执行文件的名称。  类型：字符串  默认值：无	是
-cacheFile	一个 Amazon S3 位置，其中包含一些文件可供 Hadoop 复制到本地工作目录（主要目的是提高性能）。  类型：字符串  默认值：无  约束：[URI] # [要在工作目录中创建的符号链接名称]	否
-cacheArchive	JAR要解压缩到工作目录中的文件  类型：字符串  默认值：无  约束：[URI] # [要在工作目录中创建的符号链接目录名称]	否
-combiner	合并结果  类型：字符串  默认值：无  约束：Java 类名	否

以下示例代码是写入 Python 的映射器可执行文件。此脚本是 WordCount 示例应用程序的一部分。

```
#!/usr/bin/python
import sys

def main(argv):
    line = sys.stdin.readline()
    try:
        while line:
            line = line.rstrip()
            words = line.split()
            for word in words:
                print "LongValueSum:" + word + "\t" + "1"
            line = sys.stdin.readline()
    except "end of file":
        return None
if __name__ == "__main__":
    main(sys.argv)
```

## 提交流式处理步骤

本节介绍向集群提交流式处理步骤的基本知识。Streaming 应用程序会从标准输入读取输入内容，然后针对每个输入运行脚本或可执行文件（称为映射器）。每个输入的结果都保存在本地，通常保存在 Hadoop 分布式文件系统 (HDFS) 分区上。所有输入经过映射器处理后，第二个脚本或可执行文件（名为 Reducer）会处理映射器结果。将 Reducer 的结果发送到标准输出。您可以将一系列 Streaming 步骤串联起来，让一个步骤的输出作为另一个步骤的输入。

映射器和 Reducer 都能够以文件的形式进行引用，或者您也可以提供一个 Java 类。你可以用任何支持的语言实现映射器和缩减器，包括 Ruby、Perl、Python 或 Bas PHP h。

### 使用控制台提交流式处理步骤

此示例介绍如何使用 Amazon EMR 控制台向正在运行的集群提交流式处理步骤。

### 提交流式处理步骤

1. 打开亚马逊EMR控制台，[网址为 https://console.aws.amazon.com/emr](https://console.aws.amazon.com/emr)。
2. 在 Cluster List (集群列表) 中，选择您的集群的名称。
3. 滚动到 Steps (步骤) 部分并展开它，然后选择 Add step (添加步骤)。
4. 在 Add Step (添加步骤) 对话框中：
  - 对于 Step type (步骤类型)，选择 Streaming program (流式程序)。



- 对于 Name (名称), 请接受默认名称 (流式程序) 或键入新名称。
  - 对于映射器, 键入或浏览到 Hadoop 中映射器类所在的位置或映射器可执行文件 (如 Python 程序) 所在的 S3 存储桶。路径值必须采用以下形式 *BucketName/path/MapperExecutable*。
  - 对于 Reducer, 键入或浏览到 Hadoop 中 Reducer 类所在的位置或 Reducer 可执行文件 (如 Python 程序) 所在的 S3 存储桶。路径值必须采用以下形式 *BucketName/path/MapperExecutable*。Amazon EMR 支持特殊聚合关键词。有关更多信息, 请转到 Hadoop 提供的 Aggregate 库。
  - 对于 Input S3 location (输入 S3 位置), 键入或浏览到输入数据的位置。
  - 对于 Output S3 location (输出 S3 位置), 键入或浏览到您的 Amazon S3 输出存储桶的名称。
  - 对于 Arguments (参数), 将该字段保留为空白。
  - 对于 Action on failure (出现故障时的操作), 接受默认选项 Continue (继续)。
5. 选择添加。步骤会出现在控制台中, 其状态为“Pending”。
  6. 步骤的状态会随着步骤的运行从“Pending”变为“Running”, 再变为“Completed”。要更新状态, 请选择 Actions (操作) 列上方的 Refresh (刷新) 图标。

## AWS CLI

这些示例演示了 AWS CLI 如何使用创建集群和提交流式处理步骤。

要创建集群并提交流式处理步骤, 请使用 AWS CLI

- 要创建集群并使用提交流式处理步骤 AWS CLI, 请键入以下命令并替换 *myKey* 用你的 EC2 key pair 的名字。请注意, `--files` 的实际参数应该是指向您脚本位置的 Amazon S3 路径, 并且 `-mapper` 和 `-reducer` 的实际参数应该是各自脚本文件的名称。

```
aws emr create-cluster --name "Test cluster" --release-label emr-7.2.0 --
applications Name=Hue Name=Hive Name=Pig --use-default-roles \
--ec2-attributes KeyName=myKey --instance-type m5.xlarge --instance-count 3 \
--steps Type=STREAMING,Name="Streaming Program",ActionOnFailure=CONTINUE,Args=[--
files,pathtoscripts,-mapper,mapperscript,-reducer,reducerscript,aggregate,-
input,pathtoinputdata,-output,pathtooutputbucket]
```

### Note

为了便于读取, 包含 Linux 行继续符 (\)。它们可以通过 Linux 命令删除或使用。对于 Windows, 请将它们删除或替换为脱字号 (^)。

如果不使用 `--instance-groups` 参数指定实例计数，则将启动单个主节点，其余实例将作为核心节点启动。所有节点都使用该命令中指定的实例类型。

#### Note

如果您之前没有创建默认的 Amazon EMR 服务角色和EC2实例配置文件，请在键入 `create-cluster` 子命令之前键入 `aw emr create-default-roles s` 进行创建。

有关在中使用 Amazon EMR 命令的更多信息 AWS CLI，请参阅<https://docs.aws.amazon.com/cli/latest/reference/emr>。

## 使用自定义处理数据 JAR

自定义JAR运行已编译的 Java 程序，您可以将其上传到 Amazon S3。您应该根据要启动的 Hadoop 版本编译程序，然后向您的 Amazon EMR 集群提交一个CUSTOM\_JAR步骤。有关如何编译JAR文件的更多信息，请参阅[使用 Amazon EMR 构建二进制文件](#)。

有关构建 Hadoop MapReduce 应用程序的更多信息，请参阅 Apache Hadoop 文档中的[MapReduce 教程](#)。

### 主题

- [提交自定义JAR步骤](#)

## 提交自定义JAR步骤

自定义JAR运行已编译的 Java 程序，您可以将其上传到 Amazon S3。您应该根据要启动的 Hadoop 版本编译程序，然后向您的 Amazon EMR 集群提交一个CUSTOM\_JAR步骤。有关如何编译JAR文件的更多信息，请参阅[使用 Amazon EMR 构建二进制文件](#)。

有关构建 Hadoop MapReduce 应用程序的更多信息，请参阅 Apache Hadoop 文档中的[MapReduce 教程](#)。

本节介绍在 Amazon 中提交自定义JAR步骤的基础知识EMR。通过提交自定义JAR步骤，您可以编写脚本来使用 Java 编程语言处理数据。

## 使用控制台提交自定义JAR步骤

此示例介绍如何使用 Amazon EMR 控制台向正在运行的集群提交自定义JAR步骤。

### 使用控制台提交自定义JAR步骤

1. 打开亚马逊EMR控制台，[网址为 https://console.aws.amazon.com/emr](https://console.aws.amazon.com/emr)。
2. 在 Cluster List (集群列表) 中，选择您的集群的名称。
3. 滚动到 Steps (步骤) 部分并展开它，然后选择 Add step (添加步骤)。
4. 在 Add step (添加步骤) 对话框中：
  - 对于步骤类型，选择自定义JAR。
  - 对于“名称”，接受默认名称 (自定义JAR) 或键入新名称。
  - 要查看 JARS3 位置，请键入或浏览至JAR文件所在的位置。JAR位置可能是指向 S3 的路径或类路径中的完全限定的 java 类...
  - 对于参数，以空格分隔的字符串形式键入任何所需参数，或将该字段保留为空白。
  - 对于 Action on failure (出现故障时的操作)，接受默认选项 Continue (继续)。
5. 选择添加。步骤会出现在控制台中，其状态为“Pending”。
6. 步骤的状态会随着步骤的运行从“Pending”变为“Running”，再变为“Completed”。要更新状态，请选择 Actions (操作) 列上方的 Refresh (刷新) 图标。

### 启动集群并使用提交自定义JAR步骤 AWS CLI

要启动集群并提交自定义JAR步骤，请使用 AWS CLI

要启动集群并使用提交自定义JAR步骤 AWS CLI，请键入带 `--steps` 参数的 `create-cluster` 子命令。

- 要启动集群并提交自定义JAR步骤，请键入以下命令，替换 *myKey* 用你的 EC2 key pair 的名字，然后替换 *mybucket* 使用您的存储桶名称。

```
aws emr create-cluster --name "Test cluster" --release-label emr-7.2.0 \  
--applications Name=Hue Name=Hive Name=Pig --use-default-roles \  
--ec2-attributes KeyName=myKey --instance-type m5.xlarge --instance-count 3 \  
--steps Type=CUSTOM_JAR,Name="Custom JAR \  
Step",ActionOnFailure=CONTINUE,Jar=pathtojarfile,Args=["pathtoinputdata","pathtooutputbucket"]
```

**Note**

为了便于读取，包含 Linux 行继续符 (\)。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (^)。

如果不使用 `--instance-groups` 参数指定实例计数，则将启动单个主节点，其余实例将作为核心节点启动。所有节点都使用您在命令中指定的实例类型。

**Note**

如果您之前没有创建默认的 Amazon EMR 服务角色和 EC2 实例配置文件，请在键 `aws emr create-default-roles` 入 `create-cluster` 子命令之前键入创建它们。

有关在中使用 Amazon EMR 命令的更多信息 AWS CLI，请参阅 <https://docs.aws.amazon.com/cli/latest/reference/emr>。

### 第三方依赖项

有时可能需要在 MapReduce 类路径 JARs 中包含以便与您的程序一起使用。您有两个选项来执行此操作：

- 将 `--libjars s3://URI_to_JAR` 包含在 [启动集群并使用提交自定义 JAR 步骤 AWS CLI](#) 中的过程的步骤选项中。
- 使用 `mapred-site.xml` 中修改过的 `mapreduce.application.classpath` 设置启动集群。使用 `mapred-site` 配置分类。要使用步骤创建集群 AWS CLI，如下所示：

```
aws emr create-cluster --release-label emr-7.2.0 \  
--applications Name=Hue Name=Hive Name=Pig --use-default-roles \  
--instance-type m5.xlarge --instance-count 2 --ec2-attributes KeyName=myKey \  
--steps Type=CUSTOM_JAR,Name="Custom JAR  
Step",ActionOnFailure=CONTINUE,Jar=pathtojarfile,Args=["pathtoinputdata", "pathtooutputbucket  
\  
--configurations https://s3.amazonaws.com/mybucket/myfolder/myConfig.json
```

myConfig.json:

```
[
  {
    "Classification": "mapred-site",
    "Properties": {
      "mapreduce.application.classpath": "path1,path2"
    }
  }
]
```

应将以逗号分隔的路径列表附加到每个任务的JVM类路径中。

## 读取已恢复的对象

在 Amazon 7.2.0 及更高EMR版本中，您可以使用S3A协议从表的 S3 位置读取已恢复的 Glacier 对象。先前版本中的引擎不区分 Glacier 和 Glacier Deep Archive 文件，这意味着AmazonS3Exception如果你尝试使用访问正在进行的 Glacier 文件，你就会得到一个。S3A如果存档的 Glacier 文件仍在恢复中，则此读取操作会忽略这些文件。要启用此行为，请使用设置`fs.s3a.glacier.read.restored.objects`。此设置可以是三个值：

- READ\_ALL — 此值表示亚马逊EMR不应考虑从 Amazon S3 检索到的存储类别。这是默认行为。
- SKIP\_ALL\_GLACIER — 此值表示 Amazon EMR 应忽略任何标有 Glacier 存储类的 S3 对象，并检索所有其他对象。这是 Amazon Athena 在 Glacier 对象方面的默认行为。
- READ\_RESTORED\_GLACIER\_OBJECTS — 此值表示亚马逊EMR应检查 Glacier 对象的恢复状态。如果 Amazon EMR 可以恢复该对象，则您可以像普通的 S3 对象一样读取它们。否则，亚马逊EMR会忽略 S3 Glacier 中的对象

## 示例

### Spark

要在使用 Spark 时读取已恢复的对象，请使用以下配置：

```
--conf spark.hadoop.fs.s3a.glacier.read.restored.objects=<value>
```

如果您使用 spark-sql，请改用以下配置；

```
spark-sql --conf spark.hadoop.fs.s3a.glacier.read.restored.objects=<value>
```

## Flink

如果您使用 Flink，则可以在文件中 `flink-conf.yaml` 设置配置

```
fs.s3a.glacier.read.restored.objects: <value>
```

你也可以设置 `flink-conf` 分类：

```
[
  {
    "Classification": "flink-conf",
    "Properties": {
      "fs.s3a.glacier.read.restored.objects": "<value>"
    }
  }
]
```

## Hive

如果您使用 Hive，请在 `hive-site.xml` 文件中设置配置。

```
<property>
  <name>fs.s3a.glacier.read.restored.objects</name>
  <value><value></value>
</property>
```

你也可以使用 Hive CLI 来设置属性 `--hiveconf`：

```
hive --hiveconf fs.s3a.glacier.read.restored.objects=<value>
```

## 注意事项

当您从 S3 Glacier 读取已恢复的对象时，请注意以下注意事项：

- 只有使用 S3A 方案或来访问数据时，才能读取已恢复的对象。S3AFileSystem
- 当您读取已恢复的 Glacier 对象时，Amazon EMR 不会恢复该对象本身。为此，必须使用 AWS CLI 或 AWS SDK。

## 为容器开启非均匀内存访问感知 YARN

在 Amazon 6.x 及更高EMR版本中，您可以使用非均匀内存访问权限 (NUMA) 对集群上的数据进行多处理。NUMA是一种计算机内存设计模式，在这种模式下，处理器访问自己的本地内存的速度比其他处理器上的内存或处理器之间共享的内存更快。YARN容器具有更好的性能，NUMA因为它们可以绑定到为所有后续内存分配提供服务的特定NUMA节点。这可以减少集群访问远程内存的次数。

当工作节点计算机是多节点时，您可以启用对YARN容器的NUMA支持。NUMA要确认 Worker 节点是单 NUMA 还是多 NUMA 节点，请运行以下命令。

```
lscpu | grep -i numa
NUMA node(s): 2
```

通常，大于 12x 的实例有两个 NUMA 节点。这不适用于裸机实例。

开启YARN容器NUMA感知功能

1. 在您的 Amazon EMR 6.x 集群中使用以下yarn-site配置。

```
[
  {
    "classification": "yarn-site",
    "properties": {
      "yarn.nodemanager.linux-container-executor.nonsecure-mode.local-user": "yarn",
      "yarn.nodemanager.linux-container-executor.group": "yarn",
      "yarn.nodemanager.container-executor.class": "org.apache.hadoop.yarn.server.nodemanager.LinuxContainerExecutor",
      "yarn.nodemanager.numa-awareness.enabled": "true",
      "yarn.nodemanager.numa-awareness.numactl.cmd": "/usr/bin/numactl",
      "yarn.nodemanager.numa-awareness.read-topology": "true"
    },
    "configurations": []
  }
]
```

2. 在集群中提供以下引导操作。

```
#!/bin/bash

sudo yum -y install numactl
```

```

echo 1 | sudo tee /proc/sys/kernel/numa_balancing

echo "banned.users=mapred,bin,hdfs" >> /etc/hadoop/conf/container-executor.cfg
rm -rf /var/log/hadoop-yarn/
sudo chown -R yarn:hadoop /var/log/hadoop-yarn/
sudo chmod 755 -R /var/log/hadoop-yarn/

sudo chmod 6050 /etc/hadoop/conf/container-executor.cfg

mkdir /mnt/yarn && sudo chmod 755 -R /mnt/yarn && sudo chown -R yarn:hadoop /mnt/
yarn
mkdir /mnt1/yarn && sudo chmod 755 -R /mnt1/yarn && sudo chown -R yarn:hadoop /
mnt1/yarn
mkdir /mnt2/yarn && sudo chmod 755 -R /mnt2/yarn && sudo chown -R yarn:hadoop /
mnt2/yarn

```

3. 每个容器都必须能够感知 NUMA。您可以使用NUMA标记通知每个容器中的 Java 虚拟机 (JVM)。例如，要通知JVM要在MapReduce作业NUMA中使用，请在中添加以下属性mapred-site.xml。

```

<property>
  <name>mapreduce.reduce.java.opts</name>
  <value>-XX:+UseNUMA</value>
</property>
<property>
  <name>mapreduce.map.java.opts</name>
  <value>-XX:+UseNUMA</value>
</property>

```

4. 要验证您是否已开启 NUMA，请使用以下命令搜索任何 NodeManager 日志文件。

```
grep "NUMA resources allocation is enabled," *
```

要验证是否 NodeManager 已将NUMA节点资源分配给容器，请使用以下命令搜索 NodeManager 日志，`<container_id>`替换为您自己的容器 ID。

```
grep "NUMA node" | grep <container_id>
```



## Hadoop 版本历史记录

下表列出了 Amazon 的每个发行版本中包含的 Hadoop 版本EMR，以及随应用程序一起安装的组件。有关每个发行版本中的组件版本，请参阅 [亚马逊 EMR 7.x 发布版本](#)、[亚马逊 EMR 6.x 发布版本](#) 或 [亚马逊 EMR 5.x 发布版本](#) 中的发行版“组件版本”部分。

### Hadoop 版本信息

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-7.2.0	3.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.36.2	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-7.1.0	3.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
		library, hadoop-hdfs-nameno de, hadoop-https-server, hadoop-kms-server, hadoop- mapred, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server
emr-7.0.0	3.3.6	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop- hdfs-datanode, hadoop-hdfs- library, hadoop-hdfs-nameno de, hadoop-https-server, hadoop-kms-server, hadoop- mapred, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server
emr-6.15.0	3.3.6	emrfs, emr-ddb, emr-goodi es, emr-kinesis, emr-s3-dist- cp, hadoop-client, hadoop- hdfs-datanode, hadoop-hdfs- library, hadoop-hdfs-nameno de, hadoop-https-server, hadoop-kms-server, hadoop- mapred, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn- timeline-server

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-6.14.0	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.13.0	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.12.0	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-6.11.1	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.11.0	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.10.1	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-6.10.0	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.9.1	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.9.0	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-6.8.1	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.8.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.7.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-5.36.1	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.36.0	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.6.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-5.35.0	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.5.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.4.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server



Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-6.3.1	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.3.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.2.1	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-6.2.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.1.1	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.1.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-6.0.1	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.0.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.34.0	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-5.33.1	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.33.0	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.32.1	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-5.32.0	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.31.1	2.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.31.0	2.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-5.30.2	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.30.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.30.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-5.29.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.28.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.28.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-5.27.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.27.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.26.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server



Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-5.25.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.24.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.24.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-5.23.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.23.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.22.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-5.21.2	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.21.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.21.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-5.20.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.20.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.19.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-5.19.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.18.1	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.18.0	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-5.17.2	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.17.1	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.17.0	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-5.16.1	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.16.0	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.15.1	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-5.15.0	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.14.2	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.14.1	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server



Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-5.14.0	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.13.1	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.13.0	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-5.12.3	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.12.2	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.12.1	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-5.12.0	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.11.4	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.11.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-5.11.2	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.11.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.11.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-5.10.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.10.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.9.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-5.9.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.8.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.8.2	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-5.8.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.8.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.7.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-5.7.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.6.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.6.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server



Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-5.5.4	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.5.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.5.2	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-5.5.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.5.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.4.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-5.4.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.3.2	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.3.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-5.3.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.2.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.2.2	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-5.2.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.2.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.1.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-5.1.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.0.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.0.2	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-5.0.1	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.0.0	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.9.6	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-4.9.5	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.9.4	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.9.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager



Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-4.9.2	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.9.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.8.5	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-4.8.4	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.8.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.8.2	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-4.8.1	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.8.0	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.7.4	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-4.7.3	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.7.2	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.7.1	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-4.7.0	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.6.1	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.6.0	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-4.5.0	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.4.0	2.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.3.0	2.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Amazon EMR 发布标签	Hadoop 版本	随 Hadoop 安装的组件
emr-4.2.0	2.6.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.1.0	2.6.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.0.0	2.6.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-namenode, hadoop-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

## Hadoop 发布说明 ( 按版本分类 )

### 主题

- [亚马逊 EMR 7.2.0-Hadoop 发行说明](#)
- [亚马逊 EMR 6.6.0-Hadoop 发行说明](#)

## 亚马逊 EMR 7.2.0-Hadoop 发行说明

### 亚马逊 EMR 7.2.0-Hadoop 的变化

类型	描述
新功能	<a href="#">HADOOP-18850</a> : S3A : 使用密钥启用双层服务器端加密 AWS KMS
新功能	支持 S3 Glacier 读取已恢复的对象
改进	运行 Spark Jobs 时根据内核动态扩展 S3A 的最大连接数
改进	<a href="#">HADOOP-18797</a> : 使用 S3A Magic Committer 支持并发写入
改进	<a href="#">HADOOP-19047</a> : Support InMemory 追踪 S3A Magic 的提交
改进	在 MagicV2 提交器中添加“导出”、“发布”和“”方法 AbortPublish
改进	ReadFullyIntoBuffers FileSystem API在 S3A 中实施
错误修复	<a href="#">HADOOP-18793</a> : S3A StagingCommitter 无法清理暂存上传目录
Upgrade	



类型	描述
	<a href="#">HADOOP-19008 : S3A : 将 aws-sdk 版本更新到 2.21.41</a>
Upgrade	<a href="#">HADOOP-18613</a> : 升级到 ZooKeeper 3.8.3 版本

## 亚马逊 EMR 7.2.0-Hadoop 功能

有关亚马逊EMR 7.2.0 中全新 Hadoop 功能的信息，请参阅以下列表。

- [S3 Glacier 支持读取已恢复的对象](#) — 在 Amazon 7.2.0 及更高EMR版本中，您可以使用S3A协议从表的 S3 位置读取已恢复的 Glacier 对象。如果存档的 Glacier 文件仍在恢复中，则此读取操作会忽略这些文件。
- 使用 S3A 魔法提交器支持并发写入 — Amazon EMR 7.2.0 引入了对 S3A 魔法提交器并发写入的支持。[此功能已作为 -18797 的一部分添加到 OSS Hadoop 社区。HADOOP](#)

## 亚马逊 EMR 6.6.0-Hadoop 发行说明

### 亚马逊 EMR 6.6.0-Hadoop 变更

类型	描述
错误	修复了读取BZip2文本文件时的重复记录。
逆向移植	<a href="#">HADOOP-18136</a> : 验证FileUtils.unTar() 如何处理丢失的.tar 文件
逆向移植	<a href="#">HADOOP-17627</a> : 向后移植到 branch-3.2 -17371、-17621、HADOOP -17625，将 Jetty 更新到 9.4.39 HADOOP HADOOP
逆向移植	<a href="#">HADOOP-17655</a> : 将 Jetty 升级到 9.4.40
逆向移植	<a href="#">HADOOP-17796</a> : 将码头版本升级到 9.4.43

类型	描述
逆向移植	<a href="#">HADOOP-17661</a> : <a href="#">mvn 版本</a> :set 无法解析 pom.xml
逆向移植	<a href="#">HADOOP-17236</a> : 将 snakeyaml 提高到 1.26 以缓解 -2017-18640 CVE
逆向移植	<a href="#">HADOOP-16717</a> : 移除对 Log4 的 GenericsUtil isLog 4 个依赖关系 jLogger jLoggerAdapter
逆向移植	<a href="#">HADOOP-17633</a> : 将 json-smart 提高到 2.4.2 和 9.8 , 原因是 nimbus-jose-jwt CVEs
逆向移植	<a href="#">HADOOP-17844</a> : <a href="#">将智能升级到 JSON 2.4.7</a>
逆向移植	<a href="#">HADOOP-17972</a> : branch-3.2 的向后移植 <a href="#">HADOOP -17683</a> ( 将 commons-io 更新到 2.8.0 )
逆向移植	<a href="#">HADOOP-16555</a> : <a href="#">将 commons-compress 更新到 1.19</a>
逆向移植	<a href="#">HADOOP-17370</a> : <a href="#">将 commons-compress 升级到 1.21</a>
逆向移植	<a href="#">HADOOP-17096</a> : <a href="#">修复ZStandardCompressor 输入缓冲区偏移量</a>
逆向移植	<a href="#">HADOOP-17112</a> : <a href="#">通过提交者将文件保存到 s3a 时 , 路径中不允许使用空格</a>
逆向移植	<a href="#">HADOOP-13500</a> : <a href="#">同步配置属性对象的小版本</a>
逆向移植	<a href="#">HDFS-14099</a> : <a href="#">解压缩多个帧时帧描述符未知 ZStandardDecompressor</a>
逆向移植	<a href="#">HDFS-16410</a> : <a href="#">解析中的 XML 不安全 OfflineEditsXMLLoader</a>

类型	描述
逆向移植	<a href="#">HDFS-14498</a> : LeaseManager 可以在创建失败的文件上永久循环
逆向移植	<a href="#">HDFS-15290</a> : <a href="#">NPE在启动</a> 过程中输入 HttpServer NameNode
逆向移植	<a href="#">HDFS-15293</a> : 放宽接收检查点时接受 fsimage 的条件
逆向移植	<a href="#">HDFS-12979</a> : StandbyNode 应在检查点检查后上传到 FsImage ObserverNode
逆向移植	<a href="#">YARN-10538</a> : 将重新调试节点添加到返回到 AM 的更新节点列表中
逆向移植	<a href="#">YARN-10472</a> : 向后移植 YARN -10314 ( 仅使用阴影的客户端 jar 时YarnClient抛出 NoClassDefFoundError ) 到 branch-3.2 WebSocketException
逆向移植	<a href="#">YARN-9968</a> : 由于以下原因, 公共定位器正在退出 NodeManager NullPointerException
逆向移植	<a href="#">YARN-10651</a> : CapacityScheduler 在 in 中坠毁 NPE AbstractYarnScheduler.updateNodeResource()
逆向移植	<a href="#">YARN-9339</a> : 将应用程序移至新队列后, 待处理的应用程序指标不正确
逆向移植	<a href="#">YARN-10438</a> : 在中处理空值 container Id ClientRMService#getContainerReport()
逆向移植	<a href="#">YARN-7266</a> : 如果RollingLevelDb 文件损坏或丢失, 则无法启动 ATS 1.5

类型	描述
逆向移植	<a href="#">YARN-9063</a> : 如果RollingLevelDb 文件损坏或丢失, 则无法启动 ATS 1.5
逆向移植	<a href="#">YARN-9848</a> : 还原 YARN -4946 ( COMPLETED 当日志聚合未处于最终状态时, RM 不应将应用程序视为应用程序 )。

# Apache HBase

[HBase](#)是一个开源、非关系的分布式数据库，是作为 Apache 软件基金会 Hadoop 项目的一部分而开发的。HBase在 Hadoop 分布式文件系统 (HDFS) 之上运行，为 Hadoop 生态系统提供非关系数据库功能。HBase包含在 Amazon EMR 发行版 4.6.0 及更高版本中。

HBase与 Hadoop 无缝协作，共享其文件系统，并作为 MapReduce 框架和执行引擎的直接输入和输出。HBase还与 Apache Hive 集成，支持对HBase表SQL的类似查询，与基于 Hive 的表进行联接，并支持 Java 数据库连接 (JDBC)。有关 JDBC 的更多信息，请参阅 [Apache HBase 和 Apache 网站上的 HBase 文档](#)。有关如何使用 HBase Hive 的示例，请参阅 AWS 大数据博客文章 [Combine No SQL 和在亚马逊上使用 Apache 和 Apache HBase Hive 进行大规模并行分析](#)。EMR

HBase在 Amazon 上EMR，您还可以将HBase数据直接备份到亚马逊简单存储服务 (Amazon S3) Simple Storage Service，并在启动HBase集群时从先前创建的备份中恢复。Amazon EMR 提供了与 Amazon S3 集成的其他选项，以实现数据持久性和灾难恢复。

- HBase在 Amazon S3 上——在亚马逊 5.2.0 及更高EMR版本中，您可以HBase在 Amazon S3 上使用将集群的HBase根目录和元数据直接存储到 Amazon S3 中。随后，您可以启动新集群，将其指向 Amazon S3 中的根目录位置。一次只能有一个集群使用 Amazon S3 中的HBase位置，只读副本集群除外。有关更多信息，请参阅 [HBase在亚马逊 S3 \(亚马逊 S3 存储模式\)](#) 上。
- HBase只读副本-A mazon HBase S3 上的 Amazon 5.7.0 及更高EMR版本支持只读副本集群。在只读操作中，只读副本集群提供对主集群的存储文件和元数据的只读访问权限。有关更多信息，请参阅 [使用只读副本集群](#)。
- HBase快照-作为 Amaz HBase on S3 的替代方案，在 4.0 及更高EMR版本中，您可以将HBase数据直接创建到 Amazon S3 的快照，然后使用快照恢复数据。有关更多信息，请参阅 [使用HBase快照](#)。

## Important

对于 Amazon EMR HBase 集群扩展，我们不建议对集HBase群使用[托管扩展或使用自定义策略](#)进行扩展。

下表列出了最新版本的 Amazon EMR 7.x 系列中HBase包含的版本以及亚马逊随之EMR安装的组件。HBase

有关此版本HBase中安装的组件的版本，请参阅 [7.2.0 版组件版本](#)。

## HBaseemr-7.2.0 的版本信息

亚马逊EMR发行标签	HBase版本	随 HBase 安装的组件
emr-7.2.0	HBase 2.4.17	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

下表列出了最新版本的 Amazon EMR 6.x 系列中HBase包含的版本以及亚马逊随之EMR安装的组件。HBase

有关此版本HBase中安装的组件的版本，请参阅 [6.15.0 版组件版本](#)。

## HBaseemr-6.15.0 的版本信息

亚马逊EMR发行标签	HBase版本	随 HBase 安装的组件
emr-6.15.0	HBase 2.4.17	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-

亚马逊EMR发行标签	HBase版本	随 HBase 安装的组件
		kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

**Note**

Apache HBase HBCK2 是一个单独的操作工具，用于修复HBase区域和系统表。在亚马逊 6.1.0 及更高EMR版本中，hbase-hbck2.jar 是在主节点/usr/lib/hbase-operator-tools/上提供的。有关如何构建和使用该工具的更多信息，请参阅[HBaseHBCK2](#)。

下表列出了最新版本的 Amazon EMR 5.x 系列中HBase包含的版本以及亚马逊随之EMR安装的组件。HBase

有关此版本HBase中安装的组件的版本，请参阅 [5.36.2 版组件版本](#)。

HBaseemr-5.36.2 的版本信息

亚马逊EMR发行标签	HBase版本	随 HBase 安装的组件
emr-5.36.2	HBase 1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resour

亚马逊EMR发行标签	HBase版本	随 HBase 安装的组件
		cemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

## 主题

- [使用 HBase 创建集群](#)
- [HBase在亚马逊 S3 \( 亚马逊 S3 存储模式 \) 上](#)
- [Amazon 的预写日志 \(WAL\) EMR](#)
- [使用外HBase壳](#)
- [使用 Hive 访问HBase表](#)
- [使用HBase快照](#)
- [配置 HBase](#)
- [查看HBase用户界面](#)
- [查看 HBase 日志文件](#)
- [HBase使用 Ganglia 进行监控](#)
- [从以前的HBase版本迁移](#)
- [HBase发布历史](#)

## 使用 HBase 创建集群

本节中的过程涵盖了使用 AWS Management Console 和启动集群的基础知识 AWS CLI。有关如何规划、配置和启动 Amazon EMR 集群的详细信息，请参阅 Amazon EMR 管理指南中的[规划和配置集群](#)。

### HBase使用控制台创建集群

有关使用控制台启动集群的快速步骤，请参阅《[亚马逊EMR管理指南](#)》EMR中的 [Amazon 入门](#)。

### 使用控制台启动HBase已安装的集群



1. 在 <https://console.aws.amazon.com/emr> 上打开亚马逊EMR控制台。
2. 依次选择 Create cluster (创建集群) 和 Go to advanced options (转到高级选项)。
3. 在 Software Configuration (软件配置) 中，选择 Amazon Release Version (亚马逊发行版) 4.6.0 或更高版本 (建议使用最新版本)。根据需要选择HBase其他应用程序。
4. 在 Amazon 5.2.0 及更高EMR版本中，在“HBase存储设置”下选择HDFS或 S3。有关更多信息，请参阅 [HBase在亚马逊 S3 \(亚马逊 S3 存储模式\)](#) 上。
5. 根据需要选择其它选项，然后选择 Create cluster (创建集群)。

## HBase使用创建集群 AWS CLI

使用以下命令创建HBase已安装的集群：

```
aws emr create-cluster --name "Test cluster" --release-label emr-7.2.0 \  
--applications Name=HBase --use-default-roles --ec2-attributes KeyName=myKey \  
--instance-type m5.xlarge --instance-count 3
```

### Note

为了便于读取，包含 Linux 行继续符 (\)。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (^)。

如果您HBase在 Amazon S3 上使用，请使用对JSON配置对象的引用来指定该--configurations选项。配置对象必须包含一个hbase-site分类，该分类使用该hbase.rootdir属性指定 Amazon S3 中存储HBase数据的位置。它还必须包含 hbase 分类，其中使用 hbase.emr.storageMode 属性指定 s3。以下示例演示了具有这些配置设置的JSON片段。

```
[  
  {  
    "Classification": "hbase-site",  
    "Properties": {  
      "hbase.rootdir": "s3://MyBucket/MyHBaseStore"  
    }  
  },  
  {  
    "Classification": "hbase",  
    "Properties": {
```

```
        "hbase.emr.storageMode": "s3"  
    }  
}  
]
```

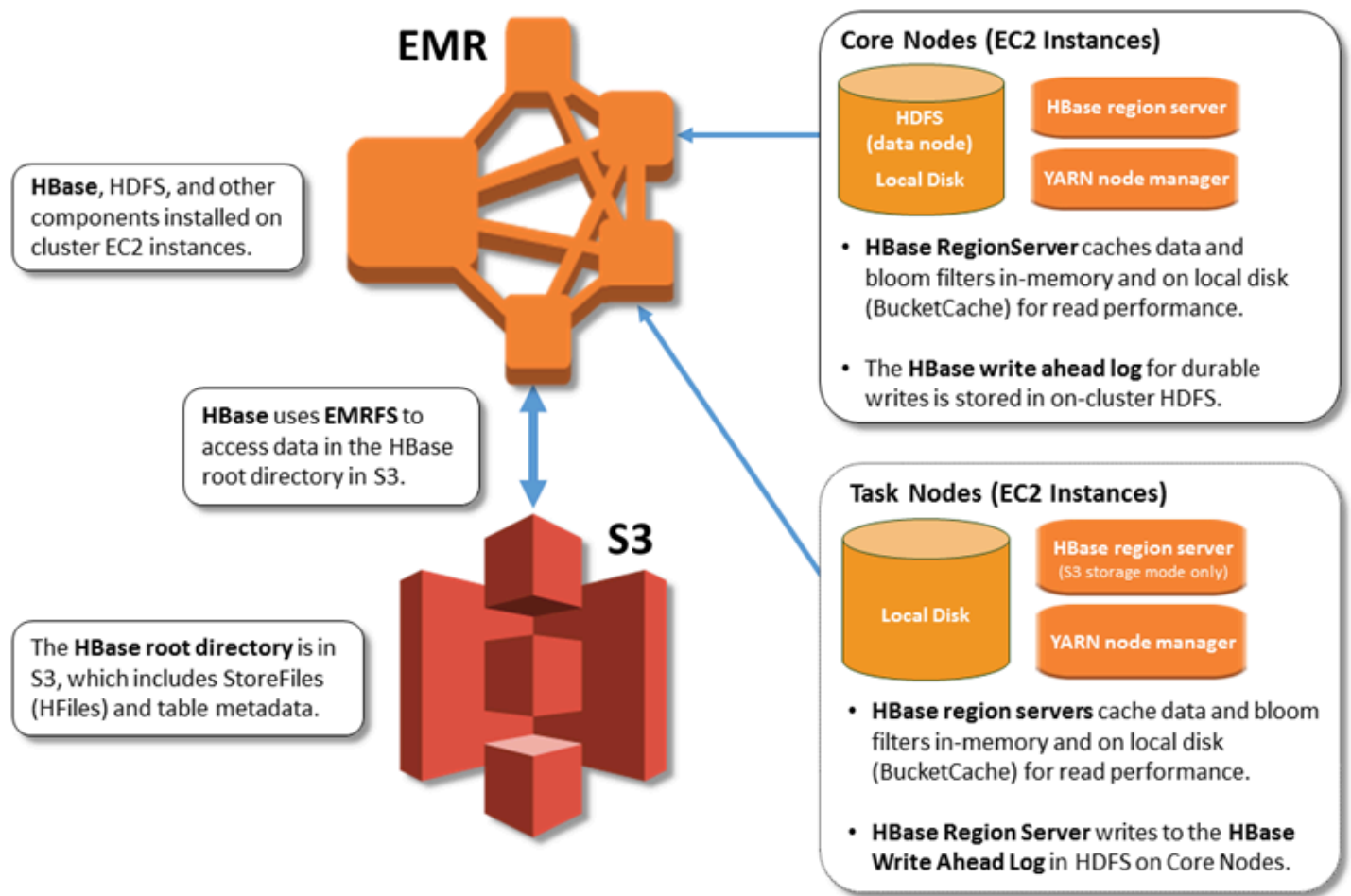
有关 Amaz HBase on S3 的更多信息，请参阅[HBase在亚马逊 S3 \( 亚马逊 S3 存储模式 \) 上](#)。有关分类的更多信息，请参阅[配置应用程序](#)。

## HBase在亚马逊 S3 ( 亚马逊 S3 存储模式 ) 上

当您在亚马逊EMR版本 5.2.0 或更高版本HBase上运行时，可以HBase在 Amazon S3 上启用，它具有以下优势：

- HBase根目录存储在 Amazon S3 中，包括HBase存储文件和表元数据。这些数据在集群之外是永久性的，可在各个 Amazon EC2 可用区域中使用，您无需使用快照或其他方法进行恢复。
- 通过在 Amazon S3 中存储文件，您可以根据自己的计算需求而不是数据要求调整亚马逊EMR集群的大小，同时还有 3 倍的HDFS复制量。
- 使用 Amazon 5.7.0 或更高EMR版本，您可以设置只读副本集群，这样您就可以在 Amazon S3 中维护数据的只读副本。如果主集群变得不可用，您可以访问只读副本集群中的数据以同时执行读取操作。
- 在 Amazon 6.2.0 及更高EMR版本中，持久HFile跟踪使用名为的HBase系统表hbase:storefile来直接跟踪用于读取操作的HFile路径。默认情况下，此功能处于启用状态，不需要执行手动迁移。

下图显示了 Amazon S3 HBase 上与相关的HBase组件。



## HBase在 Amazon S3 上启用

您可以使用亚马逊EMR控制台、或亚马逊HBase在 Amazon S3 上启用EMRAPI。AWS CLI该配置是集群创建期间的一个选项。使用控制台时，您可以通过 Advanced options (高级选项) 选择相应设置。使用时 AWS CLI，请使用 `--configurations` 选项提供JSON配置对象。配置对象的属性指定了 Amazon S3 中的存储模式和根目录位置。您指定的 Amazon S3 位置应与您的亚马逊EMR集群位于同一区域。在 Amazon S3 中，一次只能有一个活动集群使用同一个HBase根目录。有关控制台步骤和使用创建集群的详细示例 AWS CLI，请参阅。[使用 HBase 创建集群](#)以下JSON代码段显示了一个示例配置对象。

```
{
  "Classification": "hbase-site",
  "Properties": {
    "hbase.rootdir": "s3://my-bucket/my-hbase-rootdir"
  }
},
{
```

```

"Classification": "hbase",
"Properties": {
  "hbase.emr.storageMode": "s3"
}
}

```

### Note

如果您使用 Amazon S3 存储桶作 `rootdir` 为 HBase r，则必须在 Amazon S3 URI 的末尾添加一个斜杠。例如，为了避免出现问题，您必须使用 `"hbase.rootdir: s3://my-bucket/"`，而不是 `"hbase.rootdir: s3://my-bucket"`。

## 使用只读副本集群

在 Amazon S3 HBase 上使用设置主集群后，您可以创建和配置一个只读副本集群，该集群提供对与主集群相同数据的只读访问权限。当您需要同步访问权限以查询数据或在主集群变得不可用的情况下进行连续访问时，这会很有用。只读副本功能在 Amazon 5.7.0 及更高 EMR 版本中可用。

主集群和只读副本集群的设置方式相同，但有一个重要差异。两者都指向相同的 `hbase.rootdir` 位置。不过，只读副本集群的 hbase 分类包括 `"hbase.emr.readreplica.enabled": "true"` 属性。

例如，给定本主题前面所示的主群集的 JSON 分类，只读副本集群的配置如下所示：

```

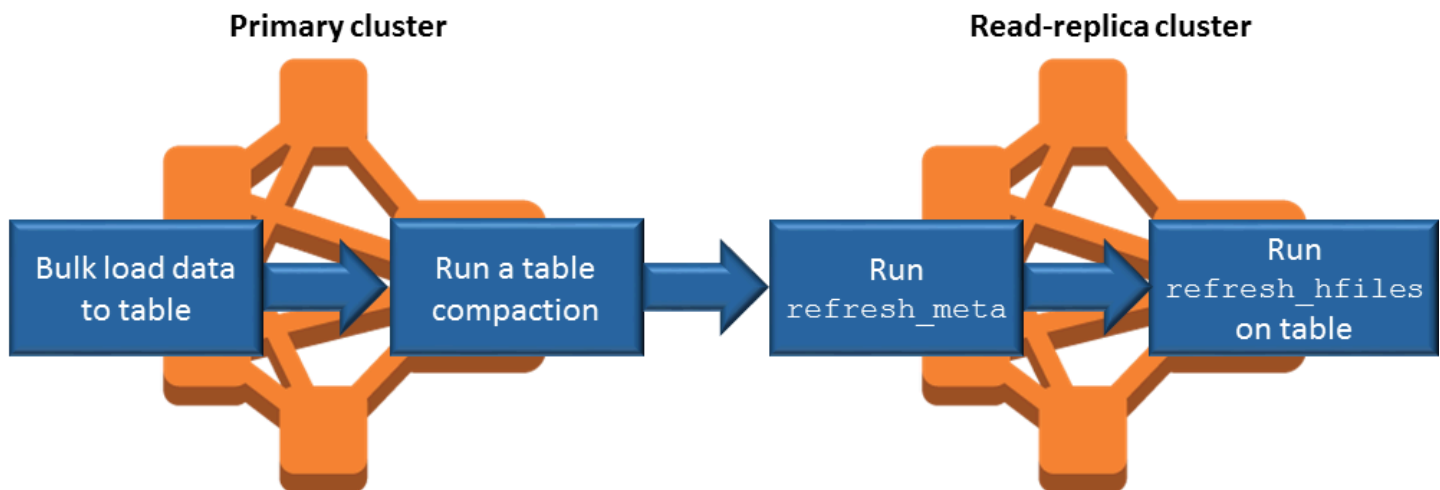
{
  "Classification": "hbase-site",
  "Properties": {
    "hbase.rootdir": "s3://my-bucket/my-hbase-rootdir"
  },
  {
    "Classification": "hbase",
    "Properties": {
      "hbase.emr.storageMode": "s3",
      "hbase.emr.readreplica.enabled": "true"
    }
  }
}

```

## 添加数据时同步只读副本

由于只读副本HBase StoreFiles 以及主集群写入 Amazon S3 的元数据，因此只读副本仅与 Amazon S3 数据存储一样最新。在写入数据时，以下指导信息可帮助您最大程度地缩短主集群和只读副本之间的滞后时间。

- 如果可能，请在主集群上批量加载数据。有关更多信息，请参阅 Apache HBase 文档中的[批量加载](#)。
- 将存储文件写入 Amazon S3 的刷新操作应在添加数据后尽快进行。手动刷新或优化刷新设置以最大限度地减少滞后时间。
- 如果压缩可能会自动运行，请运行手动压缩以避免触发压缩时出现不一致。
- 在只读副本集群上，当任何元数据发生更改时（例如，发生HBase区域拆分或压缩时，或者添加或删除表时），请运行该refresh\_meta命令。
- 在只读副本集群上，在表中添加或更改记录后，运行 refresh\_hfiles 命令。



## 持续HFile跟踪

持续HFile跟踪使用名为的HBase系统表hbase:storefile来直接跟踪用于读取操作的HFile路径。向表中添加更多数据后，会向表中添加新的HFile路径HBase。这删除了在关键写入路径操作中作为提交机制的重命名HBase操作，并通过从hbase:storefile系统表中读取而不是文件系统目录列表来缩短打开HBase区域时的恢复时间。在 Amazon 6.2.0 及更高EMR版本中，此功能默认处于启用状态，并且不需要任何手动迁移步骤。

**Note**

使用 HBase storefile 系统表进行持续HFile跟踪不支持HBase区域复制功能。有关HBase区域复制的更多信息，请参阅[时间轴一致的高可用读取](#)。

**禁用持续HFile跟踪**

从 Amazon 6.2.0 EMR 版本开始，永久HFile跟踪默认处于启用状态。要禁用持续HFile跟踪，请在启动集群时指定以下配置替代：

```
{
  "Classification": "hbase-site",
  "Properties": {
    "hbase.storefile.tracking.persist.enabled": "false",

    "hbase.hstore.engine.class": "org.apache.hadoop.hbase.regionserver.DefaultStoreEngine"
  }
}
```

**Note**

重新配置 Amazon EMR 集群时，必须更新所有实例组。

**手动同步存储文件表**

当创建新的存储文件表时，存储文件表HFiles会保持最新状态。但是，如果存储文件表由于任意原因与数据文件不同步，则可以使用以下命令手动同步数据：

同步线上区域中的存储文件表：

```
hbase org.apache.hadoop.hbase.client.example.RefreshHFilesClient <table>
```

同步离线区域中的存储文件表：

- 删除存储文件表 znode。

```
echo "ls /hbase/storefile/loaded" | sudo -u hbase hbase zkcli
[<tableName>, hbase:namespace]
# The TableName exists in the list
```

```
echo "delete /hbase/storefile/loaded/<tableName>" | sudo -u hbase hbase zkcli
# Delete the Table ZNode
echo "ls /hbase/storefile/loaded" | sudo -u hbase hbase zkcli
[hbase:namespace]
```

- 分配区域 ( 在“hbase shell”中运行 ) 。

```
hbase cli> assign '<region name>'
```

- 如果分配失败。

```
hbase cli> disable '<table name>'
hbase cli> enable '<table name>'
```

## 扩缩存储文件表

默认情况下，存储文件表可拆分为四个区域。如果存储文件表的写入负载仍然较重，之后可以手动拆分该表。

要拆分特定的热点区域，请使用以下命令 ( 在“hbase shell”中运行 ) 。

```
hbase cli> split '<region name>'
```

要拆分该表，请使用以下命令 ( 在“hbase shell”中运行 ) 。

```
hbase cli> split 'hbase:storefile'
```

## 操作注意事项

HBase区域服务器用于 BlockCache 将读取的数据存储在内存中 BucketCache ，并将读取的数据存储在本地磁盘上。此外，区域服务器用于 MemStore 将写入的数据存储在内存中，并使用预写日志在数据写入 HBase StoreFiles Amazon S3 HDFS 之前存储写入的数据。集群的读取性能与可从内存中或磁盘缓存中读取记录的频率有关。缓存失误会导致从 Amazon S3 中读取记录，与从 StoreFile 中读取相比，Amazon S3 的延迟和标准差明显更高HDFS。此外，Amazon S3 的最大请求速率低于可从本地缓存中检索内容的速率，因此对于需要进行大量读取操作的工作负载来说，缓存数据可能非常重要。有关 Amazon S3 性能的更多信息，请参阅《Amazon Simple Storage Service 用户指南》中的[性能优化](#)。

为了提高性能，我们建议您在EC2实例存储中尽可能多地缓存数据集。由于 BucketCache 使用区域服务器的EC2实例存储，因此您可以选择具有足够实例存储的EC2实例类型，然后添加 Amazon EBS 存

存储空间以适应所需的缓存大小。您还可以使用 `hbase.bucketcache.size` 属性增加附加的实例存储和EBS卷BucketCache的大小。默认设置为 8192MB。

对于写入，MemStore 刷新频率以及小规模 and 主要压缩期间的 StoreFiles 存在次数可以显著增加区域服务器的响应时间。为了获得最佳性能，请考虑增加MemStore 刷新和HRegion块乘数的大小，这会增加主要压缩之间的间隔时间，但如果您使用只读副本，也会增加一致性的延迟。在某些情况下，使用更大的文件块大小（但小于 5 GB）触发 Amazon S3 分段上传功能可能会获得更好的性能。EMRFS亚马逊EMR的区块大小默认为 128 MB。有关更多信息，请参阅 [HDFS配置](#)。在通过刷新和压缩来衡量性能时，我们很少看到有客户的数据块大小超过 1GB。此外，当 StoreFiles 需要HBase压缩的服务器较少时，压缩服务器和区域服务器的性能会达到最佳状态。

由于需要对大量目录进行重命名，因此从 Amazon S3 中删除表需要花费大量时间。请考虑禁用表而不是删除表。

有一个HBase更干净的过程可以清理旧WAL文件和存储文件。在 Amazon EMR 发行版 5.17.0 及更高版本中，清理程序将全局启用，并且可以使用以下配置属性来控制清理器行为。

配置属性	默认值	描述
<code>hbase.regionserver.hfilecleaner.large.thread.count</code>	1	分配给清理的线程数已过期 HFiles。
<code>hbase.regionserver.hfilecleaner.small.thread.count</code>	1	分配给清理的线程数过期很少 HFiles。
<code>hbase.cleaner.scan.dir.concurrent.size</code>	设置为所有可用内核的四分之一。	要扫描 oldWALs 目录的线程数。
<code>hbase.oldwals.cleaner.thread.size</code>	2	oldWALs 目录WALs下要清理的线程数。



在 Amazon EMR 5.17.0 及更早版本中，在运行繁重的工作负载时，清理操作可能会影响查询性能，因此我们建议您仅在非高峰时段启用清理器。清理器有以下 HBase shell 命令：

- `cleaner_chore_enabled` 查询是否启用了清理器。
- `cleaner_chore_run` 手动运行清理器来删除文件。
- `cleaner_chore_switch` 启用或禁用清理器并返回清理器的先前状态。例如，`cleaner_chore_switch true` 启用清理器。

## HBase在 Amazon S3 上进行性能调整的属性

当您在 Amazon S3 HBase 上使用时，可以调整以下参数以调整工作负载的性能。

配置属性	默认值	描述
<code>hbase.bucketcache.size</code>	8192	Amazon EC2 实例存储的区域服务器和 BucketCache 存储EBS卷上预留的磁盘空间量（以 MB 为单位）。此设置适用于所有区域服务器实例。较大的 BucketCache 尺寸通常对应于性能的提高
<code>hbase.hregion.memstore.flush.size</code>	134217728	触发对 Amazon S3 的 memstore 刷新的数据限制（以字节为单位）。
<code>hbase.hregion.memstore.block.multiplier</code>	4	一个乘数，用于确定阻止更新的 MemStore 上限。如果 MemStore 超出值 <code>hbase.hregion.memstore.flush.size</code> 乘以此值，则会阻止更新。MemStore 刷新和压缩可能会发生解除封锁更新的情况。
<code>hbase.hstore.blockingStoreFiles</code>	10	在阻止更新之前 StoreFiles，商店中可以存在的最大数量。

配置属性	默认值	描述
<code>hbase.hregion.max.filesize</code>	10737418240	区域被拆分之前的大小上限。

## 关闭并恢复集群而不丢失数据

要在不丢失尚未写入 Amazon S3 的数据的情况下关闭 Amazon EMR 集群，您应该将 MemStore 缓存刷新到 Amazon S3 以写入新的存储文件。首先，您需要禁用所有表格。在向集群添加步骤时，可使用以下步骤配置。有关更多信息，请参阅 Amazon EMR 管理指南中的[使用 AWS CLI 和控制台处理步骤](#)。

```
Name="Disable all tables",Jar="command-runner.jar",Args=["/bin/bash","/usr/lib/hbase/bin/disable_all_tables.sh"]
```

或者，您可以直接运行以下 bash 命令。

```
bash /usr/lib/hbase/bin/disable_all_tables.sh
```

禁用所有表后，使用 HBase shell 和以下命令刷新 `hbase:meta` 表。

```
flush 'hbase:meta'
```

然后，您可以运行 Amazon EMR 集群上提供的 shell 脚本来刷新 MemStore 缓存。您可以将它作为步骤添加，也可以使用集群中 AWS CLI 直接运行它。该脚本会禁用所有 HBase 表，这会导致每个区域 MemStore 的服务器刷新到 Amazon S3。如果此脚本成功完成，数据将保留在 Amazon S3 中，并且可以终止集群。

要使用相同 HBase 数据重启集群，请在 AWS Management Console 或中使用 `hbase.rootdir` 配置属性指定与前一个集群相同的 Amazon S3 位置。

## Amazon 的预写日志 (WAL) EMR

在 Amazon EMR 6.15 及更高版本中，你可以将 Apache HBase 预写日志 () 写 WAL 入亚马逊。EMR WAL 在 Amazon EMR 版本较低的情况下，当您使用“on Amaz HBase on S3”选项创建集群时，WAL 是唯一存储在本地磁盘中用于集群的 Apache HBase 组件，并且您可以在 Amazon S3 上存储其他组件，例如根目录、存储文件 (HFiles)、表元数据和数据。

您可以使用亚马逊EMRWAL来恢复未刷新到 Amazon S3 的数据。要完全备份您的HBase集群，请选择使用 Amazon EMR WAL 服务。在幕后，RegionServer将您的HBase预写日志 (WAL) 写入 WAL Amazon 的。EMR

如果您的集群或可用区运行状况不佳或不可用，您可以创建一个新集群，将其指向相同的 S3 根目录和 Amazon EMR WAL 工作空间，并在几分钟WAL内自动恢复数据。有关更多信息，请参阅 [从亚马逊恢复 EMR WAL](#)。

#### Note

从您创建集群之日起，Amazon 会将您的预写日志及其数据EMR保留 30 天。30 天后，亚马逊EMR会自动删除您的亚马逊EMRWAL及其数据。但是，如果您从同一 S3 根目录启动WAL已启用新集群的集群，则可以从新集群的启动时间起将您的WAL使用期限延长 30 天。在最初的 30 天期限之后，Amazon 仍EMR会清理第一个集群中的所有WAL数据。有关更多信息，请参阅 [从亚马逊恢复 EMR WAL](#)。

以下各节介绍如何在HBase已启用集EMR群时设置和EMRWAL使用 Amazon。

#### 主题

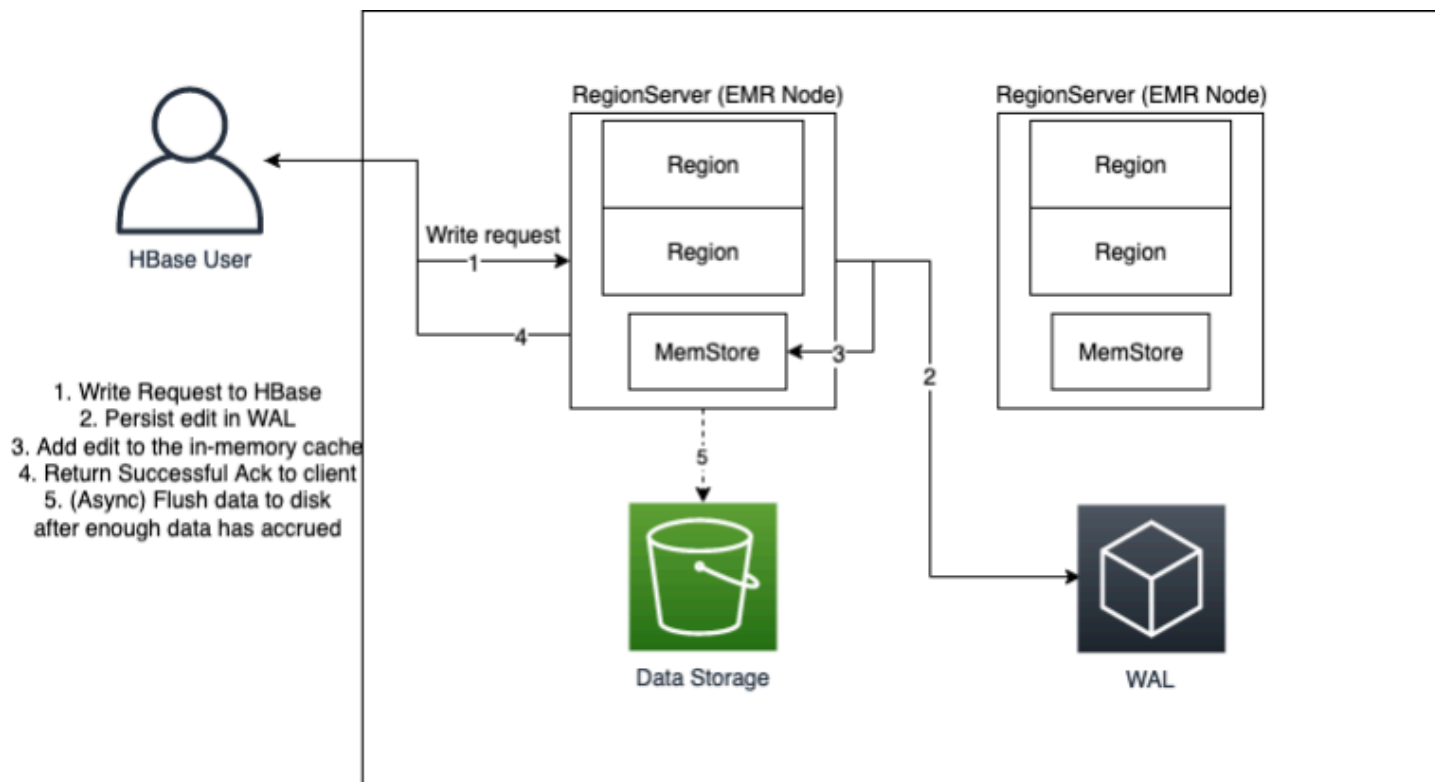
- [Amazon EMR WAL 工作空间](#)
- [Amazon 所需的权限 EMR WAL](#)
- [启用 Amazon EMR WAL](#)
- [从亚马逊恢复 EMR WAL](#)
- [在 Amazon 上使用安全配置 EMR WAL](#)
- [EMRWAL通过以下方式访问亚马逊 AWS PrivateLink](#)
- [了解 Amazon 的EMRWAL定价和指标](#)
- [为工作空间添加标签 WAL](#)
- [Amazon 的注意事项和区域 EMR WAL](#)
- [亚马逊 EMR WAL \(EMRWAL\) CLI 参考](#)

## Amazon EMR WAL 工作空间

Amazon EMR WAL 增加了WAL工作空间的概念。WAL工作空间是的逻辑容器WALs。Amazon 中的每个预写日志都EMRWAL由一个工作空间封装。WALEMR集群只写WALs入您在集群启动时配置的一个

WAL工作空间，或者defaultWALworkspace如果您未指定工作空间，则写入到该工作空间。WAL工作空间与任何现有HBase术语无关，例如命名空间。

您可以使用WAL工作空间来缩小 Amazon EMR WAL IAM 权限的范围，使其仅包含集群需要访问的工作空间。您也可以将WAL工作空间标记为基于标签的访问控制。有关标记的更多信息，请参阅[为工作空间添加标签 WAL](#)。



## Amazon 所需的权限 EMR WAL

要使您的集群连接到 Amazon EMRWAL，集群的实例配置文件需要某些IAM权限：

- Amazon EMR WAL 使用[AWSServiceRoleForEMRWAL](#)服务相关角色来检索集群状态。Amazon 会在您创建WAL工作空间时EMR自动创建此服务相关角色，或者HBase在您为 Amazon 配置工作空间EMRWAL且服务相关角色尚不存在时创建服务相关角色。

在EMRWAL为集群启用 Amazon 之前，必须将权限配置为允许自动创建 [AWSServiceRoleForEMRWAL](#)服务相关角色。有关更多信息以及添加此功能的示例语句，请参阅[使用服务相关角色进行预写](#)日志。

- 由于 Amazon EMR WAL 使用预HBase写日志 (WAL)，因此您的集群必须使用HBaseWAL。以下是您需要运行的最低IAM权限HBase。将以下内容添加到您的实例配置文件的权限策略中：

```
emrwal:DeleteWal  
emrwal:CreateWal  
emrwal:CreateWorkspace  
emrwal:AppendEdit  
emrwal:ReplayEdits  
emrwal:GetCurrentWalTime  
emrwal:CompleteWalFlush
```

### Note

如果您将 Amazon EMR WAL 的权限范围限制为最低限度，则某些 [EMRWALCLI](#) 命令将没有运行所需的权限。

## 启用 Amazon EMR WAL

使用创建集群EMRWAL时，使用以下步骤启用对 Amazon 的写入功能 AWS Command Line Interface。

### Note

您无法EMRWAL为已在运行的集群启用 Amazon，也无法启动具有相同 S3 根目录的两个集群。有关更多信息，请参阅 [Amazon 的注意事项和区域 EMR WAL](#)。

1. 在创建EMRWAL支持 Amazon 的集群之前，您必须向计划用于集群的实例配置文件添加所需的权限。有关更多信息，请参阅 [Amazon 所需的权限 EMR WAL](#)。
2. 从中创建集群 AWS CLI。使用 `--configurations` 选项提供指定 `hbase.emr.wal.enabled` 属性的 JSON 配置对象，如下例所示。
  - 在 Amazon S3 中指定存储模式和根目录位置。您指定的 Amazon S3 位置应与您的 EMR 集群位于同一区域，但一次只能有一个活动集群使用 S3 中的同一个 HBase 根目录。
  - 使用实例组配置创建您的集群。您不能将 Amazon EMR WAL 与实例队列配置一起使用。有关使用实例组创建集群的更多信息，请参阅 Amazon EMR 管理指南中的 [配置统一实例组](#)。
  - 有关创建集群的控制台步骤以及使用集群的详细 `create-cluster` 示例 AWS CLI，请参阅使用 [创建集群HBase](#)。

3. 要WAL为新集群启用，请将该`hbase.emr.wal.enabled`属性设置为`true`。以下命令包含带有示例配置对象的JSON片段。

```
aws emr create-cluster --name "hbasewal" --release-label emr-6.x.y \  
--applications Name=HBase --use-default-roles --ec2-attributes KeyName=myKey \  
--instance-type m6i.xlarge --instance-count 1 --configurations hbase.json  
$cat hbase.json  
[  
  {  
    "Classification": "hbase-site",  
    "Properties": {  
      "hbase.rootdir": "s3://MyBucket/MyHBaseStore"  
    }  
  },  
  {  
    "Classification": "hbase",  
    "Properties": {  
      "hbase.emr.storageMode": "s3",  
      "hbase.emr.wal.enabled": "true"  
    }  
  }  
]
```

在新创建HBase的集群上联机时，HBase会自动将WAL数据写入 Amazon EMR WAL 并使用 Amazon EMR WAL 进行恢复。

Example 1：创建使用 Amazon 的EMR集群 EMR WAL

```
[  
  {  
    "Classification": "hbase-site",  
    "Properties": {  
      "hbase.rootdir": "s3://MyBucket/MyHBaseStore"  
    }  
  },  
  {  
    "Classification": "hbase",  
    "Properties": {  
      "hbase.emr.storageMode": "s3",  
      "hbase.emr.wal.enabled": "true"  
    }  
  }  
]
```

]

## Example 2：使用自定义WAL工作空间创建EMR集群

```
[
  {
    "Classification": "hbase-site",
    "Properties": {
      "hbase.rootdir": "s3://MyBucket/MyHBaseStore",
      "emr.wal.workspace": "customWorkspaceName"
    }
  },
  {
    "Classification": "hbase",
    "Properties": {
      "hbase.emr.storageMode": "s3",
      "hbase.emr.wal.enabled": "true"
    }
  }
]
```

## 从亚马逊恢复 EMR WAL

由于您的原始集群EMRWAL的 Amazon 会保留 30 天，因此您可以在 30 天内WAL为新创建的集群恢复和重复使用该集群。当您从同一 S3 根目录启动新集群时，30 天的时钟将从新集群的启动时间重新开始，前提是之前的 30 天期限未到期。

使用以下步骤使用新集群恢复现有WAL集群。此过程假设您在EMRWAL启用了 Amazon 的情况下创建了原始集群。

1. 在创建WAL已启用集群的 30 天内，创建一个与原始集群 AWS 区域 相同的新集群。新集群可以位于与创建原始集群相同的区域内的同一个可用区或不同的可用区中。

配置对象属性以指定存储模式和在 Amazon S3 中的根目录位置。您指定的 Amazon S3 位置应与您的EMR集群位于同一区域，但一次只能有一个活动集群使用 S3 中的同一个HBase根目录。

有关创建集群的控制台步骤以及使用集群的详细create-cluster示例 AWS CLI，请参阅使用[创建集群HBase](#)。

2. 要将现有 Amazon EMR WAL 用于新集群，请将hbase.emr.wal.enabled属性设置为true。以下JSON片段显示了一个配置对象示例。

```
[
  {
    "Classification": "hbase-site",
    "Properties": {
      "hbase.rootdir": "s3://MyBucket/MyHBaseStore"
    }
  },
  {
    "Classification": "hbase",
    "Properties": {
      "hbase.emr.storageMode": "s3",
      "hbase.emr.wal.enabled": "true"
    }
  }
]
```

## 在 Amazon 上使用安全配置 EMR WAL

Amazon EMR 会自动加密您的集群和亚马逊EMRWAL服务之间传输的数据以及亚马逊EMRWAL中的静态数据。有关更多信息，请参阅 [Amazon 静态加密EMRWAL](#)。您还可以使用安全配置从 AWS Key Management Service (KMS) 服务中获取自己的密钥并加密存储在 Amazon 中的数据EMRWAL。

创建集群时，使用以下方法之一选择安全配置：

### Console

在中 AWS Management Console，在“安全配置和 EC2 key pair”下指定配置。

#### Security configuration and EC2 key pair - *optional* [Info](#)

##### Security configuration

Select your cluster encryption, authentication, authorization, and instance metadata service settings.



### CLI

使用 [create-cluster](#) 命令时，请从中设置 `--security-configuration` 参数。AWS CLI

有关更多信息，请参阅《[亚马逊EMR管理指南](#)》中的 [Amazon 静态加密EMRWAL](#) 和 [使用安全配置设置集群安全](#)。



有关安全的更多相关信息WAL，请参阅[使用服务相关角色进行预写](#)日志。

## EMRWAL通过以下方式访问亚马逊 AWS PrivateLink

如果您想在 AWS 网络内保持连接，Amazon 会为 EMRWAL 提供 AWS PrivateLink 支持。要进行设置 AWS PrivateLink，请使用 AWS Management Console 或 AWS Command Line Interface (AWS CLI) 创建连接到 Amazon 的接口 VPC 终端节点 EMRWAL。有关更多信息，请参阅 AWS PrivateLink 指南中的[使用接口 VPC 终端节点访问 AWS 服务](#)。

基本步骤如下所示：

1. 使用 Amazon VPC 控制台[创建 VPC 终端节点](#)。选择终端节点，然后选择创建端点。
2. 将服务类别保留为 AWS 服务。
3. 在“服务”面板的搜索栏中，键入 `emrwal`，然后选择标有标签的服务 `com.amazonaws.region.emrwal.prod`。
4. 选择您的 VPC 并保存终端节点。确保将相同的安全组连接到 EMR 集群的 VPC 终端节点上。
5. 如果您愿意，您现在可以为您的新终端节点启用私有 DNS 主机名。为您的“启用 DNS 主机名”和“启用 S DNS support”设置 `true` VPC 为。然后，选择您的终端节点 ID，从“操作”菜单中选择“编辑 VPC 设置”，然后启用私有 DNS 名称。
  - 终端节点的私有 DNS 主机名将遵循以下格式 `prod.emrwal.region.amazonaws.com`。
  - 如果您不启用私有 DNS 主机名，Amazon VPC 将按照以下格式 `endpointID.prod.emrwal.region.vpc.amazonaws.com` 为您提供 DNS 终端节点名称。
6. 要使用您的 AWS PrivateLink 终端节点，请在创建[EMRWAL 支持 Amazon 的集群](#)时修改 `emr.wal.client.endpoint` 配置，如以下示例所示：

```
[
  {
    "Classification": "hbase-site",
    "Properties": {
      "hbase.rootdir": "s3://MyBucket/MyHBaseStore",
      "emr.wal.workspace": "customWorkspaceName",
      "emr.wal.client.endpoint": "https://prod.emrwal.region.amazonaws.com"
    }
  },
  {
    "Classification": "hbase",
```

```

    "Properties": {
      "hbase.emr.storageMode": "s3",
      "hbase.emr.wal.enabled": "true"
    }
  }
]

```

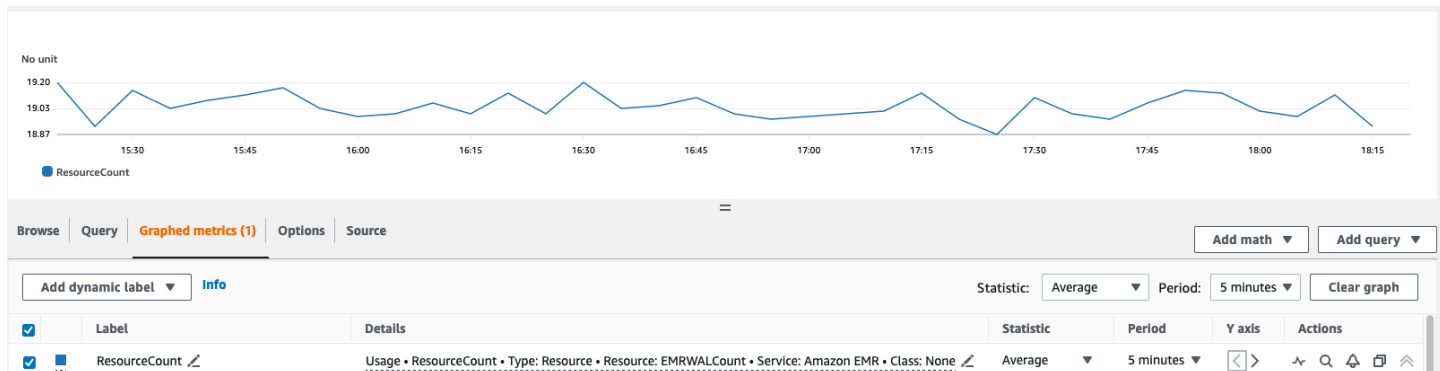
您也可以使用VPCE政策来允许或限制对Amazon的访问EMRWAL APIs。有关更多信息，请参阅AWS PrivateLink 指南中的[使用VPC终端节点策略控制对终端节点的访问](#)。

## 了解 Amazon 的EMRWAL定价和指标

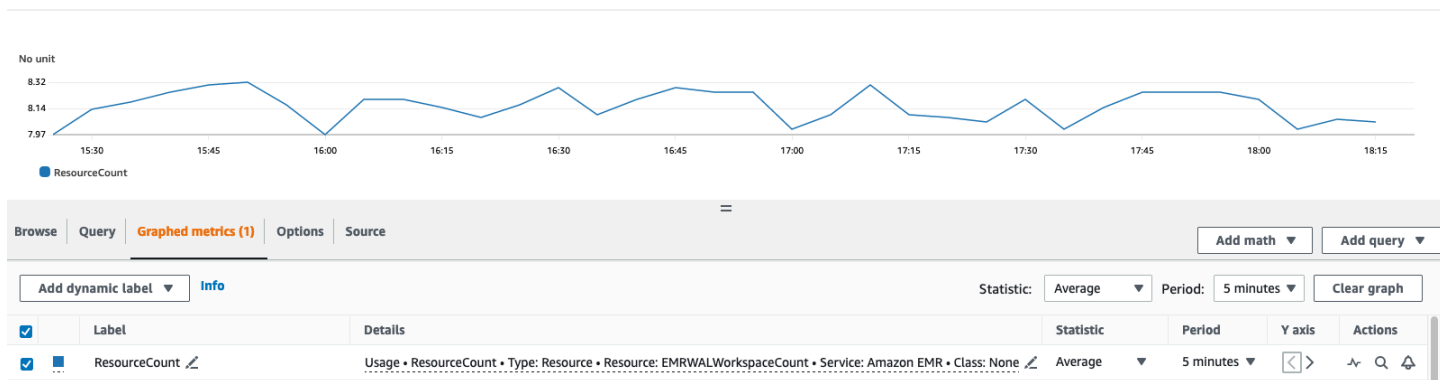
核心功能计费单位	详细信息
EMR-WAL-Read-GiB	API calls to read data from your table are billed as ReadRequestGiB. This includes <a href="#">Get and Scan</a> operations. Reads are charged based on the sizes of the read items. Amazon EMR bills at a minimum of 1 byte. For example, if you read a 1234.12 bytes item, you're charged for 1235 bytes. Reads are aggregated every hour for billing and shown as GiBs.
EMR-WAL-Write-GiB	API calls to write data from your table are billed as Write-GiB. This includes <a href="#">Put</a> operations. Writes are charged based on the sizes of the written items. Amazon EMR bills at a minimum of 1 byte. For example, if you write a 1234.12 bytes item, you're charged for 1235 bytes. Writes are aggregated every hour for billing and shown as GiBs.
EMR-WAL-WALHours	您在服务上存储WALs的数量按以下方式计费。EMR-WAL-WALHours Amazon EMR 为WAL每个HBase地区创建一个。例如，如果您创建了 20 个HBase表（包括系统表），并且每个表有两个HBase区域，那么您将使用 28,800 WAL 小时，计算公式为： <div style="border: 1px solid #ccc; border-radius: 10px; padding: 10px; margin-top: 10px;"> <math display="block">20 \text{ tables} \\ \times 2 \text{ Regions per table}</math> </div>

核心功能计费单位	详细信息
	x 1 WAL per Region x 30 days x 24 hours ----- 28,800 EMR-WAL-WALHours

### 示例EMRWALCount :



### 示例EMRWALWorkspaceCount :



## 为工作空间添加标签 WAL

您可以在创建新工作空间时向工作空间添加标签，也可以在活动工作区中为正在运行的集群添加、移除或列出标签。您无法为工作空间中的单个资源添加标签，也无法更新现有标签；相反，可以从工作区中移除不需要的标签并替换它们。

您可以从中标记工作空间。EMRWAL CLI有关为工作区添加标签的EMRWAL CLI命令列表，请参阅。[亚马逊 EMR WAL \(EMRWAL\) CLI 参考](#)

以下示例IAM策略说明了一种场景，该场景仅允许使用正确的标记

键resource\_tag\_allow\_test\_key和值resource\_tag\_allow\_test\_value进行工作空间CRUDL操作：

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Action": [
        "EMRWAL:*"
      ],
      "Effect": "Allow",
      "Resource": [
        "*",
        "*"
      ],
      "Condition": {
        "StringEquals": {
          "aws:ResourceTag/resource_tag_allow_test_key": [
            "resource_tag_allow_test_value"
          ]
        }
      }
    }
  ]
}
```

要验证现在工作区操作是否需要使用该[亚马逊 EMR WAL \(EMRWAL\) CLI 参考](#)标签，请使用tagAllowResourceTag为带有所需资源标签的工作空间调用[listTagsForResource](#)命令。如果您正确配置了条件，则命令将成功。

```
emrwal listTagsForResource -r us-east-1 -arn arn:aws:emrwal:us-east-1:arn:workspace/
tagAllowResourceTag
Tag(Key=resource_tag_allow_test_key, Value=resource_tag_allow_test_value)
```

## Amazon 的注意事项和区域 EMR WAL

### Amazon 的注意事项 EMR WAL

以下列表描述了 Amazon 的重要注意事项和限制 EMRWAL：

- 亚马逊EMRWAL可用于亚马逊 6.15.0 及更高EMR版本。

- Amazon EMR WAL 是一项可选的付费服务。您需要为实际使用量付费：读取、写入和数据存储。有关更多信息，请参阅[了解 Amazon 的EMRWAL定价和指标](#)和[Amazon EMR 定价](#)页面。
- Amazon EMR WAL 使用预HBase写日志 (WAL)。要使用 Amazon EMRWAL，您的集群必须使用 HBaseWAL。
- 要在创建集群EMRWAL时启用 Amazon，您必须拥有所需的角色权限。有关更多信息，请参阅[使用服务相关角色进行预写](#)日志。
- 使用、或创建集群EMRWAL时 AWS Management Console，AWS CLI必须启用 AmazonAPI，并且必须使用实例组配置。如果您没有使用 Amazon 创建集群，则无法EMRWAL在正在运行的集群中启用 Amazon EMR WAL。您也无法编辑hbase-site配置以在正在运行的集群EMRWAL中启用 Amazon。
- 您只能在使用 Amaz EMR WAL on S3 作为根目录的集群上启用 Amazon。
- 在 Amazon S3 中，不能在同一个HBase根目录下有多个活动集群。
- 您无法在只读副本集群EMRWAL上启用 Amazon。
- WAL在托管服务内的可用区之间进行复制。
- WAL寿命超过集群，并且仍然可用于下一个集群。
- EMRWAL在启动期间或集群运行时（处于运行状态），您无法禁用 Amazon。
- 有关工作空间限制的信息，请参阅[Amazon EMR 终端节点和配额](#)。WAL

## Amazon 的可用区域 EMR WAL

Amazon EMR WAL 服务有以下几种可供选择 AWS 区域：

- ap-northeast-1 – 亚太地区 ( 东京 )
- ap-southeast-1 – 亚太地区 ( 新加坡 )
- ap-south-1 – 亚太地区 ( 孟买 )
- ap-southeast-2 – 亚太地区 ( 悉尼 )
- eu-central-1 – 欧洲地区 ( 法兰克福 )
- eu-north-1 – 欧洲地区 ( 斯德哥尔摩 )
- eu-west-1 – 欧洲地区 ( 爱尔兰 )
- us-east-1 – 美国东部 ( 弗吉尼亚州北部 )
- us-east-2 – 美国东部 ( 俄亥俄州 )
- us-west-2 – 美国西部 ( 俄勒冈州 )
- ap-northeast-2 – 亚太地区 ( 首尔 )

## 亚马逊 EMR WAL (EMRWAL) CLI 参考

EMRWAL 命令行界面 (EMRWAL CLI) 是一个统一的工具，用于管理 Amazon 的预写日志 (WAL)。EMR 当您在创建 EMR 集群 WAL 时启用集群时，会 EMRWAL CLI 附带集群。有关启用的更多信息 WAL，请参阅 [Amazon 的预写日志 \(WAL\) EMR](#)。

EMRWAL CLI 包括以下命令：

### 主题

- [createWorkspace](#)
- [deleteWal](#)
- [deleteWorkspace](#)
- [listTagsForResource](#)
- [listWals](#)
- [listWorkspaces](#)
- [tagResource](#)
- [untagResource](#)

### createWorkspace

该 createWorkspace 命令将创建一个新的 Amazon EMR WAL 工作空间。

用法：

```
emrwal createWorkspace [-tags <tags>] [-e {endpoint}] [-r {Region}] -w {workspacename} [-h]
```

示例：

```
aws emr createWorkspace -w examples
```

### deleteWal

该 deleteWals 命令将删除您指定 EMRWAL 的 Amazon。

用法：

```
emrwal deleteWal [-e {endpoint}] [-r {Region}] [-w {workspacename}] [-p <tablePrefix>]
[-n <walName>] [-N <fullName>] [-R] [-m] [-h]
```

示例：

```
emrwal deleteWal -w examples -p hbaseTable -n exampleWal
```

## deleteWorkspace

该deleteWorkspace命令将删除您指定的 Amazon EMR WAL 工作空间。

用法：

```
emrwal deleteWorkspace [-e {endpoint}] [-r {Region}] -w {workspacename} [-h]
```

示例：

```
emrwal deleteWorkspace -w examples
```

## listTagsForResource

该listTagsForResource命令列出您为亚马逊EMRWAL工作空间指定的所有键值对标签。

用法：

```
emrwal listTagsForResource -arn {resource-arn} [-e {endpoint}] [-r {Region}] [-h]
```

示例：

```
emrwal listTagsForResource -arn arn:aws:emrwal::1234567891234:workspace/examples
```

## listWals

该listWals命令列出了您指定的工作空间EMRWALs中的所有 Amazon。

用法：

```
emrwal listWals [-nextToken {token-string}] [-pageSize {integer}] [-e {endpoint}] [-r {Region}] [-w {workspacename}] [-p <tablePrefix>] [-M {integer}] [-h]
```

示例：

```
emrwal listWals -w examplews
```

## listWorkspaces

该listWorkspaces命令列出了所有可供您使用的 Amazon EMR WAL 工作空间。

用法：

```
emrwal listWorkspaces [-nextToken {token-string}] [-pageSize {integer}] [-e {endpoint}]  
[-r {Region}] [-M {integer}] [-h]
```

示例：

```
emrwal listWorkspaces
```

## tagResource

该tagResource命令为您指定的 Amazon EMR WAL 工作空间分配一个或多个键值对标签。

用法：

```
emrwal tagResource -arn {resource-arn} -tags <tags> [-e {endpoint}] [-r {Region}] [-h]
```

示例：

```
emrwal tagResource -arn arn:aws:emrwal::1234567891234:workspace/examplews -  
tags tag_key=tag_value
```

## untagResource

该untagResource命令将一个或多个键值对标签取消分配给您指定的 Amazon EMR WAL 工作空间。

用法：

```
emrwal untagResource -arn {resource-arn} -tagKeys <tagKeys> [-e {endpoint}] [-  
r {Region}] [-h]
```



示例：

```
emrwal untagResource -arn arn:aws:emrwal::1234567891234:workspace/examples -  
tagKeys tag_key
```

## 使用外HBase壳

创建HBase集群后，下一步是连接到，HBase这样您就可以开始读取和写入数据（只读副本集群不支持数据写入）。你可以使用 [HBaseshell](#) 来测试命令。

打开外HBase壳

1. 用于SSH连接到HBase群集中的主服务器。有关如何使用连接到 Amazon EMR 主节点的信息 SSH，请参阅 [Amazon EMR 管理指南SSH中的使用连接到主节点](#)。
2. 运行 `hbase shell`。s HBase hell 打开时会出现类似于以下内容的提示。

```
hbase(main):001:0>
```

你可以从提示符下发出 HBase shell 命令。有关 shell 命令及其调用方法的更多信息，请在HBase提示符下键入 help，然后按 Enter。

## 创建表

通过以下命令可创建一个名为“t1”的表，该表拥有名为“f1”的单列系列。

```
hbase(main):001:0>create 't1', 'f1'
```

## 设置值

通过以下命令可为表“t1”和列“f1”中的行“r1”设置值“v1”。

```
hbase(main):001:0>put 't1', 'r1', 'f1:col1', 'v1'
```

## 获取值

通过以下命令可获取表“t1”中的行“r1”的值。

```
hbase(main):001:0>get 't1', 'r1'
```

## 删除表

以下命令将删除表“t1”。

```
hbase(main):001:0>drop 'ns1:t1',false
```

布尔值对应于您是否要存档表，因此如果要保存，可以将其设置为 `true`。您也可以运行不带布尔值的 `drop 'ns1:t1'` 来存档表。

## 使用 Hive 访问HBase表

HBase并[Apache Hive](#)紧密集成，允许您直接对存储在中的HBase数据运行大规模并行处理工作负载。要将 Hive 与配合使用HBase，您通常可以在同一个集群上启动它们。但是，您可以在单独的集群HBase上启动 Hive 和。在不同的集群上分别运行HBase和 Hive 可以提高性能，因为这样可以使每个应用程序更有效地使用集群资源。

以下过程说明如何使用 Hiv HBase e 连接到集群。

### Note

您只能将 Hive 集群连接到单个HBase集群。

### 将 Hive 连接到 HBase

1. 使用并HBase安装了 Hive 创建单独的集群，或者创建一个同时安装了 Hive HBase 和 Hive 的单个集群。
2. 如果您使用的是单独的集群，请修改您的安全组HBase，以便在这两个主节点之间打开 Hive 端口。
3. 用于SSH连接到安装了 Hive 的群集的主节点。有关更多信息，请参阅 [《Amazon EMR 管理指南》SSH中的“使用连接到主节点”](#)。
4. 通过以下命令启动 Hive shell。

```
hive
```

5. (可选) 如果HBase和 Hive 位于同一个集群上, 则无需执行此操作。将 Hive 集群上的HBase客户端连接到包含您的数据的HBase集群。在以下示例中, *public-DNS-name* 将替换为HBase群集主节点的公共DNS名称, 例如: `ec2-50-19-76-67.compute-1.amazonaws.com`。

```
set hbase.zookeeper.quorum=public-DNS-name;
```

6. 根据需要进行您的HBase数据运行 Hive 查询, 或者查看下一个过程。

## 从 Hive 访问HBase数据

- 在 Hive 和HBase集群之间建立连接后 (如前面的过程所示), 您可以通过在 Hive 中创建外部表来访问存储在HBase集群上的数据。

以下示例从主节点上的 Hive 提示符运行时, 会创建一个外部表, 该表引用存储在名为inputTable为的HBase表中的数据。然后, 您可以inputTable在 Hive 语句中引用以查询和修改存储在HBase集群中的数据。

```
set hbase.zookeeper.quorum=ec2-107-21-163-157.compute-1.amazonaws.com;

create external table inputTable (key string, value string)
  stored by 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
  with serdeproperties ("hbase.columns.mapping" = ":key,f1:col1")
  tblproperties ("hbase.table.name" = "t1");

select count(key) from inputTable ;
```

有关更高级的用例和组合HBase和 Hive 的示例, 请参阅 AWS 大数据博客文章 [《Combine No》](#) 和 [Apache Hive 在亚马逊上使用 Apache SQL 和 Apache HBase Hive 进行大规模并行分析](#)。EMR

## 使用HBase快照

HBase使用内置[快照](#)功能创建表的轻量级备份。在EMR集群中, 可以使用将这些备份导出到 Amazon S3 EMRFS。您可以使用 `sh HBase ell` 在主节点上创建快照。本主题向您展示如何与 shell 交互式运行这些命令, 或者通过使用或的步骤来`command-runner.jar`运行这些命令。AWS CLI AWS SDK for Java有关其他类型HBase备份的更多信息, 请参阅HBase文档中的[HBase备份](#)。

## 使用表创建快照

```
hbase snapshot create -n snapshotName -t tableName
```

使用来自 AWS CLI : `command-runner.jar`

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \  
--steps Name="HBase Shell Step",Jar="command-runner.jar",\  
Args=[ "hbase", "snapshot", "create", "-n", "snapshotName", "-t", "tableName"]
```

### AWS SDK for Java

```
HadoopJarStepConfig hbaseSnapshotConf = new HadoopJarStepConfig()  
.withJar("command-runner.jar")  
.withArgs("hbase", "snapshot", "create", "-n", "snapshotName", "-t", "tableName");
```

#### Note

如果您的快照名称不唯一，则创建操作将失败，并返回 -1 或者 255，但您可能看不到说明出现问题的错误消息。要使用相同的快照名称，请先将其删除，然后重新创建。

## 删除快照

```
hbase shell  
>> delete_snapshot 'snapshotName'
```

## 查看快照信息

```
hbase snapshot info -snapshot snapshotName
```

## 将快照导出到 Amazon S3

#### Important

如果在导出快照时未指定 `-mappers` 值，则 HBase 使用任意计算来确定映射器的数量。此值可能非常大，具体取决于表大小，这会在导出过程中对正在运行的任务产生负面影响。为此，

我们建议您指定 `-mappers` 参数或 `-bandwidth` 参数（指定每秒使用的带宽，以 MB 为单位），或同时指定这两个参数以限制导出操作使用的集群资源。或者，您可以在低使用率期间运行导出快照操作。

```
hbase snapshot export -snapshot snapshotName \
-copy-to s3://bucketName/folder -mappers 2
```

`command-runner.jar`从以下位置使用 AWS CLI：

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \
--steps Name="HBase Shell Step",Jar="command-runner.jar",\
Args=[ "hbase", "snapshot", "export", "-snapshot", "snapshotName", "-copy-to", "s3://bucketName/folder", "-mappers", "2", "-bandwidth", "50"]
```

AWS SDK for Java:

```
HadoopJarStepConfig hbaseImportSnapshotConf = new HadoopJarStepConfig()
    .withJar("command-runner.jar")
    .withArgs("hbase", "snapshot", "export",
        "-snapshot", "snapshotName", "-copy-to",
        "s3://bucketName/folder",
        "-mappers", "2", "-bandwidth", "50");
```

## 从 Amazon S3 导入快照

尽管这是导入，但此处使用的HBase选项仍然是export。

```
sudo -u hbase hbase snapshot export \
-D hbase.rootdir=s3://bucketName/folder \
-snapshot snapshotName \
-copy-to hdfs://masterPublicDNSName:8020/user/hbase \
-mappers 2
```

`command-runner.jar`从以下位置使用 AWS CLI：

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \
--steps Name="HBase Shell Step",Jar="command-runner.jar", \
Args=["sudo", "-u", "hbase", "hbase snapshot export", "-snapshot", "snapshotName", \
"-D", "hbase.rootdir=s3://bucketName/folder", \
```

```
"-copy-to", "hdfs://masterPublicDNSName:8020/user/hbase", "-mappers", "2", "-chmod", "700"]
```

## AWS SDK for Java:

```
HadoopJarStepConfig hbaseImportSnapshotConf = new HadoopJarStepConfig()
    .withJar("command-runner.jar")
    .withArgs("sudo", "-u", "hbase", "hbase", "snapshot", "export", "-D", "hbase.rootdir=s3://
path/to/snapshot",
    "-snapshot", "snapshotName", "-copy-to",
    "hdfs://masterPublicDNSName:8020/user/hbase",
    "-mappers", "2", "-chuser", "hbase");
```

## 从 s HBase hell 中的快照还原表

```
hbase shell
>> disable tableName
>> restore_snapshot snapshotName
>> enable tableName
```

HBase目前不支持 s HBase hell 中的所有快照命令。例如，没有用于恢复快照的HBase命令行选项，因此您必须在 shell 中恢复快照。这意味着 `command-runner.jar` 必须运行 Bash 命令。

### Note

由于此处使用的命令是echo，因此，即使 Amazon 运行的命令EMR返回0退出代码，您的 shell 命令仍可能失败。如果选择将 shell 命令作为步骤运行，请检查步骤日志。

```
echo 'disable tableName; \
restore_snapshot snapshotName; \
enable tableName' | hbase shell
```

以下是使用 AWS CLI 的步骤。首先，创建以下 `snapshot.json` 文件：

```
[
  {
    "Name": "restore",
    "Args": ["bash", "-c", "echo '$'disable \"tableName\"; restore_snapshot
\"snapshotName\"; enable \"tableName\"' | hbase shell"],
    "Jar": "command-runner.jar",
```

```

    "ActionOnFailure": "CONTINUE",
    "Type": "CUSTOM_JAR"
  }
]

```

```

aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \
--steps file:///./snapshot.json

```

## AWS SDK for Java:

```

HadoopJarStepConfig hbaseRestoreSnapshotConf = new HadoopJarStepConfig()
    .withJar("command-runner.jar")
    .withArgs("bash","-c","echo '$disable \"tableName\"; restore_snapshot \"snapshotName\"; enable \"snapshotName\"' | hbase shell");

```

## 配置 HBase

尽管默认HBase设置应该适用于大多数应用程序，但您可以修改HBase配置设置。为此，请使用HBase配置分类的属性。有关更多信息，请参阅 [配置应用程序](#)。

以下示例基于存储在 Amazon S3 中的配置文件创建具有备用HBase根目录的集群。myConfig.json

### Note

为了便于读取，包含 Linux 行继续符 (\)。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (^)。

```

aws emr create-cluster --release-label emr-7.2.0 --applications Name=HBase \
--instance-type m5.xlarge --instance-count 3 --configurations https://s3.amazonaws.com/mybucket/myfolder/myConfig.json

```

myConfig.json 文件指定 hbase-site 配置分类的 hbase.rootdir 属性，如以下示例中所示。Replace ( 替换 ) *ip-XXX-XX-XX-XXX.ec2.internal* 使用群集主节点的内部DNS主机名。

```

[
  {
    "Classification": "hbase-site",
    "Properties": {

```

```

    "hbase.rootdir": "hdfs://ip-XXX-XX-XX-XXX.ec2.internal:8020/user/
myCustomHBaseDir"
  }
}
]

```

### Note

在 Amazon 5.21.0 及更高EMR版本中，您可以覆盖集群配置，并为正在运行的集群中的每个实例组指定其他配置分类。为此，您可以使用 Amazon EMR 控制台、AWS Command Line Interface (AWS CLI) 或 AWS SDK。有关更多信息，请参阅[为运行的集群中的实例组提供配置](#)。

## 中对内存分配的更改 YARN

HBase不是作为YARN应用程序运行的，因此必须重新计算分配给YARN其应用程序的内存，这YARN会HBase导致安装后可用的总内存减少。在计划将YARN应用程序放在同一个集群HBase上时，应将此考虑在内。内存小于 64 GB 的实例类型拥有一半的可用内存NodeManager，然后将其分配给HBase RegionServer。对于内存大于 64 GB 的实例类型，HBase RegionServer 内存上限为 32 GB。通常，YARN设置内存是 MapReduce 缩减器任务内存的几倍。

中的表格[任务配置设置的默认值](#)显示了根据所需的内存对YARN设置所做的更改HBase。

## HBase端口号

为其选择的HBase某些端口号与默认端口号不同。以下是 Amazon HBase 上的接口和端口EMR。

### HBase端口

接口	端口	协议
HMaster	16000	TCP
HMaster用户界面	16010	HTTP
RegionServer	16020	TCP
RegionServer 信息	16030	HTTP
REST服务器	8070	HTTP



接口	端口	协议
REST用户界面	8085	HTTP
Thrift 服务器	9090	TCP
Thrift 服务器 UI	9095	HTTP

### Important

在kms-http-port亚马逊EMR发布的版本4.6.0及更高版本中，kms-admin-port是9701，是9701。

## HBase要优化的网站设置

您可以设置任何或全部HBase站点设置，以针对应用程序的工作负载优化HBase集群。建议您将以下设置作为调查的起点。

### zookeeper.session.timeout

默认超时时间为 40 秒（40000 毫秒）。如果区域服务器崩溃，则上述值是主服务器注意到区域服务器缺少及开始恢复所需的时长。要帮助更快地恢复主服务器，可以将此值减小到更短的时间。以下示例使用的是 30 秒或 30000 毫秒：

```
[
  {
    "Classification": "hbase-site",
    "Properties": {
      "zookeeper.session.timeout": "30000"
    }
  }
]
```

### hbase.regionserver.handler.count

这可定义区域服务器保持打开状态以向表提供请求的线程数。默认值 10 较低，以防止用户在使用具有大量并发客户端的大写缓冲区时终止其区域服务器。经验法则是，当每个请求的有效负载接近 MB 范

围 ( 大放置、使用大缓存进行扫描 ) 时, 将此数字保持在较低水平; 当有效负载较小 ( 获取、小看跌期权ICVs、删除 ) 时, 应将此数字保持在较低水平。以下示例将打开的线程数提高到 30 :

```
[
  {
    "Classification": "hbase-site",
    "Properties": {
      "hbase.regionserver.handler.count": "30"
    }
  }
]
```

### hbase.hregion.max.filesize

此参数控制单个区域的大小 ( 以字节为单位 )。默认情况下, 将它设置为 1073741824。如果您正在向HBase集群中写入大量数据, 并且导致频繁拆分, 则可以增加此大小以扩大单个区域。这会减少拆分, 但需要更多时间来实现从一台服务器到另一台服务器的区域负载均衡。

```
[
  {
    "Classification": "hbase-site",
    "Properties": {
      "hbase.hregion.max.filesize": "1073741824"
    }
  }
]
```

### hbase.hregion.memstore.flush.size

此参数控制 Memstore 刷新到磁盘之前的大小上限 ( 以字节为单位 )。默认为 134217728。如果您的工作负载由短时间突发的写入操作组成, 您可能希望增加此限制, 以便所有写入在突发期间都保留在内存中, 并在稍后刷新到磁盘。这可以提高突发期间的性能。

```
[
  {
    "Classification": "hbase-site",
    "Properties": {
      "hbase.hregion.memstore.flush.size": "134217728"
    }
  }
]
```

## 查看HBase用户界面

### Note

默认情况下，HBase用户界面使用不安全的HTTP连接。要启用 secure HTTP (HTTPS)，请在[HBase配置true](#)中将hbase-site分类的hbase.ssl.enabled属性设置为。有关在HBase Web 用户界面中使用 secure HTTP (HTTPS) 的更多信息，请参阅《[Apache HBase 参考指南](#)》。

HBase提供了一个基于 Web 的用户界面，可用于监控您的HBase集群。当您HBase在 Amazon 上运行时EMR，Web 界面将在主节点上运行，并且可以使用端口转发（也称为创建SSH隧道）进行查看。

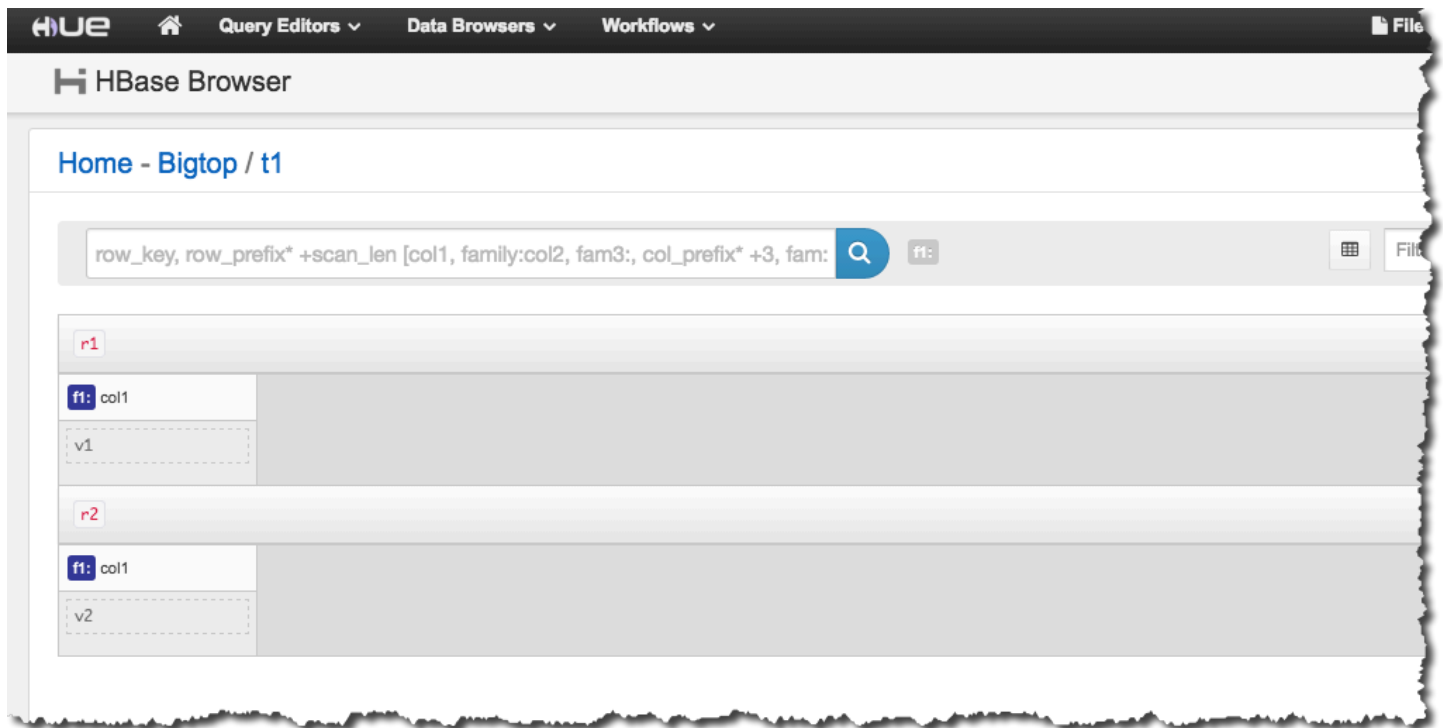
### 查看HBase用户界面

1. SSH用于通过隧道进入主节点并创建安全连接。有关更多信息，请参阅 Amazon EMR 管理指南中的[选项 2，第 1 部分：使用动态端口转发设置通往主节点的SSH隧道](#)。
2. 安装带有代理工具（例如 Firefox FoxyProxy 插件）的 Web 浏览器，为 AWS 域名创建SOCKS代理。有关更多信息，请参阅《亚马逊EMR管理指南》中的[选项 2，第 2 部分：配置代理设置以查看主节点上托管的网站](#)。
3. 设置代理并打开SSH连接后，您可以通过 `http://打开浏览器窗口来查看HBase用户界面`***master-public-dns-name***:16010/ master-status，其中 ***master-public-dns-name*** 是群集主节点的公共DNS地址。

The screenshot shows the Apache HBase web interface. At the top, there is a navigation bar with links: Home, Table Details, Local Logs, Log Level, Debug Dump, Metrics Dump, and HBase Configuration. Below the navigation bar, the page title is "Master [redacted].ec2.internal". The main content area is titled "Region Servers" and has several tabs: Base Stats (selected), Memory, Requests, Storefiles, and Compactions. Below the tabs is a table with the following data:

ServerName	Start time	Version	Requests Per Second	Num. F
[redacted].ec2.internal,16020,1461165084992	Wed Apr 20 15:11:24 UTC 2016	1.2.0	0	1
[redacted].ec2.internal,16020,1461165087881	Wed Apr 20 15:11:27 UTC 2016	1.2.0	0	2
Total:2			0	3

您也可以在 Hue HBase 中查看。例如，下面显示了[使用外HBase壳](#)中创建的表 t1：



有关 Hue 的更多信息，请参阅[Hue](#)。

## 查看 HBase 日志文件

作为其操作的一部分，HBase写入包含有关配置设置、守护程序操作和异常的详细信息的日志文件。这些日志文件可用于调试问题HBase以及跟踪性能。

如果您配置集群以将日志文件保存到 Amazon S3，您应当知道，日志会每五分钟写入 Amazon S3 一次，因此在最新的日志文件可用之前，可能会出现轻微延迟。

### 查看主节点上的HBase日志

- 您可以使用连接到主节点并导航SSH到该/var/log/hbase目录来查看当前HBase日志。除非您在集群启动时启用了针对 Amazon S3 的日志记录，否则这些日志将在集群终止后不再可用。

### 在 Amazon S3 上查看HBase日志

- 要访问 HBase Amazon S3 上的日志和其他集群日志，并使其在集群终止后可用，请在创建集群时指定一个 Amazon S3 存储桶以接收这些日志。这可以使用选项 `--log-uri` 实现。有关为集群启用日志记录的更多信息，请参阅 Amazon EMR 管理指南中的[配置日志和调试（可选）](#)。

## HBase使用 Ganglia 进行监控

Ganglia 开源项目是一个可扩展的分布式系统，旨在监控集群和网格，同时尽量减少对其性能的影响。当您在集群上启用 Ganglia 时，您可以生成报告并查看整个集群的性能，还可以检查单个节点实例的性能。有关 Ganglia 开源项目的更多信息，请参阅 <http://ganglia.info/>。有关将 Ganglia 与 Amazon EMR 集群配合使用的更多信息，请参阅[Ganglia](#)。

在配置了 Ganglia 的情况下启动集群后，您可以使用主节点上运行的图形界面访问 Ganglia 图表和报告。

Ganglia 将日志文件存储在主节点上的 `/mnt/var/lib/ganglia/rrds/` 目录中。Amazon 的早期版本 EMR 可能会将日志文件存储在 `/var/log/ganglia/rrds/` 目录中。

### 要为 Ganglia 配置集群并HBase使用 AWS CLI

- 使用类似于以下内容的 `create-cluster` 命令：

```
aws emr create-cluster --name "Test cluster" --release-label emr-7.2.0 \  
--applications Name=HBase Name=Ganglia --use-default-roles \  
--ec2-attributes KeyName=myKey --instance-type m5.xlarge \  
--instance-count 3
```

#### Note

如果默认的 Amazon EMR 服务角色和 Amazon EC2 实例配置文件不存在，则会发生错误。请使用 `aws emr create-default-roles` 命令创建它们，然后重试。

有关更多信息，请参阅[中的 Amazon EMR 命令 AWS CLI](#)。

### 在 Ganglia 网页界面中查看HBase指标

1. SSH用于通过隧道进入主节点并创建安全连接。有关更多信息，请参阅 Amazon EMR 管理指南中的[选项 2，第 1 部分：使用动态端口转发设置通往主节点的SSH隧道](#)。
2. 安装带有代理工具（例如 Firefox FoxyProxy 插件）的 Web 浏览器，为 AWS 域名创建SOCKS代理。有关更多信息，请参阅《亚马逊EMR管理指南》中的[选项 2，第 2 部分：配置代理设置以查看主节点上托管的网站](#)。

3. 设置代理并打开SSH连接后，您可以通过 <http://打开浏览器窗口来查看 Ganglia 指标>`master-public-dns-name/ganglia/`，哪里 `master-public-dns-name` 是HBase群集中主服务器的公共DNS地址。

### 在主节点上查看 Ganglia 日志文件

- 如果集群仍在运行，则可以使用连接到主节点并导航到该`/mnt/var/lib/ganglia/rrds/`目录来访问日志文件。SSH对于 EMR 3.x，请导航到该`/var/log/ganglia/rrds`目录。有关更多信息，请参阅 [《Amazon EMR 管理指南》SSH中的“使用连接到主节点”](#)。

### 查看 Amazon S3 上的 Ganglia 日志文件

- 即使您为集群启用日志记录，Ganglia 日志文件也不会自动写入 Amazon S3。要在 Amazon S3 上查看 Ganglia 日志文件，您必须手动将日志从 `/mnt/var/lib/ganglia/rrds/` 推送到 S3 存储桶。

## 从以前的HBase版本迁移

要迁移先前HBase版本的数据，请参阅《Apache HBase 参考指南》中的[升级以及HBase版本号和兼容性](#)。您可能需要特别注意从 1.0 之前版本升级的HBase要求。

## HBase发布历史

下表列出了 Amazon 的每个发行版本中HBase包含的版本EMR，以及随应用程序一起安装的组件。有关每个发行版本中的组件版本，请参阅 [亚马逊 EMR 7.x 发布版本](#)、[亚马逊 EMR 6.x 发布版本](#) 或 [亚马逊 EMR 5.x 发布版本](#) 中的发行版“组件版本”部分。

### HBase版本信息

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-7.2.0	2.4.17	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode,

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
		hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server
emr-5.36.2	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-7.1.0	2.4.17	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server



亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-7.0.0	2.4.17	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-6.15.0	2.4.17	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-6.14.0	2.4.17	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-6.13.0	2.4.17	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-6.12.0	2.4.17	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-6.11.1	2.4.15	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-6.11.0	2.4.15	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-6.10.1	2.4.15	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server



亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-6.10.0	2.4.15	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server
emr-6.9.1	2.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-6.9.0	2.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server
emr-6.8.1	2.4.12	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-6.8.0	2.4.12	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server
emr-6.7.0	2.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.36.1	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.36.0	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-6.6.0	2.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server
emr-5.35.0	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-6.5.0	2.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-6.4.0	2.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-6.3.1	2.2.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-6.3.0	2.2.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-6.2.1	2.2.6-amzn-0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-6.2.0	2.2.6-amzn-0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server



亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-6.1.1	2.2.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-6.1.0	2.2.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-6.0.1	2.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-6.0.0	2.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.34.0	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.33.1	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.33.0	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.32.1	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.32.0	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.31.1	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.31.0	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.30.2	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.30.1	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.30.0	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.29.0	1.4.10	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.28.1	1.4.10	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server



亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.28.0	1.4.10	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.27.1	1.4.10	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.27.0	1.4.10	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.26.0	1.4.10	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.25.0	1.4.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.24.1	1.4.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.24.0	1.4.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.23.1	1.4.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.23.0	1.4.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.22.0	1.4.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.21.2	1.4.8	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.21.1	1.4.8	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.21.0	1.4.8	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.20.1	1.4.8	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.20.0	1.4.8	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.19.1	1.4.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server



亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.19.0	1.4.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.18.1	1.4.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.18.0	1.4.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.17.2	1.4.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.17.1	1.4.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.17.0	1.4.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.16.1	1.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.16.0	1.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.15.1	1.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.15.0	1.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.14.2	1.4.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.14.1	1.4.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.14.0	1.4.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.13.1	1.4.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.13.0	1.4.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.12.3	1.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server



亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.12.2	1.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.12.1	1.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.12.0	1.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.11.4	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.11.3	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.11.2	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.11.1	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.11.0	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.10.1	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.10.0	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.9.1	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.9.0	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.8.3	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.8.2	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.8.1	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.8.0	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server



亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.7.1	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.7.0	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.6.1	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.6.0	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.5.4	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.5.3	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.5.2	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.5.1	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.5.0	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.4.1	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.4.0	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.3.2	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.3.1	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.3.0	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.2.3	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.2.2	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server



亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.2.1	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.2.0	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.1.1	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.1.0	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.0.3	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.0.2	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-5.0.1	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.0.0	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-4.9.6	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.9.5	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-4.9.4	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.9.3	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-4.9.2	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.9.1	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-4.8.5	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.8.4	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server



亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-4.8.3	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.8.2	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-4.8.1	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.8.0	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-4.7.4	1.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.7.3	1.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-4.7.2	1.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.7.1	1.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-4.7.0	1.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.6.1	1.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

亚马逊EMR发布标签	HBase版本	安装在一起的组件 HBase
emr-4.6.0	1.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

# Apache HCatalog

HCatalog是一款允许您访问 Pig、Spark SQL 和/或自定义 MapReduce 应用程序中的 Hive 元数据仓库的工具。HCatalog具有允许您创建表或执行其他操作的REST界面和命令行客户端。然后，您编写应用程序以使用HCatalog库访问表。有关更多信息，请参阅[使用 HCatalog](#)。HCatalog包含在 Amazon EMR 发行版 4.4.0 及更高版本中。

HCatalog在亚马逊上EMR发布版本 5.8.0 及更高版本支持使用 Glue 数据 AWS 目录作为 Hive 的元数据库。有关更多信息，请参阅[使用 AWS Glue 数据目录作为 Hive 的元数据库](#)。

下表列出了最新版本的 Amazon EMR 6.x 系列中HCatalog包含的版本以及亚马逊随之EMR安装的组件。HCatalog

有关此版本HCatalog中安装的组件的版本，请参阅 [6.15.0 版组件版本](#)。

HCatalogemr-6.15.0 的版本信息

Amazon EMR 发布标签	HCatalog版本	随 HCatalog 安装的组件
emr-6.15.0	HCatalog 3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

下表列出了最新版本的 Amazon EMR 5.x 系列中HCatalog包含的版本以及亚马逊随之EMR安装的组件。HCatalog

有关此版本HCatalog中随安装的组件的版本，请参阅 [5.36.2 版组件版本](#)。

## HCatalogemr-5.36.2 的版本信息

Amazon EMR 发布标签	HCatalog版本	随 HCatalog 安装的组件
emr-5.36.2	HCatalog 2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

### 主题

- [使用 HCatalog 创建集群](#)
- [使用 HCatalog](#)
- [示例：创建HCatalog表并使用 Pig 写入该表](#)
- [HCatalog发布历史](#)

## 使用 HCatalog 创建集群

尽管包含HCatalog在 Hive 项目中，但您必须将其作为自己的应用程序进行安装。

### 使用控制台启动HCatalog已安装的集群

以下过程创建HCatalog已安装的集群。有关使用控制台创建集群的更多信息，包括高级选项，请参阅 Amazon EMR 管理指南中的[规划和配置集群](#)。

1. 打开亚马逊EMR控制台，[网址为 https://console.aws.amazon.com/emr](https://console.aws.amazon.com/emr)。
2. 选择 Create cluster (创建集群) 以使用 Quick Create (快速创建)。



3. 对于 Software Configuration (软件配置) 字段，选择 Amazon Release Version emr-4.4.0 (Amazon 发行版 emr-4.4.0) 或更高版本。
4. 在“选择应用程序”字段中，选择“所有应用程序”或HCatalog。
5. 根据需要选择其它选项，然后选择 Create cluster (创建集群)。

## 要HCatalog使用启动集群 AWS CLI

- 使用下面的命令创建集群：

### Note

为了便于读取，包含 Linux 行继续符 ( \ )。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (^)。

```
aws emr create-cluster --name "Cluster with Hcat" --release-label emr-7.2.0 \  
--applications Name=HCatalog --ec2-attributes KeyName=myKey \  
--instance-type m5.xlarge --instance-count 3 --use-default-roles
```

## 使用 HCatalog

您可以在使用 HCatalog Hive 元数据仓库的各种应用程序中使用。本节中的示例展示了如何创建表并在 Pig 和 Spark 的上下文中使用该表SQL。

### 使用时禁用直接写入 HCatalog HStorer

每当应用程序使用[HCatStorer](#)写入存储在 Amazon S3 中的HCatalog表时，请禁用 Amazon 的直接写入功能EMR。例如，在使用 Pig STORE 命令或运行将HCatalog表写入 Amazon S3 的 Sqoop 任务时禁用直接写入。您可以通过将 `mapred.output.direct.NativeS3FileSystem` 和 `mapred.output.direct.EmrFileSystem` 配置设置为 `false` 来禁用直接写入功能。以下示例演示如何使用 Java 设置这些配置。

```
Configuration conf = new Configuration();  
conf.set("mapred.output.direct.NativeS3FileSystem", "false");  
conf.set("mapred.output.direct.EmrFileSystem", "false");
```

## 使用创建表HCatCLI并在 Pig 中使用该数据

在您的集群上创建以下脚本 impressions.q :

```
CREATE EXTERNAL TABLE impressions (
  requestBeginTime string, adId string, impressionId string, referrer string,
  userAgent string, userCookie string, ip string
)
PARTITIONED BY (dt string)
ROW FORMAT
  serde 'org.apache.hive.hcatalog.data.JsonSerDe'
  with serdeproperties ( 'paths'='requestBeginTime, adId, impressionId, referrer,
userAgent, userCookie, ip' )
LOCATION 's3://[your region].elasticmapreduce/samples/hive-ads/tables/impressions/';
ALTER TABLE impressions ADD PARTITION (dt='2009-04-13-08-05');
```

使用以下命令执行脚本 HCatCLI :

```
% hcat -f impressions.q
Logging initialized using configuration in file:/etc/hive/conf.dist/hive-
log4j.properties
OK
Time taken: 4.001 seconds
OK
Time taken: 0.519 seconds
```

打开 Grunt shell 并访问 impressions 中的数据 :

```
% pig -useHCatalog -e "A = LOAD 'impressions' USING
  org.apache.hive.hcatalog.pig.HCatLoader();
B = LIMIT A 5;
dump B;"
<snip>
(1239610346000,m9nwdo67Nx6q2kI25qt50n7peICfUM,omkxkaRpNhGPDucAiBErSh1cs0MThC,cartoonnetwork.com
  (compatible; MSIE 7.0; Windows NT 6.0; FunWebProducts; GTB6; SLCC1; .NET CLR
  2.0.50727; Media Center PC
  5.0; .NET,wcVWWTascoPbGt6bdqDbuWTPPHgOPs,69.191.224.234,2009-04-13-08-05)
(1239611000000,NjriQjd0DgWBKnkGJUP6GNTbDeK4An,AwtXPkfaWG0aNeL900sFU8Hcj6eLHt,cartoonnetwork.com
  (compatible; MSIE 7.0; Windows NT 5.1; GTB6; .NET CLR
  1.1.4322),0aMU1F2gE4CtADVHAbKjjRRks5kIgg,57.34.133.110,2009-04-13-08-05)
```

```
(1239610462000,Irpv3oiu0I5QNQiwSSTIshrLdo9cM1,i1LDq44LRSJF0hbmhB8Gk7k9gMwtBq,cartoonnetwork.com
(compatible; MSIE 6.0; Windows NT 5.2; SV1; .NET CLR 1.1.4322;
InfoPath.1),Qsb3wkLR4JAIut4Uq6FNFQIR1rCVwU,42.174.193.253,2009-04-13-08-05)
(1239611007000,q2Aawfnpe0JAvhInaIp0VGx9Kts0oP0,s3HvTf1PB8JIE0IuM6h0EebWwp0tJV,cartoonnetwork.com
(compatible; MSIE 6.0; Windows NT 5.2; SV1; .NET CLR 1.1.4322;
InfoPath.1),Qsb3wkLR4JAIut4Uq6FNFQIR1rCVwU,42.174.193.253,2009-04-13-08-05)
(1239610398000,c362vpAB0soPKGHR543cj6TRwNe0Gn,jeas5nXbQInGAgFB8jlkhnprN6cMw7,cartoonnetwork.com
(compatible; MSIE 8.0; Windows NT 5.1; Trident/4.0; GTB6; .NET CLR
1.1.4322),k96n5PnUmwHKfiUI0TFP0TNMFADgh9,51.131.29.87,2009-04-13-08-05)
7120 [main] INFO org.apache.pig.Main - Pig script completed in 7 seconds and 199
milliseconds (7199 ms)
16/03/08 23:17:10 INFO pig.Main: Pig script completed in 7 seconds and 199 milliseconds
(7199 ms)
```

## 使用 Spark 访问表格 SQL

此示例根据第一个示例中创建的表创建 Spark DataFrame，并显示前 20 行：

```
% spark-shell --jars /usr/lib/hive-hcatalog/share/hcatalog/hive-hcatalog-core-1.0.0-
amzn-3.jar
<snip>
scala> val hiveContext = new org.apache.spark.sql.hive.HiveContext(sc);
scala> val df = hiveContext.sql("SELECT * FROM impressions")
scala> df.show()
<snip>
16/03/09 17:18:46 INFO DAGScheduler: ResultStage 0 (show at <console>:32) finished in
10.702 s
16/03/09 17:18:46 INFO DAGScheduler: Job 0 finished: show at <console>:32, took
10.839905 s
+-----+-----+-----+-----+
+-----+-----+-----+-----+
|requestbegintime|          adid|    impressionid|    referrer|
|  useragent|    usercookie|          ip|          dt|
+-----+-----+-----+-----+
+-----+-----+-----+-----+
|  1239610346000|m9nwdo67Nx6q2kI25...|omkxkaRpNhGPDucAi...|cartoonnetwork.com|
Mozilla/4.0 (comp...|wcVWWTascoPbGt6bd...|69.191.224.234|2009-04-13-08-05|
|  1239611000000|NjriQjd0DgWBKnkGJ...|AWtXPkfaWG0aNeL90...|cartoonnetwork.com|
Mozilla/4.0 (comp...|0aMU1F2gE4CtADVHA...| 57.34.133.110|2009-04-13-08-05|
|  1239610462000|Irpv3oiu0I5QNQiwS...|i1LDq44LRSJF0hbmh...|cartoonnetwork.com|
Mozilla/4.0 (comp...|Qsb3wkLR4JAIut4Uq...|42.174.193.253|2009-04-13-08-05|
|  1239611007000|q2Aawfnpe0JAvhInaI...|s3HvTf1PB8JIE0IuM...|cartoonnetwork.com|
Mozilla/4.0 (comp...|Qsb3wkLR4JAIut4Uq...|42.174.193.253|2009-04-13-08-05|
```

```

| 1239610398000|c362vpAB0soPKGHRs...|jeas5nXbQInGAgFB8...|cartoonnetwork.com|
Mozilla/4.0 (comp...|k96n5PnUmwHKfiUI0...| 51.131.29.87|2009-04-13-08-05|
| 1239610600000|cjbTpruoaiEtqLuMX...|XwlohBSs8Ipxs1bRa...|cartoonnetwork.com|
Mozilla/4.0 (comp...|k96n5PnUmwHKfiUI0...| 51.131.29.87|2009-04-13-08-05|
| 1239610804000|Ms3eJHNAEItpxvimd...|4SIj4pGmgVL1625BD...|cartoonnetwork.com|
Mozilla/4.0 (comp...|k96n5PnUmwHKfiUI0...| 51.131.29.87|2009-04-13-08-05|
| 1239610872000|h5bccHX6wJReDi1jL...|EFAWiiBdVfnxwAMWP...|cartoonnetwork.com|
Mozilla/4.0 (comp...|k96n5PnUmwHKfiUI0...| 51.131.29.87|2009-04-13-08-05|
| 1239610365000|874NBpGmxNFfxEPKM...|xSvE4XtGbdTXPF2Lb...|cartoonnetwork.com|
Mozilla/5.0 (Maci...|eWDEVVUphlnRa273j...| 22.91.173.232|2009-04-13-08-05|
| 1239610348000|X8gISpUTSgh1A5reS...|TrFblGT99AgE75vuj...| corriere.it|
Mozilla/4.0 (comp...|tX1sMpnhJUhmAF7AS...| 55.35.44.79|2009-04-13-08-05|
| 1239610743000|kbKreLWB6QVueFrDm...|kVnxx9Ie2i30LTxFj...| corriere.it|
Mozilla/4.0 (comp...|tX1sMpnhJUhmAF7AS...| 55.35.44.79|2009-04-13-08-05|
| 1239610812000|9lx0SRpEi3bmEeTCu...|1B2sff99AEIwSuLVV...| corriere.it|
Mozilla/4.0 (comp...|tX1sMpnhJUhmAF7AS...| 55.35.44.79|2009-04-13-08-05|
| 1239610876000|lijjmCf2kuxfBTnjL...|AjvufgUtakUFcsIM9...| corriere.it|
Mozilla/4.0 (comp...|tX1sMpnhJUhmAF7AS...| 55.35.44.79|2009-04-13-08-05|
| 1239610941000|t8t8trgjNRPilmxuD...|agu2u2TCdqWP08rAA...| corriere.it|
Mozilla/4.0 (comp...|tX1sMpnhJUhmAF7AS...| 55.35.44.79|2009-04-13-08-05|
| 1239610490000|OGRLPVNGxiGgrCmWL...|mJg2raBUpPrC80lUm...| corriere.it|
Mozilla/4.0 (comp...|r2k96t1CNjSU9fJKN...| 71.124.66.3|2009-04-13-08-05|
| 1239610556000|OnJID12x0RXKPUgrD...|P7Pm2mPdW6w08KA3R...| corriere.it|
Mozilla/4.0 (comp...|r2k96t1CNjSU9fJKN...| 71.124.66.3|2009-04-13-08-05|
| 1239610373000|WflsvKIg0qfIE5KwR...|TJHd1VBspNcua0XPn...| corriere.it|
Mozilla/5.0 (Maci...|fj2L1ILTFGMfhdrt3...| 75.117.56.155|2009-04-13-08-05|
| 1239610768000|4MJR0XxiVCU1ueXKV...|10hGwmbvKf8ajoU8a...| corriere.it|
Mozilla/5.0 (Maci...|fj2L1ILTFGMfhdrt3...| 75.117.56.155|2009-04-13-08-05|
| 1239610832000|gWIrpDiN57i3sHatv...|RNL4C7xPi3tdar2Uc...| corriere.it|
Mozilla/5.0 (Maci...|fj2L1ILTFGMfhdrt3...| 75.117.56.155|2009-04-13-08-05|
| 1239610789000|pTne9k62kJ14QViXI...|RVxJVIQousjxUVI3r...| pixnet.net|
Mozilla/5.0 (Maci...|1bG0KiBD2xmui90kF...| 33.176.101.80|2009-04-13-08-05|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
only showing top 20 rows

scala>
```

## 示例：创建HCatalog表并使用 Pig 写入该表

您可以创建一个HCatalog表，然后使用 Apache Pig 通过使用 Amazon S3 中的数据源对其HCatStorer进行写入。HCatalog要求您禁用直接写入，否则操作会

静默失败。要同时将 `mapred.output.direct.NativeS3FileSystem` 和 `mapred.output.direct.EmrFileSystem` 配置设置为 `false`，可以使用 `mapred-site` 分类，或者通过 Grunt shell 手动操作。以下示例显示了使用创建的表 HCatCLI，然后是在 Grunt shell 中执行的命令，用于从 Amazon S3 中的示例数据文件填充表。

要运行此示例，请[使用连接到主节点SSH](#)。

创建包含以下内容的HCatalog脚本文件`wikicount.q`，该文件将创建一个名为的HCatalog表`wikicount`。

```
CREATE EXTERNAL TABLE IF NOT EXISTS wikicount(  
  col1 string,  
  col2 bigint  
)  
ROW FORMAT DELIMITED FIELDS TERMINATED BY '\001'  
STORED AS ORC  
LOCATION 's3://MyBucket/hcat/wikicount';
```

使用HCatCLI命令从文件中执行脚本。

```
hcat -f wikicount.q
```

接下来，使用 `-useHCatalog` 选项启动 Grunt shell，将配置设置为禁用直接写入，从 S3 位置加载数据，然后将结果写入 `wikicount` 表。

```
pig -useHCatalog  
SET mapred.output.direct.NativeS3FileSystem false;  
SET mapred.output.direct.EmrFileSystem false;  
A = LOAD 's3://support.elasticmapreduce/training/datasets/wikistats_tiny/' USING  
  PigStorage(' ') AS (Site:chararray, page:chararray, views:int, total_bytes:long);  
B = GROUP A BY Site;  
C = FOREACH B GENERATE group as col1, COUNT(A) as col2;  
STORE C INTO 'wikicount' USING org.apache.hive.hcatalog.pig.HCatStorer();
```

## HCatalog发布历史

下表列出了 Amazon 的每个发行版本中HCatalog包含的版本EMR，以及随应用程序一起安装的组件。有关每个发行版本中的组件版本，请参阅 [亚马逊 EMR 7.x 发布版本](#)、[亚马逊 EMR 6.x 发布版本](#) 或 [亚马逊 EMR 5.x 发布版本](#) 中的发行版“组件版本”部分。

## HCatalog版本信息

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-7.2.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-5.36.2	2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-7.1.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
		yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-7.0.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.15.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-6.14.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.13.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server



Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-6.12.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.11.1	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-6.11.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.10.1	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-6.10.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.9.1	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-6.9.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.8.1	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-6.8.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.7.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.36.1	2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-5.36.0	2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-6.6.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-5.35.0	2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-6.5.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.4.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server



Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-6.3.1	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.3.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-6.2.1	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.2.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-6.1.1	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.1.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-6.0.1	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.0.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.34.0	2.3.8	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-5.33.1	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.33.0	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-5.32.1	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.32.0	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-5.31.1	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.31.0	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-5.30.2	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server



Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.30.1	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-5.30.0	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.29.0	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.28.1	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.28.0	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.27.1	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.27.0	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.26.0	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.25.0	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.24.1	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.24.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.23.1	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.23.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.22.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.21.2	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.21.1	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server



Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.21.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.20.1	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.20.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.19.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.19.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.18.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.18.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.17.2	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.17.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.17.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.16.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.16.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.15.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.15.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.14.2	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.14.1	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server



Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.14.0	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.13.1	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.13.0	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.12.3	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.12.2	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.12.1	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.12.0	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.11.4	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.11.3	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.11.2	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.11.1	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.11.0	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.10.1	2.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.10.0	2.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.9.1	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.9.0	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server



Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.8.3	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.8.2	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.8.1	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.8.0	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.7.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.7.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.6.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.6.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.5.4	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.5.3	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.5.2	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.5.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.5.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.4.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.4.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.3.2	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server



Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.3.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.3.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.2.3	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.2.2	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.2.1	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.2.0	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.1.1	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.1.0	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.0.3	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.0.2	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-5.0.1	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.0.0	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-4.9.6	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-4.9.5	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-4.9.4	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-4.9.3	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server



Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-4.9.2	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-4.9.1	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-4.8.5	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-4.8.4	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-4.8.3	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-4.8.2	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-4.8.1	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-4.8.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-4.7.4	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-4.7.3	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-4.7.2	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-4.7.1	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-4.7.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-4.6.1	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, hive-metastore-server, mysql-server

Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-4.6.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, hive-metastore-server, mysql-server
emr-4.5.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, hive-metastore-server, mysql-server



Amazon EMR 发布标签	HCatalog版本	安装在一起的组件 HCatalog
emr-4.4.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, hive-metastore-server, mysql-server

# Apache Hive

Hive 是一种开源数据仓库和分析软件程序包，基于 Hadoop 集群运行。Hive 脚本使用一种名为 Hive QL ( 查询语言 ) 的 SQL 类似语言，它抽象编程模型并支持典型的数据仓库交互。Hive 使您能够避免基于有向无环图 (DAGs) 或使用较低级别计算机语言 ( 例如 Java ) 的 MapReduce 程序编写 Tez 作业的复杂性。

Hive 通过包含序列化格式来扩展 SQL 范式。您也可以通过创建与您的数据匹配的表架构自定义查询处理，而无需接触到数据本身。虽然 SQL 仅支持原始值类型 ( 例如日期、数字和字符串 )，但 Hive 表值是结构化元素，例如 JSON 对象、任何用户定义的数据类型或任何用 Java 编写的函数。

有关 Hive 的更多信息，请参阅 <http://hive.apache.org/>。

下表列出了最新版本的亚马逊 EMR 7.x 系列中包含的 Hive 版本，以及亚马逊随着 Hive EMR 安装的组件。

有关此版本中与 Hive 一起安装的组件版本，请参阅 [7.2.0 版本组件](#) 版本。

## emr-7.2.0 的 Hive 版本信息

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-7.2.0	Hive 3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server

下表列出了最新版本的亚马逊 EMR 6.x 系列中包含的 Hive 版本，以及亚马逊与 Hive 一起EMR安装的组件。

有关此发行版中随 Hive 安装的组件版本，请参阅[发行版 6.15.0 组件版本](#)。

#### emr-6.15.0 的 Hive 版本信息

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-6.15.0	Hive 3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server

下表列出了最新版本的亚马逊 EMR 5.x 系列中包含的 Hive 版本，以及亚马逊与 Hive 一起EMR安装的组件。

有关此版本中与 Hive 一起安装的组件的版本，请参阅[5.36.2 版组件版本](#)。

#### emr-5.36.2 的 Hive 版本信息

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.36.2	Hive 2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
		datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn

从 Amazon EMR 5.18.0 开始，您可以使用 Amazon 构建 EMR 存储库针对特定亚马逊版本中提供的库和依赖项的确切版本构建任务代码。EMR 有关更多信息，请参阅 [使用 Amazon EMR 项目存储库检查依赖项](#)。

## 主题

- [亚马逊 Hive 的区别和注意事项 EMR](#)
- [为 Hive 配置外部元存储](#)
- [使用 Hive JDBC 驱动程序](#)
- [改进 Hive 性能](#)
- [使用 Hive Live Long and Process \(\) LLAP](#)
- [Hive 中的加密](#)
- [Hive 发行历史记录](#)

## 亚马逊 Hive 的区别和注意事项 EMR

### 亚马逊上的 Apache Hive EMR 和 Apache Hive 之间的区别

本节介绍亚马逊 EMR 上的 Hive 与默认版本的 Hive 之间的区别，网址为 <http://svn.apache.org/viewvc/hive/branches/>。

## Hive 授权

亚马逊EMR支持 [Hive 授权](#)，HDFS但不支持对亚马逊 S3 EMRFS 进行授权。默认情况下，Amazon EMR 集群在禁用授权的情况下运行。

## Amazon S3 中的 Hive 文件合并操作

如果 `hive.merge.mapfiles` 为 `true`，那么 Apache Hive 会在仅 map 作业结束时合并小文件，且仅在平均的作业输出大小低于 `hive.merge.smallfiles.avgsize` 设置时才会触发合并。如果最终输出路径在中，Amazon EMR Hive 的行为完全相同。HDFS如果输出路径位于 Amazon S3 中，将忽略 `hive.merge.smallfiles.avgsize` 参数。在那种情况下，如果 `hive.merge.mapfiles` 设置为 `true`，会始终触发合并任务。

## ACID交易和 Amazon S3

Amazon EMR 6.1.0 及更高版本支持 HiveACID ( 原子性、一致性、隔离、持久性 ) 事务，因此它符合数据库的ACID属性。使用此功能，您可以使用亚马逊简单存储服务 (Amazon S3) Service 中的数据在 Hive 托管表中运行INSERTDELETE、和MERGE操作。UPDATE

## Hive 长寿并处理 () LLAP

LLAP亚马逊 EMR 5.0 版本的 Hive 2.1.0 不支持默认 Apache Hive 2.0 版中添加的@@ [功能](#)。

亚马逊EMR版本 6.0.0 及更高版本支持 Hive 的 Live Long 和 Proces LLAP s () 功能。有关更多信息，请参阅[使用 Hive。LLAP](#)

## 亚马逊EMR发布版本 4.x 和 5.x 之间的 Hive 差异

本节介绍在将 Hive 实现从亚马逊 4.x 版本上的 Hive 版本 1.0.0 迁移到EMR亚马逊 5.x 版本上的 Hive 2.x 之前需要考虑的差异。EMR

### 操作区别和注意事项

- 增加了对 [ACID \( 原子性、一致性、隔离性和持久性 \) 事务](#) 的支持：消除了EMR亚马逊 4.x 上的 Hive 1.0.0 与默认 Apache Hive 之间的差异。
- 取消了对亚马逊 S3 的直接写入：亚马逊EMR上的 Hive 1.0.0 与默认 Apache Hive 之间的差异已被消除。亚马逊 5.x EMR 版本上的 Hive 2.1.0 现在可以创建、读取和写入存储在 Amazon S3 中的临时文件。因此，要读取和写入同一个表，您就不必再在群集的本地HDFS文件系统中创建临时表作为解决方法。如果您使用受版本控制的存储桶，请确保如下所述管理这些临时文件。

- 使用 Amazon S3 受版本控制的存储桶时管理临时文件：当您在生成数据的目的地是 Amazon S3 的环境中运行 Hive 查询时，会创建许多临时文件和目录。这是新行为，如上所述。如果您使用受版本控制的 S3 存储桶，那么不删除这些临时文件会使 Amazon S3 凌乱并产生费用。请调整生命周期规则，以便包含 `/_tmp` 前缀的数据在一个短周期（例如，五天）后被删除。有关更多信息，请参阅[指定生命周期配置](#)。
- Log4j 已更新到 log4j 2：如果您使用 log4j，则可能因为此升级而需要更改您的日志记录配置。有关更多信息，请参阅[Apache log4j 2](#)。

## 性能区别和注意事项

- 与 Tez 的性能差异：在亚马逊EMR发布的 5.x 版本中，Tez 是 Hive 的默认执行引擎，而不是。MapReduceTez 为大多数工作流提供了改进的性能。
- 具有多个分区的表：生成大量动态分区的查询可能会失败，并且执行从具有多个分区的表中选择的查询可能需要比预期更长的时间。例如，从 100,000 个分区中进行选择可能需要 10 分钟或更长时间。

## 亚马逊 Hive 的其他功能 EMR

亚马逊对 Hive 进行了EMR扩展，增加了支持 Hive 与其他 AWS 服务集成的新功能，例如能够读取和写入亚马逊简单存储服务 (Amazon S3) 和 DynamoDB。

### Hive 中的变量

您可以使用美元符号和大括号在脚本中包括变量。

```
add jar ${LIB}/jsonserde.jar
```

如以下示例所示，您可以在命令行上使用 `-d` 参数将这些变量的值传递给 Hive：

```
-d LIB=s3://elasticmapreduce/samples/hive-ads/lib
```

您还可以将值传递到执行 Hive 脚本的步骤。

## 使用控制台将变量值传递到 Hive 步骤

1. 打开亚马逊EMR控制台，[网址为 https://console.aws.amazon.com/emr](https://console.aws.amazon.com/emr)。
2. 选择创建集群。
3. 在步骤部分中，对于 Add Step (添加步骤)，从列表中选择 Hive Program (Hive 程序)，然后选择 Configure and add (配置并添加)。
4. 在 Add Step (添加步骤) 对话框中，参考下表指定参数，然后选择添加。

字段	操作
脚本 S3 位置*	指定您的脚本URI在 Amazon S3 中的位置。该值必须采用以下形式 <i>BucketName /path/ScriptName</i> 。例如： <code>s3://elasticmapreduce/samples/hive-ads/libs/response-time-stats.q</code> 。
输入 S3 位置	( 可选 ) 指定您的输入文件URI在 Amazon S3 中的存储位置。该值必须采用以下形式 <i>BucketName /path/</i> 。如果指定，它将作为名INPUT为的参数传递给 Hive 脚本。例如： <code>s3://elasticmapreduce/samples/hive-ads/tables/</code> 。
输出 S3 位置	( 可选 ) 指定您希望将URI输出存储在 Amazon S3 中的位置。该值必须采用以下形式 <i>BucketName /path</i> 。如果指定，它将作为名OUTPUT为的参数传递给 Hive 脚本。例如： <code>s3://mybucket/hive-ads/output/</code> 。
参数	<p>输入要传递到 Hive 的参数列表 ( 以空格分隔的字符串 )。如果您在 Hive 脚本中定义了一个名为 <code>\${SAMPLE}</code> 的路径变量，例如：</p> <pre>CREATE EXTERNAL TABLE logs (requestBeginTime STRING, requestEndTime STRING, hostname STRING) PARTITIONED BY (dt STRING) \ ROW FORMAT serde 'com.amazon.elasticmapreduce.JsonSerde' WITH SERDEPROPERTIES ( 'paths'='requestBeginTime, requestEndTime, hostname' ) LOCATION '\${SAMPLE}/tables/impressions';</pre> <p>要传递该变量的值，请在参数窗口中键入以下内容：</p> <pre>-d SAMPLE=s3://elasticmapreduce/samples/hive-ads/</pre>

字段	操作
出现故障时的操作	<p>这决定了集群为响应任何错误而执行的操作。此设置的可能值为：</p> <ul style="list-style-type: none"> <li>• <b>Terminate cluster (终止集群)</b>：如果步骤失败，则终止集群。如果集群启用了终止保护启用了AND保持活动状态，则它不会终止。</li> <li>• <b>Cancel and wait (取消并等待)</b>：如果步骤失败，则取消剩余步骤。如果集群启用了保持活动状态，则集群不会终止。</li> <li>• <b>继续</b>：如果该步骤失败，则继续到下一个步骤。</li> </ul>

5. 根据需要选择值，然后选择创建集群。

要将变量值传递到 Hive 步骤，请使用 AWS CLI

要使用将变量值传递到 Hive 步骤 AWS CLI，请使用 `--steps` 参数并包括参数列表。

**Note**

为了便于读取，包含 Linux 行继续符 (\)。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (^)。

```
aws emr create-cluster --name "Test cluster" --release-label emr-7.2.0 \
--applications Name=Hive Name=Pig --use-default-roles --ec2-attributes
KeyName=myKey --instance-type m5.xlarge --instance-count 3 \
--steps Type=Hive,Name="Hive Program",ActionOnFailure=CONTINUE,Args=[-f,s3://
elasticmapreduce/samples/hive-ads/libs/response-time-stats.q,-d,INPUT=s3://
elasticmapreduce/samples/hive-ads/tables,-d,OUTPUT=s3://mybucket/hive-ads/output/,
-d,SAMPLE=s3://elasticmapreduce/samples/hive-ads/]
```

有关在中使用 Amazon EMR 命令的更多信息 AWS CLI，请参阅<https://docs.aws.amazon.com/cli/latest/reference/emr>。

使用 Java 将变量值传递到 Hive 步骤 SDK

- 以下示例演示如何使用将变量传递到步骤中 SDK。有关更多信息，请参阅 AWS SDK for Java API 参考 StepFactory 中的 [类](#)。



```
StepFactory stepFactory = new StepFactory();

StepConfig runHive = new StepConfig()
    .withName("Run Hive Script")
    .withActionOnFailure("TERMINATE_JOB_FLOW")
    .withHadoopJarStep(stepFactory.newRunHiveScriptStep("s3://mybucket/script.q",
        Lists.newArrayList("-d", "LIB= s3://elasticmapreduce/samples/hive-ads/lib")));
```

## Amazon EMR Hive 查询以适应部分 DynamoDB 架构

Amazon EMR Hive 允许您指定可以筛选数据的列子集，而不必要求您的查询包含所有列，从而在查询 DynamoDB 表时提供最大的灵活性。当采用稀疏数据库架构，并希望根据一些列来筛选记录（例如根据时间戳筛选）时，这种部分架构查询技术可以发挥作用。

以下示例显示了如何使用 Hive 查询执行下列操作：

- 创建 DynamoDB 表。
- 选择 DynamoDB 中的一部分项目（行）并进一步将数据范围缩小到特定列。
- 将结果数据复制到 Amazon S3。

```
DROP TABLE dynamodb;
DROP TABLE s3;

CREATE EXTERNAL TABLE dynamodb(hashKey STRING, recordTimeStamp BIGINT, fullColumn
map<String, String>)
    STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
    TBLPROPERTIES (
        "dynamodb.table.name" = "myTable",
        "dynamodb.throughput.read.percent" = ".1000",
        "dynamodb.column.mapping" = "hashKey:HashKey,recordTimeStamp:RangeKey");

CREATE EXTERNAL TABLE s3(map<String, String>)
    ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
    LOCATION 's3://bucketname/path/subpath/';

INSERT OVERWRITE TABLE s3 SELECT item fullColumn FROM dynamodb WHERE recordTimeStamp <
"2012-01-01";
```

下表显示了从 DynamoDB 中选择任意项目组合的查询语法。

查询范例	结果描述
<code>SELECT * FROM <i>table_name</i> ;</code>	从指定表选择所有项目 (行) 并包括这些项目对应的所有列的数据。
<code>SELECT * FROM <i>table_name</i> WHERE <i>field_name</i> =<i>value</i>;</code>	从指定表选择一些项目 (行) 并包括这些项目对应的所有列的数据。
<code>SELECT <i>column1_name</i> , <i>column2_name</i> , <i>column3_name</i> FROM <i>table_name</i> ;</code>	从指定表选择所有项目 (行) 并包括这些项目对应的一些列的数据。
<code>SELECT <i>column1_name</i> , <i>column2_name</i> , <i>column3_name</i> FROM <i>table_name</i> WHERE <i>field_name</i> =<i>value</i>;</code>	从指定表选择一些项目 (行) 并包括这些项目对应的一些列的数据。

## 在不同 AWS 区域的 DynamoDB 表之间复制数据

Amazon EMR Hive 提供了一个您可以为每个 DynamoDB 表设置的 `dynamodb.region` 属性。当两个表的 `dynamodb.region` 设置不同时，您在两个表之间执行的所有数据复制将自动在指定区域之间发生。

以下示例显示了如何通过用于设置 `dynamodb.region` 属性的 Hive 脚本创建 DynamoDB 表：

### Note

每个表的 `region` 属性会覆盖全局 Hive 属性。

```
CREATE EXTERNAL TABLE dynamodb(hashKey STRING, recordTimeStamp BIGINT, map<String, String> fullColumn)
  STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
  TBLPROPERTIES (
    "dynamodb.table.name" = "myTable",
    "dynamodb.region" = "eu-west-1",
    "dynamodb.throughput.read.percent" = ".1000",
    "dynamodb.column.mapping" = "hashKey:HashKey,recordTimeStamp:RangeKey");
```

## 设置每个表的 DynamoDB 吞吐量值

Amazon EMR Hive 允许您在表定义中根据每个表设置 `readThroughputPercent` DynamoDB `writeThroughputPercent` 和设置。以下 Amazon EMR Hive 脚本显示了如何设置吞吐量值。有关 DynamoDB 吞吐量值的更多信息，请参阅[指定表的读取和写入要求](#)。

```
CREATE EXTERNAL TABLE dynamodb(hashKey STRING, recordTimeStamp BIGINT, map<String,
String> fullColumn)
  STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
  TBLPROPERTIES (
    "dynamodb.table.name" = "myTable",
    "dynamodb.throughput.read.percent" = ".4",
    "dynamodb.throughput.write.percent" = "1.0",
    "dynamodb.column.mapping" = "hashKey:HashKey,recordTimeStamp:RangeKey");
```

## 为 Hive 配置外部元存储

默认情况下，Hive 在主节点文件系统上的“我的SQL数据库”中记录元存储信息。元存储包含表以及在其上构建表的基础数据的描述，包括分区名称、数据类型等。集群终止后，所有集群节点都会关闭，包括主节点。当发生此情况时，本地数据会丢失，因为节点文件系统使用的是短暂存储。如果您需要保留元存储，则必须创建一个存在于集群外部的的外部元存储。

对于外部元存储，您有两个选项：

- AWS Glue 数据目录（仅限亚马逊 5.8.0 或更高EMR版本）。

有关更多信息，请参阅[使用 AWS Glue 数据目录作为 Hive 的元数据库](#)。

- 亚马逊RDS或亚马逊 Aurora。

有关更多信息，请参阅[使用外部“我的SQL数据库”或 Amazon Aurora](#)。

### Note

如果您使用 Hive 3 并遇到 Hive 元数据仓的连接太多的问题，请将参数 `datanucleus.connectionPool.maxPoolSize` 配置为一个较小的值或增加数据库服务器可以处理的连接数。连接数的增加是由于 Hive 计算最大连接数JDBC的方式造成的。要计算可确保最佳性能的值，请参阅[Hive 配置属性](#)。

## 使用 AWS Glue 数据目录作为 Hive 的元数据库

使用亚马逊EMR版本 5.8.0 或更高版本，您可以将 Hive 配置为使用 Glue 数据目录作为 AWS 其元数据库。当您需要持久的元数据仓或由不同集群、服务、应用程序和 AWS 账户共享的元数据仓时，我们建议使用此配置。

AWS Glue 是一项完全托管的提取、转换和加载 (ETL) 服务，它使对数据进行分类、清理、丰富数据并在各种数据存储之间可靠地移动数据变得简单且经济高效。AWS Glue 数据目录提供了一个跨各种数据源和数据格式的统一元数据存储库，与亚马逊、亚马逊、亚马逊 Redshift、Redshift Spectrum RDS、Athena EMR 以及任何与 Apache Hive 元数据仓兼容的应用程序集成。AWS Glue 爬虫程序可以自动从 Amazon S3 中的源数据推断出架构，并将相关的元数据存储入数据目录中。有关数据目录的更多信息，请参阅 [AWS Glue 开发者指南中的填充 G AWS Glue 数据目录](#)。

Glue 需单独收 AWS 费。在数据目录中存储和访问元数据有月费率，Glue ETL 作业和爬虫运行时按小时费率收费，每个预配置的开发端点按每分钟计费。数据目录让您最多可免费存储一百万个对象。如果您存储的对象超过一百万个，则每超过一百万个 USD 100,000 个对象需要支付 1 美元的费用。数据目录中的对象为表、分区或数据库。有关更多信息，请参阅 [Glue 定价](#)。

### Important

如果您在 2017 年 8 月 14 日之前使用 Amazon Athena 或 Amazon Redshift Spectrum 创建了表，则数据库和表将存储在雅典娜管理的目录中，该目录与 Glue 数据目录是分开的。AWS 要将亚马逊EMR与这些表集成，您必须升级到 AWS Glue 数据目录。有关更多信息，请参阅亚马逊 Athena AWS 用户指南中的升级到 Glue [数据目录](#)。

## 将 AWS Glue 数据目录指定为元数据库

您可以使用 AWS Management Console、AWS CLI 或 Amazon EMR API 将 AWS Glue 数据目录指定为元数据仓。使用 CLI 或 API，可以使用 Hive 的配置分类来指定数据目录。此外，在 Amazon EMR 5.16.0 及更高版本中，您可以使用配置分类在不同的数据目录中指定数据目录。AWS 账户在使用控制台时，您可以使用 Advanced Options (高级选项) 或 Quick Options (快速选项) 指定数据目录。

### Console

使用控制台 AWS 将 Glue 数据目录指定为 Hive 元数据仓

1. 登录并打开亚马逊EMR控制台 AWS Management Console，[网址为 https://console.aws.amazon.com/emr/](https://console.aws.amazon.com/emr/)。

2. EC2在左侧导航窗格的“开”下EMR，选择“集群”，然后选择“创建集群”。
3. 在“应用程序包”下，选择“核心 Hadoop”或“自HBase定义”。如果您自定义集群，请确保选择 Hive 或HCatalog作为您的应用程序之一。
4. 在AWS Glue 数据目录设置下，选择用于 Hive 表元数据复选框。
5. 选择适用于集群的任何其他选项。
6. 要启动集群，选择 Create cluster ( 创建集群 )。

## CLI

要将 AWS Glue 数据目录指定为 Hive 元数据库，请使用 AWS CLI

有关使用和指定配置分类的更多信息 AWS CLI EMRAPI，请参阅[配置应用程序](#)。

- 使用 hive-site 配置分类指定 `hive.metastore.client.factory.class` 的值，如下例所示：

```
[
  {
    "Classification": "hive-site",
    "Properties": {
      "hive.metastore.client.factory.class":
      "com.amazonaws.glue.catalog.metastore.AWSGlueDataCatalogHiveClientFactory"
    }
  }
]
```

在EMR发行版 5.28.0、5.28.1、5.29.0 或 6.x 上，如果您要使用 Glue 数据目录作为元数据存储创建集群，AWS 请将设置为 `hive.metastore.schema.validation false` 这会阻止 Hive 和HCatalog根据 My 验证元存储架构。SQL如果没有此配置，则在 Hive 或上重新配置后，主实例组将处于暂停状态。HCatalog

```
[
  {
    "Classification": "hive-site",
    "Properties": {
      "hive.metastore.client.factory.class":
      "com.amazonaws.glue.catalog.metastore.AWSGlueDataCatalogHiveClientFactory",
      "hive.metastore.schema.validation": "false"
    }
  }
]
```

```

    }
  }
]

```

如果您已经拥有EMR版本为 5.28.0、5.28.1 或 5.29.0 的集群，则可以使用以下信息将主实例组设置为 `hive.metastore.schema.verificationfalse`

```

Classification = hive-site
Property       = hive.metastore.schema.verification
Value          = false

```

要在其他 AWS 账户中指定数据目录，请添加 `hive.metastore.glue.catalogid` 属性，如下示例所示。将 `acct-id` 替换为数据目录的 AWS 账户。

```

[
  {
    "Classification": "hive-site",
    "Properties": {
      "hive.metastore.client.factory.class":
"com.amazonaws.glue.catalog.metastore.AWSGlueDataCatalogHiveClientFactory",
      "hive.metastore.schema.verification": "false",
      "hive.metastore.glue.catalogid": "acct-id"
    }
  }
]

```

## IAM权限

集群的EC2实例配置文件必须具有 Glue AWS 操作的IAM权限。此外，如果您为 AWS Glue Data Catalog 对象启用加密，则还必须允许该角色加密、解密和生成 AWS KMS key 用于加密的对象。

### AWS Glue 操作的权限

如果您使用 Amazon 的默认EC2实例配置文件EMR，则无需执行任何操作。附加到 `AmazonElasticMapReduceforEC2Role` 托管策略 `EMR_EC2_DefaultRole` 允许执行所有必要的 AWS Glue 操作。但是，如果您指定自定义EC2实例配置文件和权限，则必须配置相应的 AWS Glue

操作。使用 `AmazonElasticMapReduceforEC2Role` 托管策略作为起点。有关更多信息，请参阅 Amazon EMR 管理指南中的 [集群EC2实例服务角色 \(实例配置文件\)](#)。

### 加密和解密 AWS Glue 数据目录的权限

您的实例配置文件需要使用密钥加密和解密数据的权限。如果以下语句适用，您不必配置这些权限：

- 您可以使用 Glue 的 AWS 托管密钥为 Glue 数据目录对象启用加密。AWS
- 您使用的集群与 Glue 数据目录 AWS 账户 相同。

否则，您必须将以下语句添加到您的 EC2 实例配置文件所附的权限策略中。

```
[
  {
    "Version": "2012-10-17",
    "Statement": [
      {
        "Effect": "Allow",
        "Action": [
          "kms:Decrypt",
          "kms:Encrypt",
          "kms:GenerateDataKey"
        ],
        "Resource": "arn:aws:kms:region:acct-
id:key/12345678-1234-1234-1234-123456789012"
      }
    ]
  }
]
```

有关 AWS Glue 数据目录加密的更多信息，请参阅 [Glue 开发者指南中的 AWS 加密数据目录](#)。

### 基于资源的权限

如果您在亚马逊中将 AWS Glue 与 Hive、Spark 或 Presto 结合使用，EMR Glue 支持基于资源的策略来控制对数据目录资源的访问。这些资源包括数据库、表、连接和用户定义的函数。有关更多信息，请参阅《AWS Glue 开发人员指南》中的 [AWS Glue 资源策略](#)。

使用基于资源的策略限制从 Amazon EMR 内部访问 AWS Glue 时，您在权限策略中指定的委托人必须是与创建集群时指定的 EC2 实例配置文件 ARN 关联的角色。例如，对于附加到目录的基于资源的策略，您可以 ARN 为集群 EC2 实例指定默认服务角色的角色，`EMR_EC2_DefaultRole` 正如 Principal，使用以下示例所示的格式：

```
arn:aws:iam::acct-id:role/EMR_EC2_DefaultRole
```

这些区域有：*acct-id* 可以与 AWS Glue 账号 ID 不同。这允许从不同账户的EMR集群进行访问。您可以指定多个委托人，且每个委托人都可以来自不同的账户。

## 使用 AWS Glue 数据目录时的注意事项

使用 AWS Glue 数据目录作为 Hive 的元数据仓库时，请考虑以下事项：

- 不支持JARs使用 Hive 外壳添加辅助工具。解决方法是，使用hive-site配置分类来设置hive.aux.jars.path属性，这会JARs将辅助属性添加到 Hive 类路径中。
- 不支持 [Hive 事务](#)。
- 不支持在 AWS Glue 中重命名表。
- 当您创建 Hive 表而不指定 LOCATION 时，表数据存储在与通过 hive.metastore.warehouse.dir 属性指定的位置。默认情况下，这是其中的一个位置HDFS。如果另一个集群需要访问该表，则它将失败，除非它有足够的权限访问创建该表的集群。此外，由于HDFS存储是临时性的，因此如果集群终止，表数据就会丢失，必须重新创建表。我们建议您在使用 AWS Glue 创建 Hive 表时LOCATION在 Amazon S3 中指定。此外，也可以使用 hive-site 配置分类来为 hive.metastore.warehouse.dir 指定 Amazon S3 中的位置，它适用于所有 Hive 表。如果在某个HDFS位置创建了表，而创建该表的集群仍在运行，则可以从 Glue 中将该表的位置更新为 Amazon S AWS 3。有关更多信息，请参阅 [《Glue 开发者指南》中的“在 AWS Glue 控制台AWS 上使用表格”](#)。
- 不支持包含引号和撇号的分区值，例如 PARTITION (owner="Doe's")。
- emr-5.31.0 及更高版本支持[列统计数据](#)。
- 不支持使用 [Hive 授权](#)。作为替代方案，考虑使用[基于AWS Glue 资源的策略](#)。有关更多信息，请参阅[使用基于资源的策略让 Amazon EMR 访问 AWS Glue 数据目录](#)。
- 不支持 [Hive 约束](#)。
- 不支持 [Hive 中基于成本的优化](#)。
- 不支持设置 hive.metastore.partition.inherit.table.properties。
- 不支持使用以下元存储常量：BUCKET\_COUNT, BUCKET\_FIELD\_NAME, DDL\_TIME, FIELD\_TO\_DIMENSION, FILE\_INPUT\_FORMAT, FILE\_OUTPUT\_FORMAT, HIVE\_FILTER\_FIELD\_LAST\_ACCESS, HIVE\_FILTER\_FIELD\_OWNER, HIVE\_FILTER\_FIELD\_PARAMS, IS\_ARCHIVED, META\_TABLE\_COLUMNS, META\_TABLE\_COLUMN\_TYPES, META\_TABLE\_DB, META\_TABLE\_LOCATION,



META\_TABLE\_NAME, META\_TABLE\_PARTITION\_COLUMNS, META\_TABLE\_SERDE, META\_TABLE\_STORAGE, ORIGINAL\_LOCATION。

- 使用谓词表达式时，显式值必须位于比较运算符的右侧，否则查询可能会失败。
  - 正确：SELECT \* FROM mytable WHERE time > 11
  - 错误：SELECT \* FROM mytable WHERE 11 > time
- Amazon EMR 版本 5.32.0 和 6.3.0 及更高版本支持在谓词表达式中使用用户定义的函数 (UDFs)。使用早期版本时，查询可能因 Hive 尝试优化查询执行的方式而失败。
- 不支持[临时表](#)。
- 我们建议通过 Amazon 使用应用程序创建表，EMR 而不是直接使用 AWS Glue 创建表。通过 AWS Glue 创建表可能会导致必填字段缺失并导致查询异常。
- 在 EMR 5.20.0 或更高版本中，当使用 AWS Glue Data Catalog 作为元数据存储时，Spark 和 Hive 会自动启用并行分区修剪。此更改通过并行执行多个请求来检索分区，显著缩短查询计划时间。可同时执行的分段总数介于 1 到 10 之间。默认值为 5，这是建议的设置。您可以通过以下方式更改该值：指定 hive-site 配置分类中的属性 aws.glue.partition.num.segments。如果发生节流，则可以通过将值更改为 1 来关闭此功能。有关更多信息，请参阅 [AWS Glue 分段结构](#)。

## 使用外部“我的SQL数据库”或 Amazon Aurora

要使用外部“我的SQL数据库”或 Amazon Aurora 作为 Hive 元数据仓库，可以覆盖 Hive 中元数据仓库的默认配置值，以指定外部数据库位置，无论是在 Amazon MRDS y SQL 实例上还是 Amazon Aurora P 上。ostgreSQLInstance

### Note

Hive 既不支持对元存储表的并发写入访问权限，也不阻止此权限。如果要在两个集群间共享元数据仓库信息，您必须确保不会同时写入同一元数据仓库表，除非您要写入同一元数据仓库表的不同分区。

以下步骤介绍了如何覆盖 Hive 元数据仓库位置的默认配置值和使用重新配置的元数据仓库位置启动集群。

### 创建位于集群外部的元数据仓库 EMR

1. 创建 My SQL 或 Aurora Postgre SQL 数据库。如果您使用 PostgreSQL，则必须在配置集群后对其进行配置。创建集群时 SQL 仅支持“我的”。有关 Aurora My SQL 和 Aurora Postgre 之间区别的

- 信息SQL，请参阅亚马逊 Aurora MySQL [概述SQL和使用亚马逊 Aurora PostgreSQL](#)。有关如何创建 Amazon RDS 数据库的总体信息，请参阅<https://aws.amazon.com/rds/>。
2. 修改您的安全组以允许您的数据库和 ElasticMapReduce-Master 安全组之间JDBC建立连接。有关如何修改安全组以供访问的信息，请参阅使用 [Amazon EMR 托管的安全组](#)。
  3. 在以下JDBC位置设置配置值hive-site.xml：

### Important

如果您向 Amazon EMR 配置提供诸如密码之类的敏感信息API，则会显示具有足够权限的账户的这些信息。如果您担心这些信息可能会显示给其他用户，请使用管理帐户创建集群，并通过创建一个明确拒绝elasticmapreduce:DescribeClusterAPI密钥权限的角色来限制其他IAM用户（用户或具有委派凭据的用户）访问集群上的服务。

- a. 创建一个名为 hiveConfiguration.json 的配置文件，该文件包含对 hive-site.xml 的编辑，如以下示例所示。

Replace（替换）*hostname* 使用运行数据库的 Amazon RDS 实例DNS的地址，以及 *username* 以及 *password* 以及您的数据库的凭证。有关连接到 MySQL 和 Aurora 数据库实例的更多信息，请参阅《亚马逊RDS用户指南》中的[连接到运行“我的SQL数据库引擎”的数据库实例](#)和[连接 Athena 数据库集群](#)。javax.jdo.option.ConnectionURL是JDBC元数据仓的JDBC连接字符串。javax.jdo.option.ConnectionDriverName是JDBC元数据仓的驱动程序类名。

我的SQLJDBC驱动程序由 Amazon 安装EMR。

值属性不能包含任何空格或回车。所有内容应显示在一行中。

```
[
  {
    "Classification": "hive-site",
    "Properties": {
      "javax.jdo.option.ConnectionURL": "jdbc:mysql://hostname:3306/hive?
createDatabaseIfNotExist=true",
      "javax.jdo.option.ConnectionDriverName": "org.mariadb.jdbc.Driver",
      "javax.jdo.option.ConnectionUserName": "username",
      "javax.jdo.option.ConnectionPassword": "password"
    }
  }
]
```

]

- b. 创建集群时引用该hiveConfiguration.json文件，如以下 AWS CLI 命令所示。在此命令中，此文件存储在本地，您也可将此文件上传到 Amazon S3 并在此对其进行引用，例如 `s3://DOC-EXAMPLE-BUCKET/hiveConfiguration.json`。

#### Note

为了便于读取，包含 Linux 行继续符 ( \ )。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (^)。

```
aws emr create-cluster --release-label emr-7.2.0 --instance-type m5.xlarge --
instance-count 2 \
--applications Name=Hive --configurations file://hiveConfiguration.json --use-
default-roles
```

4. 连接到集群的主节点。

有关如何连接到主节点的信息，请参阅 Amazon EMR 管理指南SSH中的[使用连接到主节点](#)。

5. 通过输入类似以下内容的命令，创建在 Amazon S3 上指定位置的 Hive 表：

```
CREATE EXTERNAL TABLE IF NOT EXISTS table_name
(
key int,
value int
)
LOCATION s3://DOC-EXAMPLE-BUCKET/hdfs/
```

6. 将 Hive 脚本添加到正在运行的集群。

您的 Hive 集群使用位于亚马逊的元数据仓运行。RDS通过指定该元数据仓位置，启动共享该元数据仓的所有其他 Hive 集群。

## 使用 Hive JDBC 驱动程序

您可以将常用商业智能工具（如 Microsoft Excel、MicroStrategy、QlikView 和 Tableau）与 Amazon EMR 结合使用来探索和显示您的数据。许多这类工具都需要 Java 数据库连接（JDBC）驱动程序或开放式数据库连接（ODBC）驱动程序。Amazon EMR 支持 JDBC 和 ODBC 连接。

以下示例演示了如何使用 SQL Workbench/J 作为 SQL 客户端，与 Amazon EMR 中的 Hive 集群连接。有关其他驱动程序的更多信息，请参阅[将业务情报工具与 Amazon EMR 结合使用](#)。

在安装和使用 SQL Workbench/J 之前，请下载驱动程序包并安装驱动程序。程序包中包含的驱动程序支持 Amazon EMR 发行版 4.0 及更高版本中提供的 Hive 版本。如需详细的发布说明和文档，请参阅程序包中的 PDF 文档。

- 最新 Hive JDBC 驱动程序包下载

<http://awssupportdatasvcs.com/bootstrap-actions/Simba/latest/>

- Hive JDBC 驱动程序的旧版本

<http://awssupportdatasvcs.com/bootstrap-actions/Simba/>

## 安装和配置 SQL Workbench

1. 从 <http://www.sql-workbench.net/downloads.html> 下载适用于您的操作系统的 SQL Workbench/J 客户端。
2. 安装 SQL Workbench/J。有关更多信息，请参阅 SQL Workbench/J 用户手册中的[安装并启动 SQL Workbench/J](#)。
3. Linux、Unix、Mac OS X 用户：在终端会话中，使用下面的命令创建到集群主节点的 SSH 隧道。将 *mmaster-public-dns-name* 替换为主节点的公有 DNS 名称，将 *path-to-key-file* 替换为您的 Amazon EC2 私有密钥 ( .pem ) 文件的位置和文件名。

```
ssh -o ServerAliveInterval=10 -i path-to-key-file -N -L 10000:localhost:10000  
hadoop@master-public-dns-name
```

Windows 用户：在 PuTTY 会话中，使用 10000 作为 Source port (源端口)、使用 *master-public-dns-name*:10000 作为 Destination (目标) 来创建到集群主节点的 SSH 隧道 (使用本地端口转发)。将 *master-public-dns-name* 替换为主节点的公有 DNS 名称。

4. 将 JDBC 驱动程序添加到 SQL Workbench。
  - a. 在 Select Connection Profile (选择连接配置文件) 对话框中，单击 Manage Drivers (管理驱动程序)。
  - b. 单击 Create a new entry (创建新条目) (空白页) 图标。
  - c. 在名称字段中，键入 **Hive JDBC**。
  - d. 对于 Library (库)，请单击 Select the JAR file(s) (选择 JAR 文件) 图标。

- e. 导航到包含提取的驱动程序的位置。选择您下载的 JDBC 驱动程序包版本中包含的驱动程序，然后单击 Open ( 打开 )。

例如，您的 JDBC 驱动程序包可能包括以下 JAR。

```
hive_metastore.jar
hive_service.jar
HiveJDBC41.jar
libfb303-0.9.0.jar
libthrift-0.9.0.jar
log4j-1.2.14.jar
ql.jar
slf4j-api-1.5.11.jar
slf4j-log4j12-1.5.11.jar
TCLIServiceClient.jar
zookeeper-3.4.6.jar
```

- f. 在 Please select one driver (请选择一个驱动程序) 对话框中，选择 `com.amazon.hive.jdbc41.HS2Driver`、确定。
5. 当您返回到 Manage Drivers (管理驱动程序) 对话框时，确认 Classname (类名) 字段已经填写，然后选择确定。
6. 当您返回到 Select Connection Profile (选择连接配置文件) 对话框时，验证驱动程序 字段是否设置为 Hive JDBC，然后在 URL 字段中提供以下 JDBC 连接字符串：`jdbc:hive2://localhost:10000/default`。
7. 选择确定进行连接。连接完成后，连接详细信息将显示在 SQL Workbench/J 窗口顶部。

有关使用 Hive 和 JDBC 界面的更多信息，请参阅 Apache Hive 文档中的 [HiveClient](#) 和 [HiveJDBCInterface](#)。

## 改进 Hive 性能

当使用 Hive 查询、读取和写入保存在亚马逊 EMR 上的 S3 中的数据时，亚马逊提供的功能可帮助优化性能。

S3 Select 可以通过将处理“下推”到 Amazon S3 来提高某些应用程序中 CSV 和 JSON 文件的查询性能。

EMRFSS3 优化的提交器是该 [OutputCommitter](#) 类的替代方案，它消除了列出和重命名操作，从而提高了 Amazon S3 使用 EMRFS 写入文件时的性能。

## 主题

- [启用 Hive EMRFS S3 优化提交器](#)
- [将 S3 Select 与 Hive 结合使用以提高查询性能](#)
- [MSCK优化](#)

## 启用 Hive EMRFS S3 优化提交器

Hive EMRFS S3 优化提交器是另一种方式，EMRHive 在使用时使用它为插入查询写入文件。EMRFS 提交程序消除了 Amazon S3 上执行的列表和重命名操作，并提高了应用程序性能。该功能从 EMR 5.34 和 EMR 6.5 开始提供。

### 启用提交程序

如果要启用 EMR Hive 作为所有 Hive 托管表和外部表的默认提交数据，请在 EMR 6.5.0 或 EMR 5.34.0 集hive-site群中使用以下配置。HiveEMRFSOptimizedCommitter

```
[
  {
    "classification": "hive-site",
    "properties": {
      "hive.blobstore.use.output-committer": "true"
    }
  }
]
```

#### Note

请勿在 `hive.exec.parallel` 设置为 `true` 时启动此功能。

## 限制

下面是适用于标签的基本限制：

- 不支持启用 Hive 自动合并小文件。即使启用了优化提交程序，也会使用默认的 Hive 提交逻辑。
- 不支持 Hive ACID 表。即使启用了优化提交程序，也会使用默认的 Hive 提交逻辑。
- 写入的文件的文件命名术语从 Hive 的 `<task_id>_<attempt_id>_<copy_n>` 更改为 `<task_id>_<attempt_id>_<copy_n>_<query_id>`。例如，文件命名为

s3://warehouse/table/partition=1/000000\_0 将更改为 s3://warehouse/table/partition=1/000000\_0-hadoop\_20210714130459\_ba7c23ec-5695-4947-9d98-8a40ef759222-1。query\_id这里是用户名、时间戳和的组合UUID。

- 当自定义分区位于不同的文件系统 ( HDFS , S3 ) 上时，此功能会自动禁用。即使启用，也会使用默认的 Hive 提交逻辑。

## 将 S3 Select 与 Hive 结合使用以提高查询性能

在亚马逊EMR发布版本 5.18.0 及更高版本中，您可以在亚马逊上[将 S3 Select](#) 与 Hive 配合使用。EMRS3 Select 可让应用程序仅从对象检索数据子集。对于亚马逊来说EMR，筛选大型数据集进行处理的计算工作从集群“向下”推送到 Amazon S3，这可以提高某些应用程序的性能并减少在 Amazon EMR 和 Amazon S3 之间传输的数据量。

基于CSV和JSON文件的 Hive 表以及在 Hive 会话true期间将s3select.filter配置变量设置为，即可支持 S3 Select。有关更多信息以及示例，请参阅 [在代码中指定 S3 Select](#)。

### S3 Select 是否适合我的应用程序？

建议您分别在使用和不使用 S3 Select 的情况下测试您的应用程序，以查看 S3 Select 是否适用于您的应用程序。

使用以下准则来确定您的应用程序是否为使用 S3 Select 的候选项：

- 您的查询将筛选掉原始数据集的一半以上的数据。
- 您的查询筛选谓词使用具有 Amazon S3 Select 支持的数据类型的列。有关更多信息，请参阅《Amazon Simple Storage Service 用户指南》中的[数据类型](#)。
- 您的 Amazon S3 和 Amazon EMR 集群之间的网络连接具有良好的传输速度和可用带宽。Amazon S3 不压缩响HTTP应，因此压缩后的输入文件的响应大小可能会增加。

### 注意事项和限制

- 不支持使用客户提供的加密密钥 (SSE-C) 的 Amazon S3 服务器端加密和客户端加密。
- 不支持 AllowQuotedRecordDelimiters 属性。如果指定该属性，则查询将失败。
- 仅支持CSV和 UTF -8 格式的JSON文件。不支持多JSON行CSVs和。
- 仅支持未压缩文件或 gzip、或 bzip2 文件。

- 不支持最后一行中的注释字符。
- 文件末尾的空行不会被处理。
- Amazon 上的 Hive EMR 支持 S3 Select 支持的原始数据类型。有关更多信息，请参阅《Amazon Simple Storage Service 用户指南》中的[数据类型](#)。

## 在代码中指定 S3 Select

要在 Hive 表中使用 S3 Select，请通过将 `com.amazonaws.emr.s3select.hive.S3SelectableTextInputFormat` 指定为 `INPUTFORMAT` 类名称以及使用 `TBLPROPERTIES` 子句为 `s3select.format` 属性指定一个值，来创建表。

默认情况下，S3 Select 在您运行查询时处于禁用状态。通过在您的 Hive 会话中将 `s3select.filter` 设置为 `true` 来启用 S3 Select，如下所示。以下示例演示了在从底层 CSV 和 JSON 文件创建表，然后使用简单的 `select` 语句查询表时，如何指定 S3 Select。

### Example CREATETABLECSV 基于表的语句

```
CREATE TABLE mys3selecttable (  
  col1 string,  
  col2 int,  
  col3 boolean  
)  
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','  
STORED AS  
INPUTFORMAT  
  'com.amazonaws.emr.s3select.hive.S3SelectableTextInputFormat'  
OUTPUTFORMAT  
  'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat'  
LOCATION 's3://path/to/mycsvfile/'  
TBLPROPERTIES (  
  "s3select.format" = "csv",  
  "s3select.headerInfo" = "ignore"  
);
```

### Example CREATETABLEJSON 基于表的语句

```
CREATE TABLE mys3selecttable (  
  col1 string,  
  col2 int,
```



```
col3 boolean
)
ROW FORMAT SERDE 'org.apache.hive.hcatalog.data.JsonSerDe'
STORED AS
INPUTFORMAT
'com.amazonaws.emr.s3select.hive.S3SelectableTextInputFormat'
OUTPUTFORMAT
'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat'
LOCATION 's3://path/to/json/'
TBLPROPERTIES (
  "s3select.format" = "json"
);
```

### Example SELECTTABLE声明

```
SET s3select.filter=true;
SELECT * FROM mys3selecttable WHERE col2 > 10;
```

## MSCK优化

Hive 在其元数据存储中存储每个表的分区列表。但是，当直接向文件系统添加分区或从文件系统中移除分区时，Hive 元数据存储不会意识到这些变化。该[MSCK命令](#)更新直接添加到文件系统或从文件系统中删除的分区的 Hive 元数据仓中的分区元数据。此命令的语法是：

```
MSCK [REPAIR] TABLE table_name [ADD/DROP/SYNC PARTITIONS];
```

Hive 将按如下方式实现此命令：

1. Hive 从元数据存储中检索表的所有分区。然后根据文件系统中不存在的分区路径列表，创建一个要从元数据存储中移除的分区列表。
2. Hive 收集文件中存在的分区路径，将其与元数据存储中的分区列表进行比较，然后生成需要添加到元数据存储的分区列表。
3. Hive 使用 ADD、DROP 或 SYNC 模式更新元数据存储。

#### Note

当元存储中有许多分区时，检查文件系统中是否不存在分区的步骤需要很长时间才能运行，因为必须对每个分区进行文件系统的existsAPI调用。

在亚马逊 EMR 6.5.0 中，Hive 引入了一个名为的标志。hive.emr.optimize.msck.fs.check启用此标志后，Hive 将从上述步骤 2 中生成的文件系统分区路径列表中检查是否存在分区，而不是进行文件系统API调用。在 Amazon EMR 6.8.0 中，Hive 默认启用了此优化，因此无需设置标志。hive.emr.optimize.msck.fs.check

## 使用 Hive Live Long and Process () LLAP

亚马逊 EMR 6.0.0 支持 Hive 的 Live Long 和 Proces LLAP s () 功能。LLAP与之前的默认 Tez 容器执行模式相比，使用具有智能内存缓存的持久守护程序来提高查询性能。

Hive LLAP 守护程序作为YARN服务进行管理和运行。由于YARN服务可以被视为长时间运行的YARN应用程序，因此您的一些集群资源专用于HiveLLAP，不能用于其他工作负载。有关更多信息，请参阅[LLAP](#)和[YARN服务API](#)。

## 在亚马逊上启用 Hive LLAP e EMR

要LLAP在 Amazon 上启用 HiveEMR，请在启动集群时提供以下配置。

```
[
  {
    "Classification": "hive",
    "Properties": {
      "hive.llap.enabled": "true"
    }
  }
]
```

有关更多信息，请参阅[配置应用程序](#)。

默认情况下，Amazon EMR 会将大约 60% 的集群YARN资源分配给 Hive LLAP 守护程序。您可以配置分配给 Hive 的群集YARN资源的百分比，LLAP以及分配给 Hive 时要考虑的任务和核心节点数量。LLAP

例如，以下配置以三个任务或核心节点上的三个守护程序启动 HiveLLAP，并将三个核心或任务节点 YARN资源的 40% 分配给 Hive 守护程序。LLAP

```
[
  {
    "Classification": "hive",
```

```

    "Properties": {
      "hive.llap.enabled": "true",
      "hive.llap.percent-allocation": "0.4",
      "hive.llap.num-instances": "3"
    }
  }
]

```

您可以在分类中使用以下hive-site配置API来覆盖默认LLAP资源设置。

属性	描述
hive.llap.daemon.yarn.container.mb	LLAP守护程序容器总大小 (以 MB 为单位)
hive.llap.daemon.memory.per.instance.mb	LLAP守护程序容器中执行程序使用的总内存 (以 MB 为单位)
hive.llap.io.memory.size	LLAP输入/输出的缓存大小
hive.llap.daemon.num.executors	每个守护程序的执行者数量 LLAP

## 在集群LLAP上手动启动 Hive

作为集群启动的一部分，使用的所有依赖项和配置LLAP都打包到 LLAP tar 存档中。如果使用启LLAP用"hive.llap.enabled": "true"，我们建议您使用 Amazon EMR 重新配置来LLAP更改配置。

否则，要对进行任何手动更改hive-site.xml，必须使用hive --service llap命令重建 LLAP tar 存档，如下例所示。

```

# Define how many resources you want to allocate to Hive LLAP

LLAP_INSTANCES=<how many llap daemons to run on cluster>
LLAP_SIZE=<total container size per llap daemon>
LLAP_EXECUTORS=<number of executors per daemon>
LLAP_XMX=<Memory used by executors>
LLAP_CACHE=<Max cache size for IO allocator>

yarn app -enableFastLaunch

```

```
hive --service llap \  
--instances $LLAP_INSTANCES \  
--size ${LLAP_SIZE}m \  
--executors $LLAP_EXECUTORS \  
--xmx ${LLAP_XMX}m \  
--cache ${LLAP_CACHE}m \  
--name llap0 \  
--auxhbase=false \  
--startImmediately
```

## 检查 Hive 状态 LLAP

使用以下命令LLAP通过 Hive 检查 Hive 的状态。

```
hive --service llapstatus
```

使用以下命令检查 Hive LLAP 的状态。YARN

```
yarn app -status (name-of-llap-service)
```

# example:

```
yarn app -status llap0 | jq
```

## 启动或停止 Hive LLAP

由于 Hive 作为永久YARN服务LLAP运行，因此您可以停止或重新启动该YARN服务以停止或重新启动 Hive LLAP。以下命令对此进行了演示。

```
yarn app -stop llap0  
yarn app -start llap0
```

## 调整 Hive 守护程序LLAP的数量

使用以下命令减少LLAP实例数量。

```
yarn app -flex llap0 -component llap -1
```

有关更多信息，请参阅 [Flex a component of a service](#)。

# Hive 中的加密

本节介绍了 Amazon EMR 支持的加密类型。

## Hive 中的 Parquet 模块化加密

Parquet 模块化加密提供列级访问控制和加密功能，以增强以 Parquet 文件格式存储的数据的隐私和数据完整性。此功能从 6.6.0 版本开始在 Amazon EMR Hive 中可用。

Amazon EMR 管理指南中的加密[选项中描述了以前支持的安全和完整性解决方案，包括加密文件或加密存储层](#)。这些解决方案可以用于 Parquet 文件，但是利用集成 Parquet 加密机制的新功能可以在列级别实现精细访问控制，并提高性能和安全性。有关此功能的更多信息，请访问 Apache github 页面[Parquet 模块化加密](#)。

用户使用 Hadoop 配置将配置传递给 Parquet 读取器和写入器。用户配置读取器和写入器以启用加密以及切换高级功能的详细配置记录在 [PARQUET-1854 : Parquet 加密管理的属性驱动接口](#)

### 用法示例

以下示例涉及使用 AWS KMS 创建加密密钥并写入 Hive 表以进行加密密钥管理。

1. 如文档 [PARQUET-1373 : 加密密钥管理工具](#)中所述，为 AWS KMS 服务实现。KmsClient 以下示例演示了一个实施片段。

```
package org.apache.parquet.crypto.keytools;

import com.amazonaws.AmazonClientException;
import com.amazonaws.AmazonServiceException;
import com.amazonaws.regions.Regions;
import com.amazonaws.services.kms.AWSKMS;
import com.amazonaws.services.kms.AWSKMSClientBuilder;
import com.amazonaws.services.kms.model.DecryptRequest;
import com.amazonaws.services.kms.model.EncryptRequest;
import com.amazonaws.util.Base64;
import org.apache.hadoop.conf.Configuration;
import org.apache.parquet.crypto.KeyAccessDeniedException;
import org.apache.parquet.crypto.ParquetCryptoRuntimeException;
import org.apache.parquet.crypto.keytools.KmsClient;
import org.slf4j.Logger;
import org.slf4j.LoggerFactory;

import java.nio.ByteBuffer;
```

```
import java.nio.charset.Charset;
import java.nio.charset.StandardCharsets;

public class AwsKmsClient implements KmsClient {

    private static final AWSKMS_AWSKMS_CLIENT = AWSKMSClientBuilder
        .standard()
        .withRegion(Regions.US_WEST_2)
        .build();
    private static final Logger LOG = LoggerFactory.getLogger(AwsKmsClient.class);

    private String kmsToken;
    private Configuration hadoopConfiguration;

    @Override
    public void initialize(Configuration configuration, String kmsInstanceID, String
kmsInstanceURL, String accessToken) throws KeyAccessDeniedException {
        hadoopConfiguration = configuration;
        kmsToken = accessToken;
    }

    @Override
    public String wrapKey(byte[] keyBytes, String masterKeyIdentifier) throws
KeyAccessDeniedException {
        String value = null;
        try {
            ByteBuffer plaintext = ByteBuffer.wrap(keyBytes);

            EncryptRequest req = new
EncryptRequest().withKeyId(masterKeyIdentifier).withPlaintext(plaintext);
            ByteBuffer ciphertext = AWSKMS_CLIENT.encrypt(req).getCiphertextBlob();

            byte[] base64EncodedValue = Base64.encode(ciphertext.array());
            value = new String(base64EncodedValue, Charset.forName("UTF-8"));
        } catch (AmazonClientException ae) {
            throw new KeyAccessDeniedException(ae.getMessage());
        }
        return value;
    }

    @Override
    public byte[] unwrapKey(String wrappedKey, String masterKeyIdentifier) throws
KeyAccessDeniedException {
```

```

byte[] arr = null;
try {
    ByteBuffer ciphertext =
ByteBuffer.wrap(Base64.decode(wrappedKey.getBytes(StandardCharsets.UTF_8)));
    DecryptRequest request = new
DecryptRequest().withKeyId(masterKeyIdIdentifier).withCiphertextBlob(ciphertext);
    ByteBuffer decipheredtext =
AWSKMS_CLIENT.decrypt(request).getPlaintext();
    arr = new byte[decipheredtext.remaining()];
    decipheredtext.get(arr);
} catch (AmazonClientException ae) {
    throw new KeyAccessDeniedException(ae.getMessage());
}
return arr;
}
}

```

- 按照《AWS Key Management Service 开发者指南》中的创建密钥中所述，为页脚以及您的IAM角色具有访问权限的列[创建 AWS KMS 加密密钥](#)。默认IAM角色是 EMR\_ECS\_default。
- 在 Amazon EMR 集群上的 Hive 应用程序上，使用上面的ADD JAR语句添加上面的客户端，如[Apache Hive](#) 资源文档中所述。下面是一个示例语句：

```
ADD JAR 's3://location-to-custom-jar';
```

另一种方法是使用引导操作将添加JAR到 Hive 中。auxlib以下是要添加到引导操作的示例行：

```
aws s3 cp 's3://location-to-custom-jar' /usr/lib/hive/auxlib
```

- 设置下列配置值：

```

set
parquet.crypto.factory.class=org.apache.parquet.crypto.keytools.PropertiesDrivenCryptoFactory;
set
parquet.encryption.kms.client.class=org.apache.parquet.crypto.keytools.AwsKmsClient;

```

- 创建一个 Parquet 格式的 Hive 表，在其中指定 AWS KMS 键SERDEPROPERTIES并向其中插入一些数据：

```

CREATE TABLE my_table(name STRING, credit_card STRING)
ROW FORMAT SERDE 'org.apache.hadoop.hive ql.io.parquet.serde.ParquetHiveSerDe'

```

```

WITH SERDEPROPERTIES (
  'parquet.encryption.column.key'=<aws-kms-key-id-for-column-1>: credit_card',
  'parquet.encryption.footer.key'='<aws-kms-key-id-for-footer>')
STORED AS parquet
LOCATION "s3://<bucket>/<warehouse-location>/my_table";

INSERT INTO my_table SELECT
java_method ('org.apache.commons.lang.RandomStringUtils','randomAlphabetic',5) as
  name,
java_method ('org.apache.commons.lang.RandomStringUtils','randomAlphabetic',10) as
  credit_card
from (select 1) x lateral view posexplode(split(space(100),' ')) pe as i,x;

select * from my_table;

```

6. 验证当您在同一位置创建无法访问 AWS KMS 密钥的外部表时（例如，IAM角色访问被拒绝），您是否无法读取数据。

```

CREATE EXTERNAL TABLE ext_table (name STRING, credit_card STRING)
ROW FORMAT SERDE 'org.apache.hadoop.hive ql.io.parquet.serde.ParquetHiveSerDe'
STORED AS parquet
LOCATION "s3://<bucket>/<warehouse-location>/my_table";

SELECT * FROM ext_table;

```

7. 最后一条语句应触发以下异常：

```

Failed with exception
java.io.IOException:org.apache.parquet.crypto.KeyAccessDeniedException: Footer key:
access denied

```

## 2 中的传输中 HiveServer加密

从 Amazon EMR 版本 6.9.0 开始，作为安全配置的一部分，HiveServer2 (HS2) SSL 已启用TLS/。[2 中的传输中 HiveServer加密](#)这会影响到您连接到启用传输中加密的 Amazon EMR 集群上HS2运行的方式。要连接HS2，必须修改中的TRUSTSTORE\_PATH和TRUSTSTORE\_PASSWORD参数值JDBCURL。以下URL是HS2带有所需参数的JDBC连接示例：

```

jdbc:hive2://HOST_NAME:10000/
default;ssl=true;sslTrustStore=TRUSTSTORE_PATH;trustStorePassword=TRUSTSTORE_PASSWORD

```



使用下面的相应说明进行集群内或集群外 HiveServer 2 加密。

### On-cluster HS2 access

如果您在访问主节点之后使用 Beeline 客户端访问 HiveServer 2，则使用配置 `ssl.server.truststore.location` 和 `/etc/hadoop/conf/ssl-server.xml` 用查找 `TRUSTSTORE_PATH` 和 `TRUSTSTORE_PASSWORD` 参数值。SSH `ssl.server.truststore.password`

以下示例命令可帮助您检索这些配置：

```
TRUSTSTORE_PATH=$(sed -n '/ssl.server.truststore.location/,+2p' /etc/hadoop/conf/ssl-server.xml | awk -F "[><]" '/value/{print $3}')
TRUSTSTORE_PASSWORD=$(sed -n '/ssl.server.truststore.password/,+2p' /etc/hadoop/conf/ssl-server.xml | awk -F "[><]" '/value/{print $3}')
```

### Off-cluster HS2 access

如果您从 Amazon EMR 集群之外的客户端访问 HiveServer 2。您可以使用以下方法之一来获取和：`TRUSTSTORE_PATH``TRUSTSTORE_PASSWORD`

- 将[安全配置](#)期间创建 PEM 的文件转换为 JKS 文件，并在 JDBC 连接中使用该文件 URL。例如，对于 `openssl` 和 `keytool`，请使用以下命令：

```
openssl pkcs12 -export -in trustedCertificates.pem -inkey privateKey.pem -out trustedCertificates.p12 -name "certificate"
keytool -importkeystore -srckeystore trustedCertificates.p12 -srcstoretype pkcs12 -destkeystore trustedCertificates.jks
```

- 或者，参考 `/etc/hadoop/conf/ssl-server.xml` 以使用配置 `ssl.server.truststore.location` 和 `ssl.server.truststore.password` 来查找 `TRUSTSTORE_PATH` 和 `TRUSTSTORE_PASSWORD` 参数值。将信任存储库文件下载到客户端计算机，并使用客户端计算机上的路径作为 `TRUSTSTORE_PATH`。

有关从 Amazon EMR 集群之外的客户端访问应用程序的更多信息，请参阅[使用 Hive JDBC 驱动程序](#)。

## Hive 发行历史记录

下表列出了 Amazon 的每个发行版本中包含的 Hive 版本EMR，以及随应用程序一起安装的组件。有关每个发行版本中的组件版本，请参阅 [亚马逊 EMR 7.x 发布版本](#)、[亚马逊 EMR 6.x 发布版本](#) 或 [亚马逊 EMR 5.x 发布版本](#) 中的发行版“组件版本”部分。

### Hive 版本信息

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-7.2.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server
emr-5.36.2	2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client,

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
		hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn
emr-7.1.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-7.0.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server
emr-6.15.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-6.14.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server
emr-6.13.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-6.12.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server
emr-6.11.1	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-6.11.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server
emr-6.10.1	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-6.10.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server
emr-6.9.1	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server



Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-6.9.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-6.8.1	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-6.8.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-6.7.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.36.1	2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn
emr-5.36.0	2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-6.6.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-5.35.0	2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-6.5.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-6.4.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-6.3.1	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-6.3.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-6.2.1	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-6.2.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-6.1.1	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-6.1.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server



Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-6.0.1	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-6.0.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.34.0	2.3.8	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn
emr-5.33.1	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.33.0	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn
emr-5.32.1	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.32.0	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn
emr-5.31.1	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.31.0	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn
emr-5.30.2	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.30.1	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn
emr-5.30.0	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.29.0	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mysql-server, tez-on-yarn
emr-5.28.1	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mysql-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.28.0	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mysql-server, tez-on-yarn
emr-5.27.1	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn



Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.27.0	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.26.0	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.25.0	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.24.1	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.24.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.23.1	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.23.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.22.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.21.2	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.21.1	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.21.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.20.1	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.20.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpserver, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.19.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpserver, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.19.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftp-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.18.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftp-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn



Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.18.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.17.2	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.17.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.17.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.16.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.16.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.15.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.15.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.14.2	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.14.1	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.14.0	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.13.1	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.13.0	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.12.3	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.12.2	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.12.1	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn



Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.12.0	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.11.4	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.11.3	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.11.2	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.11.1	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.11.0	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.10.1	2.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.10.0	2.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.9.1	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.9.0	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.8.3	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.8.2	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.8.1	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.8.0	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.7.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.7.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn



Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.6.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.6.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.5.4	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.5.3	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.5.2	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.5.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.5.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.4.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.4.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.3.2	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.3.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn
emr-5.3.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.2.3	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn
emr-5.2.2	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.2.1	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn
emr-5.2.0	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn



Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.1.1	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn
emr-5.1.0	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.0.3	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn
emr-5.0.2	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-5.0.1	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn
emr-5.0.0	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-4.9.6	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.9.5	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-4.9.4	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.9.3	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-4.9.2	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.9.1	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-4.8.5	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.8.4	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-4.8.3	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.8.2	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server



Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-4.8.1	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.8.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-4.7.4	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.7.3	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-4.7.2	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.7.1	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-4.7.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.6.1	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-4.6.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server
emr-4.5.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server
emr-4.4.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-4.3.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server
emr-4.2.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server
emr-4.1.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server

Amazon EMR 发布标签	Hive 版本	随 Hive 安装的组件
emr-4.0.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server

## Hive 发布说明 ( 按版本分类 )

### 主题

- [亚马逊 EMR 7.1.0-Hive 发行说明](#)
- [亚马逊 EMR 6.15.0-Hive 发行说明](#)
- [亚马逊 EMR 6.14.0-Hive 发行说明](#)
- [亚马逊 EMR 6.13.0-Hive 发行说明](#)
- [亚马逊 EMR 6.12.0-Hive 发行说明](#)
- [亚马逊 EMR 6.11.0-Hive 发行说明](#)
- [亚马逊 EMR 6.10.0-Hive 发行说明](#)
- [亚马逊 EMR 6.9.0-Hive 发行说明](#)
- [亚马逊 EMR 6.8.0-Hive 发行说明](#)
- [亚马逊 EMR 6.7.0-Hive 发行说明](#)
- [亚马逊 EMR 6.6.0-Hive 发行说明](#)

## 亚马逊 EMR 7.1.0-Hive 发行说明

### 亚马逊 EMR 7.1.0-Hive 的变化

类型	描述
错误修复	<a href="#">HIVE-24381</a> -如果包含跳过页眉/页脚，则压缩文本输入返回 0 行。
错误修复	<a href="#">HIVE-24190</a> —LLAP: EXCEPTION 根据-4233， ShuffleHandler 可能会返回 DISK ERROR __。 TEZ
错误修复	<a href="#">HIVE-23073</a> — Shade Netty。
错误修复	<a href="#">HIVE-23073</a> — Shade Netty 然后升级到 netty 4.1.48.Final。
错误修复	<a href="#">HIVE-23148</a> — 由于 netty 阴影，Llap 外部客户端流中断。
错误修复	<a href="#">HIVE-25180</a> — 升级 Netty。
错误修复	<a href="#">HIVE-24524</a> — LLAP ShuffleHandler：升级到 Netty4 并在可能的情况下从 Hive 中删除 Netty3 依赖关系。
错误修复	<a href="#">HIVE-28000</a> — Hive QL：当无法进行类型强制时，“not in”子句给出的结果不正确。
错误修复	<a href="#">HIVE-27993</a> — Netty4 ShuffleHandler 应该使用 1 个 boss 线程。
Upgrade	将 Netty 升级到 4.1.100.Final
Upgrade	将 Jetty 升级到 9.4.53.v20231009
Upgrade	将 Zookeeper 升级到 3.9.1



## 亚马逊 EMR 7.1.0-Hive 的变化

- 亚马逊 EMR 7.1 将 Hive 升级到 Netty 4.1.100.Final，以解决 Netty3 中的安全漏洞。由于 hive-druid-handler 依赖于 netty3，因此 Hive 在 Amazon 7.1 hive-druid-handler JAR 中没有 Hive 的类路径中。EMR 一旦 Druid 处理程序支持 4.1.100.Final 或更高 EMR 版本的 Netty，即将发布的亚马逊版本就会将其包含在 Hive 的类路径中。如果您需要 Amazon 7.1 或更高 EMR 版本 hive-druid-handler JAR 中的，请联系 AWS 支持人员。

## 亚马逊 EMR 6.15.0-Hive 发行说明

### 亚马逊 EMR 6.15.0-Hive 变更

类型	描述
功能	支持 <a href="#">TEZ-4397</a> — 对于 Tez 异步拆分打开，Hive 现在支持中描述的 Tez 配置。 <a href="#">Tez 异步拆分打开</a>
错误修复	<a href="#">HIVE-25400</a> — 将更新的偏移量移至 <code>BytesColumnVector setValPre allocated</code>
错误修复	<a href="#">HIVE-25190</a> — 修复了中的许多小额分配。 <code>BytesColumnVector</code>
Upgrade	将 Apache Hadoop 升级到 3.3.6。
Upgrade	<a href="#">HIVE-26684</a> — 升级到 <code>maven-shade-plugin 3.4.1</code> 。
改进	要缩短 Amazon EMR 集群的启动时间，请从 HCatalog 启动脚本中删除 15 秒的睡眠时间。

## 亚马逊 EMR 6.14.0-Hive 发行说明

### 亚马逊 EMR 6.14.0-Hive 的变化

类型	描述
改进	<a href="#">HIVE-26762</a> : 移除操作数修剪 HiveFilterSetOpTransposeRule
错误修复	<a href="#">HIVE-27582</a> : 不要将HBase表输入格式缓存为 FetchOperator
错误修复	<a href="#">HIVE-26452</a> : NPE当转换JOIN到MAPJOIN和 JOIN列被多次引用时
错误修复	<a href="#">HIVE-26416</a> : AcidUtils.isRawFormatFile() 向文件投掷 InvalidProtocolBufferException non-ORC
错误修复	<a href="#">HIVE-26105</a> : 如果“注释”列包含特定的汉字，则显示列会显示额外值
错误修复	<a href="#">HIVE-25864</a> : Hive 查询优化使用窗口功能为谓词下推创建了错误的计划
错误修复	<a href="#">HIVE-25224</a> : 涉及表的多个INSERT语句出错结果不同 bucketing_versions
错误修复	<a href="#">HIVE-24151</a> : 如果字符串包含字符，则MultiDelimitSerDe 移动数据 non-ASCII
错误修复	<a href="#">HIVE-23606</a> : (LLAP) 延迟清理 DirectByteBuffer EncodedReaderImpl
错误修复	<a href="#">HIVE-22165</a> : <a href="#">HIVE-14296</a> 引入的同步SessionManager.closeSession 会导致繁忙的 Hive 服务器出现高延迟
错误修复	<a href="#">HIVE-21304</a> : 让存储桶版本的使用更强大

## 亚马逊 EMR 6.13.0-Hive 发行说明

### 亚马逊 EMR 6.13.0-Hive 的变化

类型	描述
改进	升级 Python 脚本以支持 Python3
改进	<a href="#">HIVE-27097</a> : 改进客户端和服务器的重试策略 MetaStore
错误修复	<a href="#">HIVE-21778</a> :CBO: “Struct 不为空” 被评估为可为空总是导致查询中过滤器失误
错误修复	<a href="#">HIVE-21009</a> : 添加用户设置绑定用户的功能
错误修复	<a href="#">HIVE-22661</a> : 在路径中加载了数据的非存储桶表上压缩失败
错误修复	<a href="#">HIVE-19718</a> : 批量添加分区还会为每个分区提取表
错误修复	<a href="#">HIVE-22173</a> : 在编译过程中使用多个横向视图的查询挂起
错误修复	<a href="#">HIVE-27088</a> : 合并带有后联接过滤器的内部和外部联接时结果不正确
错误修复	<a href="#">HIVE-21935</a> : Hive 矢量化 : 使用矢量化会降低性能 UDF
错误修复	<a href="#">HIVE-25299</a> : 对于非时区, 将时间戳转换为数字数据类型是不正确的 UTC
错误修复	<a href="#">HIVE-24626</a> :LLAP: 如果所有 IO 电梯线程都忙着排队到其他队列已满的读取器, 则读取器线程可能会被饿死
错误修复	<a href="#">HIVE-27029</a> : Hive 查询失败并显示文件系统关闭错误, -26352 已完成重做 HIVE

类型	描述
错误修复	<a href="#">HIVE-26352</a> : Tez 队列访问检查失败，压缩时出现异常 GSS
错误修复	<a href="#">HIVE-24590</a> : 操作日志仍然会泄露 log4j 附加程序
错误修复	<a href="#">HIVE-24552</a> : 可能的HMS连接泄漏或积聚在里面 loadDynamicPartitions
错误修复	<a href="#">HIVE-27069</a> : <a href="#">加入存储桶地图时结果不正确</a>
错误修复	<a href="#">HIVE-27344</a> : 在 #close 中添加空值检查 RecordReaderImpl
错误修复	<a href="#">HIVE-27439</a> : Support 十进制空格
错误修复	<a href="#">HIVE-27267</a> : 使用子查询对十进制分区列进行存储桶映射联接时结果不正确
错误修复	<a href="#">HIVE-21986</a> : HiveServer Web UI : 在默认响应标头中设置 Strict-Transport-Security
错误修复	<a href="#">HIVE-22148</a> : S3A 委托令牌未添加到 Compactor 的任务配置中。
错误修复	<a href="#">HIVE-22622</a> : Hive 允许创建具有重复属性名称的结构
错误修复	<a href="#">HIVE-22008</a> : LIKE运算符应匹配多行输入
错误修复	<a href="#">HIVE-23144</a> :LLAP: 开始清理吧 QueryTracker serviceStop
错误修复	<a href="#">HIVE-22391</a> : NPE在检查 Hive 查询结果缓存时
错误修复	<a href="#">HIVE-23305</a> : LlapTaskSchedulerService addNode 由于比赛条件而 NullPointerException 入场

类型	描述
错误修复	<a href="#">HIVE-22178</a> : Parq FilterPredicate uet 在后面扔了 CastException SchemaEvolution
错误修复	<a href="#">HIVE-21517</a> : 修复 AggregateStatsCache
错误修复	<a href="#">HIVE-21825</a> : 改善启用主动/被动 HA 时的客户端错误消息
错误修复	<a href="#">HIVE-23389</a> : <u>可能导致</u> FilterMergeRule AssertionError
错误修复	<a href="#">HIVE-22767</a> : Beeline 无法正确解析注释中的分号
错误修复	<a href="#">HIVE-22996</a> : BasicStats 解析时应主动检查空字符串或空字符串
错误修复	<a href="#">HIVE-22808</a> : <u>无法处理</u> HiveRelFieldTrimmer HiveTableFunctionScan
错误修复	<a href="#">HIVE-22437</a> : 锁定元数据NPE时缓存LLAP元数据。
错误修复	<a href="#">HIVE-22606</a> : 在级别下 AvroSerde 记录 avro.schema.literal INFO
错误修复	<a href="#">HIVE-22713</a> : 不对 Join-Fil (*)-RS 结构进行持续传播
错误修复	<a href="#">HIVE-21624</a> :LLAP: 线程级别的 Cpu 指标已损坏
错误修复	<a href="#">HIVE-22815</a> : 减少中不必要的文件系统对象创建 MROutput
错误修复	<a href="#">HIVE-23060</a> : 查询失败并显示错误“分组集表达式不在 BY 键中。GROUP在令牌附近遇到错误”

类型	描述
错误修复	<a href="#">HIVE-22236</a> : 无法创建视图选择包含子查询的视图 NOT IN
错误修复	<a href="#">HIVE-19886</a> : 如果使用 <code>—hiveconf hive.log.file</code> , 则日志可能会被定向到 2 个文件
错误修复	<a href="#">HIVE-20620</a> : 使用动态分区插入存储桶排序 MM 表时会出现冲突
错误修复	<a href="#">HIVE-14557</a> : 同时启用 MapJoin 时空指针 SkewJoin
错误修复	<a href="#">HIVE-20471</a> : 获取默认数据库路径时出现问题
错误修复	<a href="#">HIVE-20598</a> : 修复计算中的错别字 HiveAlgorithmsUtil
错误修复	<a href="#">HIVE-14737</a> : 在 Kerberized Hive Server 2 Web 用户界面中访问 /logs 时出现问题
错误修复	<a href="#">HIVE-20733</a> : GenericUDFOPEqual NS 不得在计划描述中使用 =
错误修复	<a href="#">HIVE-20848</a> : 设置后 UpdateInputAccessTimeHook 查询失败并显示未找到表。
错误修复	<a href="#">HIVE-18929</a> : HiveStringUtils.java humanReadableInt 中的方法存在争用条件。
错误修复	<a href="#">HIVE-20841</a> :LLAP: 将动态端口设置为可配置
错误修复	<a href="#">HIVE-20930</a> : VectorCoalesce 在FILTER模式下不生效
错误修复	<a href="#">HIVE-21007</a> : 半加入 + 联盟可能会导致计划错误

类型	描述
错误修复	<a href="#">HIVE-21074</a> : Hive 存储桶表查询修剪不适用于 IS 条件 NOT NULL
错误修复	<a href="#">HIVE-21223</a> : 分区不存在时 CachedStore 返回空分区
错误修复	<a href="#">HIVE-19625</a> : 在 Hive 中潜在NPE并隐藏实际异常# copyFiles
错误修复	<a href="#">HIVE-17020</a> : 激进的 RS 重复数据删除可能会错误地移除 OP 树枝
错误修复	<a href="#">HIVE-20168</a> : <a href="#">隐藏日志</a> 记录 ReduceSinkOperator
错误修复	<a href="#">HIVE-20879</a> : 在投影表达式中使用空值会导致 CastException
错误修复	<a href="#">HIVE-20888</a> : TxnHandler: 在不可变列表上调用 sort ()
错误修复	<a href="#">HIVE-19948</a> : 如果字符串内 HiveCli 有引号 , 则无法正确用分号分隔命令
错误修复	<a href="#">HIVE-20621</a> : 在 resultset.n GetOperationStatus ext 中调用会导致增量缓慢
错误修复	<a href="#">HIVE-20854</a> : Sensible 默认值 : Hive 的 Zookeeper 心跳间隔为 20 分钟 , 改为 2
错误修复	<a href="#">HIVE-20330</a> : HCatLoader无法为具有多个 InputJobInfo 输入的作业处理多个对象
错误修复	<a href="#">HIVE-20787</a> : MapJoinBytesTableContainer dummyRow 手机壳无法处理重用

类型	描述
错误修复	<a href="#">HIVE-20331</a> : 使用 union all、横向视图和 Join 进行查询失败，并显示“无法在子运算符中找到父运算符”
错误修复	<a href="#">HIVE-19968</a> : UDF异常不会被抛出
错误修复	<a href="#">HIVE-20410</a> : 中止在事务表上插入覆盖会导致“没有足够的历史记录可用于...” 错误
错误修复	<a href="#">HIVE-20059</a> : Hive 直播应在例外情况下无条件尝试阴影前缀
错误修复	<a href="#">HIVE-19424</a> : 在 NPE MetaDataFormatters
错误修复	<a href="#">HIVE-20355</a> : 的清理参数。 HiveConnection setSchema
错误修复	<a href="#">HIVE-20858</a> : 未使用实用程序中的配置正确初始化序列化器。 createEmptyBuckets
错误修复	<a href="#">HIVE-20424</a> : schematool 不得污染直线历史记录
错误修复	<a href="#">HIVE-20338</a> :LLAP: 对于具有突变语义协议含义的文件系统强制使用合成文件 ID HDFS POSIX
错误修复	<a href="#">HIVE-11708</a> : 逻辑运算符提升为 ClassCast Exceptions NULL
错误修复	<a href="#">HIVE-21082</a> : 在HPL/中SQL，声明语句不支持字符类型的变量
错误修复	<a href="#">HIVE-16690</a> : <a href="#">根据集群大小配置 Tez 笛卡尔乘积边缘 LLAP</a>
错误修复	<a href="#">HIVE-21296</a> : 删除 varchar 分区会引发异常



类型	描述
错误修复	<a href="#">HIVE-14516</a> : 。 OrclInputFormat SplitGenerator。 callInternal
错误修复	<a href="#">HIVE-20981</a> : 直播/泄露 AbstractRecordWriter HeapMemoryMonitor
错误修复	<a href="#">HIVE-20043</a> : HiveServer2 : SessionState 周围有一个静态同步块 AtomicBoolean
错误修复	<a href="#">HIVE-20191</a> : 如果 PreCommit 补丁为空，补丁应用程序不会失败
错误修复	<a href="#">HIVE-20400</a> : 创建表应始终使用完全限定的路径，以避免潜在的 FS 歧义
错误修复	skewedInfo 在访问倾斜的列之前添加空值检查

## 亚马逊 EMR 6.12.0-Hive 发行说明

### 亚马逊 EMR 6.12.0-Hive 的变化

类型	描述
改进	添加了对 JDK 11 和 JDK 17 运行时的 Support
改进	添加了使用 S3 Select 时对查询区分大小写和保留关键字列名的支持。要使用它，请以“s3select.column.mapping”= " <i>column1:fieldName1</i> , <i>column2:fieldName2</i> , ...” 的格式定义表格属性
改进	<a href="#">HIVE-23133</a> : 不同硬件架构的数值运算可能有不同的结果
改进	<a href="#">HIVE-27145</a> : StrictMath 用于剩余的数学函数，作为 -23133 的后续函数 HIVE

类型	描述
错误修复	<a href="#">修复在 Hive 6.4.0 中移植 -22900 而导致的 <code>get_partitions_by_filter</code> 和 <code>get_num_partitions_by_filter</code> 中的通配符不兼容问题</a> <a href="#">HMS APIs</a> <a href="#">HIVE</a> EMR
错误修复	<a href="#">HIVE-26736</a> : 包含子句的嵌套视图授权失败 WITH
错误修复	<a href="#">HIVE-22416</a> : 启用并行执行时缺少与 MR 相关的操作日志
错误修复	<a href="#">HIVE-19653</a> : 带有分组集的 <code>groupby</code> 的谓词下推不正确
错误修复	<a href="#">HIVE-22094</a> : <a href="#">查询失败并显示</a> : <code>hive.ql.exec.vector.ClassCastException r.DecimalColumnVector</code> 无法转换到 <code>Hive.ql.exec.vector.decimal64 ColumnVector</code>
错误修复	<a href="#">HIVE-26340</a> : 如果查询具有大写窗口函数，则向量化PTF运算符失败
错误修复	<a href="#">HIVE-26184</a> : COLLECT_ 当某些GROUP按键高度倾斜时，SET使用 BY 的速度非常慢
错误修复	<a href="#">HIVE-26373</a> : 从HBase带有 Avro 数据的表中读取时间戳 <code>ClassCastException</code> 时
错误修复	<a href="#">HIVE-26388</a> : <code>ClassCastException</code> 当CTAS 查询源表中存在非字符串类型的列时升级- <a href="#">HIVE26172</a> : <a href="#">Hive-由于 -2021-36373 和 -2021-36374</a> ，将 Ant 升级到 1.10.11 CVE CVE
错误修复	<a href="#">HIVE-26114</a> : 使用带前缀空间的 <code>dfs</code> 命令修复 <code>jdbc</code> 连接 <code>hiveserver2</code> 会导致异常

类型	描述
错误修复	<a href="#">HIVE-26396</a> : trunc 函数在精度拦截方面存在问题，结果有很多 0
错误修复	<a href="#">HIVE-26446</a> : HiveProtoLoggingHook 无法填充分区表的 TablesWritten 字段。
错误修复	<a href="#">HIVE-26639</a> : ConstantVectorExpression 并且 ExplainTask 不应依赖默认字符集
错误修复	<a href="#">HIVE-22670</a> : ArrayIndexOutOfBoundsException 当使用矢量化阅读器读取镶木地板文件时
错误修复	<a href="#">HIVE-23607</a> : 权限问题：在另一个视图上创建视图成功但更改视图失败
错误修复	<a href="#">HIVE-25498</a> : 包含超过 31 个计数不同函数的查询返回错误结果
错误修复	<a href="#">HIVE-25780</a> : DistinctExpansion 创建超过 64 个分组集 II
错误修复	<a href="#">HIVE-23868</a> : 窗口功能规范：支持 0 之前/以下
错误修复	<a href="#">HIVE-24539</a> : OrcInputFormat 架构生成应遵循列分隔符
错误修复	<a href="#">HIVE-23476</a> :LLAP: 还要为 mmap 案例预先分配竞技场
错误修复	<a href="#">HIVE-25806</a> : 可能泄漏-Parquet , I LlapCache AwareFs O LLAP
错误修复	<a href="#">HIVE-23498</a> : 开启禁用 HTTP Trace 方法 ThriftHttpCliService
错误修复	<a href="#">HIVE-25729</a> : ThriftUnionObjectInspector 应在完全启动后收到通知

类型	描述
错误修复	<a href="#">HIVE-23846</a> : 避免对位向量进行不必要的序列化和反序列化
错误修复	<a href="#">HIVE-24233</a> : 除非子查询在禁用 cbo 的情况下抛出空指针
错误修复	<a href="#">HIVE-24276</a> : HiveServer2 loggerconf.jsp 跨站脚本 () 漏洞 XSS
错误修复	<a href="#">HIVE-25721</a> : 外部联接结果错误
错误修复	<a href="#">HIVE-25223</a> : 带限制的选择不返回非本机表上的任何行
错误修复	<a href="#">HIVE-25794</a> : CombineHiveRecordReader: 循环中的日志语句会导致内存压力
错误修复	<a href="#">HIVE-23602</a> : 使用 Java Concurrent Package 作为操作句柄集
错误修复	<a href="#">HIVE-24045</a> : 没有与创建默认数据库的时间相关的日志
错误修复	<a href="#">HIVE-24305</a> : 如果将值括在引号中, avro 十进制架构无法正确填充比例/精度
错误修复	<a href="#">HIVE-25844</a> : 异常反序列化错误可能会导致 beeline 立即终止
错误修复	<a href="#">HIVE-25040</a> : 删除数据库级联无法删除永久函数
错误修复	<a href="#">HIVE-23501</a> : AOOB VectorDeserializeRow 当复杂类型转换为原始类型时
错误修复	<a href="#">HIVE-23704</a> : Thrift HTTP Server 无法正确处理身份验证处理

类型	描述
错误修复	<a href="#">HIVE-23529</a> : 当 row_deserial CTAS ize 时 uniontype 中断了
错误修复	<a href="#">HIVE-24144</a> : 中的 getIdentifierQuote字符串 HiveDatabaseMetaData 返回的值不正确
错误修复	<a href="#">HIVE-23850</a> : PPD当主题不是存在分组集的列时允许
错误修复	<a href="#">HIVE-24036</a> : 序列化通话计划时出现了 Kryo 异常 getSplits UDF
错误修复	<a href="#">HIVE-25919</a> : ClassCastException 当推入布尔列谓词时 HBaseStorageHandler
错误修复	<a href="#">HIVE-25261</a> : R etryingHMSHandler 应该用目标的简短描述来 MetaException 包裹
错误修复	<a href="#">HIVE-24792</a> : <a href="#">操作</a> 中可能出现线程泄漏
错误修复	<a href="#">HIVE-23409</a> : 如果由于时间轴服务关闭而导致 TezSession 应用程序重新打开失败 , SessionPool 则在重试后将关闭默认 TezSession 从
错误修复	<a href="#">HIVE-23615</a> : 不要尊重 Beeline 命令类中的空指针
错误修复	<a href="#">HIVE-24849</a> : 当位置有大量文件时创建外部表套接字超时 ( 影响 3.1.2 )
错误修复	<a href="#">HIVE-24193</a> : 在重命名的 hive acid 表上选择查询不会产生任何输出
错误修复	<a href="#">HIVE-25209</a> : 使用生成意外SELECT结果的SUM函数进行查询

类型	描述
错误修复	<a href="#">HIVE-23666</a> : 当 groupby 运算符没有设置分组时, 会跳过 checkHashMode效率
错误修复	<a href="#">HIVE-23873</a> : 当关闭时, 查询 Hive JDBCStorageHandler 表失败 NPE CBO
错误修复	<a href="#">HIVE-24149</a> : HiveStreamingConnection 无法关闭连接 HMS
错误修复	<a href="#">HIVE-25561</a> : 被终止的任务不应提交文件。(影响 2.x 和 3.x 版本)
错误修复	<a href="#">HIVE-25683</a> : 关闭读者进来。AcidUtils isRawFormat文件
错误修复	<a href="#">HIVE-24294</a> : TezSessionPool 会话可以抛出 AssertionError
错误修复	<a href="#">HIVE-24182</a> : 永久游骑兵授权问题 UDFs
错误修复	<a href="#">HIVE-22805</a> : 未实现使用条件数组或地图的矢量化并引发错误
错误修复	<a href="#">HIVE-22828</a> : Decimal64 : NVL并且CASE语句会隐式地将 decimal64 转换为 128
错误修复	<a href="#">HIVE-21398</a> : 不应将包含估计统计数据的列视为唯一键
错误修复	<a href="#">HIVE-22490</a> : 添加路径中包含特殊字符的 jar 会引发错误
错误修复	<a href="#">HIVE-22700</a> : 未经授权的压缩可能会泄漏内存
错误修复	<a href="#">HIVE-22053</a> : 创建函数时函数名称未标准化
错误修复	<a href="#">HIVE-22595</a> : 在带有外部架构的 Avro 表上进行动态分区插入失败

类型	描述
错误修复	<a href="#">HIVE-21795</a> : 在分区表上进行 mapjoin 时，汇总摘要行可能会丢失
错误修复	<a href="#">HIVE-22987</a> : in when when ClassCast Exception 为空 VectorCoalesce DataTypePhysicalVariation
错误修复	<a href="#">HIVE-22814</a> : ArrayIndexOutOfBounds 在向量化中 getData Type PhysicalVariation
错误修复	<a href="#">HIVE-22523</a> : 如果队列已满，中的错误处理程序 LlapRecordReader 可能会阻塞
错误修复	<a href="#">HIVE-21796</a> : ArrayWritableObjectInspector.equals 可能需要 $O(2^{\text{nesting\_depth}})$ 时间
错误修复	<a href="#">HIVE-22929</a> : 性能：带引号的标识符解析通过字符串使用一次性正则表达式。replaceAll()
错误修复	<a href="#">HIVE-21641</a> : 与 beeline 相比，Llap 外部客户端返回精度/比例不同的十进制列
错误修复	<a href="#">HIVE-22207</a> : Tez : 当集群上的“dfs.block.size”为“128m” NumberFormatException 时 SplitGenerator 抛出
错误修复	<a href="#">HIVE-22114</a> : 当所有存储桶都为空时，插入仅限分区插入表的查询失败
错误修复	<a href="#">HIVE-22165</a> : -14296 之后引入了HIVE同步。SessionManager closeSession 在繁忙的 Hive 服务器中导致高延迟
错误修复	<a href="#">HIVE-22744</a> : TezTask 对于具有多个外边的顶点，应具有比例排序内存

类型	描述
错误修复	<a href="#">HIVE-22072</a> : 更改表格以进行列更改不会更新约束引用
错误修复	<a href="#">HIVE-22075</a> : 修复从 -14200 开始的 max-reducers=1 回归 HIVE
错误修复	<a href="#">HIVE-22527</a> : Hive on Tez : 合并小文件的 Job 将提交到另一个队列 ( 默认队列 )
错误修复	<a href="#">HIVE-22816</a> : QueryCache: 使用视图的查询可以在扩展后对其进行缓存 CTE
错误修复	<a href="#">HIVE-22733</a> : 在 Hive 中禁用操作日志属性后, 仍在HS2保存操作日志
错误修复	<a href="#">HIVE-22699</a> : 掩码UDFs应掩盖数值 0
错误修复	<a href="#">HIVE-23356</a> : 使用分组集表达式处理查询时, 哈希聚合始终处于禁用状态。
错误修复	<a href="#">HIVE-21568</a> : 。 HiveRelOptUtil isRowFiltering 计划应跳过项目
错误修复	<a href="#">HIVE-21760</a> : <a href="#">联接</a> 时应绕过共享工作优化 SMB
错误修复	<a href="#">HIVE-22712</a> : 无论用户定义的队列如何, ReExec 驱动程序都会在默认队列中执行提交查询
错误修复	<a href="#">HIVE-21397</a> : BloomFilter 对于 Hive 来说, Managed [ACID] 表无法按预期工作
错误修复	<a href="#">HIVE-23011</a> : 共享工作优化器在比较联接时应检查剩余谓词
错误修复	<a href="#">HIVE-21412</a> : PostExecOrcFileDump 不适用于表格 ACID



类型	描述
错误修复	<a href="#">HIVE-22201</a> : ArrayIndexOutOfBoundsException 如果未选择大牌桌，则checkShuffleSizeForLargeTable 抛出 ConvertJoinMapJoin #
错误修复	<a href="#">HIVE-21971</a> : 由于 `:ReflectionUtils: CONSTRUCTOR_CACHE` 使用临时函数 + 泛型，HS2泄漏了类加载器 UDF
错误修复	<a href="#">HIVE-21368</a> : 矢量化：不必要的十进制 64-> 转换 HiveDecimal
错误修复	<a href="#">HIVE-25416</a> : Hive 元存储内存泄漏是因为错误 datanucleus-api-jdo
错误修复	<a href="#">HIVE-22219</a> : 关闭节点管理器会阻止服务的重启 LLAP
错误修复	<a href="#">HIVE-21793</a> : 即使 hive.stats.fetch.column.stats 设置为 false，也会CBO检索列统计信息
错误修复	<a href="#">HIVE-22163</a> :CBO: 启用开CBO启统计估计，即使禁用了估计
错误修复	<a href="#">HIVE-18735</a> : 像丢失交易属性一样创建表
错误修复	<a href="#">HIVE-22433</a> : Hive JDBC 存储处理器：从数据源和从数据源提取的结果不正确 BOOLEAN TIMESTAMP DataType JDBC
错误修复	<a href="#">HIVE-19430</a> : 。 ObjectStore cleanNotificationEvents OutOfMemory 关于大量待处理的事件
错误修复	<a href="#">HIVE-20785</a> : 中的密钥名称错误。JDBC DatabaseMetaData getPrimaryKeys 方法

类型	描述
错误修复	<a href="#">HIVE-16116</a> : 当 beeline.hiveconfvariables= {} 在 beeline.properties 中NPE时 Beeline 会抛出
错误修复	<a href="#">HIVE-20066</a> : 将 hive.load.data.owner 与全额本金进行比较
错误修复	<a href="#">HIVE-20489</a> : 解释查询计划挂起
错误修复	<a href="#">HIVE-21033</a> : 忘记关闭操作会切断更多的 2 个输出 HiveServer
错误修复	<a href="#">HIVE-19888</a> : 来自的误导性 “METASTORE_FILTER_HOOK 将被忽略” 警告 SessionState
错误修复	<a href="#">HIVE-20303</a> : INSERTOVERWRITETABLED PARTITION b.table (...) IF NOT EXISTS 投掷 InvalidTableException
错误修复	<a href="#">HIVE-16144</a> : CompactionInfo 没有等于/但在 Set hashCode 中使用
错误修复	<a href="#">HIVE-20818</a> : 使用子查询创建的视图会将 WHERE子查询中引用的视图视为直接输入
错误修复	<a href="#">HIVE-21005</a> :LLAP: 每次拆分读取更多条纹泄漏 ZlibCodecs
错误修复	<a href="#">HIVE-20771</a> : 在空结构上 LazyBinarySerDe 失败。
错误修复	<a href="#">HIVE-18852</a> : <a href="#">变更表验证</a> 中出现误导性错误消息
错误修复	<a href="#">HIVE-21124</a> : HPL/SQL不支持该语句 CREATE TABLE LIKE
错误修复	<a href="#">HIVE-20935</a> : llap 包压缩包上传失败导致服务启动失败 EC2 LLAP

类型	描述
错误修复	<a href="#">HIVE-20409</a> : HiveACID : Update/Delete/Merge 无法清理 hdfs 暂存目录
错误修复	<a href="#">HIVE-20570</a> : 与 hive.optimize.union.remove=true 的联盟计划不正确
错误修复	<a href="#">HIVE-20421</a> : hive-default.xml.template 中的非法字符实体 '\b'
错误修复	<a href="#">HIVE-19133</a> : WebUI 分阶段性能HS2指标显示不正确
错误修复	<a href="#">HIVE-18977</a> : 列出分区使用和直接返回不同的结果 JDO SQL
错误修复	<a href="#">HIVE-20034</a> : 回滚 MetaStore 异常处理更改以实现向后兼容
错误修复	<a href="#">HIVE-20672</a> : 登录的线程 LlapTaskSchedulerService 应每隔一段固定时间间隔报告一次
错误修复	<a href="#">HIVE-12812</a> : 默认启用 mapred.input.dir.recursive 以支持与聚合函数的并集
错误修复	<a href="#">HIVE-20147</a> : Hive 直播摄取满足于同步日志记录
错误修复	<a href="#">HIVE-19203</a> : 中的线程安全问题 HiveMetaStore
错误修复	<a href="#">HIVE-20091</a> : Tez : 为输出添加安全证书 FileSinkOperator
错误修复	<a href="#">HIVE-16906</a> : Hive ATSHook 应该在连接之前检查 yarn.timeline-service.enabled ATS
错误修复	<a href="#">HIVE-20714</a> : 单个SHOW属性的 tblproperties 返回名称列中的值

类型	描述
错误修复	<a href="#">HIVE-24730</a> : Shims 类以静默方式覆盖来自 hive-site.xml 和 tez-site.xml 的值
错误修复	<a href="#">HIVE-22055</a> : 从文本文件加载数据后，选择计数给出的结果不正确

## 亚马逊 EMR 6.11.0-Hive 发行说明

### 亚马逊 EMR 6.11.0-Hive 的变化

类型	描述
改进	增加了对多线程删除分区的支持，以提高删除分区的性能
改进	支持读取编码的 Hive 查询文件
改进	默认情况下为 Hive on Tez 作业启用 Tez Shuffle Handler
错误	<a href="#">添加了一个选项，允许在启用 hive.groupby.skewindata 时向简化器分配密钥以修复错误的结果（在 -20220 中报告）HIVE</a>
错误	修复了配置默认分区名称时统计数据计算失败的问题
错误	尊重在启用传输中加密的集群中开箱即用配置 SSL 为 HiveServer 2 时传递的任何自定义 SSL 分类参数
逆向移植	<a href="#">HIVE-23617</a> : 修复了存储 api 问题 FindBug
逆向移植	<a href="#">HIVE-26408</a> : 矢量化 : 修复暂存列的取消分配，不要重复使用子列作为输出 ConstantVectorExpression

类型	描述
逆向移植	<a href="#">HIVE-23614</a> : 始终传递给 HiveConfig removeTempOr DuplicateFiles
逆向移植	<a href="#">HIVE-23354</a> : 从中删除文件大小健全性检查 compareTempOr DuplicateFiles
逆向移植	<a href="#">HIVE-20344</a> : 已修复 PrivilegeSynchronizer 抛问题。SBA AccessControlException还引入了 hive.privilege.synchronizer 属性来禁用权限同步器
逆向移植	<a href="#">HIVE-15826</a> : 支持为所有人配置 “serialization.encoding” SerDes
逆向移植	<a href="#">HIVE-18284</a> : 修复在 dynpart NPE 排序优化中插入带有 “分配依据” 子句的数据时的问题
逆向移植	<a href="#">HIVE-24930</a> : 操作员。 setDone() 在矢量化代码路径中不使用子操作的短路 ( 如果 == 1 ) childSize
逆向移植	<a href="#">HIVE-24523</a> : 的矢量化读取路径 LazySimpleSerde 不支持时间戳 SERDEPROPERTIES
逆向移植	<a href="#">HIVE-23265</a> : 在设置了限制和偏移的情况下返回重复的行集
逆向移植	<a href="#">HIVE-21492</a> : VectorizedParquetRecordReader 无法读取使用节俭/自定义工具生成的镶木地板文件
逆向移植	<a href="#">HIVE-22540</a> : 矢量化 : Decimal64 列不适用。 VectorizedBatchUtil makeLikeColumn向量 ()
逆向移植	<a href="#">HIVE-22588</a> : 切换向量 groupby 模式时，刷新其余分组集的剩余行

类型	描述
逆向移植	<a href="#">HIVE-22551</a> : BytesColumnVector initBuffer 应始终如一地清理向量和长度
逆向移植	<a href="#">HIVE-22448</a> :CBO: 使用分组按键扩展不同的多个计数
逆向移植	<a href="#">HIVE-22248</a> : 修复持续存在的统计问题
逆向移植	<a href="#">HIVE-22210</a> : 矢量化可能会重用过滤中涉及的计算输出列
逆向移植	<a href="#">HIVE-21531</a> : 矢量化 : 所有NULL哈希码都不是使用 Murmur3 计算的
逆向移植	<a href="#">HIVE-20419</a> : 矢量化 : 防止在哈希映射键中使用 VectorPartitionDesc 后发生变异
逆向移植	<a href="#">HIVE-19388</a> : ClassCastException 在初始化期间 VectorMapJoinCommonOperator
逆向移植	<a href="#">HIVE-21584</a> : Java 11 准备工作 : 系统类加载器不是 URLClassLoader
逆向移植	<a href="#">HIVE-25107</a> : Classpath 日志记录应该处于级别 (#2271) DEBUG
逆向移植	<a href="#">HIVE-22097</a> : java.util 不兼容。 ArrayList适用于 java 11
逆向移植	<a href="#">HIVE-23938</a> :LLAP: JDK11-某些 GC 日志文件轮换相关的 jvm 参数不能再使用了
逆向移植	<a href="#">HIVE-26226</a> : 在升级时将 jdk.tools dep 从 hive-metastore 中排除
逆向移植	<a href="#">HIVE-17879</a> : 升级 Datanucleus Maven 插件

类型	描述
逆向移植	<a href="#">HIVE-27004</a> : DateTimeFormatterBuilder # 在高于 8 的 Java 版本中appendZoneText 无法解析 UTC '+'
逆向移植	<a href="#">HIVE-16812</a> : VectorizedOrcAcidRowBatchReader不筛选删除事件
逆向移植	<a href="#">HIVE-17917</a> : 。 VectorizedOrcAcidRowBatchReader computeOffsetAnd存储桶优化
逆向移植	<a href="#">HIVE-19985</a> :ACID: 跳过解码只读查询的 ROW__ID 部分
逆向移植	<a href="#">HIVE-20635</a> : VectorizedOrcAcidRowBatchReader不筛选原始文件的删除事件
Upgrade	将 Javadoc 升级到 3.3.1
Upgrade	将 Javassist 升级到 3.24.1-GA
Upgrade	更新 apache-directory-server 到 2.0.0-M14

## 新配置

名称	分类	描述
hive.metastore.fs.drop.partition.threads	hive-site	删除分区线程池中的核心线程数。
hive.metastore.fs.drop.partition.keepalive.time	hive-site	空闲的丢弃分区异步线程（来自线程池）在终止之前等待新任务到达的时间（以秒为单位）。

名称	分类	描述
hive.metastore.fs.drop.partition.threadpool.max.queue.size	hive-site	线程池中用于从文件系统中删除分区的最大队列大小。
hive.groupby.enable.deterministic.distribution	hive-site	启用向 Reducer 的密钥分配确定性。它将在调用用于随机分区的 rand 函数时传递一个恒定的种子值。
hive.privilege.synchronizer	hive-site	是否在 HiveServer 2 中定期同步来自外部授权者的权限。
hive.cli.query.file.encoding	hive-site	cli 参数中提供的所有类型的查询文件（查询文件、init 查询文件、rc 文件等）的文件编码。
hive.emr.tez.shuffle.enabled	hive-site	Hive on Tez 作业现在默认使用 tez_shuffle 而不是 mapreduce_shuffle 作为默认 Shuffle 处理程序。

## 已弃用的配置

由于 [HIVE-23354](#)，以下配置属性已被弃用，亚马逊EMR版本 6.11.0 及更高版本不再支持这些属性。

名称	默认值
hive.mapred.reduce.tasks.speculative.execution	false
tez.am.speculation.enabled	false



## 亚马逊 EMR 6.10.0-Hive 发行说明

### 亚马逊 EMR 6.10.0-Hive 变更

类型	描述
功能	<a href="#">通过 P IAM assthrough ( Hive /Steps ) 为 Apache Hive 查询 ( 写入 ) 启用 AWS Lake Formation 基于访问控制。CLI API</a>
改进	默认情况下禁用配置 <code>hive.log.explain.output</code> 以减小日志大小
逆向移植	<a href="#">HIVE-26408</a> : 矢量化 : 修复暂存列的取消分配, 不要重复使用子列作为输出 ConstantVectorExpression
逆向移植	<a href="#">HIVE-22269</a> : <a href="#">修复由于-20703导致统计数据丢失而导致的动态分区插入查询中错误的缩减器计数。HIVE</a>
逆向移植	<a href="#">HIVE-22891</a> : 在 CombineHiveRecord 非执行模式下跳过 PartitionDesc 提取 LLAP
逆向移植	<a href="#">HIVE-23804</a> : 在 Hive 元数据仓库架构中为列统计数据特定表添加默认数据库, 使其向后兼容
逆向移植	<a href="#">HIVE-25277</a> : 云对象存储的 Hive 分区删除速度很慢, 价格昂贵 ListFiles
逆向移植	<a href="#">HIVE-19202</a> : 由于进入而CBO失败。NullPointerException HiveAggregate isBucketedInput()
逆向移植	<a href="#">HIVE-19048</a> : 修复 beeline Initscript 错误被忽略的问题
逆向移植	<a href="#">HIVE-21085</a> : 实例化视图注册表启动非外部 tez 会话

类型	描述
逆向移植	<a href="#">HIVE-21675</a> : CREATEVIEW如果视图NOT EXISTS已经存在，IF 会返回错误而不是“确定”。这是 Hive 2 的回归。
逆向移植	<a href="#">HIVE-21646</a> : Tez : TezTasks 防止逃避线程日志上下文
逆向移植	<a href="#">HIVE-22054</a> : 避免使用递归列出来检查目录是否为空
逆向移植	<a href="#">HIVE-16587</a> : 插入带有嵌NPE套空值的复杂类型时
逆向移植	<a href="#">HIVE-22647</a> : <a href="#">默认</a> 启用会话池
逆向移植	<a href="#">HIVE-13288</a> : 中出现了令人困惑的异常消息。DagUtils localizeResource
逆向移植	<a href="#">HIVE-23870</a> : 优化中的多个文本转换。WritableHiveCharObjectInspector getPrimitiveJava对象/ HiveCharWritable
逆向移植	<a href="#">HIVE-21498</a> : 将 Thrift 升级到 0.13.0
逆向移植	<a href="#">HIVE-24378</a> : 在转换十进制之前，不会删除前导空格和尾随空格
逆向移植	<a href="#">HIVE-21341</a> : 合理的默认值 : hive.server2.idle.operation.timeout 和 hive.server2.idle.session.timeout 太高了
逆向移植	<a href="#">HIVE-22465</a> : 在中添加 ssl conf TezConfigurationFactory
逆向移植	<a href="#">HIVE-24710</a> : 优化计数 (*) 的PTF迭代以降低IO 成本 CPU

类型	描述
逆向移植	<a href="#">HIVE-15406</a> : 考虑对新的“trunc”函数进行矢量化
逆向移植	<a href="#">HIVE-21541</a> : <a href="#">修复 -15406</a> 中缺少的 asf 标题 HIVE
逆向移植	<a href="#">HIVE-24808</a> : 缓存已解析的日期
逆向移植	<a href="#">HIVE-24746</a> :PTF: TimestampValueBoundaryScanner 可以在范围计算期间进行优化
逆向移植	<a href="#">HIVE-25059</a> : 复制期间更改事件转换为重命名
逆向移植	<a href="#">HIVE-25142</a> : 在 map join 快速哈希表中重新哈希会导致大密钥损坏
逆向移植	<a href="#">HIVE-23756</a> : 在 package.jdo 文件中添加了更多限制
逆向移植	<a href="#">HIVE-25150</a> : 在十进制转换之前不会删除制表符，类似于空格字符，后者已作为-24378的一部分修复 HIVE
逆向移植	<a href="#">HIVE-25093</a> : date_format () UDF 仅按时区返回输出 UTC
逆向移植	<a href="#">HIVE-25268</a> : 如果本地时区不是 1900 年之前的日期，date_format udf 会返回错误的结果 UTC
逆向移植	<a href="#">HIVE-25338</a> : 如果输入为空，则AIOBE在 conv UDF 中
逆向移植	<a href="#">HIVE-22400</a> : <a href="#">带</a> 时间返回的UDF分钟 NULL
逆向移植	<a href="#">HIVE-25058</a> :PTF: TimestampValueBoundaryScanner 可以在距离计算期间进行优化 pt2-isDistanceGreater

类型	描述
逆向移植	<a href="#">HIVE-25449</a> : 在一些非时区的 tez 任务中运行时 datediff () 给出了错误的输出 UTC
逆向移植	<a href="#">HIVE-23688</a> : 矢量化 : IndexArrayOutOfBoundsException 适用于包含空值的地图类型列
逆向移植	<a href="#">HIVE-22247</a> : FileNotFoundException 当分区的任务输出为空时 HiveHFileOutputFormat 抛出
逆向移植	<a href="#">HIVE-25570</a> : Hive 应发送完整URL路径以获得命令插入覆盖位置的授权
逆向移植	<a href="#">HIVE-22903</a> : 如果分区子句中有常量表达式 , 矢量化的 row_number () 会在一批之后重置行号
逆向移植	<a href="#">HIVE-25549</a> : 表达式在 BY 或 BY 子句中的 PARTITION窗口函数的结果错误 ORDER
逆向移植	<a href="#">HIVE-25579</a> : LOAD覆盖追加而不是覆盖
逆向移植	<a href="#">HIVE-25659</a> : 应根据数据库允许的最大参数拆分带有 IN/ (NOTIN) 的 Metastore 直接 sql 查询 SQL
逆向移植	<a href="#">HIVE-20502</a> : 修复了在运行 skewjoin_mapjoin10.q NPE 时使用列统计数据时的问题。
逆向移植	<a href="#">HIVE-25765</a> : 当文件大小较大时 , skip.header.line.count 属性会跳过每个块的行 FetchOperator
错误	当hive.stats.column.autogather 和hive.groupby.skewindata 都启用时 , 修复NPE了在某些情况下插入的问题

类型	描述
错误	修复NPE未设置mapred.tasktracker.expiry.interval 值的情况

## 亚马逊 EMR 6.9.0-Hive 发行说明

### 亚马逊 EMR 6.9.0-Hive 的变化

类型	描述
Upgrade	将 Jetty 升级到 <a href="#">9.4.48.v20220622</a>
Upgrade	对于 Hadoop 3.3.3 的支持
功能	Amazon EMR Hive 与 Lake Formation 集成，用于交互式GCSCAPI工作负载。
功能	亚马逊 EMR Hive 与 Iceberg 集成。
改进	如果使用 Amazon EMR 安全配置启用 <a href="#">传输SSL中加密</a> ，则在 HiveServer 2 中启用。
改进	默认启用 Hive EMRFS Amazon S3 优化提交器。有关更多信息，请参阅 <a href="#">启用 Hive EMRFS S3 优化提交器</a> 。
改进	<a href="#">添加仅继HiveHBaseTableInputFormatV2 承的映射版本 InputFormat 以修复 SPARK -34210</a> 。将 hive.hbase.inputformat.v2 设置为 true 以使用它。
改进	等待 TezaM 在后台使用 <a href="#">hive.cli.tez.session.async</a> 启动，而不是终止它后立即启动新版本。使用 hive.emr.cli.tez.session.open.timeout 以秒为单位设置此超时。

类型	描述
改进	添加选项 <a href="#">hive.conf.restricted.list.append</a> ，以将逗号分隔的配置附加到现有的受限配置列表 <code>hive.conf.restricted.list</code> 中。
改进	由于未为数据库定义位置而导致 Hive 查询失败时，会出现更清晰的错误消息。
逆向移植	<a href="#">HIVE-24484</a> ：将 Hadoop 升级到 3.3.1，将 Tez 升级到 0.10.2
逆向移植	<a href="#">HIVE-22398</a> ：通过删除YARN队列管理。ShimLoader
逆向移植	<a href="#">HIVE-23190</a> :LLAP: 修改 IndexCache 以将文件系统对象传递给。TezSpillRecord
逆向移植	<a href="#">HIVE-22185</a> ：HADOOP-15832 会导致使用集群进行测试时出现问题。MiniYarn
逆向移植	<a href="#">HIVE-21670</a> ：替换mockito-all 为依赖关系。mockito-core
逆向移植	<a href="#">HIVE-24542</a> ：为升级做好番石榴准备。
逆向移植	<a href="#">HIVE-23751:QTest: 在 -1658 2 之后重写#mkdirs() 方法ProxyFileSystem 以对齐。HADOOP</a>
逆向移植	<a href="#">HIVE-21603</a> ：Java 11 准备工作：更新powermock 版本。
逆向移植	<a href="#">HIVE-24083</a> ：Hadoop 3.0 中的hcatalog错误：需要身份验证类型。
逆向移植	<a href="#">HIVE-24282</a> ：除非明确提及，否则显示列不应 对输出列进行排序。

类型	描述
逆向移植	<a href="#">HIVE-20656</a> : 合理的默认值 : 地图聚合内存配置过于激进。
逆向移植	<a href="#">HIVE-25443</a> : 当复杂的数据类型超过 1024 个值时, Arrow SerDe 无法序列化/反序列化复杂的数据类型
逆向移植	<a href="#">HIVE-19792</a> : 将兽人升级到 1.5.2 并启用 decimal_64 架构演化测试。
逆向移植	<a href="#">HIVE-20437</a> : 处理从浮点型、双精度型和十进制开始的架构演变。
逆向移植	<a href="#">HIVE-21987</a> : Hive 无法读取用十进制注释的 Parquet int32。
逆向移植	<a href="#">HIVE-20038</a> : 更新对非存储桶表和分区表的查询。NPE

## 亚马逊 EMR 6.9.0-Hive 已知问题

- 在 Amazon EMR 6.0 到 6.9.x 中, 具有动态分区和 BY 或 BY 子句 ORDER 的 INSERT 查询将始终有 SORT 两个缩减器。此问题是由 OSS 更改 [HIVE-20703](#) 引起的, 该更改将动态排序分区优化置于基于成本的决策之下。如果您的工作负载不需要对动态分区进行排序, 建议将 `hive.optimize.sort.dynamic.partition.threshold` 属性设置为 -1 以禁用新功能并获得计算正确的 Reducer 数量。此问题已在 OSS Hive 中作为 [HIVE-22269](#) 的一部分修复, 并已在亚马逊 6.10.0 中修复。EMR

## 亚马逊 EMR 6.8.0-Hive 发行说明

### 亚马逊 EMR 6.8.0-Hive 的变化

类型	描述
改进	减少 msck 命令中的文件系统调用。性能改进 ( 在 10k 以上的分区上提高约 15-20 倍 )
逆向移植	<a href="#">HIVE-20678</a> : iveHBaseTableOutputFormat 应实现 H 以确保兼容性 HiveOutputFormat
逆向移植	<a href="#">HIVE-21040</a> : msck 在目录树的最后一层列出了不必要的文件
逆向移植	<a href="#">HIVE-21460</a> : 加载数据后面加上 select * 查询会导致结果不正确
逆向移植	<a href="#">HIVE-21660</a> : 使用合并所有视图和带有爆炸的后续视图时结果错误
逆向移植	<a href="#">HIVE-22505</a> : 由错误的矢量化 ClassCast Exception 运算符选择引起
逆向移植	<a href="#">HIVE-22513</a> : 筛选操作中投射列的持续传播可能会导致结果不正确
逆向移植	<a href="#">HIVE-23435</a> : 完全外部联接结果缺少行
逆向移植	<a href="#">HIVE-24209</a> : 启用矢量化时, NOTBETWEEN 操作的搜索参数转换不正确
逆向移植	<a href="#">HIVE-24934</a> : G Ch VectorizedExpressions eck 中不需要注释 enericUDFSQCount
逆向移植	<a href="#">HIVE-25278</a> : HiveProjectJoinTransposeRule 可能使用窗口表达式进行无效的转换
逆向移植	<a href="#">HIVE-25505</a> : 标题结果不正确。如果第一行为空, 则跳过.header.line.count



类型	描述
逆向移植	<a href="#">HIVE-26080</a> : 将 accumulo-core 升级到 1.10.1
逆向移植	<a href="#">HIVE-26235</a> : OR 二进制列上的条件返回空结果
错误	修复启动期间 stderr 中的多个SLF4J绑定警告日志
错误	修复分区和表位于不同的文件系统上时，SHOWTABLEEXTENDED查询失败并出现 FS 错误错误。

## 亚马逊 EMR 6.8.0-Hive 已知问题

- 在 Amazon EMR 6.0 到 6.9.x 中，具有动态分区和 BY 或 BY 子句ORDER的INSERT查询将始终有SORT两个缩减器。此问题是由OSS更改 [HIVE-20703](#) 引起的，该更改将动态排序分区优化置于基于成本的决策之下。如果您的工作负载不需要对动态分区进行排序，建议将 `hive.optimize.sort.dynamic.partition.threshold` 属性设置为 -1 以禁用新功能并获得计算正确的 Reducer 数量。此问题已在 OSS Hive 中作为 [HIVE-22269](#) 的一部分修复，并已在亚马逊 6.10.0 中修复。EMR

## 亚马逊 EMR 6.7.0-Hive 发行说明

### 亚马逊 EMR 6.7.0-Hive 变更

类型	描述
功能	<a href="#">Amazon EMR Hive 与集成。LakeFormation</a>
功能	Hive EMRFS Amazon S3 优化提交者的额外审核日志。Hive config : <code>hive.blobstore.output-committer.logging</code> , 默认值 : <code>false</code>
功能	如果在插入覆盖时删除目标目录，选择结果为空，会导致未分区的表/静态分区的行为与 Hive 2.x 类似。Hive config : <code>hive.emr.</code>

类型	描述
	<code>iow.clean.target.dir</code> , 默认值 : <code>false</code>
错误	修复了使用 Hive EMRFS Amazon S3 优化提交程序进行分区存储桶排序时出现的间歇性查询失败问题。
Upgrade	已将 Hive 升级到版本 3.1.3。请参阅 <a href="#">Apache Hive 3.1.3 发布说明</a> 以了解更多详细信息。
Upgrade	已将 Parquet 升级到 <a href="#">1.12.2</a> 。
逆向移植	<a href="#">HIVE-20065</a> : Metastore 不应该依赖 jackson 1.x
逆向移植	<a href="#">HIVE-20071</a> : 迁移到 jackson 2.x 并禁止使用
逆向移植	<a href="#">HIVE-20607</a> : TxnHandler 应使用 PreparedStatement 来执行直接查询 SQL
逆向移植	<a href="#">HIVE-20740</a> : 移除全局锁定。 ObjectStore setConf 方法
逆向移植	<a href="#">HIVE-20961</a> : 停用实施 NVL
逆向移植	<a href="#">HIVE-22059</a> : hive-exec jar 不包含 (fasterxml) jackson 库
逆向移植	<a href="#">HIVE-22351</a> : 修复中不正确的线程用法 ObjectStore TestObjectStore
逆向移植	<a href="#">HIVE-23534</a> : NPE在 RetryingMetaStoreClient #invoke 中捕捉时没有留言 MetaException
逆向移植	<a href="#">HIVE-24048</a> : 将 Jackson 组件协调到版本 2.10. 最新版本-Hive

类型	描述
逆向移植	<a href="#">HIVE-24768</a> : 到处使用 jackson-bom 进行版本替换
逆向移植	<a href="#">HIVE-24816</a> : 由于 -2020-25649, 将 jackson 升级到 2.10.5.1 或 2.11.0+ CVE
逆向移植	<a href="#">HIVE-25971</a> : 由于缓存的线程池未关闭, Tez 任务关闭延迟
逆向移植	<a href="#">HIVE-26036</a> : 由 getMTable () NPE 引起 ObjectStore

## 亚马逊 EMR 6.7.0-Hive 已知问题

- 与 join 位于同一列上的窗口函数的查询可能会导致无效的转换 ( 如 [HIVE-25278](#) 中所述 ) , 并导致错误的结果或查询失败。解决方法是在查询CBO级别禁用此类查询。该修复程序将在 6.7.0 之后的 Amazon EMR 版本中推出。有关更多信息, 请联系 AWS 支持人员。
- 在 Amazon EMR 6.0 到 6.9.x 中, 具有动态分区和 BY 或 BY 子句ORDER的INSERT查询将始终有SORT两个缩减器。此问题是由OSS更改 [HIVE-20703](#) 引起的, 该更改将动态排序分区优化置于基于成本的决策之下。如果您的工作负载不需要对动态分区进行排序, 建议将 `hive.optimize.sort.dynamic.partition.threshold` 属性设置为 -1 以禁用新功能并获得计算正确的 Reducer 数量。此问题已在 OSS Hive 中作为 [HIVE-22269](#) 的一部分修复, 并已在亚马逊 6.10.0 中修复。EMR

## 亚马逊 EMR 6.6.0-Hive 发行说明

### 亚马逊 EMR 6.6.0-Hive 的变化

类型	描述
Upgrade	将 Parquet 升级到 <a href="#">1.12.1</a> 。
Upgrade	将 jetty jars 版本升级到 9.4.43.v20210629

类型	描述
错误	修复了在 Hive 集群上启用 Hive 时LLAP导致在所有任务/核心节点上安装 Hive 的问题。
逆向移植	<a href="#">HIVE-25942</a> : 由于 <a href="#">-2021-294</a> 25 , 将 commons-io 升级到 2.8.0 CVE
逆向移植	<a href="#">HIVE-25726</a> : 由于-2020-13936 , 将速度升级到 2.3 CVE
逆向移植	<a href="#">HIVE-25680</a> : 授权 #get_table_meta HiveMetastore Server API 使用任何授权模型。 HiveMetastore
逆向移植	<a href="#">HIVE-25554</a> : 将箭头版本升级到 0.15
逆向移植	<a href="#">HIVE-25242</a> : 使用矢量化.adaptor = 已选择时 , 查询执行速度非常慢
逆向移植	<a href="#">HIVE-25085</a> : MetaStore 客户端不再跨会话共享。
逆向移植	<a href="#">HIVE-24827</a> : 对于非文本文件 , Hive 聚合查询返回的结果不正确。
逆向移植	<a href="#">HIVE-24683</a> : Hadoop getFileId 23Shims 容易出现不存在的路径 NPE
逆向移植	<a href="#">HIVE-24656</a> : 地图和数组类型为空的查询CBO失败
逆向移植	<a href="#">HIVE-24556</a> : 针对没有孙子的案例 DefaultGraphWalker 进行优化
逆向移植	<a href="#">HIVE-24408</a> : 将 Parquet 升级到 1.11.1
逆向移植	<a href="#">HIVE-24391</a> : <a href="#">修复FIX TestOrcFile 分支 3.1 中的故障</a>

类型	描述
逆向移植	<a href="#">HIVE-24362</a> : 对于具有大量节点的AST树，树处理效果不佳
逆向移植	<a href="#">HIVE-24316</a> : 在 branch-3.1 ORC 中从 1.5.6 升级到 1.5.8
逆向移植	<a href="#">HIVE-24307</a> : 带有属性文件和-e 参数的 Beeline 失败了
逆向移植	<a href="#">HIVE-24245</a> : PTF用计数向量化，在分区上区分会产生错误的结果。
逆向移植	<a href="#">HIVE-24224</a> : 修复 Tez 上的 Hive 在压缩文件上跳过页眉/页脚的问题
逆向移植	<a href="#">HIVE-24157</a> : 严格模式在CAST时间戳上失败 ↔ 数字
逆向移植	<a href="#">HIVE-24113</a> : NPE在 G 中 enericUDFTo UnixTimeStamp
逆向移植	<a href="#">HIVE-23987</a> : 将箭头版本升级到 0.11.0
逆向移植	<a href="#">HIVE-23972</a> : 将外部客户端 ID 添加到外部客户端 LLAP
逆向移植	<a href="#">HIVE-23806</a> : 避免在扩展架构时清除所有分区中的列统计状态。这提高了 alter table add columns 语句的运行时间。
逆向移植	<a href="#">HIVE-23779</a> : BasicStatsTask 信息未在 beeline 控制台中打印
逆向移植	<a href="#">HIVE-23306</a> : 如果系统设置了配置，则RESET 命令不起作用。 getProperty
逆向移植	<a href="#">HIVE-23164</a> : 由于非守护程序线程，服务器未正确终止

类型	描述
逆向移植	<a href="#">HIVE-22967</a> : 在 Tez 上支持 Hive.reloadable.aux.jars.path for Hive
逆向移植	<a href="#">HIVE-22934</a> : 错误流的 Hive 服务器交互式日志计数器
逆向移植	<a href="#">HIVE-22901</a> : 变量替换可能导致循环引用失效 OOM
逆向移植	<a href="#">HIVE-22769</a> : 查询结果不正确, 压缩文本文件的拆分生成过程中查询失败
逆向移植	<a href="#">HIVE-22716</a> : Reading to 被 ByteBuffer 破坏了 ParquetFooterInputFromCache
逆向移植	<a href="#">HIVE-22648</a> : 将 Parquet 升级到 1.11.0
逆向移植	<a href="#">HIVE-22640</a> : Decimal64ColumnVector : ClassCastException 当分区类型为十进制时
逆向移植	<a href="#">HIVE-22621</a> : 不稳定的测试用例 : 。 TestLlapSignerImpl testSigning
逆向移植	<a href="#">HIVE-22533</a> : 修复可能的LLAP守护程序 Web 用户界面漏洞
逆向移植	<a href="#">HIVE-22532</a> : PTFPPD可能通过 Rank/ 函数错误地突破极限 DenseRank
逆向移植	<a href="#">HIVE-22514</a> : HiveProtoLoggingHook 可能会消耗大量内存
逆向移植	<a href="#">HIVE-22476</a> : 当 hive.fetch.task.conversion 设置为无时, Hive datediff 函数提供的结果不一致
逆向移植	<a href="#">HIVE-22429</a> : 在 Hive 3 上使用 bucketing _version 1 迁移的聚簇表使用 bucketing_version 2 进行插入

类型	描述
逆向移植	<a href="#">HIVE-22412</a> : StatsUtils 解释时投掷 NPE
逆向移植	<a href="#">HIVE-22360</a> : 当加载的文件中的列数多于表架构中的列数时，最后一列 MultiDelimitSerDe 返回错误的结果
逆向移植	<a href="#">HIVE-22332</a> : 自 -540 以来，Hive 应确保有效的架构演化设置 ORC
逆向移植	<a href="#">HIVE-22331</a> : 不带参数的 unix_timestamp 返回以毫秒而不是秒为单位的时间戳
逆向移植	<a href="#">HIVE-22275</a> : 。 OperationManager queryIdOperation无法正确清理多个 queryIds
逆向移植	<a href="#">HIVE-22273</a> : 删除临时目录时访问检查失败
逆向移植	<a href="#">HIVE-22270</a> : 将 commons-io 升级到 2.6
逆向移植	<a href="#">HIVE-22241</a> : 实现使用日期/时间戳的内部表示 UDF形式和格里高利-朱利安混合日历来解释日期/时间戳
逆向移植	<a href="#">HIVE-22241</a> : 实现使用日期/时间戳的内部表示形式和 G UDF regorian-Julian 混合来解释日期/时间戳
逆向移植	<a href="#">HIVE-22232</a> : NPE当 hive.order.columnalign 设置为 false 时
逆向移植	<a href="#">HIVE-22231</a> : 通过 Knox 进行大尺寸的 Hive 查询失败，管道损坏写入失败
逆向移植	<a href="#">HIVE-22221</a> : Llap 外部客户端-需要减少 # LlapBaseInputFormat getSplits
逆向移植	<a href="#">HIVE-22208</a> : 重写包含带掩码列的表上的联接的查询时，带有保留关键字的列名不会被转义

类型	描述
逆向移植	<a href="#">HIVE-22197</a> : Common Merge 加入会抛出类施法异常。
逆向移植	<a href="#">HIVE-22170</a> : from_unixtime 和 unix_timestamp 应使用用户会话时区
逆向移植	<a href="#">HIVE-22169</a> : Tez : SplitGenerator 正在尝试查找 Tez 不存在的计划文件
逆向移植	<a href="#">HIVE-22168</a> : 从 llap 缓存热路径中删除非常昂贵的日志记录
逆向移植	<a href="#">HIVE-22161</a> :UDF: 在 org.apache.hadoop.hive ql.udf 上 FunctionRegistry 同步。UDFType 班级
逆向移植	<a href="#">HIVE-22120</a> : 修复特定边界条件下左外地图连接中的错误结果/ ArrayOutOfBound 异常
逆向移植	<a href="#">HIVE-22115</a> : 如果属性设置为 false , 则阻止创建查询路由附加器
逆向移植	<a href="#">HIVE-22113</a> : 防止相关设备LLAP关机 AMReporter RuntimeException
逆向移植	<a href="#">HIVE-22106</a> : 移除分区评估的跨查询同步
逆向移植	<a href="#">HIVE-22099</a> : 自 20007 年以来 , 几个与日期相关的日期UDFs无法正确处理儒略日期 HIVE
逆向移植	<a href="#">HIVE-22037</a> : 由于以下HS2原因而关闭时应该会记录 OOM
逆向移植	<a href="#">HIVE-21976</a> : 方解石中的偏移量应为空而不是零 HiveSortLimit
逆向移植	<a href="#">HIVE-21924</a> : 即使页眉/页脚存在也要拆分文本文件



类型	描述
逆向移植	<a href="#">HIVE-21913</a> : GenericUDTFGet Splits 处理用户名的方式应与处理用户名的方式相同 LLAP
逆向移植	<a href="#">HIVE-21905</a> : 泛型在课堂上有所改进 FetchOperator
逆向移植	<a href="#">HIVE-21902</a> : HiveServer2 用户界面 : jetty 响应标头需要 X-Frame-Options
逆向移植	<a href="#">HIVE-21888</a> : 将 hive.parquet.timestamp.skip.conversion 默认设置为 true
逆向移植	<a href="#">HIVE-21868</a> : 矢量化... CAST FORMAT
逆向移植	<a href="#">HIVE-21864</a> : # LlapBaselInputFormat closeAll
逆向移植	<a href="#">HIVE-21863</a> : 改进表达式的向量化器类型转换 WHEN
逆向移植	<a href="#">HIVE-21862</a> : ORCpdp 生成带有时间戳的错误结果
逆向移植	<a href="#">HIVE-21846</a> : 在 TezaM 中创建一个定期获取指标的话题 LlapDaemon
逆向移植	<a href="#">HIVE-21837</a> : 当所选列的值完全 MapJoin 为空时会抛出异常
逆向移植	<a href="#">HIVE-21834</a> : 避免不必要的调用以简化过滤条件
逆向移植	<a href="#">HIVE-21832</a> : 用于获取平均队列/服务/响应时间的新指标
逆向移植	<a href="#">HIVE-21827</a> : 多次调用 SemanticAnalyzer不通过方法 getTableObject ByName

类型	描述
逆向移植	<a href="#">HIVE-21822</a> : 通过新方法公开 LlapDaemon 指标 API
逆向移植	<a href="#">HIVE-21818</a> :CBO: 复制 TableRelOptHiveTable 有元数据仓流量
逆向移植	<a href="#">HIVE-21815</a> : ORC文件中的统计数据被解析两次
逆向移植	<a href="#">HIVE-21805</a> : HiveServer2 : 使用快速 ShutdownHookManager APIs
逆向移植	<a href="#">HIVE-21799</a> : NullPointerException 在 DynamicPartitionPruningOptimization , 当联接键位于聚合列上时
逆向移植	<a href="#">HIVE-21794</a> : 将物化视图参数添加到 sqlStdAuth SafeVarNameRegexes
逆向移植	<a href="#">HIVE-21768</a> :JDBC: 删除未封闭查询的默认联合前缀 UNION
逆向移植	<a href="#">HIVE-21746</a> : ArrayIndexOutOfBoundsException 在动态分区哈希连接期间 , 禁用 CBO
逆向移植	<a href="#">HIVE-21717</a> : 重命名移动任务中的目录失败。
逆向移植	<a href="#">HIVE-21685</a> : 使用多个 IN 子句的查询简化错误
逆向移植	<a href="#">HIVE-21681</a> : 描述格式化显示多个主键的错误信息
逆向移植	<a href="#">HIVE-21651</a> : 将 protobuf serde 移至 hive-exec 。
逆向移植	<a href="#">HIVE-21619</a> : 在 explain extended 中打印不精确的时间戳类型 SQL

类型	描述
逆向移植	<a href="#">HIVE-21592</a> : 当 OptimizedSql 表达式包含时不显示 CONCAT
逆向移植	<a href="#">HIVE-21576</a> : 介绍... CAST FORMAT以及有限的:2016 日期SQL时间格式列表
逆向移植	<a href="#">HIVE-21573</a> : 如果身份验证设置为 , 二进制传输将忽略主体 delegationToken
逆向移植	<a href="#">HIVE-21550</a> : TestObjectStore 测试不稳定-无法在请求的时间内获得锁
逆向移植	<a href="#">HIVE-21544</a> : 在折叠过程中, 恒定传播会破坏合并/大小写/当表达式时
逆向移植	<a href="#">HIVE-21539</a> : GroupBy + where 同一列上的子句会导致查询重写不正确
逆向移植	<a href="#">HIVE-21538</a> : Beeline : 尽管控制台读取器没有传递给连接参数, 但密码来源
逆向移植	<a href="#">HIVE-21509</a> : LLAP可能会缓存损坏的列向量并返回错误的查询结果
逆向移植	<a href="#">HIVE-21499</a> : 如果创建命令失败, 则不应从注册表中删除该函数 AlreadyExistsException
逆向移植	<a href="#">HIVE-21496</a> : 自动调整无序缓冲区大小可能会溢出
逆向移植	<a href="#">HIVE-21468</a> : <a href="#">存储处理</a> 程序的标识符名称区分大小写 JDBC
逆向移植	<a href="#">HIVE-21467</a> : 移除已弃用的 junit.framework.assert 导入

类型	描述
逆向移植	<a href="#">HIVE-21435</a> : 在构建时 LlapBaseInputFormat 应从 TASK _ ATTEMPT _ ID conf 获取任务编号 ( 如果存在 ) SubmitWorkRequestProto
逆向移植	<a href="#">HIVE-21389</a> : Hive 分发给 <a href="#">-21247</a> 之后错过 javax.ws.rs-api.jar HIVE
逆向移植	<a href="#">HIVE-21385</a> : 允许禁用向下推不可拆分的计算到源 JDBC
逆向移植	<a href="#">HIVE-21383</a> : JDBC存储处理程序 : 如果指定 , 则使用目录和架构检索表
逆向移植	<a href="#">HIVE-21382</a> : 按键减少分组优化——query23 中不会减少密钥
逆向移植	<a href="#">HIVE-21362</a> : 添加输入格式和 serde 以从 protobuf 文件中读取。
逆向移植	<a href="#">HIVE-21340</a> :CBO: 修剪输入到 a 中的非键列 SemiJoin
逆向移植	<a href="#">HIVE-21332</a> : 清除未锁定的缓冲区而不是锁定的缓冲区
逆向移植	<a href="#">HIVE-21329</a> : 自定义 Tez 运行时无序输出缓冲区大小取决于操作员管道
逆向移植	<a href="#">HIVE-21295</a> : StorageHandler 应使用 Hive 惯例将日期转换为字符串
逆向移植	<a href="#">HIVE-21294</a> : 矢量化 : 1-reducer Shuffle 可以跳过对象哈希函数
逆向移植	<a href="#">HIVE-21255</a> : 移入 QueryConditionBuilder JdbcStorageHandler

类型	描述
逆向移植	<a href="#">HIVE-21253</a> : Support in DB2 JDBC StorageHandler
逆向移植	<a href="#">HIVE-21232</a> :LLAP: 添加缓存失误友好拆分亲和力和提供者
逆向移植	<a href="#">HIVE-21214</a> : MoveTask : 使用文件重复数据 attemptId 删除代替文件大小 compareTempOr DuplicateFiles
逆向移植	<a href="#">HIVE-21184</a> : 添加带有成本信息的解释和解释格式化CBO计划
逆向移植	<a href="#">HIVE-21182</a> : 在计划期间跳过设置 Hive 暂存目录的步骤
逆向移植	<a href="#">HIVE-21171</a> : 如果已启用, 则跳过为 tez 创建临时目录的操作 RPC
逆向移植	<a href="#">HIVE-21126</a> : 允许在 # 中进行会话级别查询 LlapBaseInputFormat getSplit
逆向移植	<a href="#">HIVE-21177</a> : 动态分区哈希联接期间出现“找不到字段”错误
逆向移植	<a href="#">HIVE-21061</a> : CTAS查询失败, 源为空 IllegalStateException
逆向移植	<a href="#">HIVE-21041</a> :NPE , ParseException 在从逻辑计划中获取架构时
逆向移植	<a href="#">HIVE-21013</a> : JdbcStorageHandler 无法在 Oracle 中找到分区列
逆向移植	<a href="#">HIVE-21006</a> : 扩展 SharedWorkOptimizer 以在有重复利用机会时移除半联接

类型	描述
逆向移植	<a href="#">HIVE-20992</a> : 将配置 hive.metastore.dbaccess.ssl.properties 拆分为更有意义的配置
逆向移植	<a href="#">HIVE-20989</a> : JDBC- GetOperationStatus + 日志可以通过睡眠阻止查询进度
逆向移植	<a href="#">HIVE-20988</a> : 多列上有主键的分组查询结果错误
逆向移植	<a href="#">HIVE-20985</a> : 如果选择运算符输入是临时列, 则矢量化可能会将其中一些作为输出重复使用
逆向移植	<a href="#">HIVE-20978</a> : “hive.jdbc.*” 应该添加到 sqlStdAuth SafeVarNameRegexes
逆向移植	<a href="#">HIVE-20953</a> : 如果在创建函数时无法将其添加到元存储中, 则将其从函数注册表中删除。
逆向移植	<a href="#">HIVE-20952</a> : 清理.java VectorizationContext
逆向移植	<a href="#">HIVE-20951:LLAP</a> : <b>始终</b> 将 Xms 设置为 50%
逆向移植	<a href="#">HIVE-20949</a> : 改善物理规划中的PKFK基数估计
逆向移植	<a href="#">HIVE-20944</a> : <b>在查询</b> 编译期间不验证统计信息
逆向移植	<a href="#">HIVE-20940</a> : Bridge 案例, 其中方解石的类型分辨率比 Hive 更严格。
逆向移植	<a href="#">HIVE-20937</a> : Postgres jdbc 查询失败并显示“不得为负数” LIMIT
逆向移植	<a href="#">HIVE-20926</a> : 当布隆过滤器条目较高或没有统计数据时, 半加入减少提示失败
逆向移植	<a href="#">HIVE-20920</a> : 使用SQL约束来改进联接重新排序算法

类型	描述
逆向移植	<a href="#">HIVE-20918</a> : 用于启用/禁用将计算从方解石推送到连接的标志 JDBC
逆向移植	<a href="#">HIVE-20915</a> : <a href="#">向</a> HO 和 MR 提供动态排序分区优化
逆向移植	<a href="#">HIVE-20910</a> : 由于动态分区排序优化，在分区表中插入操作失败
逆向移植	<a href="#">HIVE-20899</a> : Keytab for URI S LLAP YARN ervice 仅限于支持 HDFS
逆向移植	<a href="#">HIVE-20898</a> : 对于时间相关函数，不能将参数转换为不可为空的类型
逆向移植	<a href="#">HIVE-20881</a> : 恒定传播过于简化了预测
逆向移植	<a href="#">HIVE-20880</a> : 更新 hive.stats.filter.in.min.ratio 的默认值
逆向移植	<a href="#">HIVE-20873</a> : 使用 Murmur 哈希来减少哈希冲突 VectorHashKeyWrapperTwoLong
逆向移植	<a href="#">HIVE-20868</a> : 当有子操作进入时 TezDummyOperator，SMB加入会间歇性失败 getFinalOp MapRecordProcessor
逆向移植	<a href="#">HIVE-20853</a> : 曝光。 ShuffleHandler registerDag在 llap 守护程序中 API
逆向移植	<a href="#">HIVE-20850</a> : 如果可能，将有条件的案例从投影推送到尺寸表
逆向移植	<a href="#">HIVE-20842</a> : 修复 HIVE -20660 中引入的逻辑来估计分组依据的统计数据
逆向移植	<a href="#">HIVE-20839</a> : 动态分区哈希联接期间出现“找不到字段”错误

类型	描述
逆向移植	<a href="#">HIVE-20835</a> : 约束和 MV 重写之间的交互可能会在方解石计划器中创建循环
逆向移植	<a href="#">HIVE-20834</a> : 保留缓存查询中引用的 Hive QueryResultCache 条目 SemanticAnalyzer
逆向移植	<a href="#">HIVE-20830</a> : <b>某些</b> 情况下 JdbcStorageHandler 范围查询断言失败
逆向移植	<a href="#">HIVE-20829</a> : JdbcStorageHandler 射程分割投掷 NPE
逆向移植	<a href="#">HIVE-20827</a> : <b>空数组</b> 的结果不一致
逆向移植	<a href="#">HIVE-20826</a> : 增强将左侧的加入 + 分组转换为左半联接的 HiveSemiJoin 规则
逆向移植	<a href="#">HIVE-20821</a> : 将 0 重写SUM为 + 组合 SUM COALESCE
逆向移植	<a href="#">HIVE-20815</a> : JdbcRecordReader.next 不能吃异常
逆向移植	<a href="#">HIVE-20813</a> : udf to_epoch_milli 还需要支持不带时区的时间戳。
逆向移植	<a href="#">HIVE-20804</a> : 进一步改进了带约束条件的按优化分组
逆向移植	<a href="#">HIVE-20792</a> : 插入带区域的时间戳会截断数据
逆向移植	<a href="#">HIVE-20788</a> : <b>创建筛选器</b> 时，延长 SJ 缩减可能会错误地回溯列
逆向移植	<a href="#">HIVE-20778</a> : 如果计划中的所有联接都是通过关联逻辑创建的，则可能不会触发联接重新排序



类型	描述
逆向移植	<a href="#">HIVE-20772</a> : 在中记录每个任务的计数器 CPU LLAP
逆向移植	<a href="#">HIVE-20768</a> : 添加翻滚窗口 UDF
逆向移植	<a href="#">HIVE-20767</a> : 联接运算符之间的多个项目可能会影响使用约束进行联接重新排序
逆向移植	<a href="#">HIVE-20762</a> : NOTIFICATION_LOG 清理间隔被硬编码为 60 秒 , 而且太小了
逆向移植	<a href="#">HIVE-20761</a> : 在 notification_sequence 表上选择更新有重试间隔且重试次数太小
逆向移植	<a href="#">HIVE-20751</a> : 将箭头版本升级到 0.10.0
逆向移植	<a href="#">HIVE-20746</a> : HiveProtoHookLogger 不在一天结束时关闭文件。
逆向移植	<a href="#">HIVE-20744</a> : 使用SQL约束来改进联接重新排序算法
逆向移植	<a href="#">HIVE-20740</a> : 移除全局锁定。 ObjectStore setConf 方法。这款 cherrypick 将适用于 Hive 3.2 和 4.x 的 HIVE -20740 向后移植到 3.1.x
逆向移植	<a href="#">HIVE-20734</a> : Beeline : 当 beeline-site.xml 处于并且 Hive CLI 重定向到 beeline 时 , 它应该使用系统用户名/虚拟密码而不是提示输入用户名/虚拟密码
逆向移植	<a href="#">HIVE-20731</a> : <a href="#">应</a> 授权中的密钥库文件 JdbcStorageHandler
逆向移植	<a href="#">HIVE-20720</a> : <a href="#">向处理</a> 程序添加分区列选项 JDBC

类型	描述
逆向移植	<a href="#">HIVE-20719</a> : 在 hive.optimize.sort.sort.dynamic.partition UPDATE 优化和矢量化开启后 SELECT 语句失败
逆向移植	<a href="#">HIVE-20718</a> : 添加带约束的 perf cli 驱动程序
逆向移植	<a href="#">HIVE-20716</a> : 将 hive.cbo.stats.stats.correlated.multi.key.joins 的默认值设置为 true
逆向移植	<a href="#">HIVE-20712</a> : HivePointLookupOptimizer 应该提取深层案例
逆向移植	<a href="#">HIVE-20710</a> : 如果没有类型, 常量折叠可能不会创建空常量
逆向移植	<a href="#">HIVE-20706</a> : external_jdbc_table2.q 间歇性失败
逆向移植	<a href="#">HIVE-20704</a> : 扩展 HivePreFilteringRule 以支持其他功能
逆向移植	<a href="#">HIVE-20703</a> : 将动态排序分区优化置于基于成本的决策之下
逆向移植	<a href="#">HIVE-20702</a> : <a href="#">考虑选择</a> 映射连接期间数据结构感知估计产生的开销
逆向移植	<a href="#">HIVE-20692</a> : 启用 NOT x IS (NOT) [TRUE] 表达式的折叠 FALSE
逆向移植	<a href="#">HIVE-20691</a> : 修复 org.apache.hadoop.hive.cli。TestMiniLlapCliDriver。testCliDriver[cttl]
逆向移植	<a href="#">HIVE-20682</a> : 如果主线程关闭了共享, sessionHive 则异步查询执行可能会失败

类型	描述
逆向移植	<a href="#">HIVE-20676</a> : HiveServer2 : 未设置 Privilege Synchronizer为守护程序状态
逆向移植	<a href="#">HIVE-20660</a> : 通过将总行数绑定到源表 , 可以改进按统计数据分组的估计
逆向移植	<a href="#">HIVE-20652</a> : JdbcStorageHandler 将两个不同数据源的连接推送到 jdbc 驱动程序
逆向移植	<a href="#">HIVE-20651</a> : <a href="#">应对</a> JdbcStorageHandler 密码进行加密
逆向移植	<a href="#">HIVE-20649</a> : 兽人作家的LLAP感知内存管理器
逆向移植	<a href="#">HIVE-20648</a> :LLAP: 按运算符分组的向量应使用每个执行器的内存
逆向移植	<a href="#">HIVE-20646</a> : 如果分区筛选条件有 IS , 则不会将其推送到元数据仓查询 NOT NULL
逆向移植	<a href="#">HIVE-20644</a> : 避免通过 Hive 运行时异常暴露敏感信息
逆向移植	<a href="#">HIVE-20636</a> : 改进外连接后的空值数量估计
逆向移植	<a href="#">HIVE-20632</a> : 如果在查询的表上创建了物化视图 , 则使用 get_splits 进行查询UDF将失败
逆向移植	<a href="#">HIVE-20627</a> : 并发异步查询间歇性失败并导致内存泄漏 LockException
逆向移植	<a href="#">HIVE-20623</a> : 共享工作 : 扩展地图联接缓存条目的共享 LLAP
逆向移植	<a href="#">HIVE-20619</a> : <a href="#">默认情况下</a> 包含 MultiDelimitSerDe 在 2 中 HiveServer

类型	描述
逆向移植	<a href="#">HIVE-20618</a> : 在联接选择期间， BucketMap Join可以为非存储桶表选择此选项
逆向移植	<a href="#">HIVE-20617</a> : 修复 IN 表达式中常量的类型，使其类型正确
逆向移植	<a href="#">HIVE-20612</a> : 为创建新的联接多键关联标志 CBO
逆向移植	<a href="#">HIVE-20603</a> : 更改表位置文件系统后插入分区时出现“错误的 FS” 错误
逆向移植	<a href="#">HIVE-20601</a> : <code>_</code> 中的 EnvironmentContext ALTER事件为空 PARTITION DbNotificationList ener
逆向移植	<a href="#">HIVE-20583</a> : 仅在 kerberos 身份验证中使用规范主机名 HiveConnection
逆向移植	<a href="#">HIVE-20582</a> : 将 hive 原型日志中的 hflush 设置为可配置
逆向移植	<a href="#">HIVE-20563</a> : 矢量化 : 当THEN/ELSE类型和结果类型不同时，CASEWHEN表达式失败
逆向移植	<a href="#">HIVE-20558</a> : 将 hive.hashtable.key.count.adjustment 的默认值更改为 0.99
逆向移植	<a href="#">HIVE-20552</a> : 更快地从中获取架构 LogicalPlan
逆向移植	<a href="#">HIVE-20550</a> : 切换 W 使用直线ebHCat 提交 Hive 查询
逆向移植	<a href="#">HIVE-20537</a> : 多列联接估计值，其中不相关的列与 Hive 不同 CBO

类型	描述
逆向移植	<a href="#">HIVE-20524</a> : 从 Hive 版本 2 到版本 3 再到 Schema Evolution 检查已中断 ALTER TABLE VARCHAR DECIMAL
逆向移植	<a href="#">HIVE-20522</a> : 由于字段可为空， HiveFilterSetOpTransposeRule 可能会引发断言错误
逆向移植	<a href="#">HIVE-20521</a> : HS2 doAs =true 在 hadoop.tmp.dir、MR 和 S3A 文件系统中存在权限问题
逆向移植	<a href="#">HIVE-20515</a> : 使用结果缓存和查询临时目录时查询结果为空，结果缓存目录位于不同的文件系统中
逆向移植	<a href="#">HIVE-20508</a> : Hive 不支持 “user @realm” 类型的用户名
逆向移植	<a href="#">HIVE-20507</a> : Beeline : 添加一个实用程序命令来从 beeline-site.xml 检索所有 uri
逆向移植	<a href="#">HIVE-20505</a> : 将 <a href="#">org.openjdk.jmh: jmh-core</a> 升级到 1.21
逆向移植	<a href="#">HIVE-20503</a> : 在 <a href="#">mapJoin</a> 选择期间使用可感知数据结构的估计
逆向移植	<a href="#">HIVE-20498</a> : 支持列统计数据自动收集的日期类型
逆向移植	<a href="#">HIVE-20496</a> : <a href="#">矢量化</a> : <a href="#">矢量化</a> PTF IllegalStateException
逆向移植	<a href="#">HIVE-20494</a> : <a href="#">-19440</a> 之后 GenericUDFRestrict InformationSchema 被打破了 HIVE
逆向移植	<a href="#">HIVE-20477</a> : 如果表达式 OptimizedSql 包含，则不显示 INs

类型	描述
逆向移植	<a href="#">HIVE-20467</a> : 在创建/删除资源计划中允许 NOT EXISTS IF /IF EXISTS
逆向移植	<a href="#">HIVE-20462</a> : CREATEVIEW如果视图已经存在, 则 “I NOT EXISTS F” 失败
逆向移植	<a href="#">HIVE-20455</a> : 来自 security.authorization 的日志。PrivilegeSynchronizer.run
逆向移植	<a href="#">HIVE-20439</a> : 在选择加入时使用膨胀的内存限制进行 llap
逆向移植	<a href="#">HIVE-20433</a> : 隐式字符串到时间戳的转换速度很慢
逆向移植	<a href="#">HIVE-20432</a> : 将整数类型重写为 I BETWEEN N 以进行统计估计
逆向移植	<a href="#">HIVE-20423</a> : 设置NULLSLAST为默认的空顺序
逆向移植	<a href="#">HIVE-20418</a> : 对于未选择任何列的查询, LLAPI O 可能无法正确处理禁用行索引的ORC文件
逆向移植	<a href="#">HIVE-20412</a> : 在 NPE HiveMetaHook
逆向移植	<a href="#">HIVE-20406</a> : 嵌套合并给出的结果不正确
逆向移植	<a href="#">HIVE-20399</a> : 使用未完全限CTAS定的自定义表位置对于 MM 表失败
逆向移植	<a href="#">HIVE-20393</a> : 半连接减少 : 用于行为不一致 markSemiJoin DPP
逆向移植	<a href="#">HIVE-20391</a> : 分解聚合函数时 HiveAggregateReduceFunctionsRule可能会推断出错误的返回类型

类型	描述
逆向移植	<a href="#">HIVE-20383</a> : 队列名称无效和 Hive proto 事件挂钩存在同步问题。
逆向移植	<a href="#">HIVE-20367</a> : Vectorization : 支持直播、 、 、 PTF AVG MAX MIN SUM
逆向移植	<a href="#">HIVE-20366</a> : TPC-DS query78 统计数据估计值已关闭，因为过滤器为空
逆向移植	<a href="#">HIVE-20364</a> : 更新 hive.map.aggr.hash.min.reduction 的默认值
逆向移植	<a href="#">HIVE-20352</a> : 矢量化 : Support 分组功能
逆向移植	<a href="#">HIVE-20347</a> : hive.optimize.sort.dynamic.partition 应该适用于分区和 MV CTAS
逆向移植	<a href="#">HIVE-20345</a> : 如果从其他调用中删除表，则删除数据库可能会挂起
逆向移植	<a href="#">HIVE-20343</a> : <a href="#">Hive 3</a> : CTAS不尊重 transactional_properties
逆向移植	<a href="#">HIVE-20340</a> : Druid 需要CASTs从时间戳到将时间戳函数的输出用作 Strin STRING 时显式显示
逆向移植	<a href="#">HIVE-20339</a> : 矢量化 : 解除不必要的限制，导致某些不必要的限制无法矢量化 PTF RANK
逆向移植	<a href="#">HIVE-20337</a> : CachedStore: getPartitionsBy Expr 未正确填充分区列表
逆向移植	<a href="#">HIVE-20336</a> : 物化视图的屏蔽和筛选策略
逆向移植	<a href="#">HIVE-20326</a> : 使用RELY默认值而不是“否”创建约束 RELY

类型	描述
逆向移植	<a href="#">HIVE-20321</a> : 矢量化 : 将 1 列的内存大小缩小到 <1 VectorHashKeyWrapper CacheLine
逆向移植	<a href="#">HIVE-20320</a> : 打开 hive.optimize.remove_sq_count_check 标志
逆向移植	<a href="#">HIVE-20315</a> : 矢量化 : 修复更多 NULL /错误结果问题并避免不必要的魔法/转换
逆向移植	<a href="#">HIVE-20314</a> : 在实例化视图重写中包括分区修剪
逆向移植	<a href="#">HIVE-20312</a> : 允许 arrow 客户端使用自己的箭头客户端 BufferAllocator LlapOutputFormatService
逆向移植	<a href="#">HIVE-20302</a> :LLAP: IO 中的非向量化执行会忽略虚拟列, 包括 __ID ROW
逆向移植	<a href="#">HIVE-20300</a> : VectorFileSinkArrowOperator
逆向移植	<a href="#">HIVE-202999</a> : 签名者单元测试中的LLAP潜在竞赛
逆向移植	<a href="#">HIVE-20296</a> : 改进后 HivePointLookupOptimizerRule 能够从更复杂的上下文中提取内容
逆向移植	<a href="#">HIVE-20294</a> : 矢量化 : 修复 NULL /中的错误结果问题// COALESCE ELT
逆向移植	<a href="#">HIVE-20292</a> : 定义了主约束的 tpcds query93 中连接顺序不正确
逆向移植	<a href="#">HIVE-2020</a> : <a href="#">延迟初始化 ArrowColumnarBatch SerDe</a> , 因此它不会在此期间分配缓冲区 GetSplits



类型	描述
逆向移植	<a href="#">HIVE-20281</a> : SharedWorkOptimizer 失败并显示“操作员缓存内容和实际计划不同”
逆向移植	<a href="#">HIVE-20277</a> : 矢量化 : 不支持返回的大小写表达式 BOOLEAN FILTER
逆向移植	<a href="#">HIVE-20267</a> : 扩展 WebUI 以包含表单以动态配置日志级别
逆向移植	<a href="#">HIVE-20263</a> : 变量中有错字 HiveReduceExpressionsWithStatsRule
逆向移植	<a href="#">HIVE-20260</a> : 当另一列NDV的筛选器更改行数时, 不应缩放一列的行数
逆向移植	<a href="#">HIVE-20252</a> : 减少半连接 : 如果小桌边上游有地图连接, 则可能无法检测到由于半连接分支而导致的周期。
逆向移植	<a href="#">HIVE-20245</a> : 矢量化 : 修复 NULL /IN 中的/错误结果问题 BETWEEN
逆向移植	<a href="#">HIVE-20241</a> : <a href="#">在语句中支持 Support</a> 分区规范 CTAS
逆向移植	<a href="#">HIVE-20240</a> : 半连接减少 : 使用局部变量检查外部表条件
逆向移植	<a href="#">HIVE-20226</a> : 当请求 maxEvents 超过表的最大行数时HMS getNextNotification 将引发异常
逆向移植	<a href="#">HIVE-20225</a> : 支持 Terad SerDe ata 二进制格式
逆向移植	<a href="#">HIVE-20213</a> : 将方解石升级到 1.17.0

类型	描述
逆向移植	<a href="#">HIVE-20212</a> : Hiveserver2 在 http 模式下发出默认指标。General.open_Connections 错误地发出指标
逆向移植	<a href="#">HIVE-20210</a> : Simple Fetch 优化器应该在对非分区列进行筛选并且转换量最小 MapReduce 时导致
逆向移植	<a href="#">HIVE-20209</a> : 首次尝试 repl 转储时 Metastore 连接失败
逆向移植	<a href="#">HIVE-20207</a> : 矢量化 : 修复过滤器/比较中 NULL /错误的结果问题
逆向移植	<a href="#">HIVE-20204</a> : 输入期间的类型转换
逆向移植	<a href="#">HIVE-20203</a> : Arrow 泄露 SerDe 了 DirectByteBuffer
逆向移植	<a href="#">HIVE-20197</a> : 矢量化 : 添加 DECIMAL_64 测试, 添加日期/间隔/时间戳算法, 并通过聚合添加更多算法 GROUP
逆向移植	<a href="#">HIVE-20193</a> : 解释计划中 cboInfo 不存在 json
逆向移植	<a href="#">HIVE-20192</a> : 嵌入HS2式元数据仓正在泄漏对象 JDOPersistenceManager
逆向移植	<a href="#">HIVE-20183</a> : 如果源表包含空桶, 则从存储桶表中插入数据可能会导致数据丢失
逆向移植	<a href="#">HIVE-20177</a> : 矢量化 : 减少 KeyWrapper直播模式下的分配 GroupBy
逆向移植	<a href="#">HIVE-20174</a> : 矢量化 : 修复 BY 聚合函数中的 NULL /错误结果问题 GROUP

类型	描述
逆向移植	<a href="#">HIVE-20172</a> : 尝试连接到 StatsUpdater 远程元 GSS数据仓库时因异常而失败
逆向移植	<a href="#">HIVE-20153</a> : 在 Hive 2+ 中, 计数和求和UDF 消耗更多内存
逆向移植	<a href="#">HIVE-20152</a> : 当 repl 转储失败时重置数据库状态, 因此可以重命名表
逆向移植	<a href="#">HIVE-20149</a> : TestHiveCli 失败/超时
逆向移植	<a href="#">HIVE-20130</a> : 更好地记录信息架构同步器
逆向移植	<a href="#">HIVE-20129</a> : 兽人表恢复到基于位置的架构演变
逆向移植	<a href="#">HIVE-20118</a> : 。 SessionStateUserAuthentication or getGroupNames
逆向移植	<a href="#">HIVE-20116</a> : TezTask 正在使用父记录器
逆向移植	<a href="#">HIVE-20115</a> : 酸性表格不应使用页脚扫描进行分析
逆向移植	<a href="#">HIVE-20103</a> : WM : 仅当至少使用了一个聚合计数器时才使用聚合DAG计数器
逆向移植	<a href="#">HIVE-20101</a> : BloomKFilter : 避免完全使用本地字节 [] 数组
逆向移植	<a href="#">HIVE-21000</a> : OpTraits : 选择 Optraits 应在检测到不匹配时停止
逆向移植	<a href="#">HIVE-20098</a> : 统计信息 : 获取日期列分区统计信息NPE时
逆向移植	<a href="#">HIVE-20095</a> : 修复了将计算推送到 jdbc 外部表的功能

类型	描述
逆向移植	<a href="#">HIVE-20093</a> : LlapOutputFormatService: ArrowBuf 与 Netty 一起使用进行会计
逆向移植	<a href="#">HIVE-20090</a> : 延长半联接简化过滤器的创建时间, 以便能够发现新的机会
逆向移植	<a href="#">HIVE-20088</a> : Beeline 配置位置路径组合不正确
逆向移植	<a href="#">HIVE-20082</a> : HiveDecimal 转换为字符串的转换无法正确格式化十进制
逆向移植	<a href="#">HIVE-20069</a> : 修复半连接优化和半连接优化时的重新优化 DPP
逆向移植	<a href="#">HIVE-20051</a> : 跳过临时表的授权
逆向移植	<a href="#">HIVE-20044</a> : Arrow Serde 应该填充字符值并正确处理空字符串
逆向移植	<a href="#">HIVE-20028</a> : 元存储客户端缓存配置使用不正确
逆向移植	<a href="#">HIVE-20025</a> : 清理由创建的事件文件 HiveProto LoggingHook
逆向移植	<a href="#">HIVE-20020</a> : Hive contrib jar 不应该在库中
逆向移植	<a href="#">HIVE-20013</a> : 在 to_date 函数的日期类型中添加隐式强制转换
逆向移植	<a href="#">HIVE-20011</a> : 在 proto 日志挂钩中退出追加模式
逆向移植	<a href="#">HIVE-20005</a> : acid_table_stats、acid_no_buckets 等——分支上的查询结果发生了变化
逆向移植	<a href="#">HIVE-20004</a> : ConvertDecimal64 使用错误的比例会ToDecimal 导致结果不正确

类型	描述
逆向移植	<a href="#">HIVE-19995</a> : <a href="#">ac id table</a> 的汇总行流量
逆向移植	<a href="#">HIVE-19993</a> : 不可能使用同时显示为列名的表别名
逆向移植	<a href="#">HIVE-19992</a> : 矢量化 : -19951 的后续内容 HIVE--> 添加调用。SchemaEvolution isOnlyImplicitORC仅当数据类型转换不是隐式时，才会进行转换以禁用编码的 LLAP I/O
逆向移植	<a href="#">HIVE-19989</a> : Metastore 使用错误的应用程序名称作为指标 HADOOP2
逆向移植	<a href="#">HIVE-19981</a> : 应将该 HiveStrictManagedMigration 实用程序转换为外部表的托管表设置为在删除表时删除数据
逆向移植	<a href="#">HIVE-19967</a> : SMB加入 : Need Optraits for ala Op PTFOperator GBY
逆向移植	<a href="#">HIVE-19935</a> : Hive WM 会话已终止 : 更新任务计数失败 LLAP
逆向移植	<a href="#">HIVE-19924</a> : 标记 Repl Load 运行的 distcp 作业
逆向移植	<a href="#">HIVE-19891</a> : 使用自定义分区目录插入外部表可能会导致数据丢失
逆向移植	<a href="#">HIVE-19850</a> : Tez 中的动态分区修剪会导致“找不到用于表格扫描的工作”错误
逆向移植	<a href="#">HIVE-19806</a> : 对 qtests 输出进行排序以避免测试结果不稳定
逆向移植	<a href="#">HIVE-19770</a> : CBO支持选择中包含多个相同列的查询

类型	描述
逆向移植	<a href="#">HIVE-19769</a> : 为数据库和表名创建专用对象
逆向移植	<a href="#">HIVE-19765</a> : 将 Parquet 特定测试添加到 BlobstoreCliDriver
逆向移植	<a href="#">HIVE-19759</a> : 片状测试 : # TestRpc testServerPort
逆向移植	<a href="#">HIVE-19711</a> : 重构 Hive 架构工具
逆向移植	<a href="#">HIVE-19701</a> : getDelegationTokenFromMetaStore不需要同步
逆向移植	<a href="#">HIVE-19694</a> : 创建物化视图语句应在运行 MV 的语句之前检查 MV 名称是否存在冲突。SQL
逆向移植	<a href="#">HIVE-19674</a> : 按十进制常量分组向下推到 Druid 表
逆向移植	<a href="#">HIVE-19668</a> : 重复的 org.antlr.runtime 浪费了超过 30% 的堆量。CommonToken's 和重复的字符串
逆向移植	<a href="#">HIVE-19663</a> : 重构 LLAP IO 报告生成
逆向移植	<a href="#">HIVE-19661</a> : 将 Hive 切换UDFs为使用 Re2J 正则表达式引擎
逆向移植	<a href="#">HIVE-19628</a> : <u>可能</u> 在 NPE LLAP testSigning
逆向移植	<a href="#">HIVE-19568</a> : 主动/被动 HS2 HA : 不允许直接连接到被动实例 HS2
逆向移植	<a href="#">HIVE-19564</a> : 矢量化 : 修复算术中的 NULL / 错误结果问题
逆向移植	<a href="#">HIVE-1952</a> : 启用 #druidkafkamini_basic .q TestMiniDruidKafkaCliDriver

类型	描述
逆向移植	<a href="#">HIVE-19432</a> : GetTablesOperation 如果蜂巢的数据库和表太多，则速度太慢
逆向移植	<a href="#">HIVE-19360</a> :CBO: 向对象添加“优化SQL” QueryPlan
逆向移植	<a href="#">HIVE-19326</a> : stats auto gather : 查询期间的聚合不正确 UNION
逆向移植	<a href="#">HIVE-19313</a> : TestJdbcWithDBTokenStoreNoDoAs测试失败了
逆向移植	<a href="#">HIVE-19285</a> : 将日志添加到的子类中 MetadataOperation
逆向移植	<a href="#">HIVE-19235</a> : 更新 Minimr 测试的黄金文件
逆向移植	<a href="#">HIVE-19104</a> : 当通过重试开始测试时，实例应该 MetaStore 是独立的
逆向移植	<a href="#">HIVE-18986</a> : 表重命名将运行 java.lang.StackOverflowError dataNucleus 如果表包含大量列，则输入
逆向移植	<a href="#">HIVE-18920</a> :CBO: 在第一次查询之前初始化 Janino 提供者
逆向移植	<a href="#">HIVE-18873</a> : 静默地跳过 MR 的谓词下推 HiveInputFormat 可能会导致存储处理程序产生错误的结果
逆向移植	<a href="#">HIVE-18871</a> : 由于将 hive.aux.jars.path 设置为 hdfs 而导致的 hive on tez 执行错误://
逆向移植	<a href="#">HIVE-18725</a> : 改进列引用错误时子查询的错误处理

类型	描述
逆向移植	<a href="#">HIVE-18696</a> : 如果出现以下情况, 则可能无法在 HiveMetaStore .add_partitions_core 方法中正确清理分区文件夹
逆向移植	<a href="#">HIVE-18453</a> :ACID: 添加“CREATETRANSACTONALTABLE”语法以统一ACIDORC并支持 Parquet
逆向移植	<a href="#">HIVE-18201</a> : 禁用 XPROD _ for sq_count_EDGE chec
逆向移植	<a href="#">HIVE-18140</a> : 在基本统计数据混合大小写下, 分区表统计数据可能会出错
逆向移植	<a href="#">HIVE-17921</a> : 带结构的聚合会产生错误的结果 LLAP
逆向移植	<a href="#">HIVE-17896</a> : TopNKey : 创建一个独立的可量化 T 运算符 opNKey
逆向移植	<a href="#">HIVE-17840</a> : <b>如果出现</b> 异常 HiveMetaStore 。transactionallListeners notifyEvent 失败
逆向移植	<a href="#">HIVE-17043</a> : 如果以后未引用, 则从按键分组中移除非唯一列
逆向移植	<a href="#">HIVE-17040</a> : 在 FK 关系存在的情况下加入淘汰赛
逆向移植	<a href="#">HIVE-16839</a> : <b>同时</b> 更改同一个分区commitTransaction 时对 openTransaction /的调用不平衡
逆向移植	<a href="#">HIVE-16100</a> : 动态排序分区优化器丢失同级运算符
逆向移植	<a href="#">HIVE-15956</a> : StackOverflowError 当丢弃大量分区时



类型	描述
逆向移植	<a href="#">HIVE-15177</a> : 当 kerberos 身份验证类型设置为且主体包含 _ 时 Hive 身份验证失败 fromSubject HOST
逆向移植	<a href="#">HIVE-14898</a> : HS2不应为空身份验证标头错误记录调用堆栈
逆向移植	<a href="#">HIVE-14493</a> : 对物化视图的分区支持
逆向移植	<a href="#">HIVE-14431</a> : 识别为 COALESCE CASE
逆向移植	<a href="#">HIVE-13457</a> : <a href="#">为监控</a> 信息创建HS2RESTAPI端点
逆向移植	<a href="#">HIVE-12342</a> : 将 hive.optimize.index.filter 的默认值设置为 true
逆向移植	<a href="#">HIVE-10296</a> : 当 Hive 在元存储上运行多联接查询时观察到强制转换异常
逆向移植	<a href="#">HIVE-6980</a> : 使用直接 sql 删除表

## 亚马逊 EMR 6.6.0-Hive 配置更改

- 作为OSS变更 [HIVE-20703](#) 的一部分，用于对动态分区进行排序的属性已hive.optimize.sort.dynamic.partition替换为。hive.optimize.sort.dynamic.partition.threshold

hive.optimize.sort.dynamic.partition.threshold 配置具有以下潜在值：

值	描述
0 (默认值)	在使用ORC文件时，将动态分区排序的优化作为基于成本的决定。INSERT查询中允许的最大写入器数量是根据 ( 执行者/容器内存 ) * ( 兽人占用的内存百分比 ) 除以单个写入器占用的最大内存 ( 条带大小 ) 计算得出的。

值	描述
-1	禁用优化以对动态分区进行完全排序。
1	对动态分区启用全局排序。这样可以使 Reducer 中的每个分区值只打开一个记录写入器，从而减小 Reducer 的内存压力。
2  ( 或更大的整数 )	告知 Hive 使用指定的整数作为最大写入器数的阈值。

### 亚马逊 EMR 6.6.0-Hive 已知问题

- 与 join 位于同一列上的窗口函数的查询可能会导致无效的转换（如 [HIVE-25278](#) 中所述），并导致错误的结果或查询失败。解决方法是，您可以在查询CBO级别禁用此类查询。如需更多信息，请联系 AWS 支持人员。
- 亚马逊 EMR 6.6.0 包含 Hive 软件版本 3.1.2。[Hive 3.1.2 引入了一项功能，如果文本文件包含页眉和页脚，则可将其拆分 \(HIVE-21924\)](#)。Apache Tez App Master 读取您的每个文件以确定数据范围内的偏移点。如果您的查询读取大量小型文本文件，这些行为综合起来可能会影响性能。解决方法是使用 CombineHiveInputFormat 并通过配置以下属性调整最大拆分大小：

```
SET hive.tez.input.format=org.apache.hadoop.hive.q1.io.CombineHiveInputFormat;
SET mapreduce.input.fileinputformat.split.maxsize=16777216;
```

- 在 Amazon EMR 6.0 到 6.9.x 中，具有动态分区和 BY 或 BY 子句ORDER的INSERT查询将始终有SORT两个缩减器。此问题是由OSS更改 [HIVE-20703](#) 引起的，该更改将动态排序分区优化置于基于成本的决策之下。如果您的工作负载不需要对动态分区进行排序，建议将 hive.optimize.sort.dynamic.partition.threshold 属性设置为 -1 以禁用新功能并获得计算正确的 Reducer 数量。此问题已在 OSS Hive 中作为 [HIVE-22269](#) 的一部分修复，并已在亚马逊 6.10.0 中修复。EMR

# Hudi

[Apache Hudi](#) 是一种开源数据管理框架，用于通过提供记录级插入、更新、更新插入和删除功能来简化增量数据处理和数据管道开发工作。更新插入指的是将记录插入到现有数据集中（如果它们不存在）或对数据集进行更新（如果它们存在）的功能。通过高效地管理数据在 Amazon S3 中的布局方式，Hudi 允许近乎实时地摄取和更新数据。Hudi 仔细维护对数据集执行的操作的元数据，以帮助确保操作是原子级且是一致的。

Hudi 集成了 [Apache Spark](#)、[Apache Hive](#) 和 [Presto](#)。在亚马逊EMR发布的6.1.0及更高版本中，Hudi 还与 [Trino \( Presto \)](#) 集成。SQL

在亚马逊EMR发布的 5.28.0 及更高版本中，在EMR安装 Spark、Hive、Presto 或 Flink 时，会默认安装 Hudi 组件。你可以使用 Spark 或 Hudi DeltaStreamer 实用程序来创建或更新 Hudi 数据集。您可以使用 Hive、Spark、Presto 或 Flink 以交互方式查询 Hudi 数据集，或使用增量拉取功能构建数据处理管道。增量拉取是指仅拉取两个操作之间更改的数据的功能。

这些功能使得 Hudi 适用于以下使用案例：

- 处理来自传感器和其它需要特定数据插入和更新事件的物联网 (IoT) 设备的流数据。
- 在用户可能会选择被忘记或修改其对数据使用方式的同意的应用程序中，遵守数据隐私法规。
- 实现[变更数据捕获 \(CDC\) 系统](#)，允许您随着时间的推移将更改应用于数据集。

下表列出了最新版本的亚马逊 EMR 7.x 系列中包含的 Hudi 版本，以及亚马逊与 Hudi 一起EMR安装的组件。

有关此版本中与 Hudi 一起安装的组件版本，请参阅 [7.2.0 版本组件](#) 版本。

emr-7.2.0 的 Hudi 版本信息


Amazon EMR 发布标签	Hudi 版本	随 Hudi 安装的组件
emr-7.2.0	Hudi 0.14.1-amzn-1	Not available.

下表列出了最新版本的亚马逊 EMR 6.x 系列中包含的 Hudi 版本，以及亚马逊与 Hudi 一起EMR安装的组件。

有关此发行版中随 Hudi 安装的组件版本，请参阅[发行版 6.15.0 组件版本](#)。

## emr-6.15.0 的 Hudi 版本信息

Amazon EMR 发布标签	Hudi 版本	随 Hudi 安装的组件
emr-6.15.0	Hudi 0.14.0-amzn-0	Not available.

 Note

亚马逊EMR版本6.8.0附带A [pache Hudi 0.11.1](#)；但是，EMR亚马逊6.8.0集群也与Hudi 0.12.0的开源集群兼容。hudi-spark3.3-bundle\_2.12

下表列出了最新版本的亚马逊 EMR 5.x 系列中包含的 Hudi 版本，以及亚马逊与 Hudi 一起EMR安装的组件。

有关此版本中与 Hudi 一起安装的组件的版本，请参阅[版本 5.36.2 组件版本](#)。

## emr-5.36.2 的 Hudi 版本信息

Amazon EMR 发布标签	Hudi 版本	随 Hudi 安装的组件
emr-5.36.2	Hudi 0.10.1-amzn-1	Not available.

## 主题

- [Hudi 的工作原理](#)
- [在亚马逊上使用 Hudi 的注意事项和限制 EMR](#)
- [创建安装了 Hudi 的集群](#)
- [使用 Hudi 数据集](#)
- [使用 Hudi CLI](#)
- [Hudi 发行版历史记录](#)

## Hudi 的工作原理

在亚马逊上使用 Hudi 时EMR，您可以使用 Spark 数据源API或 Hudi DeltaStreamer 实用程序向数据集写入数据。Hudi 将数据集整理到 *basepath* 下一个分区的目录结构中，类似于传统的 Hive 表。如

何将数据布局为这些目录中的文件的具体细节取决于您选择的数据集类型。您可以选择“写入时复制 (CoW)”或“读取时合并 (MOM)”。

无论数据集类型如何，数据集中的每个分区都由其相对于 basepath 的 partitionpath 唯一标识。在每个分区中，记录分布到多个数据文件中。有关更多信息，请参阅 Apache Hudi 文档中的[文件管理](#)。

Hudi 中的每个操作都有一个相应的提交，由一个单调递增的时间戳标识，称为 Instant。Hudi 将对数据集执行的一系列操作保留为时间轴。Hudi 依靠此时间轴提供读取器和写入器之间的快照隔离，并支持回滚到前一个时间点。有关 Hudi 记录的操作和操作状态的更多信息，请参阅 Apache Hudi 文档中的[Timeline](#)。

## 了解数据集存储类型：写入时复制与读取时合并

创建 Hudi 数据集时，可以指定数据集在写入时复制或读取时合并。

- 写入时复制 (CoW) – 数据以列状格式存储 (Parquet)，并且每次更新都会在写入过程中创建一个新版本的文件。CoW 是默认存储类型。
- 读取时合并 (MOR) – 数据使用列式 (Parquet) 和基于行 (Avro) 的格式的组合进行存储。更新记录到基于行的增量文件中，并根据需要进行压缩以创建新版本的列式文件。

对于 CoW 数据集，每次更新记录时，包含该记录的文件都会使用更新后的值进行重写。对于 MoR 数据集，每次进行更新时，Hudi 仅写入已更改记录对应的行。MoR 更适合写入或更改繁重而读取量较少的工作负载。CoW 更适合更改频率较低但读取量繁重的工作负载。

Hudi 为访问数据提供三个逻辑视图：

- 读取优化视图 – 提供来自 CoW 表的最新提交数据集和来自 MOR 表的最新压缩数据集。
- 增量视图 — 在 CoW 数据集中的两个操作之间提供更改流，以提供下游作业和提取、转换、加载 (ETL) 工作流程。
- 实时视图 – 通过内联合并列式和基于行的文件，从 MOR 表中提供最新提交的数据。

当您查询读取优化的视图时，查询将返回所有压缩数据，但不包括最新的增量提交。查询此数据可提供良好的读取性能，但忽略最新的数据。当您查询实时视图时，Hudi 会在读取时将压缩的数据与增量提交合并。最新的数据可用于查询，但合并的计算开销使查询性能降低。通过查询压缩数据或实时数据的功能，您可以在查询时在性能和灵活性之间进行选择。

有关在存储类型之间权衡的更多信息，请参阅 Apache Hudi 文档中的[存储类型和视图](#)。

在 MoR 的 Hive 元数据仓中创建两个表：一个具有您指定的名称的表（即读取优化视图）和一个附加了 `_rt` 的同名表（即实时视图）。您可以查询这两个表。

## 将 Hudi 数据集注册到您的元数据仓

当您在 Hive 元存储中注册 Hudi 表时，您可以像使用任何其他表一样使用 Hive、Spark SQL 或 Presto 查询 Hudi 表。此外，您可以将 Hive 和 Spark 配置为使用 AWS Glue 数据目录作为元数据库，从而将 Hudi 与 AWS Glue 集成。对于 MoR 表，Hudi 将数据集注册为元数据仓中的两个表：一个具有您指定的名称的表（即读取优化视图）和一个附加了 `_rt` 的同名表（即实时视图）。

当您使用 Spark 创建 Hudi 数据集时，您可以通过将 `HIVE_SYNC_ENABLED_OPT_KEY` 选项设置为 `"true"` 并提供其它必需的属性来向 Hive 元数据仓注册 Hudi 表。有关更多信息，请参阅[使用 Hudi 数据集](#)。此外，您可以使用 `hive_sync_tool` 命令行实用程序将 Hudi 数据集单独注册为元数据仓中的表。

## 在亚马逊上使用 Hudi 的注意事项和限制 EMR

- 记录键字段不能为 Null 或空 – 您指定为记录键字段的字段不能具有 null 或空值。
- 默认情况下，架构会在 upsert 和 insert 时更新 — Hudi 提供了一个接口 `HoodieRecordPayload`，用于确定如何合并输入 DataFrame 和现有 Hudi 数据集以生成新的、更新的数据集。Hudi 提供了该类的默认实现 `OverwriteWithLatestAvroPayload`，它会覆盖现有记录并更新输入中指定的架构。DataFrame 要自定义此逻辑以实现合并和部分更新，您可以使用 `DataSourceWriteOptions.PAYLOAD_CLASS_OPT_KEY` 参数提供 `HoodieRecordPayload` 接口的实现。
- 删除需要架构 – 删除时，必须指定记录键、分区键和预组合键字段。其它列可以成为 null 或空，但需要完整的架构。
- MoR 表限制 – MoR 表不支持保存点。您可以使用读取优化视图或来自 Spark SQL、Presto 或 Hive 的实时视图 (`tableName_rt`) 查询 moR 表。使用读取优化视图仅公开基本文件数据，不会公开基本数据和日志数据的合并视图。
- Hive
  - 要在 Hive 元数据仓中注册表，Hudi 需要 Hive Thrift 服务器在默认端口 10000 上运行。如果使用自定义端口覆盖此端口，请传递 `HIVE_URL_OPT_KEY` 选项，如以下示例所示。

```
.option(DataSourceWriteOptions.HIVE_URL_OPT_KEY, "jdbc:hive2://localhost:override-port-number)
```

- Spark 中的 `timestamp` 数据类型在 Hive 中注册为 `long` 数据类型，而不注册为 Hive 的 `timestamp` 类型。

- Presto

- 在版本低于 0.6.0 的 Hudi 中，Presto 不支持读取 MoR 实时表。
- Presto 仅支持快照查询。
- 要使 Presto 正确解释 Hudi 数据集列，请将 `hive.parquet_use_column_names` 值设置为 `true`。
- 要设置会话的值，请在 Presto shell 中运行以下命令：

```
set session hive.parquet_use_column_names=true
```

- 要在集群级别设置值，请使用 `presto-connector-hive` 配置分类将 `hive.parquet.use_column_names` 设置为 `true`，如以下示例所示。有关更多信息，请参阅[配置应用程序](#)。

```
[
  {
    "Classification": "presto-connector-hive",
    "Properties": {
      "hive.parquet.use-column-names": "true"
    }
  }
]
```

- HBase索引

- 用于构建 Hudi 的 HBase 版本可能与《EMR发布指南》中列出的版本不同。要为 Spark 会话提取正确的依赖项，请运行以下命令。

```
spark-shell \  
--jars /usr/lib/spark/external/lib/spark-avro.jar,/usr/lib/hudi/cli/lib/*.jar \  
--conf "spark.serializer=org.apache.spark.serializer.KryoSerializer" \  
--conf "spark.sql.hive.convertMetastoreParquet=false"
```

## 创建安装了 Hudi 的集群

在亚马逊EMR发布的5.28.0及更高版本中，当EMR安装Spark、Hive或Presto时，亚马逊会默认安装Hudi组件。要在 Amazon 上使用 HudiEMR，请创建一个安装了以下一个或多个应用程序的集群：

- Hadoop
- Hive

- Spark
- Presto
- Flink

您可以使用 AWS Management Console AWS CLI、或 Amazon 创建集群EMRAPI。

## 要使用 Hudi 创建集群，请使用 AWS Management Console

1. 导航到新的 Amazon EMR 控制台，然后从侧面导航栏中选择“切换到旧主机”。有关切换到旧控制台后预期情况的更多信息，请参阅 [Using the old console](#)。
2. 依次选择 Create cluster (创建集群)、Go to advanced options (转到高级选项)。
3. 在 Software Configuration (软件配置) 下，对于 Release (发行版)，选择 emr-5.28.0 或更高版本，然后选择 Hadoop、Hive、Spark、Presto、Tez 以及集群需要的其他应用程序。
4. 根据应用程序的需要配置其它选项，然后选择下一步。
5. 根据需要配置 Hardware (硬件) 和 General cluster settings (常规集群设置) 的选项。
6. 对于安全选项，我们建议您选择一个EC2密钥对，您可以使用该密钥对连接到主节点命令行SSH。这允许你运行本指南中描述的 Spark shell CLI 命令、Hive 命令和 Hudi CLI 命令。
7. 根据需要选择其它安全选项，然后选择 Create cluster (创建集群)。

## 使用 Hudi 数据集

Hudi 支持通过 Spark 在 Hudi 数据集中插入、更新和删除数据。有关更多信息，请参阅 Apache Hudi 文档中的[写入 Hudi 表格](#)。

以下示例演示如何启动交互式 Spark 外壳、使用 Spark 提交或使用亚马逊EMR笔记本在亚马逊EMR上使用 Hudi。您也可以使用 Hudi DeltaStreamer 实用程序或其他工具写入数据集。在本节中，示例演示了在使用默认hadoop用户连接到主节点时使用 SSH Spark shell 处理数据集。

### 使用亚马逊 EMR 6.7 及更高版本启动 Spark 外壳

运行spark-shellspark-submit、或spark-sql使用 Amazon EMR 6.7.0 或更高版本时，请传递以下命令。



**Note**

亚马逊 EMR 6.7.0 使用 [Apache Hudi 0.11.0-amzn-0](#)，与之前的 Hudi 版本相比，它包含了重大改进。有关更多信息，请参阅 [Apache Hudi 0.11.0 Migration Guide](#)（《Apache Hudi 0.11.0 迁移指南》）。此选项卡上的示例反映了这些更改。

在主节点上打开 Spark Shell

1. 使用连接到主节点SSH。有关更多信息，请参阅 [《Amazon EMR 管理指南》SSH中的“使用连接到主节点”](#)。
2. 输入以下命令以启动 Spark shell。要使用 PySpark 外壳，请更换 `spark-shell` 替换为 `pyspark`。

```
spark-shell --jars /usr/lib/hudi/hudi-spark-bundle.jar \  
--conf "spark.serializer=org.apache.spark.serializer.KryoSerializer" \  
--conf \  
"spark.sql.catalog.spark_catalog=org.apache.spark.sql.hudi.catalog.HoodieCatalog" \  
\  
--conf "spark.sql.extensions=org.apache.spark.sql.hudi.HoodieSparkSessionExtension"
```

使用亚马逊 EMR 6.6 及更早版本启动 Spark 外壳

运行 `spark-shell`、`spark-submit`、或 `spark-sql` 使用 Amazon EMR 6.6.x 或更早版本时，请传递以下命令。

**Note**

- 亚马逊 EMR 6.2 和 5.31 及更高版本（Hudi 0.6.x 及更高版本）可以从配置中省略 `spark-avro.jar`
- 亚马逊 EMR 6.5 和 5.35 及更高版本（Hudi 0.9.x 及更高版本）可以从配置中省略 `spark.sql.hive.convertMetastoreParquet=false`。
- 亚马逊 EMR 6.6 和 5.36 及更高版本（Hudi 0.10.x 及更高版本）必须包含 [版本 : 0.10.0 Spark 指南中所述的HoodieSparkSessionExtension配置](#)：

```
--conf
"spark.sql.extensions=org.apache.spark.sql.hudi.HoodieSparkSessionExtension"
\
```

## 在主节点上打开 Spark Shell

1. 使用连接到主节点SSH。有关更多信息，请参阅 [《Amazon EMR 管理指南》SSH中的“使用连接到主节点”](#)。
2. 输入以下命令以启动 Spark shell。要使用 PySpark 外壳，请更换 *spark-shell* 替换为 *pyspark*。

```
spark-shell \
--conf "spark.serializer=org.apache.spark.serializer.KryoSerializer" \
--conf "spark.sql.hive.convertMetastoreParquet=false" \
--jars /usr/lib/hudi/hudi-spark-bundle.jar,/usr/lib/spark/external/lib/spark-avro.jar
```

## 使用 Hudi 搭载亚马逊 EMR 6.7 及更高版本的亚马逊EMR笔记本电脑

要将 Hudi 与 Amazon EMR Notebooks 配合使用，必须先将 Hudi jar 文件从本地文件系统复制到笔记本集群的主节点HDFS上。然后，您可以使用笔记本编辑器将EMR笔记本配置为使用 Hudi。

### 将 Hudi 与 Amazon EMR 笔记本电脑配合使用

1. 为 Amazon EMR 笔记本创建并启动集群。有关更多信息，请参阅 [《亚马逊EMR管理指南》中的为笔记本创建亚马逊EMR集群](#)。
2. 使用连接到群集的主节点，SSH然后将 jar 文件从本地文件系统复制到，HDFS如以下示例所示。在示例中，HDFS为了便于文件管理，我们在中创建了一个目录。如果需要HDFS，您可以选择自己的目的地。

```
hdfs dfs -mkdir -p /apps/hudi/lib
```

```
hdfs dfs -copyFromLocal /usr/lib/hudi/hudi-spark-bundle.jar /apps/hudi/lib/hudi-spark-bundle.jar
```

3. 打开笔记本编辑器，输入以下示例中的代码，然后运行它。

```
%%configure
{ "conf": {
    "spark.jars":"hdfs:///apps/hudi/lib/hudi-spark-bundle.jar",
    "spark.serializer":"org.apache.spark.serializer.KryoSerializer",
    "spark.sql.catalog.spark_catalog":
    "org.apache.spark.sql.hudi.catalog.HoodieCatalog",

    "spark.sql.extensions":"org.apache.spark.sql.hudi.HoodieSparkSessionExtension"
  }}

```

## 使用 Hudi 搭载亚马逊 EMR 6.6 及更早版本的亚马逊EMR笔记本电脑

要将 Hudi 与 Amazon EMR Notebooks 配合使用，必须先将 Hudi jar 文件从本地文件系统复制到笔记本集群的主节点HDFS上。然后，您可以使用笔记本编辑器将EMR笔记本配置为使用 Hudi。

将 Hudi 与 Amazon EMR 笔记本电脑配合使用

1. 为 Amazon EMR 笔记本创建并启动集群。有关更多信息，请参阅 [《亚马逊EMR管理指南》中的为笔记本创建亚马逊EMR集群](#)。
2. 使用连接到群集的主节点，SSH然后将 jar 文件从本地文件系统复制到，HDFS如以下示例所示。在示例中，HDFS为了便于文件管理，我们在中创建了一个目录。如果需要HDFS，您可以选择自己的目的地。

```
hdfs dfs -mkdir -p /apps/hudi/lib
```

```
hdfs dfs -copyFromLocal /usr/lib/hudi/hudi-spark-bundle.jar /apps/hudi/lib/hudi-spark-bundle.jar
```

```
hdfs dfs -copyFromLocal /usr/lib/spark/external/lib/spark-avro.jar /apps/hudi/lib/spark-avro.jar
```

3. 打开笔记本编辑器，输入以下示例中的代码，然后运行它。

```
{ "conf": {
    "spark.jars":"hdfs:///apps/hudi/lib/hudi-spark-bundle.jar,hdfs:///apps/hudi/lib/spark-avro.jar",
    "spark.serializer":"org.apache.spark.serializer.KryoSerializer",
    "spark.sql.hive.convertMetastoreParquet":"false"
  }}

```

```
}}
```

## 初始化 Hudi 的 Spark 会话

使用 Scala 时，您必须在 Spark 会话中导入以下类。这需要在每个 Spark 会话中完成一次。

```
import org.apache.spark.sql.SaveMode
import org.apache.spark.sql.functions._
import org.apache.hudi.DataSourceWriteOptions
import org.apache.hudi.DataSourceReadOptions
import org.apache.hudi.config.HoodieWriteConfig
import org.apache.hudi.hive.MultiPartKeyValueExtractor
import org.apache.hudi.hive.HiveSyncConfig
import org.apache.hudi.sync.common.HoodieSyncConfig
```

## 写入 Hudi 数据集

以下示例说明如何创建 DataFrame 并将其写为 Hudi 数据集。

### Note

要将代码示例粘贴到 Spark shell 中，请在提示符处键入 **:paste**，粘贴示例，然后按 **CTRL + D**。

每次向 Hudi 数据集写入时，都必须指定 `DataSourceWriteOptions`。DataFrame 这些选项中的许多选项在写入操作之间可能是相同的。以下示例使用 `hudiOptions` 变量指定常用选项，随后的示例使用这些选项。

在 Amazon EMR 6.7 及更高版本中使用 Scala 进行编写

### Note

亚马逊 EMR 6.7.0 使用 [Apache Hudi 0.11.0-amzn-0](#)，与之前的 Hudi 版本相比，它包含了重大改进。有关更多信息，请参阅 [Apache Hudi 0.11.0 Migration Guide](#) (《Apache Hudi 0.11.0 迁移指南》)。此选项卡上的示例反映了这些更改。

```
// Create a DataFrame
```

```

val inputDF = Seq(
  ("100", "2015-01-01", "2015-01-01T13:51:39.340396Z"),
  ("101", "2015-01-01", "2015-01-01T12:14:58.597216Z"),
  ("102", "2015-01-01", "2015-01-01T13:51:40.417052Z"),
  ("103", "2015-01-01", "2015-01-01T13:51:40.519832Z"),
  ("104", "2015-01-02", "2015-01-01T12:15:00.512679Z"),
  ("105", "2015-01-02", "2015-01-01T13:51:42.248818Z")
).toDF("id", "creation_date", "last_update_time")

//Specify common DataSourceWriteOptions in the single hudiOptions variable
val hudiOptions = Map[String,String](
  HoodieWriteConfig.TBL_NAME.key -> "tableName",
  DataSourceWriteOptions.TABLE_TYPE.key -> "COPY_ON_WRITE",
  DataSourceWriteOptions.RECORDKEY_FIELD_OPT_KEY -> "id",
  DataSourceWriteOptions.PARTITIONPATH_FIELD_OPT_KEY -> "creation_date",
  DataSourceWriteOptions.PRECOMBINE_FIELD_OPT_KEY -> "last_update_time",
  DataSourceWriteOptions.HIVE_SYNC_ENABLED_OPT_KEY -> "true",
  DataSourceWriteOptions.HIVE_TABLE_OPT_KEY -> "tableName",
  DataSourceWriteOptions.HIVE_PARTITION_FIELDS_OPT_KEY -> "creation_date",
  HoodieSyncConfig.META_SYNC_PARTITION_EXTRACTOR_CLASS.key ->
"org.apache.hudi.hive.MultiPartKeysValueExtractor",
  HoodieSyncConfig.META_SYNC_ENABLED.key -> "true",
  HiveSyncConfig.HIVE_SYNC_MODE.key -> "hms",
  HoodieSyncConfig.META_SYNC_TABLE_NAME.key -> "tableName",
  HoodieSyncConfig.META_SYNC_PARTITION_FIELDS.key -> "creation_date"
)

// Write the DataFrame as a Hudi dataset
(inputDF.write
  .format("hudi")
  .options(hudiOptions)
  .option(DataSourceWriteOptions.OPERATION_OPT_KEY,"insert")
  .mode(SaveMode.Overwrite)
  .save("s3://DOC-EXAMPLE-BUCKET/myhudidataset/"))

```

## 在 Amazon EMR 6.6 及更早版本中使用 Scala 进行编写

```

// Create a DataFrame
val inputDF = Seq(
  ("100", "2015-01-01", "2015-01-01T13:51:39.340396Z"),
  ("101", "2015-01-01", "2015-01-01T12:14:58.597216Z"),
  ("102", "2015-01-01", "2015-01-01T13:51:40.417052Z"),
  ("103", "2015-01-01", "2015-01-01T13:51:40.519832Z"),

```

```

("104", "2015-01-02", "2015-01-01T12:15:00.512679Z"),
("105", "2015-01-02", "2015-01-01T13:51:42.248818Z")
).toDF("id", "creation_date", "last_update_time")

//Specify common DataSourceWriteOptions in the single hudiOptions variable
val hudiOptions = Map[String,String](
  HoodieWriteConfig.TABLE_NAME -> "tableName",
  DataSourceWriteOptions.TABLE_TYPE_OPT_KEY -> "COPY_ON_WRITE",
  DataSourceWriteOptions.RECORDKEY_FIELD_OPT_KEY -> "id",
  DataSourceWriteOptions.PARTITIONPATH_FIELD_OPT_KEY -> "creation_date",
  DataSourceWriteOptions.PRECOMBINE_FIELD_OPT_KEY -> "last_update_time",
  DataSourceWriteOptions.HIVE_SYNC_ENABLED_OPT_KEY -> "true",
  DataSourceWriteOptions.HIVE_TABLE_OPT_KEY -> "tableName",
  DataSourceWriteOptions.HIVE_PARTITION_FIELDS_OPT_KEY -> "creation_date",
  DataSourceWriteOptions.HIVE_PARTITION_EXTRACTOR_CLASS_OPT_KEY ->
  classOf[MultiPartKeysValueExtractor].getName
)

// Write the DataFrame as a Hudi dataset
(inputDF.write
  .format("org.apache.hudi")
  .option(DataSourceWriteOptions.OPERATION_OPT_KEY,
DataSourceWriteOptions.INSERT_OPERATION_OPT_VAL)
  .options(hudiOptions)
  .mode(SaveMode.Overwrite)
  .save("s3://DOC-EXAMPLE-BUCKET/myhudidataset/"))

```

## 使用写作 PySpark

```

# Create a DataFrame
inputDF = spark.createDataFrame(
  [
    ("100", "2015-01-01", "2015-01-01T13:51:39.340396Z"),
    ("101", "2015-01-01", "2015-01-01T12:14:58.597216Z"),
    ("102", "2015-01-01", "2015-01-01T13:51:40.417052Z"),
    ("103", "2015-01-01", "2015-01-01T13:51:40.519832Z"),
    ("104", "2015-01-02", "2015-01-01T12:15:00.512679Z"),
    ("105", "2015-01-02", "2015-01-01T13:51:42.248818Z"),
  ],
  ["id", "creation_date", "last_update_time"]
)

# Specify common DataSourceWriteOptions in the single hudiOptions variable

```

```

hoodieOptions = {
'hoodie.table.name': 'tableName',
'hoodie.datasource.write.recordkey.field': 'id',
'hoodie.datasource.write.partitionpath.field': 'creation_date',
'hoodie.datasource.write.precombine.field': 'last_update_time',
'hoodie.datasource.hive_sync.enable': 'true',
'hoodie.datasource.hive_sync.table': 'tableName',
'hoodie.datasource.hive_sync.partition_fields': 'creation_date',
'hoodie.datasource.hive_sync.partition_extractor_class':
'org.apache.hudi.hive.MultiPartKeysValueExtractor'
}

# Write a DataFrame as a Hudi dataset
inputDF.write \
.format('org.apache.hudi') \
.option('hoodie.datasource.write.operation', 'insert') \
.options(**hoodieOptions) \
.mode('overwrite') \
.save('s3://DOC-EXAMPLE-BUCKET/myhudidataset/')

```

### Note

您可能会在代码示例和通知中看到“hoodie”而不是 Hudi。Hudi 代码库广泛使用旧的“hoodie”拼写。

## DataSourceWriteOptions Hudi 的参考资料

选项	描述
TABLE_NAME	要在其中注册数据集的表名称。
TABLE_TYPE_OPT_KEY	可选。指定数据集是创建为 "COPY_ON_WRITE" 还是 "MERGE_ON_READ"。默认值为 "COPY_ON_WRITE"。
RECORDKEY_FIELD_OPT_KEY	其值将用作 HoodieKey 的 recordKey 组件的记录键字段。实际值将通过调用 <code>.toString()</code> 来获得。可使用点表示法指定嵌套字段，例如 <code>a.b.c</code> 。

选项	描述
PARTITIONPATH_FIELD_OPT_KEY	其值将用作 HoodieKey 的 partition Path 组件的分区路径字段。实际值将通过对该字段值调用 <code>.toString()</code> 来获得。
PRECOMBINE_FIELD_OPT_KEY	在实际写入之前在预合并中使用的字段。如果两个记录具有相同的键值，Hudi 为预合并选择字段值最大的记录（由 <code>Object.compareTo()</code> 确定）。

仅在元数据仓中注册 Hudi 数据集表时才需要以下选项。如果您未将 Hudi 数据集注册为 Hive 元数据仓中的表，则不需要这些选项。

#### DataSourceWriteOptions Hive 的参考资料

选项	描述
HIVE_DATABASE_OPT_KEY	要同步到的 Hive 数据库。默认值为 "default"。
HIVE_PARTITION_EXTRACTOR_CLASS_OPT_KEY	用于将分区字段值提取到 Hive 分区列中的类。
HIVE_PARTITION_FIELDS_OPT_KEY	数据集中用于确定 Hive 分区列的字段。
HIVE_SYNC_ENABLED_OPT_KEY	设置为 "true" 时，将向 Apache Hive 元数据仓注册数据集。默认值为 "false"。
HIVE_TABLE_OPT_KEY	必需。Hive 中要同步到的表的名称。例如，"my_hudi_table_cow"。
HIVE_USER_OPT_KEY	可选。同步时要使用的 Hive 用户名。例如，"hadoop"。
HIVE_PASS_OPT_KEY	可选。由 HIVE_USER_OPT_KEY 指定的用户的 Hive 密码。
HIVE_URL_OPT_KEY	Hive 元数据仓 URL。



## 更新插入数据

以下示例演示如何通过编写 a DataFrame 来更新插入数据。与之前的插入示例不同，OPERATION\_OPT\_KEY 值设置为 UPSERT\_OPERATION\_OPT\_VAL。此外，还指定 .mode(SaveMode.Append) 以指示应追加记录。

在 Amazon EMR 6.7 及更高版本中使用 Scala 进行更新

### Note

亚马逊 EMR 6.7.0 使用 [Apache Hudi 0.11.0-amzn-0](#)，与之前的 Hudi 版本相比，它包含了重大改进。有关更多信息，请参阅 [Apache Hudi 0.11.0 Migration Guide](#) (《Apache Hudi 0.11.0 迁移指南》)。此选项卡上的示例反映了这些更改。

```
// Create a new DataFrame from the first row of inputDF with a different creation_date value
val updateDF = inputDF.limit(1).withColumn("creation_date", lit("new_value"))

(updateDF.write
  .format("hudi")
  .options(hudiOptions)
  .option(DataSourceWriteOptions.OPERATION_OPT_KEY, "upsert")
  .mode(SaveMode.Append)
  .save("s3://DOC-EXAMPLE-BUCKET/myhuidataset/"))
```

在 Amazon EMR 6.6 及更早版本中使用 Scala 进行更新

```
// Create a new DataFrame from the first row of inputDF with a different creation_date value
val updateDF = inputDF.limit(1).withColumn("creation_date", lit("new_value"))

(updateDF.write
  .format("org.apache.hudi")
  .option(DataSourceWriteOptions.OPERATION_OPT_KEY,
    DataSourceWriteOptions.UPSERT_OPERATION_OPT_VAL)
  .options(hudiOptions)
  .mode(SaveMode.Append)
  .save("s3://DOC-EXAMPLE-BUCKET/myhuidataset/"))
```

## 使用 Upsert PySpark

```
from pyspark.sql.functions import lit

# Create a new DataFrame from the first row of inputDF with a different creation_date
value
updateDF = inputDF.limit(1).withColumn('creation_date', lit('new_value'))

updateDF.write \
    .format('org.apache.hudi') \
    .option('hoodie.datasource.write.operation', 'upsert') \
    .options(**hudiOptions) \
    .mode('append') \
    .save('s3://DOC-EXAMPLE-BUCKET/myhudidataset/')
```

## 删除记录

要硬删除记录，您可以更新插入一个空的负载。在这种情况下，PAYLOAD\_CLASS\_OPT\_KEY 选项指定 EmptyHoodieRecordPayload 类。该示例使用 upsert 示例中使用的相同方法来指定相同的记录。

DataFrame updateDF

在 Amazon EMR 6.7 及更高版本中使用 Scala 进行删除

### Note

亚马逊 EMR 6.7.0 使用 [Apache Hudi 0.11.0-amzn-0](#)，与之前的 Hudi 版本相比，它包含了重大改进。有关更多信息，请参阅 [Apache Hudi 0.11.0 Migration Guide](#) (《Apache Hudi 0.11.0 迁移指南》)。此选项卡上的示例反映了这些更改。

```
(updateDF.write
    .format("hudi")
    .options(hudiOptions)
    .option(DataSourceWriteOptions.OPERATION_OPT_KEY, "delete")
    .mode(SaveMode.Append)
    .save("s3://DOC-EXAMPLE-BUCKET/myhudidataset/"))
```

在 Amazon EMR 6.6 及更早版本中使用 Scala 进行删除

```
(updateDF.write
    .format("org.apache.hudi")
```

```

.option(DataSourceWriteOptions.OPERATION_OPT_KEY,
DataSourceWriteOptions.UPSERT_OPERATION_OPT_VAL)
.option(DataSourceWriteOptions.PAYLOAD_CLASS_OPT_KEY,
"org.apache.hudi.common.model.EmptyHoodieRecordPayload")
.mode(SaveMode.Append)
.save("s3://DOC-EXAMPLE-BUCKET/myhuidataset/")

```

## 使用删除 PySpark

```

updateDF.write \
    .format('org.apache.hudi') \
    .option('hoodie.datasource.write.operation', 'upsert') \
    .option('hoodie.datasource.write.payload.class',
'org.apache.hudi.common.model.EmptyHoodieRecordPayload') \
    .options(**hudiOptions) \
    .mode('append') \
    .save('s3://DOC-EXAMPLE-BUCKET/myhuidataset/')

```

您还可以通过以下方式硬删除数据：将 `OPERATION_OPT_KEY` 设置为 `DELETE_OPERATION_OPT_VAL` 来删除您提交的数据集中的所有记录。有关执行软删除的说明，以及有关删除 Hudi 表中存储的数据的详细信息，请参阅 Apache Hudi 文档中的 [Deletes](#)。

## 从 Hudi 数据集读取

要在当前时间点检索数据，Hudi 默认情况下执行快照查询。以下是查询在 [写入 Hudi 数据集](#) 中写入 S3 的数据集的示例。Replace（替换）`s3://DOC-EXAMPLE-BUCKET/myhuidataset` 添加表路径，为每个分区级别添加通配符星号，再加上一个星号。在此示例中，有一个分区级别，因此我们添加了两个通配符。

阅读在 Amazon EMR 6.7 及更高版本中使用 Scala

### Note

亚马逊 EMR 6.7.0 使用 [Apache Hudi 0.11.0-amzn-0](#)，与之前的 Hudi 版本相比，它包含了重大改进。有关更多信息，请参阅 [Apache Hudi 0.11.0 Migration Guide](#)（《Apache Hudi 0.11.0 迁移指南》）。此选项卡上的示例反映了这些更改。

```

val snapshotQueryDF = spark.read
    .format("hudi")
    .load("s3://DOC-EXAMPLE-BUCKET/myhuidataset")

```

```
.show()
```

## 在 Amazon EMR 6.6 及更早版本中使用 Scala 阅读

```
(val snapshotQueryDF = spark.read
  .format("org.apache.hudi")
  .load("s3://DOC-EXAMPLE-BUCKET/myhudidataset" + "/*/*"))

snapshotQueryDF.show()
```

## 使用阅读方式 PySpark

```
snapshotQueryDF = spark.read \
  .format('org.apache.hudi') \
  .load('s3://DOC-EXAMPLE-BUCKET/myhudidataset' + '/*/*')

snapshotQueryDF.show()
```

## 递增查询

您还可以使用 Hudi 执行增量查询，以获取自给定提交时间戳以来已更改的记录流。为此，请将 `QUERY_TYPE_OPT_KEY` 字段设置为 `QUERY_TYPE_INCREMENTAL_OPT_VAL`。然后，为 `BEGIN_INSTANTTIME_OPT_KEY` 添加一个值，以获取自指定时间以来写入的所有记录。递增查询的效率通常是批处理查询的十倍，因为它们只处理更改的记录。

执行增量查询时，请使用根（基）表路径，而不需要用于快照查询的通配符星号。

### Note

Presto 不支持递增查询。

## 使用 Scala 进行增量查询

```
(val incQueryDF = spark.read
  .format("org.apache.hudi")
  .option(DataSourceReadOptions.QUERY_TYPE_OPT_KEY,
    DataSourceReadOptions.QUERY_TYPE_INCREMENTAL_OPT_VAL)
  .option(DataSourceReadOptions.BEGIN_INSTANTTIME_OPT_KEY, <beginInstantTime>)
  .load("s3://DOC-EXAMPLE-BUCKET/myhudidataset" ))
```

```
incQueryDF.show()
```

## 使用增量查询 PySpark

```
readOptions = {  
    'hoodie.datasource.query.type': 'incremental',  
    'hoodie.datasource.read.begin.instanttime': <beginInstantTime>,  
}  
  
incQueryDF = spark.read \  
    .format('org.apache.hudi') \  
    .options(**readOptions) \  
    .load('s3://DOC-EXAMPLE-BUCKET/myhudidataset')  
  
incQueryDF.show()
```

有关从 Hudi 数据集读取的更多信息，请参阅 Apache Hudi 文档中的 [查询 Hudi 表](#)。

## 使用 Hudi CLI

您可以使用 Hudi CLI 来管理 Hudi 数据集，以查看有关提交、文件系统、统计信息等的信息。您还可以使用手动执行压缩、安排压实或取消预定的压缩。CLI 有关更多信息，请参阅 Apache Hudi 文档 CLI 中的 [通过交互](#)。

### 启动 Hudi CLI 并连接到数据集

1. 使用连接到主节点 SSH。有关更多信息，请参阅 [《Amazon EMR 管理指南》SSH 中的“使用连接到主节点”](#)。
2. 在命令行中，键入 `/usr/lib/hudi/cli/bin/hudi-cli.sh`。

命令提示符更改为 `hudi->`。

3. 键入以下代码以连接到数据集。Replace（替换）`s3://DOC-EXAMPLE-BUCKET/myhudidataset` 其中包含您要使用的数据集的路径。我们使用的值与前面示例中建立的值相同。

```
connect --path s3://DOC-EXAMPLE-BUCKET/myhudidataset
```

命令提示符将更改以包括您连接到的数据集，如以下示例所示。

```
hudi:myhudidataset->
```

## Hudi 发行版历史记录

下表列出了 Amazon 的每个发行版本中包含的 Hudi 版本EMR，以及随应用程序一起安装的组件。有关每个发行版本中的组件版本，请参阅 [亚马逊 EMR 7.x 发布版本](#)、[亚马逊 EMR 6.x 发布版本](#) 或 [亚马逊 EMR 5.x 发布版本](#) 中的发行版“组件版本”部分。

Hudi 版本信息。

Amazon EMR 发布标签	Hudi 版本	随 Hudi 安装的组件
emr-7.2.0	0.14.1-amzn-1	Not available.
emr-5.36.2	0.10.1-amzn-1	Not available.
emr-7.1.0	0.14.1-amzn-0	Not available.
emr-7.0.0	0.14.0-amzn-1	Not available.
emr-6.15.0	0.14.0-amzn-0	Not available.
emr-6.14.0	0.13.1-amzn-2	Not available.
emr-6.13.0	0.13.1-amzn-1	Not available.
emr-6.12.0	0.13.1-amzn-0	Not available.
emr-6.11.1	0.13.0-amzn-0	Not available.
emr-6.11.0	0.13.0-amzn-0	Not available.
emr-6.10.1	0.12.2-amzn-0	Not available.
emr-6.10.0	0.12.2-amzn-0	Not available.
emr-6.9.1	0.12.1-amzn-0	Not available.
emr-6.9.0	0.12.1-amzn-0	Not available.

Amazon EMR 发布标签	Hudi 版本	随 Hudi 安装的组件
emr-6.8.1	0.11.1-amzn-0	Not available.
emr-6.8.0	0.11.1-amzn-0	Not available.
emr-6.7.0	0.11.0-amzn-0	Not available.
emr-5.36.1	0.10.1-amzn-1	Not available.
emr-5.36.0	0.10.1-amzn-1	Not available.
emr-6.6.0	0.10.1-amzn-0	Not available.
emr-5.35.0	0.9.0-amzn-2	Not available.
emr-6.5.0	0.9.0-amzn-1	Not available.
emr-6.4.0	0.8.0-amzn-0	Not available.
emr-6.3.1	0.7.0-amzn-0	Not available.
emr-6.3.0	0.7.0-amzn-0	Not available.
emr-6.2.1	0.6.0-amzn-1	Not available.
emr-6.2.0	0.6.0-amzn-1	Not available.
emr-6.1.1	0.5.2-incubating-amzn-2	Not available.
emr-6.1.0	0.5.2-incubating-amzn-2	Not available.
emr-6.0.1	0.5.0-incubating-amzn-1	Not available.
emr-6.0.0	0.5.0-incubating-amzn-1	Not available.
emr-5.34.0	0.9.0-amzn-0	Not available.
emr-5.33.1	0.7.0-amzn-1	Not available.
emr-5.33.0	0.7.0-amzn-1	Not available.

Amazon EMR 发布标签	Hudi 版本	随 Hudi 安装的组件
emr-5.32.1	0.6.0-amzn-0	Not available.
emr-5.32.0	0.6.0-amzn-0	Not available.
emr-5.31.1	0.6.0-amzn-0	Not available.
emr-5.31.0	0.6.0-amzn-0	Not available.
emr-5.30.2	0.5.2-incubating	Not available.
emr-5.30.1	0.5.2-incubating	Not available.
emr-5.30.0	0.5.2-incubating	Not available.
emr-5.29.0	0.5.0-incubating	Not available.
emr-5.28.1	0.5.0-incubating	Not available.
emr-5.28.0	0.5.0-incubating	Not available.



# Hue

Hue ( Hadoop 用户体验 ) 是一个开源、基于 Web 的图形用户界面，适用于亚马逊EMR和 Apache Hadoop。Hue 将多个不同的 Hadoop 生态系统项目组合在一起，形成一个可配置界面。亚马逊EMR 还在亚马逊中添加了专门针对Hue的自定义设置。EMRHue 充当在您的集群上运行的应用程序的前端，使您可以使用可能更加熟悉或对用户更友好的界面与应用程序进行交互。通过 Hue 中的应用程序 (如 Hive 和 Pig 编辑器)，无需登录集群即可使用各应用程序相应的 Shell 交互式运行脚本。在集群启动后，您可以使用 Hue 或类似界面与应用程序进行完全交互。有关 Hue 的更多信息，请参阅 <http://gethue.com>。

当您使用 Amazon EMR 控制台启动集群时，系统会默认安装 Hue。您可以选择不安装 Hue，方法是在启动集群时使用 Amazon EMR 控制台中的高级 **--applications** 选项，或者在中使用 `create-cluster` 时明确指定选项并省略 Hue。AWS CLI

## 主题

- [Hue 版本信息](#)
- [Amazon 上支持和不支持的 Hue 功能 EMR](#)
- [注意事项](#)
- [连接到 Hue Web 用户界面](#)
- [在 Amazon 中将 Hue 与远程数据库配合使用 RDS](#)
- [Hue 的高级配置](#)
- [Hue 发行版历史记录](#)

## Hue 版本信息

### 适用于 7.2.0 的 Hue 版本

下表列出了最新版本的亚马逊 EMR 7.x 系列中包含的 Hue 版本，以及亚马逊随着 Hue EMR 安装的组件。

有关此版本中与 Hue 一起安装的组件版本，请参阅 [7.2.0 版组件版本](#)。

## emr-7.2.0 的 Hue 版本信息

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-7.2.0	Hue 4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

## Hue 的 6.15.0 版本

下表列出了最新版本的亚马逊 EMR 6.x 系列中包含的 Hue 版本，以及亚马逊随着 Hue EMR 安装的组件。

有关此发行版中随 Hue 安装的组件版本，请参阅[发行版 6.15.0 组件版本](#)。

## emr-6.15.0 的 Hue 版本信息

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-6.15.0	Hue 4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

## 适用于 5.36.2 的 Hue 版本

下表列出了最新版本的亚马逊 EMR 5.x 系列中包含的 Hue 版本，以及亚马逊随着 Hue EMR 安装的组件。

有关此版本中与 Hue 一起安装的组件的版本，请参阅 [5.36.2 版组件版本](#)。

### emr-5.36.2 的 Hue 版本信息

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-5.36.2	Hue 4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

## Amazon 上支持和不支持的 Hue 功能 EMR

在亚马逊 EMR 7.0.0 及更高版本中，Hue 需要 Python 3.9 或更高版本。使用传输中加密时，您必须使用 2048 位密码密钥。

- Amazon S3 和 Hadoop File System (HDFS) Browser
  - 有了相应的权限，您就可以在属于您账户的临时存储和 S3 HDFS 存储桶之间浏览和移动数据。
  - 默认情况下，Hue 中的超级用户可以访问允许 Amazon EMR IAM 角色访问的所有文件。新建用户不会自动拥有对 Amazon S3 filebrowser 的访问权限，并且必须为其组启用 `filebrowser.s3_access` 权限。
- Hive — 对数据运行交互式查询。此外，这也是编程或批处理查询原型的一种有用方法。
- Pig — 对数据运行脚本或发出交互式命令。
- Oozie — 创建并监控 Oozie 工作流。
- 元存储管理器 — 可用于查看和操作 Hive 元存储的内容（导入/创建、删除等）。

- 任务浏览器 — 查看您提交的 Hadoop 作业的状态。
- 用户管理-管理 Hue 帐户并将LDAP用户与 Hue 集成。
- AWS 示例 — 有几个“ready-to-run”示例使用 Hue 中的应用程序处理来自各种 AWS 服务的示例数据。登录 Hue 后，您将转到 Hue 应用程序主页，其中预安装了示例。
- 只有亚马逊 5.9.0 及更高EMR版本支持 Livy Server。
- 要使用 Hue Notebook for Spark，您必须在 Hue 中安装 Livy 和 Spark。
- Trino/Presto — Hue 支持连接到 Trino 和 Presto 数据服务器。在 Amazon EMR 7.0.0 及更高版本中，使用和连接器可以简化这种连接trino-python-client。presto-python-client低于 7.0.0 的版本使用通过网关的JDBC连接器。JDBC
- 不支持 Hue 控制面板。
- 不支持 Post SQL gre。

## 注意事项

在亚马逊上使用 [Hue](#) 时，请考虑以下限制EMR。

### 使用大型 Hue 元数据表时的性能

如果 Hue 元数据数据库变得太大，性能可能会降低。要检查表的大小，请先使用连接到 Amazon on Amaz EMR on EC2 集群的主节点，SSH然后运行命令sudo mysql -u root启动 My SQL CLI。要获取表的大小，请运行查询SELECT COUNT(\*) FROM hue.<table\_name>。有关内容，请参阅以下内容 <table\_name> 可以是：

- 桌面文档
- 桌面\_document2
- oozie\_job
- beeswax\_savedQuery
- beeswax\_session
- beeswax\_queryHistory

如果运行该查询返回的计数超过 100000，则应运行以下清理命令来删除旧记录。

```
cd /opt/cloudera/parcels/CDH/lib/hue # Hue home directory ./build/env/bin/hue
desktop_document_cleanup
```

有关清理数据库的更多信息，请参阅 [Hue 文档中的参考页面](#)。

Hue 不会自动清理表格，但是 Amazon 5.12.0 及更高EMR版本提供了一种删除表格中旧文档的方法。创建以下 shell 脚本并将其作为一个步骤在 Amazon EMR 集群中运行，其整数参数表示在元数据数据库中保存文档的最大天数。

```
#!/bin/bash
if grep isMaster /mnt/var/lib/info/instance.json | grep false;
then
    echo "This is not the primary node; do nothing, exiting"
    exit 0
fi
while [ ! -f /usr/lib/hue/desktop/core/src/desktop/management/commands/
desktop_document_cleanup.py ]
do
    sleep 1
done
sudo systemctl stop hue.service
sudo sed -i 's+ LOG.warn+ # LOG.warn+g' /usr/lib/hue/desktop/core/src/desktop/
management/commands/desktop_document_cleanup.py
sudo /usr/lib/hue/build/env/bin/hue desktop_document_cleanup --keep-days $1
sudo systemctl start hue.service
```

## Hue 版本之间的不兼容

如果您在多个启用 Hue 的集群中使用相同的 Hue 元数据数据库，我们建议这些集群运行相同版本的 Hue。不同版本的 Hue 可以有不同的 Hue 元数据数据库架构。对不同的版本使用同一个数据库可能会导致 Hue 安装失败。例如，尝试对安装了 4.10.0 和 4.11.0 的两个集群使用同一个数据库可能会导致尝试登录 4.10.0 集群的用户出现登录错误。

有关 Amazon EMR 发行标签和相应的 Hue 已安装版本的列表，请参阅 [Hue 版本历史记录](#)。

## 连接到 Hue Web 用户界面

连接到 Hue Web 用户界面与连接到集群主节点上托管的任何HTTP接口相同。以下过程将介绍如何访问 Hue 用户界面。有关更多信息，请参阅 Amazon EMR 管理指南中的[查看EMR集群上托管的 Web 界面](#)。

查看 Hue Web 用户界面

1. 按照 Amazon EMR 管理指南中的以下说明[使用动态端口转发设置通往主节点的SSH隧道](#)。

2. 在浏览器中键入以下地址以打开 Hue Web 界面：<http://master public DNS:8888>其中 *master public dns* 例如，是您的集群主节点的公共DNS名称 `ec2-11-22-333-44.compute-1.amazonaws.com`。
3. 在 Hue 登录屏幕上，如果您是首次登录的管理员，请输入用户名和密码以创建您的 Hue 超级用户账户，然后选择创建账户。否则，请键入您的用户名和密码，然后选择 Create account (创建账户)，或输入管理员提供的凭证。

## 在 Amazon 中将 Hue 与远程数据库配合使用 RDS

默认情况下，Hue 用户信息和查询历史记录存储在主节点的本地“我的SQL数据库”中。或者，您可以使用存储在 Amazon S3 中的配置和存储在 Amazon Relational Database Service (Amazon) 中的“我的SQL数据库”来创建一个或多个支持 Hue 的集群。RDS 这样，您就可以在不保持 Amazon EMR 集群运行的情况下保存 Hue 创建的用户信息和查询历史记录。我们建议使用 Amazon S3 服务器端加密来存储配置文件。

首先为 Hue 创建远程数据库。

创建外部“我的SQL数据库”

1. 打开 Amazon RDS 控制台，网址为 <https://console.aws.amazon.com/rds/>。
2. 点击 Launch a DB Instance (启动数据库实例)。
3. 选择“我的”SQL，然后单击“选择”。
4. 保留默认选择“多可用区部署和预配置IOPS存储”，然后单击“下一步”。
5. 保留 Instance Specifications (实例规格) 的默认值，指定设置，然后单击 Next (下一步)。
6. 在 Configure Advanced Settings (配置高级设置) 页面上，选择相应的安全组和数据库名称。您使用的安全组必须至少允许从集群的主节点对端口 3306 进行TCP入口访问。如果此时您尚未创建集群，则可以允许所有主机连接到端口 3306 并在启动集群之后调整安全组。单击 Launch DB Instance (启动数据库实例)。
7. 在RDS控制面板中，选择实例，然后选择您刚刚创建的实例。当您的数据库可用时，记下数据库名称、用户名、密码和RDS实例主机名。您将在创建和配置集群时用到此信息。

要在启动集群时使用 Hue 指定外部“我的SQL数据库” AWS CLI

要在使用启动集群时为 Hue 指定外部 My SQL 数据库 AWS CLI，请使用您在创建RDS实例时记下的信息来配置hue.ini配置配置对象

**Note**

您可以创建使用同一个外部数据库的多个集群，但是每个集群将共享查询历史记录和用户信息。

- 使用 AWS CLI，创建安装了 Hue 的集群，使用您创建的外部数据库，并引用具有指定数据库属性的 Hue 配置分类的配置文件。以下示例创建一个安装了 Hue 的集群，引用了 Amazon S3 中的配置文件 `myConfig.json`，该文件指定数据库配置。

**Note**

为了便于读取，包含 Linux 行继续符 (`\`)。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (`^`)。

```
aws emr create-cluster --release-label emr-7.2.0 --applications Name=Hue Name=Spark
Name=Hive \
--instance-type m5.xlarge --instance-count 3 \
--configurations https://s3.amazonaws.com/mybucket/myfolder/myConfig.json --use-
default-roles
```

下面显示的是 `myConfig.json` 文件的内容示例。Replace ( 替换 ) *dbname*, *username*, *password*，以及 *RDS instance hostname* 使用您之前在 RDS 控制面板中记下的值。

```
[{
  "Classification": "hue-ini",
  "Properties": {},
  "Configurations": [
    {
      "Classification": "desktop",
      "Properties": {},
      "Configurations": [
        {
          "Classification": "database",
          "Properties": {
            "name": "dbname",
            "user": "username",
            "password": "password",
            "host": "RDS instance hostname",
```

```
        "port": "3306",
        "engine": "mysql"
    },
    "Configurations": []
}
]
}
]
}]
```

## 故障排除

在 Amazon 出现RDS故障转移时

由于 Hue 数据库实例无响应或正在进行故障转移，因此用户可能会在运行查询时遇到延迟。以下是有关此问题的一些事实和准则：

- 如果您登录到 Amazon RDS 控制台，则可以搜索故障转移事件。例如，要查看是否正在进行或已发生故障转移，请查找诸如“多可用区实例故障转移已启动”和“多可用区实例故障转移已完成”之类的事件。
- RDS实例完成故障转移大约需要 30 秒。
- 如果您在 Hue 中遇到查询的 longer-than-normal 响应，请尝试重新执行查询。

## Hue 的高级配置

本节包括以下主题。

主题

- [为LDAP用户配置 Hue](#)

### 为LDAP用户配置 Hue

与集成LDAP允许用户使用存储在LDAP目录中的现有凭据登录 Hue。将 Hue 与集成时LDAP，无需在 Hue 中独立管理用户信息。以下信息演示了 Hue 与 Microsoft Active Directory 的集成，但配置选项类似于任何LDAP目录。

LDAP身份验证首先绑定到服务器并建立连接。然后，建立的连接将用于任何后续查询，以搜索LDAP用户信息。除非您的 Active Directory 服务器允许匿名连接，否则需要使用绑定可分辨名称和密码建立



连接。绑定可分辨名称 ( DN ) 由 `bind_dn` 配置设置定义。绑定密码由 `bind_password` 配置设置定义。Hue 有两种绑定LDAP请求的方法：搜索绑定和直接绑定。在 Amazon 上使用 Hue 的首选方法 EMR是搜索绑定。

在 Active Directory 中使用搜索绑定时，Hue 会通过用户名属性 ( 由 `user_name_attr` config 定义 ) 来查找需要从基本可分辨名称 ( 或 DN ) 中检索的属性。当 Hue 用户不知道完整 DN 时，搜索绑定非常有用。

例如，您可能已将 `user_name_attr` config 设置为使用通用名称 ( CN )。在这种情况下，Active Directory 服务器使用登录期间提供的 Hue 用户名在目录树中搜索匹配的通用名称，从基本可分辨名称开始。如果找到 Hue 用户的通用名称，则服务器返回用户的可分辨名称。然后，Hue 构造一个可分辨名称，用于通过执行绑定操作对用户进行身份验证。

#### Note

搜索绑定可搜索所有目录子树中的用户名，从基本可分辨名称开始。在 Hue LDAP 配置中指定的基本可分辨名称应是用户名最接近的父级，否则您的LDAP身份验证性能可能会受到影响。

在 Active Directory 中使用直接绑定时，必须使用精确的 `nt_domain` 或 `ldap_username_pattern` 进行身份验证。当使用直接绑定时，如果 `nt` 域 ( 由 `nt_domain` 配置设置定义 ) 属性已定义，则使用以下形式创建用户可分辨名称模板：`<login username>@nt_domain`。此模板用于搜索以基本可分辨名称开始的所有目录子树。如果未配置 `nt` 域，Hue 会为用户搜索精确的可分辨名称模式 ( 由 `ldap_username_pattern` 配置设置定义 )。在这种情况下，服务器在所有目录子树中搜索匹配的 `ldap_username_pattern` 值，从基本可分辨名称开始。

要启动具有 Hue LDAP 属性的集群，请使用 AWS CLI

- 要为指定LDAP属性hue-ini，请创建一个安装了 Hue 的集群，并引用一个带有配置属性的 json 文件LDAP。下面显示了一个示例命令，此命令引用 Amazon S3 中存储的配置文件 `myConfig.json`。

```
aws emr create-cluster --release-label emr-7.2.0 --applications Name=Hue Name=Spark
Name=Hive \
--instance-type m5.xlarge --instance-count 3 --configurations https://
s3.amazonaws.com/mybucket/myfolder/myConfig.json.
```

下面显示的是 `myConfig.json` 的内容示例。

```
[
  {
    "Classification": "hue-ini",
    "Properties": {},
    "Configurations": [
      {
        "Classification": "desktop",
        "Properties": {},
        "Configurations": [
          {
            "Classification": "ldap",
            "Properties": {},
            "Configurations": [
              {
                "Classification": "ldap_servers",
                "Properties": {},
                "Configurations": [
                  {
                    "Classification": "yourcompany",
                    "Properties": {
                      "base_dn":
"DC=yourcompany,DC=hue,DC=com",
                      "ldap_url": "ldap://ldapurl",
                      "search_bind_authentication": "true",
                      "bind_dn":
"CN=hue,CN=users,DC=yourcompany,DC=hue,DC=com",
                      "bind_password": "password"
                    },
                    "Configurations": []
                  }
                ]
              }
            ]
          }
        ],
        "Properties": {
          "backend": "desktop.auth.backend.LdapBackend"
        }
      }
    ]
  }
]
```

```

    ]
  }
]

```

### Note

在 Amazon 5.21.0 及更高EMR版本中，您可以覆盖集群配置，并为正在运行的集群中的每个实例组指定其他配置分类。为此，您可以使用 Amazon EMR 控制台、AWS Command Line Interface (AWS CLI) 或 AWS SDK。有关更多信息，请参阅[为运行的集群中的实例组提供配置](#)。

### 在 Hue 中查看LDAP设置

1. 确认您有与 Amazon EMR 集群主节点的有效VPN连接或SSH隧道。然后，在浏览器中键入 `master-public-dns:8888` 打开 Hue 网页界面。
2. 使用 Hue 管理员凭证登录。如果 Did you know? (您知道吗?) 窗口打开，单击 Got it, prof! (明白了，教授!) 可关闭该窗口。
3. 在工具栏中，单击 Hue 图标。
4. 在关于 Hue 页面上，选择配置。
5. 在 Configuration Sections and Variables (配置部分和变量) 部分，单击 Desktop (桌面)。
6. 滚动到 Idap 部分以查看您的设置。

## Hue 发行版历史记录

下表列出了 Amazon 的每个发行版本中包含的 Hue 版本EMR，以及随应用程序一起安装的组件。有关每个发行版本中的组件版本，请参阅 [亚马逊 EMR 7.x 发布版本](#)、[亚马逊 EMR 6.x 发布版本](#) 或 [亚马逊 EMR 5.x 发布版本](#) 中的发行版“组件版本”部分。

### Hue 版本信息

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-7.2.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode,

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
		hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-5.36.2	4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-7.1.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-7.0.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.15.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.14.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-6.13.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.12.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.11.1	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-6.11.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.10.1	4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.10.0	4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-6.9.1	4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.9.0	4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.8.1	4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server



Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-6.8.0	4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.7.0	4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-5.36.1	4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-5.36.0	4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.6.0	4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-5.35.0	4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-6.5.0	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.4.0	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.3.1	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-6.3.0	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.2.1	4.8.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.2.0	4.8.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-6.1.1	4.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.1.0	4.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.0.1	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-6.0.0	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-5.34.0	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-5.33.1	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-5.33.0	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-5.32.1	4.8.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-5.32.0	4.8.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-5.31.1	4.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-5.31.0	4.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-5.30.2	4.6.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server



Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-5.30.1	4.6.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-5.30.0	4.6.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-5.29.0	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-5.28.1	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.28.0	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.27.1	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-5.27.0	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.26.0	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.25.0	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-5.24.1	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.24.0	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.23.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-5.23.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.22.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.21.2	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-5.21.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.21.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.20.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-5.20.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.19.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.19.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-5.18.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.18.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.17.2	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server



Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-5.17.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.17.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.16.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-5.16.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.15.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.15.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-5.14.2	4.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.14.1	4.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.14.0	4.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-5.13.1	4.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.13.0	4.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.12.3	4.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-5.12.2	4.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.12.1	4.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.12.0	4.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-5.11.4	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.11.3	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.11.2	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-5.11.1	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.11.0	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.10.1	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-5.10.0	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.9.1	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.9.0	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server



Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-5.8.3	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.8.2	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.8.1	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-5.8.0	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.7.1	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.7.0	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-5.6.1	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.6.0	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.5.4	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-5.5.3	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-5.5.2	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-5.5.1	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-5.5.0	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-5.4.1	3.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-5.4.0	3.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-5.3.2	3.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-5.3.1	3.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-5.3.0	3.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-5.2.3	3.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-5.2.2	3.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-5.2.1	3.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-5.2.0	3.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-5.1.1	3.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-5.1.0	3.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server



Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-5.0.3	3.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server
emr-5.0.2	3.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server
emr-5.0.1	3.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-5.0.0	3.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server
emr-4.9.6	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-4.9.5	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-4.9.4	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-4.9.3	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-4.9.2	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-4.9.1	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-4.8.5	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-4.8.4	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-4.8.3	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-4.8.2	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-4.8.1	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-4.8.0	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-4.7.4	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-4.7.3	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-4.7.2	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-4.7.1	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server
emr-4.7.0	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server

Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-4.6.1	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server
emr-4.6.0	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server
emr-4.5.0	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server
emr-4.4.0	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server



Amazon EMR 发布标签	Hue 版本	随 Hue 安装的组件
emr-4.3.0	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server
emr-4.2.0	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server
emr-4.1.0	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server

# Iceberg

[Apache Iceberg](#) 是 Amazon Simple Storage Service ( Amazon S3 ) 中适用于大型数据集的开放表格式。它为大型表、原子提交、并发写入和SQL兼容表演提供了快速的查询性能。从亚马逊 EMR 6.5.0 开始，你可以在亚马逊EMR集群上使用 Iceberg 表格式的 Apache Spark 3。

下表列出了最新版本的亚马逊 EMR 7.x 系列中包含的 Iceberg 版本，以及亚马逊与 Iceberg 一起EMR 安装的组件。

有关此版本中与 Iceberg 一起安装的组件的版本，请参阅 [7.2.0 版本的组件](#) 版本。

emr-7.2.0 的 Iceberg 版本信息

Amazon EMR 发布标签	Iceberg 版本	随 Iceberg 安装的组件
emr-7.2.0	Iceberg 1.5.0-amzn-0	Not available.

下表列出了最新版本的亚马逊 EMR 6.x 系列中包含的 Iceberg 版本，以及亚马逊与 Iceberg 一起EMR 安装的组件。

有关此发行版中随 Iceberg 安装的组件版本，请参阅[发行版 6.15.0 组件版本](#)。

emr-6.15.0 的 Iceberg 版本信息

Amazon EMR 发布标签	Iceberg 版本	随 Iceberg 安装的组件
emr-6.15.0	Iceberg 1.4.0-amzn-0	Not available.

## 主题

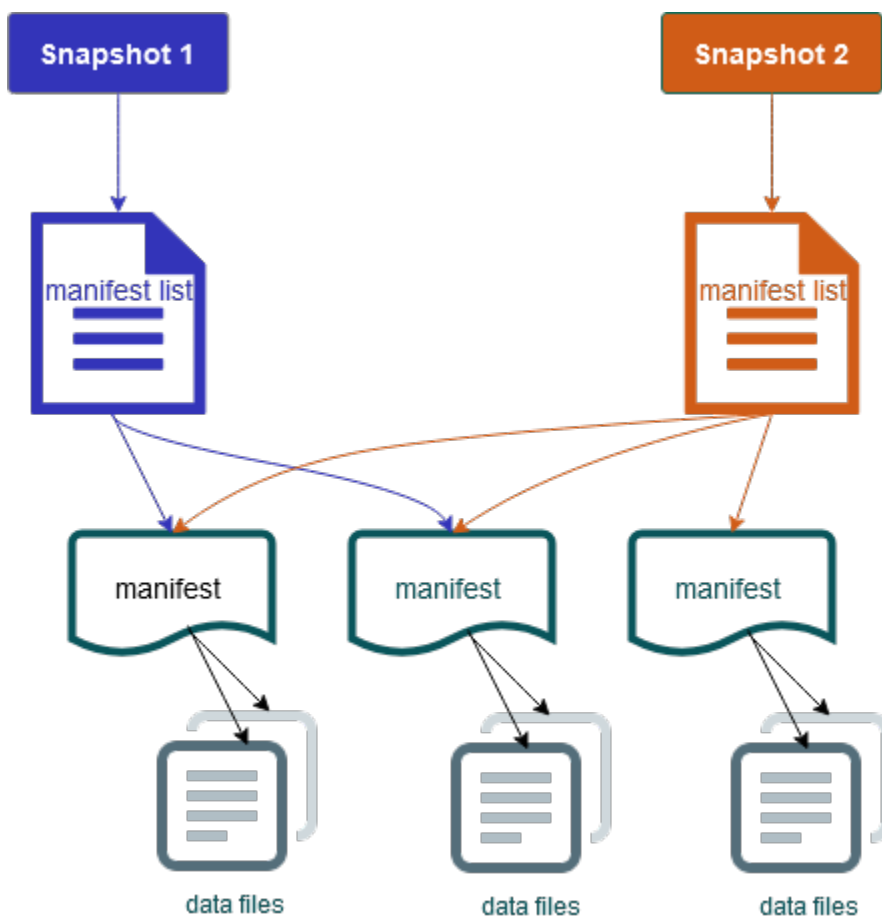
- [Iceberg 的工作原理](#)
- [使用安装有 Iceberg 的集群](#)
- [在亚马逊上使用 Iceberg 的注意事项和限制 EMR](#)
- [Iceberg 发布历史记录](#)

## Iceberg 的工作原理

Iceberg 跟踪表中而非目录中的单个数据文件。这样，写入器便可在原位置创建数据文件（文件不会移动或更改）。此外，写入器只能在显式提交时将文件添加到表中。表状态在元数据文件中维护。对表状态的所有更改都会创建一个新的元数据文件，该文件会原子式替换旧的元数据。表元数据文件跟踪表架构、分区配置和其他属性。

其还包括表内容的快照。每个快照都是表中数据文件在某个时间点的完整集合。快照在元数据文件中列出，但快照中的文件存储在单独的清单文件中。通过在表元数据文件之间的原子转换来实现快照隔离。读取器使用加载表元数据时最新的快照。读取器在刷新并选取新的元数据位置之前不会受到更改的影响。快照中的数据文件存储在一个或多个清单文件中，其中包含表中每个数据文件、分区数据及其指标一行。快照是清单中所有文件的并集。清单文件还可以在快照之间共享，以避免重写不常更改的元数据。

Iceberg 快照图



Iceberg 提供以下功能：

- 支持 Amazon S3 数据湖中的ACID交易和时空旅行。
- 提交重试次数受益于[乐观并发](#)的性能优势。
- 由于解决了文件级冲突问题，因此具有高并发能力。
- 通过元数据中每列的最小最大统计数据，您可以跳过文件，从而提高选择性查询的性能。
- 您可以通过分区发展将表整理为灵活的分区布局，以便能够更新分区架构。然后，查询和数据量可以在不依赖物理目录的情况下进行更改。
- 支持[架构发展](#)和强制执行。
- Iceberg 表充当幂等性数据汇和可重放的源。其能够通过一次精确的管道支持流式处理和批处理。幂等性数据汇会跟踪过去成功的写入操作。因此，数据汇可以在失败时再次请求数据，并可在数据已多次发送时丢弃数据。
- 查看历史记录和谱系，包括表发展、操作历史记录和每次提交的统计数据。
- 从现有数据集迁移，可选择数据格式 ( Parquet、ORC、Avro ) 和分析引擎 ( Spark、Trino、PrestoDB、Flink、Hive ) 。

## 使用安装有 Iceberg 的集群

本节包含将 Iceberg 与 Spark、Trino、Flink 和 Hive 结合使用的信息。

### 将 Iceberg 集群与 Spark 结合使用

从 Amazon EMR 版本 6.5.0 开始，您可以将 Iceberg 与 Spark 集群配合使用，无需包含引导操作。对于 Amazon 6.4.0 及更早EMR版本，您可以使用引导操作来预安装所有必要的依赖项。

在本教程中，您将使用在 Amazon EMR Spark 集群上使用 Iceberg。AWS CLI 要使用控制台创建安装了 Iceberg 的集群，请按照使用 Amazon Athena、Amazon 和 Glue 构建 Apache Iceberg 数据湖中的步骤进行操作。EMR AWS

### 创建 Iceberg 集群

您可以使用 AWS Management Console、AWS CLI 或 Amazon EMR API 创建安装了 Iceberg 的集群。在本教程中，您将使用在 AWS CLI Amazon EMR 集群上使用 Iceberg。要使用控制台创建安装了 Iceberg 的集群，请按照使用 Amazon Athena、Amazon 和 Glue 构建 Apache Iceberg 数据湖中的步骤进行操作。EMR AWS

要将 Amazon 上的 Iceberg EMR 与一起使用 AWS CLI，请先按照以下步骤创建一个集群。有关使用指定 Iceberg 分类的信息 AWS CLI，请参阅[创建集群 AWS CLI 时使用提供配置](#)或[创建集群SDK时使用 Java 提供配置](#)。

1. 创建 `configurations.json` 文件并输入以下内容：

```
[{
  "Classification":"iceberg-defaults",
  "Properties":{"iceberg.enabled":"true"}
}]
```

2. 接下来，使用以下配置创建集群。将实例 Amazon S3 桶路径和子网 ID 替换为您自己的值。

```
aws emr create-cluster --release-label emr-6.5.0 \
--applications Name=Spark \
--configurations file://iceberg_configurations.json \
--region us-east-1 \
--name My_Spark_Iceberg_Cluster \
--log-uri s3://DOC-EXAMPLE-BUCKET/ \
--instance-type m5.xlarge \
--instance-count 2 \
--service-role EMR_DefaultRole_V2 \
--ec2-attributes
InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-1234567890abcdef0
```

或者，您可以创建包含 Spark 应用程序的 Amazon EMR 集群，并将该文件 `/usr/share/aws/iceberg/lib/iceberg-spark3-runtime.jar` 作为 JAR 依赖项包含在 Spark 任务中。有关更多信息，请参阅[提交应用程序](#)。

要将 jar 作为 Spark 作业中的依赖项包含在内，请将以下配置属性添加到 Spark 应用程序中：

```
--conf "spark.jars=/usr/share/aws/iceberg/lib/iceberg-spark3-runtime.jar"
```

有关 Spark 作业依赖项的更多信息，请参阅 Apache Spark 文档 [Running Spark on Kubernetes](#)（在 Kubernetes 上运行 Spark）中的 [Dependency Management](#)（依赖项管理）。

## 为 Iceberg 初始化 Spark 会话

以下示例演示如何启动交互式 Spark 外壳、使用 Spark 提交或使用亚马逊 EMR 笔记本在亚马逊 EMR 上使用 Iceberg。

spark-shell

1. 使用连接到主节点 SSH。有关更多信息，请参阅 [《Amazon EMR 管理指南》SSH 中的“使用连接到主节点”](#)。

2. 输入以下命令以启动 Spark shell。要使用 PySpark 外壳，请spark-shell替换为pyspark。

```
spark-shell \  
--conf  
  "spark.sql.extensions=org.apache.iceberg.spark.extensions.IcebergSparkSessionExtensions"  
 \  
--conf "spark.sql.catalog.dev=org.apache.iceberg.spark.SparkCatalog" \  
--conf "spark.sql.catalog.dev.type=hadoop" \  
--conf "spark.sql.catalog.dev.warehouse=s3://DOC-EXAMPLE-BUCKET/example-prefix/"
```

## spark-submit

1. 使用连接到主节点SSH。有关更多信息，请参阅 [《Amazon EMR 管理指南》SSH中的“使用连接到主节点”](#)。
2. 输入以下命令以为 Iceberg 启动 Spark 会话。

```
spark-submit \  
--conf  
  "spark.sql.extensions=org.apache.iceberg.spark.extensions.IcebergSparkSessionExtensions"  
 \  
--conf "spark.sql.catalog.dev=org.apache.iceberg.spark.SparkCatalog" \  
--conf "spark.sql.catalog.dev.type=hadoop" \  
--conf "spark.sql.catalog.dev.warehouse=s3://DOC-EXAMPLE-BUCKET/example-prefix/"
```

## EMR Studio notebooks

要使用 EMR Studio 笔记本初始化 Spark 会话，请使用亚马逊EMR笔记本中的%%configure神奇命令配置 Spark 会话，如下例所示。有关更多信息，请参阅 [《Amazon EMR 管理指南》中的“使用EMR笔记本魔法”](#)。

```
%%configure -f  
{  
  "conf":{  
  
    "spark.sql.extensions":"org.apache.iceberg.spark.extensions.IcebergSparkSessionExtensions",  
    "spark.sql.catalog.dev":"org.apache.iceberg.spark.SparkCatalog",  
    "spark.sql.catalog.dev.type":"hadoop",  
    "spark.sql.catalog.dev.warehouse":"s3://DOC-EXAMPLE-BUCKET/example-prefix/"  
  }  
}
```

```
}
```

## 写入 Iceberg 表

以下示例说明如何创建 DataFrame 并将其写为 Iceberg 数据集。这些示例演示了以默认 hadoop 用户 SSH 身份连接到主节点时使用 Spark shell 处理数据集。

### Note

要将代码示例粘贴到 Spark Shell 中，请在提示符处键入 `:paste`，粘贴示例，然后按 `CTRL +D`。

## PySpark

Spark 包含一个基于 Python 的 Shell `pyspark`，您可以用它来设计以 Python 编写的 Spark 程序的原型。在主节点上调用 `pyspark`。

```
## Create a DataFrame.
data = spark.createDataFrame([
    ("100", "2015-01-01", "2015-01-01T13:51:39.340396Z"),
    ("101", "2015-01-01", "2015-01-01T12:14:58.597216Z"),
    ("102", "2015-01-01", "2015-01-01T13:51:40.417052Z"),
    ("103", "2015-01-01", "2015-01-01T13:51:40.519832Z")
], ["id", "creation_date", "last_update_time"])

## Write a DataFrame as a Iceberg dataset to the Amazon S3 location.
spark.sql("""CREATE TABLE IF NOT EXISTS dev.db.iceberg_table (id string,
creation_date string,
last_update_time string)
USING iceberg
location 's3://DOC-EXAMPLE-BUCKET/example-prefix/db/iceberg_table'""")

data.writeTo("dev.db.iceberg_table").append()
```

## Scala

```
import org.apache.spark.sql.SaveMode
import org.apache.spark.sql.functions._
```

```
// Create a DataFrame.
val data = Seq(
  ("100", "2015-01-01", "2015-01-01T13:51:39.340396Z"),
  ("101", "2015-01-01", "2015-01-01T12:14:58.597216Z"),
  ("102", "2015-01-01", "2015-01-01T13:51:40.417052Z"),
  ("103", "2015-01-01", "2015-01-01T13:51:40.519832Z")
).toDF("id", "creation_date", "last_update_time")

// Write a DataFrame as a Iceberg dataset to the Amazon S3 location.
spark.sql("""CREATE TABLE IF NOT EXISTS dev.db.iceberg_table (id string,
creation_date string,
last_update_time string)
USING iceberg
location 's3://DOC-EXAMPLE-BUCKET/example-prefix/db/iceberg_table'""")

data.writeTo("dev.db.iceberg_table").append()
```

## 从 Iceberg 表读取

### PySpark

```
df = spark.read.format("iceberg").load("dev.db.iceberg_table")
df.show()
```

### Scala

```
val df = spark.read.format("iceberg").load("dev.db.iceberg_table")
df.show()
```

### Spark SQL

```
SELECT * from dev.db.iceberg_table LIMIT 10
```

## 将 Spark 属性配置为使用 Glue AWS 数据目录作为 Iceberg 表元存储

要使用 Glue AWS Catalog 作为 Iceberg 表格的元数据仓，请按如下方式设置 Spark 配置属性：

```
spark-submit \  
  --conf spark.sql.catalog.my_catalog=org.apache.iceberg.spark.SparkCatalog \  
  --conf spark.sql.catalog.my_catalog.warehouse=s3://<bucket>/<prefix> \  
  ...
```



```
--conf spark.sql.catalog.my_catalog.catalog-impl=org.apache.iceberg.aws.glue.GlueCatalog \  
--conf spark.sql.catalog.my_catalog.io-impl=org.apache.iceberg.aws.s3.S3FileIO \  
--conf spark.sql.catalog.my_catalog.lock-impl=org.apache.iceberg.aws.dynamodb.DynamoDbLockManager \  
--conf spark.sql.catalog.my_catalog.lock.table=myGlueLockTable
```

## 将 Iceberg 集群与 Trino 结合使用

从亚马逊EMR版本 6.6.0 开始，您可以将 Iceberg 与 Trino 集群一起使用。

在本教程中，您将使用在 Amazon EMR Trino 集群上使用 Iceberg。AWS CLI 要使用控制台创建安装了 Iceberg 的集群，请按照使用 Amazon Ath [ena](#)、[Amazon](#) 和 [Glue 构建 Apache Iceberg 数据湖](#) 中的步骤进行操作。EMR AWS

### 创建 Iceberg 集群

要将 Amazon 上的 Iceberg EMR 与一起使用 AWS CLI，请先按照以下步骤创建一个集群。有关使用指定 Iceberg 分类的信息 AWS CLI，请参阅[创建集群 AWS CLI 时使用提供配置](#)或[创建集群SDK时使用 Java 提供配置](#)。

1. 创建 `iceberg.properties` 文件，然后为您选择的目录设置一个值。例如，假设您想将 Hive 元存储作为目录使用，则您的文件应包含以下内容。

```
connector.name=iceberg  
hive.metastore.uri=thrift://localhost:9083
```

如果您想使用 AWS Glue 数据目录作为存储，则您的文件应包含以下内容。

```
connector.name=iceberg  
iceberg.catalog.type=glue
```

2. 创建一个会将 `iceberg.properties` 从 Amazon S3 复制到 `/etc/trino/conf/catalog/iceberg.properties` 的引导操作，如下例所示。有关引导操作的信息，请参阅[创建引导操作以安装其他软件](#)。

```
set -ex  
sudo aws s3 cp s3://DOC-EXAMPLE-BUCKET/iceberg.properties /etc/trino/conf/catalog/  
iceberg.properties
```

3. 使用以下配置创建一个集群，将示例引导操作脚本路径和密钥名称替换为您自己的值。

```
aws emr create-cluster --release-label emr-6.7.0 \  
--applications Name=Trino \  
--region us-east-1 \  
--name My_Trino_Iceberg_Cluster \  
--bootstrap-actions '[{"Path":"s3://DOC-EXAMPLE-BUCKET", "Name":"Add  
iceberg.properties"}]' \  
--instance-groups InstanceGroupType=MASTER,InstanceCount=1,InstanceType=c3.4xlarge  
InstanceGroupType=CORE,InstanceCount=3,InstanceType=c3.4xlarge \  
--use-default-roles \  
--ec2-attributes KeyName=<key-name>
```

## 为 Iceberg 初始化 Trino 会话

要初始化 Trino 会话，请运行以下命令。

```
trino-cli --catalog iceberg
```

## 写入 Iceberg 表

使用以下SQL命令创建并写入您的表。

```
trino> SHOW SCHEMAS;  
trino> CREATE TABLE default.iceberg_table (  
    id int,  
    data varchar,  
    category varchar)  
WITH (  
    format = 'PARQUET',  
    partitioning = ARRAY['category', 'bucket(id, 16)'],  
    location = 's3://DOC-EXAMPLE-BUCKET/<prefix>')  
  
trino> INSERT INTO default.iceberg_table VALUES (1,'a','c1'), (2,'b','c2'),  
(3,'c','c3');
```

## 从 Iceberg 表读取

要从 Iceberg 表读取，请运行以下命令。

```
trino> SELECT * from default.iceberg_table;
```

## 将 Iceberg 集群与 Flink 结合使用

从 Amazon 6.9.0 EMR 版本开始，在使用开源 Iceberg Flink 集成时，无需执行设置步骤，即可将 Iceberg 与 Flink 集群配合使用。

### 创建 Iceberg 集群

您可以使用 AWS Management Console、或 Amazon EMR API 创建安装了 Iceberg 的 AWS CLI 集群。在本教程中，您将使用在 AWS CLI Amazon EMR 集群上使用 Iceberg。要使用控制台创建安装了 Iceberg 的集群，请按照使用 Amazon Athena、Amazon 和 Glue 构建 Apache Iceberg 数据湖中的步骤进行操作。EMR AWS

要将 Amazon 上的 Iceberg EMR 与一起使用 AWS CLI，请先按照以下步骤创建一个集群。有关使用指定 Iceberg 分类的信息 AWS CLI，请参阅[创建集群 AWS CLI 时使用提供配置](#)或[创建集群 SDK 时使用 Java 提供配置](#)。使用以下内容创建名为 configurations.json 的文件：

```
[{
  "Classification": "iceberg-defaults",
  "Properties": {"iceberg.enabled": "true"}
}]
```

接下来，使用以下配置创建集群，将示例 Amazon S3 桶路径和子网 ID 替换为您自己的值：

```
aws emr create-cluster --release-label emr-6.9.0 \
--applications Name=Flink \
--configurations file://iceberg_configurations.json \
--region us-east-1 \
--name My_flink_Iceberg_Cluster \
--log-uri s3://DOC-EXAMPLE-BUCKET/ \
--instance-type m5.xlarge \
--instance-count 2 \
--service-role EMR_DefaultRole \
--ec2-attributes InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-1234567890abcdef
```

或者，您可以创建一个包含 Flink 应用程序的 Amazon EMR 6.9.0 集群，并将该文件 /usr/share/aws/iceberg/lib/iceberg-flink-runtime.jar 用作 Flink 作业中的 JAR 依赖项。

### 使用 Flink 客户端 SQL

SQL 客户端脚本位于下方 /usr/lib/flink/bin。您可以使用以下命令运行脚本：

```
flink-yarn-session -d # starting the Flink YARN Session in detached mode
./sql-client.sh
```

这将启动 Flink SQL 外壳。

## Flink 示例

### 创建 Iceberg 表

#### Flink SQL

```
CREATE CATALOG glue_catalog WITH (
  'type'='iceberg',
  'warehouse'='<WAREHOUSE>',
  'catalog-impl'='org.apache.iceberg.aws.glue.GlueCatalog',
  'io-impl'='org.apache.iceberg.aws.s3.S3FileIO',
  'lock-impl'='org.apache.iceberg.aws.dynamodb.DynamoDbLockManager',
  'lock.table'='myGlueLockTable'
);

USE CATALOG glue_catalog;

CREATE DATABASE IF NOT EXISTS <DB>;

USE <DB>;

CREATE TABLE IF NOT EXISTS `glue_catalog`.`<DB>`.`sample` (id int, data string);
```

#### 桌子 API

```
EnvironmentSettings settings =
    EnvironmentSettings.newInstance().inBatchMode().build();

TableEnvironment tEnv = TableEnvironment.create(settings);

String warehouse = "<WAREHOUSE>";
String db = "<DB>";

tEnv.executeSql(
    "CREATE CATALOG glue_catalog WITH (\n"
    + "  'type'='iceberg',\n"
    + "  'warehouse'='"
    + warehouse
```

```

        + "' ,\n"
        + "    'catalog-impl'='org.apache.iceberg.aws.glue.GlueCatalog',
\n"
        + "    'io-impl'='org.apache.iceberg.aws.s3.S3FileIO'\n"
        + " );");

tEnv.executeSql("USE CATALOG glue_catalog;");
tEnv.executeSql("CREATE DATABASE IF NOT EXISTS " + db + ";");
tEnv.executeSql("USE " + db + ";");
tEnv.executeSql(
    "CREATE TABLE `glue_catalog`.`" + db + "`.`sample` (id bigint, data string);");

```

## 写入 Iceberg 表

### Flink SQL

```
INSERT INTO `glue_catalog`.`<DB>`.`sample` values (1, 'a'),(2,'b'),(3,'c');
```

### 桌子 API

```

tEnv.executeSql(
    "INSERT INTO `glue_catalog`.`"
    + db
    + "`.`sample` values (1, 'a'),(2,'b'),(3,'c');");

```

### 数据流 API

```

final StreamExecutionEnvironment env =
    StreamExecutionEnvironment.getExecutionEnvironment();

final StreamTableEnvironment tableEnv = StreamTableEnvironment.create(env);

String db = "<DB Name>";

String warehouse = "<Warehouse Path>";

GenericRowData rowData1 = new GenericRowData(2);
rowData1.setField(0, 1L);
rowData1.setField(1, StringData.fromString("a"));

DataStream<RowData> input = env.fromElements(rowData1);

```

```

Map<String, String> props = new HashMap<>();
props.put("type", "iceberg");
props.put("warehouse", warehouse);
props.put("io-impl", "org.apache.iceberg.aws.s3.S3FileIO");

CatalogLoader glueCatalogLoader =
    CatalogLoader.custom(
        "glue",
        props,
        new Configuration(),
        "org.apache.iceberg.aws.glue.GlueCatalog");

TableLoader tableLoader =
    TableLoader.fromCatalog(glueCatalogLoader, TableIdentifier.of(db, "sample"));

DataStreamSink<Void> dataStreamSink =
    FlinkSink.forRowData(input).tableLoader(tableLoader).append();

env.execute("Datastream Write");

```

## 从 Iceberg 表读取

### Flink SQL

```
SELECT * FROM `glue_catalog`.`<DB>`.`sample`;
```

### 桌子 API

```
Table result = tEnv.sqlQuery("select * from `glue_catalog`.`" + db + "`.`sample`");
```

### 数据流 API

```

final StreamExecutionEnvironment env =
    StreamExecutionEnvironment.getExecutionEnvironment();

final StreamTableEnvironment tableEnv = StreamTableEnvironment.create(env);

String db = "<DB Name>";

String warehouse = "<Warehouse Path>";

Map<String, String> props = new HashMap<>();
props.put("type", "iceberg");

```

```
props.put("warehouse", warehouse);
props.put("io-impl", "org.apache.iceberg.aws.s3.S3FileIO");

CatalogLoader glueCatalogLoader =
    CatalogLoader.custom(
        "glue",
        props,
        new Configuration(),
        "org.apache.iceberg.aws.glue.GlueCatalog");

TableLoader tableLoader =
    TableLoader.fromCatalog(glueCatalogLoader, TableIdentifier.of(db, "sample"));

DataStream<RowData> batch =
    FlinkSource.forRowData().env(env).tableLoader(tableLoader).streaming(false).build();

batch.print().name("print-sink");
```

## 使用 Hive 目录

确保如 [使用 Hive 元存储和 Glue 目录配置 Flink](#) 中所述解析 Flink 和 Hive 依赖项。

## 运行 Flink 作业

向 Flink 提交作业的一种方法是使用每项作业 Flink YARN 会话。这可以通过以下命令启动：

```
sudo flink run -m yarn-cluster -p 4 -yjm 1024m -ytm 4096m $JAR_FILE_NAME
```

## 将 Iceberg 集群与 Hive 结合使用

在 Amazon 6.9.0 及更高EMR版本中，您可以将 Iceberg 与 Hive 集群配合使用，而无需执行开源 Iceberg Hive 集成所需的设置步骤。对于 Amazon 6.8.0 及更早EMR版本，您可以使用引导操作安装 iceberg-hive-runtime jar 来配置 Hive 以支持 Iceberg。

亚马逊 EMR 6.9.0 包含 [Hive 3.1.3 与 Iceberg 0.14.1 集成](#)的所有功能，还包括亚马逊EMR添加的功能，例如在运行时自动选择支持的执行引擎（亚马逊 6.9.0 版）。EMR EKS

## 创建 Iceberg 集群

您可以使用 AWS Management Console、AWS CLI 或 Amazon EMR API 创建安装了 Iceberg 的集群。在本教程中，您将使用在 AWS CLI Amazon EMR 集群上使用 Iceberg。要使用控制台创建安装

了 Iceberg 的集群，请按照使用 Amazon Athena、Amazon Glue 和 Glue 构建冰山数据湖中的步骤进行操作。EMR AWS

要将 Amazon 上的 Iceberg EMR 与一起使用 AWS CLI，请先按照以下步骤创建一个集群。有关使用 AWS CLI 或 Java 指定 Iceberg 分类的信息 SDK，请参阅 [创建集群 AWS CLI 时使用提供配置](#) 或 [创建集群 SDK 时使用 Java 提供配置](#)。使用以下内容创建名为 `configurations.json` 的文件：

```
[{
  "Classification": "iceberg-defaults",
  "Properties": {"iceberg.enabled": "true"}
}]
```

接下来，使用以下配置创建集群，将示例 Amazon S3 桶路径和子网 ID 替换为您自己的值：

```
aws emr create-cluster --release-label emr-6.9.0 \
--applications Name=Hive \
--configurations file://iceberg_configurations.json \
--region us-east-1 \
--name My_hive_Iceberg_Cluster \
--log-uri s3://DOC-EXAMPLE-BUCKET/ \
--instance-type m5.xlarge \
--instance-count 2 \
--service-role EMR_DefaultRole \
--ec2-attributes InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-1234567890abcdef
```

Hive Iceberg 集群执行以下操作：

- 在 Hive 中加载 Iceberg Hive 运行时 jar 并为 Hive 引擎启用 Iceberg 相关配置。
- 启用 Amazon EMR Hive 的动态执行引擎选择，以防止用户为兼容 Iceberg 而设置支持的执行引擎。

#### Note

Hive Iceberg 集群目前不支持 Glue AWS 数据目录。默认 Iceberg 目录为 HiveCatalog，它对应于为 Hive 环境配置的元存储。有关目录管理的更多信息，请参阅 [Apache Hive 文档 HCatalog 中的使用](#)。



## 功能支持

亚马逊 EMR 6.9.0 支持 Hive 3.1.3 和 Iceberg 0.14.1。该功能支持仅限于 Hive 3.1.2 和 3.1.3 的 Iceberg 兼容功能。支持以下命令：

- 在亚马逊EMR发布6.9.0至6.12.x版本时，您必须将 libfb303 jar包含在Hive目录中。auxlib使用以下命令将其包含在内：

```
sudo /usr/bin/ln -sf /usr/lib/hive/lib/libfb303-*.jar /usr/lib/hive/auxlib/libfb303.jar
```

在 Amazon 6.13 及更高EMR版本中，libfb303jar 会自动符号链接到 Hive auxlib 目录。

- 创建表

- 非分区表 - 可以通过提供存储处理程序在 Hive 中创建外部表，如下所示：

```
CREATE EXTERNAL TABLE x (i int) STORED BY  
'org.apache.iceberg.mr.hive.HiveIcebergStorageHandler'
```

- 分区表 - 可以在 Hive 中创建外部分区表，如下所示：

```
CREATE EXTERNAL TABLE x (i int) PARTITIONED BY (j int) STORED BY  
'org.apache.iceberg.mr.hive.HiveIcebergStorageHandler'
```

### Note

Hive 3 PARQUET 不支持AVRO/ORC/STORED AS的文件格式。默认且唯一的选项是 Parquet。

- 删除表 - DROP TABLE 命令用于删除表，如以下示例中所示：

```
DROP TABLE [IF EXISTS] table_name [PURGE];
```

- 读取表 - SELECT 语句可用于读取 Hive 中的 Iceberg 表，如以下示例中所示。支持的执行引擎为 MR 和 Tez。

```
SELECT * FROM table_name
```

有关 Hive 的选择语法的信息，请参阅[LanguageManual 选择](#)。有关在 Hive 中使用 Iceberg 表的选择语句的信息，请参阅 [Apache Iceberg Select](#)。

- 插入到表中 - HiveQL 的 INSERT INTO 语句仅适用于支持 Map Reduce 执行引擎的 Iceberg 表。亚马逊EMR用户无需明确设置执行引擎，因为 Amazon EMR Hive 会在运行时为 Iceberg Tables 选择引擎。

- 单表插入 - 例如：

```
INSERT INTO table_name VALUES ('a', 1);
INSERT INTO table_name SELECT...;
```

- 多表插入 - 支持在语句中插入非原子多表。例如：

```
FROM source
INSERT INTO table_1 SELECT a, b
INSERT INTO table_2 SELECT c,d;
```

## 在亚马逊上使用 Iceberg 的注意事项和限制 EMR

本节包含将 Iceberg 与 Spark、Trino、Flink 和 Hive 结合使用的注意事项和限制。

### 将 Iceberg 与 Spark 结合使用的注意事项

- 默认情况下，亚马逊 EMR 6.5.0 不支持EMR在亚马逊上运行 Iceberg EKS。Amazon EMR 6.5.0 自定义映像可用，因此您可以--jars local:///usr/share/aws/iceberg/lib/iceberg-spark3-runtime.jar作为spark-submit参数传递，以便在亚马逊EMR上创建 Iceberg 表。EKS有关更多信息，请参阅《[亚马逊EKS开发指南](#)》中的[EMR使用自定义图像在亚马逊中提交 Spark 工作负载](#)。EMR您也可以联系 AWS Support 获取帮助。从亚马逊 EMR 6.6.0 开始，亚马逊支持 Iceberg。EMR EKS
- 使用 AWS Glue 作为 Iceberg 的目录时，请确保要在其中创建表的数据库存在于 Glue 中 AWS。如果您正在使用诸如的服务 AWS Lake Formation，但无法加载目录，请确保您拥有对服务的正确访问权限以执行命令。

### 将 Iceberg 与 Trino 结合使用的注意事项

- 亚马逊 EMR 6.5 本身不支持 Iceberg 的 Trino Iceberg Catalog。Trino 需要 Iceberg v0.11，因此我们建议在 Trino 上启动与 Spark EMR 集群分开的亚马逊集群，并在该集群上包含 Iceberg v0.11。

- 使用 AWS Glue 作为 Iceberg 的目录时，请确保要在其中创建表的数据库存在于 Glue 中 AWS。如果您正在使用诸如的服务 AWS Lake Formation，但无法加载目录，请确保您拥有对服务的正确访问权限以执行命令。

## 将 Iceberg 与 Flink 结合使用的注意事项

使用 AWS Glue 作为 Iceberg 的目录时，请确保要在其中创建表的数据库存在于 Glue 中 AWS。如果您正在使用诸如的服务 AWS Lake Formation，但无法加载目录，请确保您拥有对服务的正确访问权限以执行命令。

## 将 Iceberg 与 Hive 结合使用的注意事项

- Iceberg 支持以下查询类型：
  - 创建表
  - 删除表
  - 插入到表中
  - 读取表
- ( 数据操作语言MapReduce ) 操作仅支持 MR DML () 执行引擎，Hive 3.1.3 中已弃用 MR。
- AWS 带有 Hive 的 Iceberg 目前不支持 Glue 数据目录。
- 错误处理不够强大。在配置错误的情况下，插入查询可能会成功完成。但是，无法更新元数据可能会导致数据丢失。

## Iceberg 发布历史记录

下表列出了 Amazon 的每个发行版本中包含的 Iceberg 版本EMR，以及随应用程序一起安装的组件。有关每个发行版本中的组件版本，请参阅 [亚马逊 EMR 7.x 发布版本](#)、[亚马逊 EMR 6.x 发布版本](#) 或 [亚马逊 EMR 5.x 发布版本](#) 中的发行版“组件版本”部分。

## Iceberg 版本信息

Amazon EMR 发布标签	Iceberg 版本	随 Iceberg 安装的组件
emr-7.2.0	1.5.0-amzn-0	Not available.
emr-7.1.0	1.4.3-amzn-0	Not available.
emr-7.0.0	1.4.2-amzn-0	Not available.
emr-6.15.0	1.4.0-amzn-0	Not available.
emr-6.14.0	1.3.1-amzn-0	Not available.
emr-6.13.0	1.3.0-amzn-1	Not available.
emr-6.12.0	1.3.0-amzn-0	Not available.
emr-6.11.1	1.2.0-amzn-0	Not available.
emr-6.11.0	1.2.0-amzn-0	Not available.
emr-6.10.1	1.1.0-amzn-0	Not available.
emr-6.10.0	1.1.0-amzn-0	Not available.
emr-6.9.1	0.14.1-amzn-0	Not available.
emr-6.9.0	0.14.1-amzn-0	Not available.
emr-6.8.1	0.14.0-amzn-0	Not available.
emr-6.8.0	0.14.0-amzn-0	Not available.
emr-6.7.0	0.13.1-amzn-0	Not available.
emr-6.6.0	0.13.1	Not available.
emr-6.5.0	0.12.0	Not available.

## Iceberg 发布说明 ( 按版本分类 )

- [亚马逊 EMR 6.9.0-Iceberg 发行说明](#)

### 亚马逊 EMR 6.9.0-Iceberg 发行说明

#### 亚马逊 EMR 6.9.0-Iceberg 变更

类型	描述
功能	亚马逊 EMR Flink 与 Iceberg 集成。
功能	亚马逊 EMR Hive 与 Iceberg 集成。
功能	支持在 Amazon FSx for Lustre 上缓存 Iceberg 元数据文件，以缩短查询规划时间。
逆向移植	<a href="#">PR 5050</a> : Flink 1.15 : 支持行内插入注释中的写入选项。SQL
逆向移植	<a href="#">PR 5282</a> : AWS: 通过打开新的数据文件流修复 PUT重试失败问题。
逆向移植	<a href="#">PR 5318</a> : Flink 1.15 : 弥合 FlinkSource 和 IcebergSource (FLIP-27) 之间的差距，并添加了可选配置以在 Flink 中使用 FLIP -27 来源。SQL
逆向移植	<a href="#">PR 5344</a> : Flink 1.14 : 弥合 FlinkSource 和 IcebergSource (FLIP-27) 之间的差距，并添加了可选配置以在 Flink 中使用 FLIP -27 来源。SQL
逆向移植	<a href="#">PR 5393</a> : Flink 1.14、1.15 : 避免在 -27 来源阅读器中将 Iceberg 指标转换 MetricContext 为 Flink 指标。FLIP

类型	描述
逆向移植	<a href="#">PR 5401</a> : Flink 1.14、1.15 : 在 PR #5393 中错过了 FLIP -27 个来源读者指标的 IcebergSourceReader 群组。
逆向移植	<a href="#">PR 5679</a> : Spark 3.2、3.3 : 修复了节点的可空性传播问题。 MergeRows
逆向移植	<a href="#">PR 5860</a> : Spark 3.3 : 修复 RewriteManifestProcedure 在 Date 分区表上运行 QueryFailure 时的问题。
逆向移植	<a href="#">PR 5880</a> : Spark 3.3 : 修复了投影中的可空性。 merge-on-read
逆向移植	<a href="#">PR 5917</a> : Spark 3.2 : 修复了投影中的可空性。 merge-on-read

# Amazon EMR 上的 Jupyter 笔记本

[Jupyter 笔记本](#)是一个开源 Web 应用程序，可用于创建和共享包含实时代码、公式、可视化效果和叙述性文本的文档。Amazon EMR 为您提供了三个使用 Jupyter 笔记本的选项：

主题

- [EMR Studio](#)
- [基于 Jupyter 笔记本的 Amazon EMR Notebook](#)
- [JupyterHub](#)

## EMR Studio

Amazon EMR Studio 是一个基于 Web 的集成开发环境 (IDE)，适用于依托 Amazon EMR 集群运行的完全托管式 [Jupyter 笔记本](#)。您可以为团队设置 EMR Studio，以开发、可视化和调试用 R、Python、Scala 和 PySpark 编写的应用程序。

我们建议在 Amazon EMR 上使用 Jupyter 笔记本时使用 EMR Studio。详情请参见《Amazon EMR 管理指南》中的 [EMR Studio](#)。

## 基于 Jupyter 笔记本的 Amazon EMR Notebook

EMR Notebooks 是 Amazon EMR 控制台中内置的一个 [Jupyter 笔记本](#)环境，您可以在该环境中快速创建 Jupyter 笔记本，将它们连接到 Spark 集群，然后在控制台中打开 Jupyter 笔记本编辑器，以便远程运行查询和代码。EMR 笔记本独立于集群保存在 Amazon S3 中，可实现持久性存储、快速访问和灵活性。您可以打开多个笔记本、将多个笔记本连接到单个集群，以及在不同集群上重新使用笔记本。

详情请参见《Amazon EMR 管理指南》中的 [EMR Notebooks](#)。

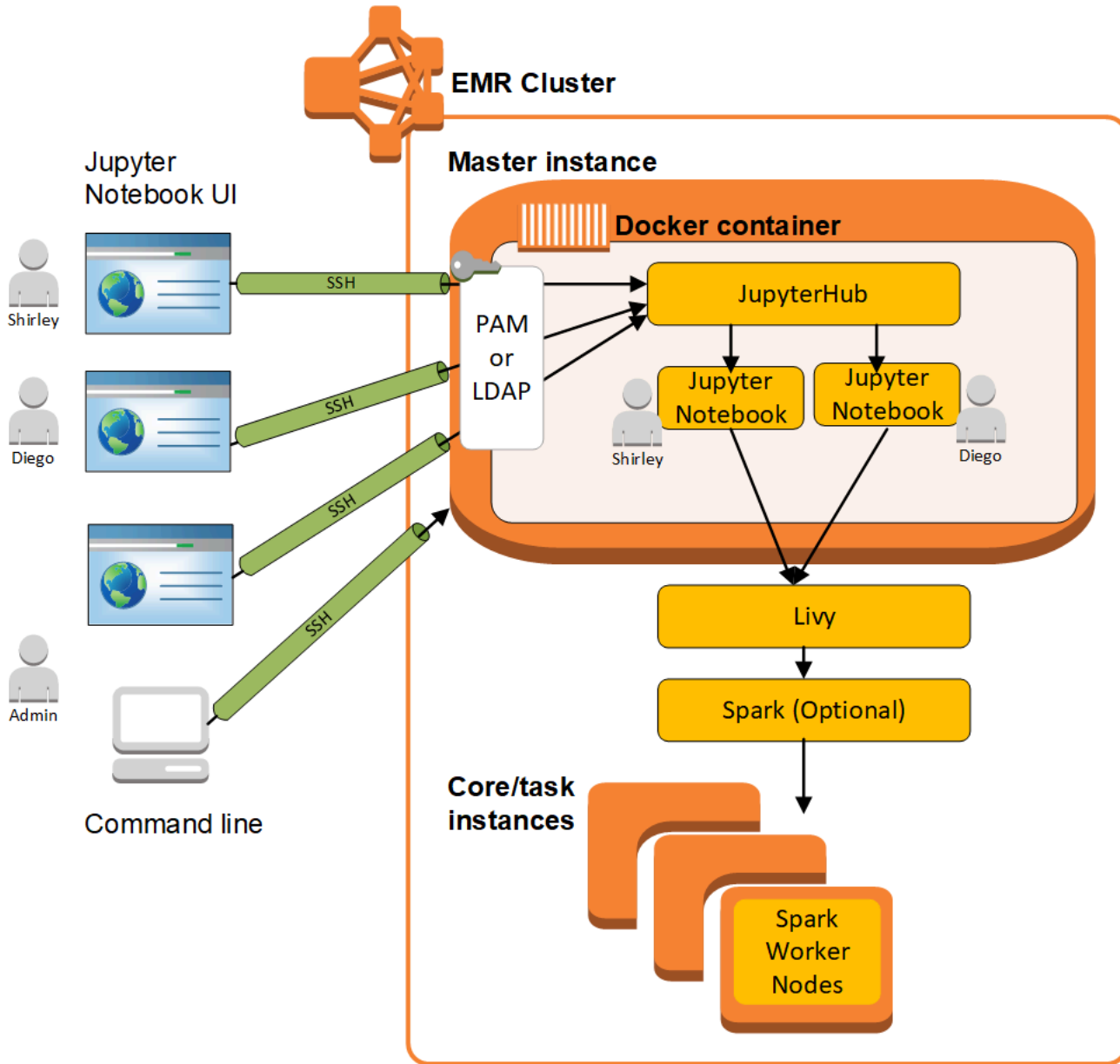
## JupyterHub

[Jupyter notebook](#) 是一个开源 Web 应用程序，可用于创建和共享包含实时代码、公式、可视化效果和叙述性文本的文档。[JupyterHub](#)允许您托管单用户 Jupyter 笔记本服务器的多个实例。当您使用创建集群时 JupyterHub，Amazon EMR 会在集群的主节点上创建一个 Docker 容器。JupyterHub、Jupyter 和 [Sparkmagic](#) 所需的所有组件都在容器内运行。

Sparkmagic 是一个内核库，允许 Jupyter 笔记本电脑EMR通过它与在亚马逊上[运行的 Apache Spark 进行交互](#)[Apache Livy](#)，[Ap](#)ache Spark 是 Spark 的服务器。REST当你使用创建集群时，Spark 和

Apache Livy 会自动安装。JupyterHubJupyter 的默认 Python 3 内核与 Sparkmagic 提供的 PySpark 3 PySpark、和 Spark 内核一起可用。你可以使用这些内核通过 Python 和 S SQL cala 运行临时的 Spark 代码和交互式查询。可以在 Docker 容器内手动安装其它内核。有关更多信息，请参阅[安装其它内核和库](#)。

下图描述了 on Amaz JupyterHub EMR on 的组件以及适用于笔记本用户和管理员的相应身份验证方法。有关更多信息，请参阅 [添加 Jupyter notebook 用户和管理员](#)。



下表列出了最新版本的 Amazon EMR 7.x 系列中 JupyterHub 包含的版本以及亚马逊随之EMR安装的组件。 JupyterHub

有关此版本 JupyterHub 中安装的组件的版本，请参阅 [7.2.0 版组件版本](#)。



## JupyterHub emr-7.2.0 的版本信息

亚马逊EMR发行标签	JupyterHub 版本	安装的组件 JupyterHub
emr-7.2.0	JupyterHub 1.5.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

下表列出了最新版本的 Amazon EMR 6.x 系列中 JupyterHub 包含的版本以及亚马逊随之EMR安装的组件。 JupyterHub

有关此版本 JupyterHub 中安装的组件的版本，请参阅 [6.15.0 版组件版本](#)。

## JupyterHub emr-6.15.0 的版本信息

亚马逊EMR发行标签	JupyterHub 版本	安装的组件 JupyterHub
emr-6.15.0	JupyterHub 1.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server,

亚马逊EMR发行标签	JupyterHub 版本	安装的组件 JupyterHub
		spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

下表列出了最新版本的 Amazon EMR 5.x 系列中 JupyterHub 包含的版本以及亚马逊随之EMR安装的组件。 JupyterHub

有关此版本 JupyterHub 中安装的组件的版本，请参阅 [5.36.2 版组件版本](#)。

JupyterHub emr-5.36.2 的版本信息

亚马逊EMR发行标签	JupyterHub 版本	安装的组件 JupyterHub
emr-5.36.2	JupyterHub 1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

亚马逊 JupyterHub 上包含的 Python 3 内核EMR是 3.6.4。

jupyterhub容器中安装的库可能因亚马逊EMR发行版本和亚马逊EC2AMI版本而异。

使用 **conda** 列出已安装的库

- 在主节点命令行上运行以下命令：

```
sudo docker exec jupyterhub bash -c "conda list"
```

## 使用 **pip** 列出已安装的库

- 在主节点命令行上运行以下命令：

```
sudo docker exec jupyterhub bash -c "pip freeze"
```

### 主题

- [使用创建集群 JupyterHub](#)
- [JupyterHub 在 Amazon 上使用时的注意事项 EMR](#)
- [正在配置 JupyterHub](#)
- [在 Amazon S3 中配置笔记本的持久性](#)
- [连接到主节点和笔记本服务器](#)
- [JupyterHub 配置和管理](#)
- [添加 Jupyter notebook 用户和管理员](#)
- [安装其它内核和库](#)
- [JupyterHub 发布历史](#)

## 使用创建集群 JupyterHub

您可以 JupyterHub 使用 AWS Management Console、AWS Command Line Interface 或 Amazon 创建 Amazon EMR 集群 EMRAPI。确保不使用在完成步骤后自动终止的选项 (AWS CLI 中的 `--auto-terminate` 选项) 创建此集群。此外，确保管理员和笔记本用户可以访问创建集群时使用的密钥对。有关更多信息，请参阅《Amazon EMR 管理指南》中的 [“使用密钥对作为SSH证书”](#)。

### JupyterHub 使用控制台创建集群

使用以下步骤在 Amazon EMR 控制台中使用高级选项创建 JupyterHub 已安装的集群。

使用亚马逊EMR控制台创建 JupyterHub 已安装的 Amazon EMR 集群

1. 导航到新的 Amazon EMR 控制台，然后从侧面导航栏中选择“切换到旧主机”。有关切换到旧控制台后预期情况的更多信息，请参阅 [Using the old console](#)。
2. 依次选择 Create cluster (创建集群)、Go to advanced options (转到高级选项)。
3. 在 Software Configuration (软件配置) 下：

- 对于“版本”，选择 emr-5.36.2，然后选择。 JupyterHub
  - 如果你使用 Spark，要使用 AWS Glue 数据目录作为 Spark 的元数据库SQL，请选择“用于 Spark 表元数据”。有关更多信息，请参阅 [使用 AWS Glue 数据目录作为 Spark 的元数据库 SQL](#)。
  - 对于编辑软件设置，选择输入配置并指定值，或者选择JSON从 S3 加载并指定JSON配置文件。有关更多信息，请参阅[正在配置 JupyterHub](#)。
4. 在 Add steps (添加步骤) (可选) 下，配置创建集群后要运行的步骤，确保 Auto-terminate cluster after the last step is completed (完成最后的步骤后，自动终止集群) 未选中，然后选择 Next (下一步)。
  5. 选择 Hardware Configuration (硬件配置) 选项、Next (下一步)。有关更多信息，请参阅 Amazon EMR 管理指南中的[配置集群硬件和联网](#)。
  6. 选择 General Cluster Settings (常规集群设置) 和 Next (下一步) 选项。
  7. 选择 Security Options (安全选项) 以指定密钥对，然后选择 Create Cluster (创建集群)。

## JupyterHub 使用创建集群 AWS CLI

要使用启动集群 JupyterHub，请使用 `aws emr create-cluster` 命令，并在 `--applications` 选项中指定 `Name=JupyterHub`。以下示例在 Amazon 上启动 EMR 具有两个 EC2 实例（一个主实例和一个核心实例）的 JupyterHub 集群。此外，已启用调试，日志存储在 `--log-uri` 所指定的 Amazon S3 位置中。指定的密钥对提供对集群中的 Amazon EC2 实例的访问权限。

### Note

为了便于读取，包含 Linux 行继续符 (`\`)。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (`^`)。

```
aws emr create-cluster --name="MyJupyterHubCluster" --release-label emr-5.36.2 \  
--applications Name=JupyterHub --log-uri s3://MyBucket/MyJupyterClusterLogs \  
--use-default-roles --instance-type m5.xlarge --instance-count 2 --ec2-attributes \  
KeyName=MyKeyPair
```

## JupyterHub 在 Amazon 上使用时的注意事项 EMR

在 Amazon JupyterHub 上使用，请考虑以下几点 EMR。

### ⚠ Warning

用户笔记本和文件将保存到主节点上的文件系统中。这是短暂存储，在集群终止后将不复存在。集群终止后，此数据如果未备份的化将丢失。建议使用 cron 作业或其它适用于应用程序的方式安排定期备份。

此外，如果容器重启，在容器内进行的配置更改可能不复存在。建议为容器配置编写脚本或以其它方式实现容器配置的自动化，以便可以更轻松地重现自定义。

- 不支持使用 Amazon EMR 安全配置设置的 Kerberos 身份验证。
- 不支持 [OAuthenticator](#)。

## 正在配置 JupyterHub

您可以通过连接到集群主节点EMR并编辑配置文件来自定义 Amaz JupyterHub on 和个人用户笔记本的配置。在更改值之后，重启 jupyterhub 容器。

修改以下文件中的属性以配置 JupyterHub 和各个 Jupyter 笔记本：

- `jupyterhub_config.py` – 默认情况下，此文件保存在主节点上的 `/etc/jupyter/conf/` 目录中。有关更多信息，请参阅 JupyterHub 文档中的[配置基础知识](#)。
- `jupyter_notebook_config.py` – 默认情况下，该文件保存在 `/etc/jupyter/` 目录中，并作为默认值复制到 jupyterhub 容器中。有关更多信息，请参阅 Jupyter notebook 文档中的[配置文件和命令行选项](#)。

您也可以使用 `jupyter-sparkmagic-conf` 配置分类自定义 Sparkmagic，这会更新 Sparkmagic 的 `config.json` 文件中的值。有关可用设置的更多信息，请参阅上的 [e\\_xample\\_c onfig.json](#)。GitHub 有关在 Amazon 中对应用程序使用配置分类的更多信息EMR，请参阅[配置应用程序](#)。

以下示例使用启动集群 AWS CLI，引用 Sparkmagic 配置分类设置MyJupyterConfig.json的文件。

### 📘 Note

为了便于读取，包含 Linux 行继续符 (`\`)。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (`^`)。

```
aws emr create-cluster --use-default-roles --release-label emr-5.14.0 \
--applications Name=Jupyter --instance-type m4.xlarge --instance-count 3 \
--ec2-attributes KeyName=MyKey,SubnetId=subnet-1234a5b6 --configurations file://
MyJupyterConfig.json
```

MyJupyterConfig.json 的示例内容如下所示：

```
[
  {
    "Classification": "jupyter-sparkmagic-conf",
    "Properties": {
      "kernel_python_credentials" : "{\"username\":\"diego\",\"base64_password\":
\"mypass\",\"url\":\"http://localhost:8998\",\"auth\":\"None\"}"
    }
  }
]
```

#### Note

在 Amazon 5.21.0 及更高EMR版本中，您可以覆盖集群配置，并为正在运行的集群中的每个实例组指定其他配置分类。为此，您可以使用 Amazon EMR 控制台、AWS Command Line Interface (AWS CLI) 或 AWS SDK。有关更多信息，请参阅[为运行的集群中的实例组提供配置](#)。

## 在 Amazon S3 中配置笔记本的持久性

您可以在 Amazon 中配置 JupyterHub 集群，EMR以使用户保存的笔记本保留在 Amazon S3 中，而不是集群实例上的临时存储空间。EC2

在创建集群时使用 `jupyter-s3-conf` 配置分类指定 Amazon S3 持久性。有关更多信息，请参阅[配置应用程序](#)。

除了使用 `s3.persistence.enabled` 属性启用 Amazon S3 持久性之外，还请使用 `s3.persistence.bucket` 属性在保存笔记本的 Amazon S3 中指定存储桶。每个用户的笔记本将保存到指定存储桶中的 `jupyter/jupyterhub-user-name` 文件夹。存储桶必须已存在于 Amazon S3 中，并且您在创建集群时为EC2实例配置文件指定的角色必须拥有该存储桶的权限（默认情况下，角色为 `EMR_EC2_DefaultRole`）。有关更多信息，请参阅[配置IAM角色以获得 Amazon AWS 服务的 EMR权限](#)。

当您使用相同的配置分类属性启动新集群时，用户可以打开内容来自自己保存位置的笔记本。

请注意，当您启用了 Amazon S3 后，如果将文件作为模块导入到笔记本中，这会导致文件上传到 Amazon S3。当您在不启用 Amazon S3 持久性的情况下导入文件时，它们会上传到您的 JupyterHub 容器中。

以下示例启用 Amazon S3 持久性。用户保存的笔记本保存在每个用户的 `s3://MyJupyterBackups/jupyter/jupyterhub-user-name` 文件夹中，其中 *jupyterhub-user-name* 是一个用户名（如 `diego`）。

```
[
  {
    "Classification": "jupyter-s3-conf",
    "Properties": {
      "s3.persistence.enabled": "true",
      "s3.persistence.bucket": "MyJupyterBackups"
    }
  }
]
```

## 连接到主节点和笔记本服务器

JupyterHub 管理员和笔记本用户必须使用 SSH 隧道连接到群集主节点，然后连接到主节点 JupyterHub 上提供服务的 Web 界面。有关配置 SSH 隧道和使用隧道代理 Web 连接的更多信息，请参阅 Amazon EMR 管理指南中的 [Connect 到集群](#)。

默认情况下，Amazon JupyterHub 上 EMR 可通过主节点上的端口 9443 进行访问。内部 JupyterHub 代理还通过端口 9443 为笔记本实例提供服务。JupyterHub 而且 Jupyter 的 Web 界面可以使用以下 URL 模式进行访问：

`https://MasterNodeDNS:9443`

可以使用 `c.JupyterHub.port` 文件中的 `jupyterhub_config.py` 属性指定不同的端口。有关更多信息，请参阅 JupyterHub 文档中的 [网络基础知识](#)。

默认情况下，Amazon JupyterHub on EMR 使用自签名证书进行 SSL 加密。HTTPS 用户连接时，系统将提示用户信任自签名凭证。可以使用自己的受信任凭证和密钥。将主节点上 `server.crt` 目录中的默认凭证文件 `server.key` 和密钥文件 `/etc/jupyter/conf/` 更换为自己的凭证和密钥文件。使用 `jupyterhub_config.py` 文件中的 `c.JupyterHub.ssl_key` 和 `c.JupyterHub.ssl_cert` 属性来指定您的 SSL 材质。有关更多信息，请参阅 JupyterHub 文档中的 [安全设置](#)。在更新 `jupyterhub_config.py` 之后，重启容器。

## JupyterHub 配置和管理

JupyterHub 相关组件在运行 Ubuntu 操作系统的名为 `jupyterhub` Docker 容器中运行。有多种方法可用于管理此容器内运行的组件。

### Warning

在此容器内执行的自定义将在此容器重启后不复存在。建议为容器配置编写脚本或以其它方式实现容器配置的自动化，以便可以更轻松地重现自定义。

## 使用命令行管理

使用连接到主节点时SSH，您可以使用 Docker 命令行接口 (CLI) 发出命令，并通过名称 (`jupyterhub`) 或 ID 指定容器。例如，`sudo docker exec jupyterhub command` 将运行容器内运行的操作系统或应用程序识别的命令。可以使用此方法将用户添加到操作系统和在 Docker 容器内安装其它应用程序和库。例如，默认容器映像包括用于安装软件包的 Conda，因此可能在主节点命令行上运行以下命令以在容器内安装应用程序 Keras：

```
sudo docker exec jupyterhub conda install keras
```

## 通过提交步骤管理

步骤是将工作提交到集群的一种方式。可以在启动集群时提交步骤，也可以将步骤提交给正在运行的集群。可以使用 `command-runner.jar` 将在命令行上运行的命令作为步骤提交。有关更多信息，请参阅 [《Amazon EMR 管理指南》中的使用CLI和控制台处理步骤](#)和 [在 Amazon EMR 集群上运行命令和脚本](#)。

例如，您可以在本地计算机上使用以下 AWS CLI 命令来安装 Keras，方法与在前面的示例中通过主节点的命令行所做的方式相同：

```
aws emr add-steps --cluster-id MyClusterID --steps Name="Command Runner",Jar="command-runner.jar",Args="/usr/bin/sudo","/usr/bin/docker","exec","jupyterhub","conda","install","keras"
```

此外，可以为步骤的序列编写脚本，并将脚本上传到 Amazon S3，然后在创建集群或将脚本作为步骤添加时使用 `script-runner.jar` 运行脚本。有关更多信息，请参阅 [在 Amazon EMR 集群上运行命令和脚本](#)。有关示例，请参阅 [the section called “示例：用于添加多个用户的清除脚本”](#)。



## 使用“管理” REST APIs

Jupyter JupyterHub、，以及可用于发送REST APIs请求的 JupyterHub 提供的HTTP代理。要向发送请求 JupyterHub，您必须在请求中传递API令牌。您可以使用主节点curl命令行中的命令来执行REST命令。有关更多信息，请参阅以下资源：

- REST API在@@ [JupyterHub的文档中使用](#) JupyterHub，其中包括生成API令牌的说明
- [Jupyter 笔记本服务器](#)已开启 API GitHub
- [configurable-http-proxy](#) on GitHub

以下示例演示如何使用 for 获取用户列表。REST API JupyterHub 该命令传递先前生成的管理员令牌，并使用默认端口 9443，将输出传送到 [jq 以便于查看：JupyterHub](#)

```
curl -XGET -s -k https://$HOST:9443/hub/api/users \
-H "Authorization: token $admin_token" | jq .
```

## 添加 Jupyter notebook 用户和管理员

您可以使用两种方法中的一种让用户进行身份验证， JupyterHub 以便他们可以创建笔记本并可以选择进行管理 JupyterHub。最简单的方法是使用可插入 JupyterHub的身份验证模块 () PAM。此外，亚马逊 JupyterHub 上还EMR支持 [JupyterHub用于从服务器（例如微软 Active Directory LDAP 服务器）获取用户身份的身份LDAP验证器插件](#)。此部分提供了通过每种身份验证方法添加用户的说明和示例。

JupyterHub Amazon 上EMR有一个具有管理员权限的默认用户。此用户名为 jovyan，密码为 jupyter。强烈建议将此用户替换为另一个具有管理员权限的用户。您可以在创建集群时使用一个步骤以执行该操作，也可以在集群运行时连接到主节点以执行该操作。

### 主题

- [使用 PAM 身份验证](#)
- [使用 LDAP 身份验证](#)
- [用户模拟](#)

## 使用 PAM 身份验证

在 Amazon JupyterHub 上创建PAM用户EMR是一个分为两个步骤的过程。第一步是，将用户添加到在主节点上的 jupyterhub 容器中运行的操作系统，以及为每个用户添加一个相应的用户主目录。第二

步是将这些操作系统用户添加为 JupyterHub用户，此过程称为白名单。JupyterHub添加 JupyterHub 用户后，他们可以连接到 JupyterHub URL并提供其操作系统凭据进行访问。

当用户登录时，会 JupyterHub 打开该用户的笔记本服务器实例，该实例保存在主节点上用户的主目录中，即`/var/lib/jupyter/home/username`。如果笔记本服务器实例不存在，则会在用户的主目录中 JupyterHub 生成一个笔记本实例。以下各节演示如何将用户单独添加到操作系统和操作系统中 JupyterHub，接下来是添加多个用户的基本 bash 脚本。

## 将操作系统用户添加到容器

以下示例先在容器内使用 `useradd` 命令添加单个用户 `diego` 并为该用户创建一个主目录。第二个命令使用 `chpasswd` 为此用户设置密码 `diego`。使用连接时，命令在主节点命令行上运行SSH。还可以使用步骤运行这些命令，如之前的[通过提交步骤管理](#)中所述。

```
sudo docker exec jupyterhub useradd -m -s /bin/bash -N diego
sudo docker exec jupyterhub bash -c "echo diego:diego | chpasswd"
```

## 添加 JupyterHub 用户

您可以使用 JupyterHub 或中的管理面板RESTAPI来添加用户和管理员，或者只添加用户。

### 使用中的管理面板添加用户和管理员 JupyterHub

1. 使用连接主节点SSH并登录 `https://MasterNodeDNS:9443` 具有管理员权限的身份。
2. 选择 Control Panel (控制面板)、Admin (管理员)。
3. 选择 User (用户)、Add Users (添加用户)，或选择 Admin (管理员)、Add Admins (添加管理员)。

### 要添加用户，请使用 REST API

1. 使用SSH并在主节点上使用以下命令连接到主节点，或者将该命令作为步骤运行。
2. 获取管理令牌以发出API请求，然后替换 `AdminToken` 在接下来的步骤中使用该令牌。
3. 使用以下命令，替换 `UserName` 使用在容器内创建的操作系统用户。

```
curl -XPOST -H "Authorization: token AdminToken" "https://$(hostname):9443/hub/api/users/UserName"
```

**Note**

首次登录 JupyterHub Web 界面时，系统会自动将您添加为 JupyterHub 非管理员用户。

示例：用于添加多个用户的清除脚本

以下示例 bash 脚本结合了本节中前面的步骤，创建了多个 JupyterHub 用户。此脚本可以直接在主节点上运行，也可上载到 Amazon S3 并在之后作为步骤运行。

该脚本首先建立用户名数组，然后使用该 jupyterhub token 命令为默认管理员 jovyan 创建 API 令牌。然后，它在 jupyterhub 容器中为每个用户创建一个操作系统用户，以为每个用户分配一个与其用户名相同的初始密码。最后，它调用该 REST API 操作在中创建每个用户 JupyterHub。它会传递脚本前面生成的令牌，并将 REST 响应传送到管道上，jq 以便于查看。

```
# Bulk add users to container and JupyterHub with temp password of username
set -x
USERS=(shirley diego ana richard li john mary anaya)
TOKEN=$(sudo docker exec jupyterhub /opt/conda/bin/jupyterhub token jovyan | tail -1)
for i in "${USERS[@]}";
do
    sudo docker exec jupyterhub useradd -m -s /bin/bash -N $i
    sudo docker exec jupyterhub bash -c "echo $i:$i | chpasswd"
    curl -XPOST --silent -k https://$(hostname):9443/hub/api/users/$i \
    -H "Authorization: token $TOKEN" | jq
done
```

将此脚本保存到 Amazon S3 中的位置（如 s3://mybucket/createjupyterusers.sh）。然后，可以使用 script-runner.jar 将此脚本作为步骤运行。

示例：创建集群时运行脚本（AWS CLI）

**Note**

为了便于读取，包含 Linux 行继续符（\）。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号（^）。

```
aws emr create-cluster --name="MyJupyterHubCluster" --release-label emr-5.36.2 \
```

```
--applications Name=JupyterHub --log-uri s3://MyBucket/MyJupyterClusterLogs \  
--use-default-roles --instance-type m5.xlarge --instance-count 2 --ec2-attributes  
  KeyName=MyKeyPair \  
--steps Type=CUSTOM_JAR,Name=CustomJAR,ActionOnFailure=CONTINUE,\  
Jar=s3://region.elasticmapreduce/libs/script-runner/script-runner.jar,Args=["s3://  
mybucket/createjupyterusers.sh"]
```

在现有集群上运行脚本 ( AWS CLI )

### Note

为了便于读取，包含 Linux 行继续符 ( \ )。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (^)。

```
aws emr add-steps --cluster-id j-XXXXXXXX --steps Type=CUSTOM_JAR,\  
Name=CustomJAR,ActionOnFailure=CONTINUE,\  
Jar=s3://region.elasticmapreduce/libs/script-runner/script-runner.jar,Args=["s3://  
mybucket/createjupyterusers.sh"]
```

## 使用 LDAP 身份验证

轻型目录访问协议 (LDAP) 是一种应用程序协议，用于查询和修改与存储在LDAP兼容目录服务提供商 (例如 Active Directory 或 Open LDAP 服务器) 中的资源 (例如用户和计算机) 相对应的对象。您可以在 Amazon JupyterHub 上使用 [LDAPwith 的 JupyterHub身份验证器插件](#)EMR来LDAP进行用户身份验证。该插件处理LDAP用户的登录会话，并向 Jupyter 提供用户信息。这允许用户使用存储在LDAP兼容服务器中的身份凭据来连接 JupyterHub 和笔记本。

本节中的步骤将引导您完成以下步骤来设置和启LDAP用 LDAP Authenticator 插 JupyterHub件。在连接到主节点命令行时执行这些步骤。有关更多信息，请参阅[连接到主节点和笔记本服务器](#)。

1. 创建包含LDAP服务器相关信息的LDAP配置文件，例如主机 IP 地址、端口、绑定名称等。
2. 修改/etc/jupyter/conf/jupyterhub\_config.py为启用LDAP身份验证器插件。 JupyterHub
3. 创建并运行在jupyterhub容器LDAP内配置脚本。
4. LDAP查询用户，然后在容器中为每个用户创建主目录。 JupyterHub 需要主目录来托管笔记本电脑。
5. 运行可重新启动的脚本 JupyterHub

### ⚠ Important

在设置之前LDAP，请测试您的网络基础架构，以确保LDAP服务器和群集主节点可以根据需要进行通信。TLS通常通过普通TCP连接使用端口 389。如果您的LDAP连接使用SSL，则众所周知的TCP端口SSL是 636。

## 创建LDAP配置文件

下方的示例使用以下占位符配置值。将这些值替换为与您的实施匹配的参数。

- LDAP服务器运行的是版本 3，可在端口 389 上使用。这是的标准非SSL端口LDAP。
- 基本可分辨名称 (DN) 为 `dc=example, dc=org`。

使用文本编辑器创建内容与下类似的 [ldap.conf](#) 文件。使用适合您的LDAP实施的值。Replace ( 替换 ) *host* 使用LDAP服务器的 IP 地址或可解析的主机名。

```
base dc=example,dc=org
uri ldap://host
ldap_version 3
binddn cn=admin,dc=example,dc=org
bindpw admin
```

## 启用LDAP身份验证器插件 JupyterHub

使用文本编辑器修改 `/etc/jupyter/conf/jupyterhub_config.py` 文件并添加与下类似的 [ldapauthenticator](#) 属性。Replace ( 替换 ) *host* 使用LDAP服务器的 IP 地址或可解析的主机名。该示例假设用户对象位于名为的组织单位 (ou) 中 *people*，并使用您之前使用建立的可分辨名称组件 `ldap.conf`。

```
c.JupyterHub.authenticator_class = 'ldapauthenticator.LDAPAuthenticator'
c.LDAPAuthenticator.use_ssl = False
c.LDAPAuthenticator.server_address = 'host'
c.LDAPAuthenticator.bind_dn_template = 'cn={username},ou=people,dc=example,dc=org'
```

## 在容器LDAP内配置

使用文本编辑器创建包含以下内容的清除脚本：

```
#!/bin/bash
```

```
# Uncomment the following lines to install LDAP client libraries only if
# using Amazon EMR release version 5.14.0. Later versions install libraries by default.
# sudo docker exec jupyterhub bash -c "sudo apt-get update"
# sudo docker exec jupyterhub bash -c "sudo apt-get -y install libnss-ldap libpam-ldap
  ldap-utils nscd"

# Copy ldap.conf
sudo docker cp ldap.conf jupyterhub:/etc/ldap/
sudo docker exec jupyterhub bash -c "cat /etc/ldap/ldap.conf"

# configure nss switch
sudo docker exec jupyterhub bash -c "sed -i 's/\(^passwd.*\)/\1 ldap/g' /etc/
nsswitch.conf"
sudo docker exec jupyterhub bash -c "sed -i 's/\(^group.*\)/\1 ldap/g' /etc/
nsswitch.conf"
sudo docker exec jupyterhub bash -c "sed -i 's/\(^shadow.*\)/\1 ldap/g' /etc/
nsswitch.conf"
sudo docker exec jupyterhub bash -c "cat /etc/nsswitch.conf"

# configure PAM to create home directories
sudo docker exec jupyterhub bash -c "echo 'session required          pam_mkhomedir.so
  skel=/etc/skel umask=077' >> /etc/pam.d/common-session"
sudo docker exec jupyterhub bash -c "cat /etc/pam.d/common-session"

# restart nscd service
sudo docker exec jupyterhub bash -c "sudo service nscd restart"

# Test
sudo docker exec jupyterhub bash -c "getent passwd"

# Install ldap plugin
sudo docker exec jupyterhub bash -c "pip install jupyterhub-ldapauthenticator"
```

将脚本保存到主节点，然后从主节点命令行运行它。例如，对于另存为 `configure_ldap_client.sh` 的脚本，使此文件成为可执行文件：

```
chmod +x configure_ldap_client.sh
```

并运行此脚本：

```
./configure_ldap_client.sh
```

## 将属性添加到 Active Directory

要查找每个用户并在数据库中创建相应的条目，JupyterHub docker 容器需要 Active Directory 中相应用户对象的以下属性。有关更多信息，请参阅“既然 Unix 属性插件不再适用于 Active Directory 用户和计算机 MMC”管理单元，我该如何继续编辑 GID/UID RFC2307 属性一节？在文章《[澄清 Unix 的身份管理状态](#)》(IDMU) 和 [Windows Server 2016 技术预览版及以后的版本中的 NIS 服务器角色](#)中。

- homeDirectory

这是用户主目录的位置，通常是 `/home/username`。

- gidNumber

这是一个大于 60000 的值，尚未被其它用户使用。检查 `etc/passwd` 文件中是否有正在使用的 GID。

- uidNumber

这是一个大于 60000 的值，尚未被其它组使用。检查 `etc/group` 文件中是否有正在使用的 UID。

- uid

这与 `username`。

### 创建用户主目录

JupyterHub 需要容器内的主目录来对 LDAP 用户进行身份验证和存储实例数据。以下示例演示了 LDAP 目录中的两个用户 `shirley` 和 `diego`。

第一步是使用 `ldapsearch` 向 LDAP 服务器查询每个用户的用户 ID 和群组 ID 信息，如以下示例所示，替换 `host` 使用 LDAP 服务器的 IP 地址或可解析的主机名：

```
ldapsearch -x -H ldap://host \  
-D "cn=admin,dc=example,dc=org" \  
-w admin \  
-b "ou=people,dc=example,dc=org" \  
-s sub \  
"(objectclass=*)" uidNumber gidNumber
```

对于用户 `shirley` 和 `diego`，该 `ldapsearch` 命令返回的响应与以下内容类似。

```
# extended LDIF
```

```
# LDAPv3
# base <ou=people,dc=example,dc=org> with scope subtree
# filter: (objectclass=*)
# requesting: uidNumber gidNumber sn

# people, example.org
dn: ou=people,dc=example,dc=org

# diego, people, example.org
dn: cn=diego,ou=people,dc=example,dc=org
sn: B
uidNumber: 1001
gidNumber: 100

# shirley, people, example.org
dn: cn=shirley,ou=people,dc=example,dc=org
sn: A
uidNumber: 1002
gidNumber: 100

# search result
search: 2
result: 0 Success

# numResponses: 4
# numEntries: 3
```

通过使用响应中的信息，在容器内运行命令以为每个用户公用名 (cn) 创建一个主目录。使用 `uidNumber` 和 `gidNumber` 确定用户对主目录的所有权。以下示例命令可为用户执行此操作 *shirley*。

```
sudo docker container exec jupyterhub bash -c "mkdir /home/shirley"
sudo docker container exec jupyterhub bash -c "chown -R $uidNumber /home/shirley"
sudo docker container exec jupyterhub bash -c "sudo chgrp -R $gidNumber /home/shirley"
```

### Note

LDAP的身份验证器 JupyterHub 不支持创建本地用户。有关更多信息，请参阅[有关创建本地用户的LDAP身份验证器配置说明](#)。

要手动创建本地用户，请使用以下命令。



```
sudo docker exec jupyterhub bash -c "echo 'shirley:x:$uidNumber:$gidNumber::/home/shirley:/bin/bash' >> /etc/passwd"
```

## 重启 JupyterHub 容器

运行以下命令重新启动 jupyterhub 容器：

```
sudo docker stop jupyterhub
sudo docker start jupyterhub
```

## 用户模拟

在 Jupyter 笔记本中运行的 Spark 作业在亚马逊上执行期间会遍历多个应用程序。EMR 例如，Sparkmagic 接收到用户在 Jupyter 中运行的 PySpark 3 个代码，Sparkmagic 使用 HTTPPOST 请求将其提交给 Livy，然后使用创建一个 Spark 作业在集群上执行。YARN

默认情况下，以这种方式提交的 YARN 作业都以用户身份运行 livy，无论启动作业的用户是谁。通过设置用户模拟，您可以将笔记本用户的用户 ID 也设置为与作业关联的用户。YARN 每个用户启动的作业分别与 shirley 和 diego 相关联，而不是由与 livy 用户关联的 shirley 和 diego 同时启动作业。这有助于审核 Jupyter 使用情况以及在组织中管理应用程序。

只有在从 Sparkmagic 到 Livy 的调用未进行身份验证时，才支持该配置。不支持在 Hadoop 应用程序和 Livy 之间提供身份验证或代理层的应用程序（如 Apache Knox Gateway）。本节中配置用户模拟的步骤假设 JupyterHub 和 Livy 在同一个主节点上运行。如果您的应用程序有单独的集群，则[步骤 3：为用户创建 HDFS 主目录](#)需要进行修改，以便在 Livy 主节点上创建 HDFS 目录。

### 配置用户模拟的步骤

- [步骤 1：配置 Livy](#)
- [步骤 2：添加用户](#)
- [步骤 3：为用户创建 HDFS 主目录](#)

### 步骤 1：配置 Livy

在创建集群时，您可以使用 livy-conf 和 core-site 配置分类启用 Livy 用户模拟，如以下示例所示。将配置分类另存为 JSON 然后在创建集群时引用它，或者以内联方式指定配置分类。有关更多信息，请参阅[配置应用程序](#)。

```
[
  {
    "Classification": "livy-conf",
    "Properties": {
      "livy.impersonation.enabled": "true"
    }
  },
  {
    "Classification": "core-site",
    "Properties": {
      "hadoop.proxyuser.livy.groups": "*",
      "hadoop.proxyuser.livy.hosts": "*"
    }
  }
]
```

## 步骤 2：添加用户

使用PAM或添加 JupyterHub 用户LDAP。有关更多信息，请参阅 [使用 PAM 身份验证](#) 和 [使用 LDAP 身份验证](#)。

## 步骤 3：为用户创建HDFS主目录

您已连接到主节点以创建用户。在仍连接到主节点时，复制以下内容并将其保存到脚本文件中。该脚本为主节点上的每个 JupyterHub 用户创建HDFS主目录。该脚本假设您使用的是默认管理员用户 ID，*jovyan*。

```
#!/bin/bash

CURL="curl --silent -k"
HOST=$(curl -s http://169.254.169.254/latest/meta-data/local-hostname)

admin_token() {
  local user=jovyan
  local pwd=jupyter
  local token=$(($CURL https://$HOST:9443/hub/api/authorizations/token \
    -d "{\"username\":\"$user\", \"password\":\"$pwd\"}" | jq ".token")
  if [[ $token != null ]]; then
    token=$(echo $token | sed 's/"//g')
  else
    echo "Unable to get Jupyter API Token."
    exit 1
  fi
}
```

```
fi
echo $token
}

# Get Jupyter Admin token
token=$(admin_token)

# Get list of Jupyter users
users=$(curl -XGET -s -k https://$HOST:9443/hub/api/users \
-H "Authorization: token $token" | jq '.[].name' | sed 's/"//g')

# Create HDFS home dir
for user in ${users[@]};
do
echo "Create hdfs home dir for $user"
hadoop fs -mkdir /user/$user
hadoop fs -chmod 777 /user/$user
done
```

## 安装其它内核和库

当你 JupyterHub 在亚马逊上创建集群时EMR，Jupyter 的默认 Python 3 内核以及 Sparkmagic 的 Spark 内核将 PySpark 安装在 Docker 容器上。可以安装其它内核。还可以安装其它库和软件包，然后将它们导入相应的 shell。

### 安装内核

内核安装在 Docker 容器中。安装内核最简单的方式是，创建包含安装命令的清除脚本，将脚本保存到主节点，然后使用 `sudo docker exec jupyterhub script_name` 命令以在 jupyterhub 容器内运行脚本。以下示例脚本安装内核，然后在主节点上安装内核的一些库，以便之后在 Jupyter 中使用内核时可以导出库。

```
#!/bin/bash

# Install Python 2 kernel
conda create -n py27 python=2.7 anaconda
source /opt/conda/envs/py27/bin/activate
apt-get update
apt-get install -y gcc
/opt/conda/envs/py27/bin/python -m pip install --upgrade ipykernel
/opt/conda/envs/py27/bin/python -m ipykernel install
```

```
# Install libraries for Python 2
/opt/conda/envs/py27/bin/pip install paramiko nltk scipy numpy scikit-learn pandas
```

要在容器内安装内核和库，请打开至主节点的终端连接，将脚本保存到 `/etc/jupyter/install_kernels.sh`，然后在主节点命令行上运行以下命令：

```
sudo docker exec jupyterhub bash /etc/jupyter/install_kernels.sh
```

## 使用库和安装其它库

Ama EMR zon JupyterHub 上预装了一组适用于 Python 3 的核心机器学习和数据科学库。可以使用 `sudo docker exec jupyterhub bash -c "conda list"` 和 `sudo docker exec jupyterhub bash -c "pip freeze"`。

如果 Spark 作业需要 Worker 节点上的库，建议使用引导操作运行脚本以在创建集群时安装库。集群创建过程中，引导操作将在所有集群节点上运行，这将简化安装。如果于集群运行后在核心/Worker 节点上安装库，则操作更复杂。我们在此部分中提供了示例 Python 程序以演示如何安装这些库。

此部分中演示的引导操作和 Python 程序示例都使用保存到 Amazon S3 的清除脚本在所有节点上安装库。

以下示例中引用的脚本将通过 pip 安装适用于 Python 3 内核的 paramiko、nltk、scipy、scikit-learn 和 pandas：

```
#!/bin/bash

sudo python3 -m pip install boto3 paramiko nltk scipy scikit-learn pandas
```

创建脚本后，将其上传到 Amazon S3 中的位置（例如，`s3://mybucket/install-my-jupyter-libraries.sh`）。有关更多信息，请参阅《Amazon Simple Storage Service 用户指南》中的[上传对象](#)，以便可以在引导操作或 Python 程序中使用此操作。

要指定在创建集群时在所有节点上安装库的引导操作，请使用 AWS CLI

1. 创建与之前的示例类似的脚本并将脚本保存在 Amazon S3 中的位置。我们将使用示例 `s3://mybucket/install-my-jupyter-libraries.sh`。
2. 使用选项创建集群 JupyterHub 并使用 `--bootstrap-actions` 选项的 Path 参数来指定脚本位置，如以下示例所示：

**Note**

为了便于读取，包含 Linux 行继续符 (\)。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (^)。

```
aws emr create-cluster --name="MyJupyterHubCluster" --release-label emr-5.36.2 \
--applications Name=JupyterHub --log-uri s3://MyBucket/MyJupyterClusterLogs \
--use-default-roles --instance-type m5.xlarge --instance-count 2 --ec2-attributes
KeyName=MyKeyPair \
--bootstrap-actions Path=s3://mybucket/install-my-jupyter-
libraries.sh,Name=InstallJupyterLibs
```

指定将在使用控制台创建集群时在所有节点上安装库的引导操作

1. 导航到新的 Amazon EMR 控制台，然后从侧面导航栏中选择“切换到旧主机”。有关切换到旧控制台后预期情况的更多信息，请参阅 [Using the old console](#)。
2. 依次选择 Create cluster (创建集群)、Go to advanced options (转到高级选项)。
3. 根据应用程序的情况，指定 Software and Steps (软件和步骤) 和 Hardware (硬件) 的设置。
4. 在 General Cluster Settings (常规集群设置) 屏幕上，展开 Bootstrap Actions (引导操作)。
5. 对于 Add bootstrap action (添加引导操作)，选择 Custom action (自定义操作)、Configure and add (配置和添加)。
6. 对于名称，输入一个易于理解的名称。对于脚本位置，请输入脚本在 Amazon S3 中的位置 (我们使用的示例是 s3://mybucket/ install-my-jupyter-libraries .sh )。保留 Optional arguments (可选参数) 为空，然后选择 Add (添加)。
7. 指定集群的其它设置，然后选择 Next (下一步)。
8. 指定安全设置，然后选择 Create cluster (创建集群)。

Example 在运行集群的核心节点上安装库

在 Jupyter 内的主节点上安装库之后，可以通过不同的方式将库安装在运行的核心节点上。以下示例显示了编写为在本地计算机上运行的 Python 程序。当您在本地运行 Python 程序时，它会使用 of AWS Systems Manager 来运行示例脚本，如本节前面所示，该脚本将在集群的核心节点上安装库。AWS-RunShellScript

```
import argparse
import time
import boto3

def install_libraries_on_core_nodes(cluster_id, script_path, emr_client, ssm_client):
    """
    Copies and runs a shell script on the core nodes in the cluster.

    :param cluster_id: The ID of the cluster.
    :param script_path: The path to the script, typically an Amazon S3 object URL.
    :param emr_client: The Boto3 Amazon EMR client.
    :param ssm_client: The Boto3 AWS Systems Manager client.
    """
    core_nodes = emr_client.list_instances(
        ClusterId=cluster_id, InstanceGroupTypes=["CORE"]
    )["Instances"]
    core_instance_ids = [node["Ec2InstanceId"] for node in core_nodes]
    print(f"Found core instances: {core_instance_ids}.")

    commands = [
        # Copy the shell script from Amazon S3 to each node instance.
        f"aws s3 cp {script_path} /home/hadoop",
        # Run the shell script to install libraries on each node instance.
        "bash /home/hadoop/install_libraries.sh",
    ]
    for command in commands:
        print(f"Sending '{command}' to core instances...")
        command_id = ssm_client.send_command(
            InstanceIds=core_instance_ids,
            DocumentName="AWS-RunShellScript",
            Parameters={"commands": [command]},
            TimeoutSeconds=3600,
        )["Command"]["CommandId"]
        while True:
            # Verify the previous step succeeded before running the next step.
            cmd_result = ssm_client.list_commands(CommandId=command_id)["Commands"][0]
            if cmd_result["StatusDetails"] == "Success":
                print(f"Command succeeded.")
                break
            elif cmd_result["StatusDetails"] in ["Pending", "InProgress"]:
                print(f"Command status is {cmd_result['StatusDetails']}, waiting...")
                time.sleep(10)
```

```

        else:
            print(f"Command status is {cmd_result['StatusDetails']}, quitting.")
            raise RuntimeError(
                f"Command {command} failed to run. "
                f"Details: {cmd_result['StatusDetails']}"
            )

def main():
    parser = argparse.ArgumentParser()
    parser.add_argument("cluster_id", help="The ID of the cluster.")
    parser.add_argument("script_path", help="The path to the script in Amazon S3.")
    args = parser.parse_args()

    emr_client = boto3.client("emr")
    ssm_client = boto3.client("ssm")

    install_libraries_on_core_nodes(
        args.cluster_id, args.script_path, emr_client, ssm_client
    )

if __name__ == "__main__":
    main()

```

## JupyterHub 发布历史

下表列出了 Amazon 的每个发行版本中 JupyterHub 包含的版本EMR，以及随应用程序一起安装的组件。有关每个发行版本中的组件版本，请参阅 [亚马逊 EMR 7.x 发布版本](#)、[亚马逊 EMR 6.x 发布版本](#) 或 [亚马逊 EMR 5.x 发布版本](#) 中的发行版“组件版本”部分。

### JupyterHub 版本信息

亚马逊EMR发布标签	JupyterHub 版本	安装的组件 JupyterHub
emr-7.2.0	1.5.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager,

亚马逊EMR发布标签	JupyterHub 版本	安装的组件 JupyterHub
		hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.36.2	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-7.1.0	1.5.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub



亚马逊EMR发布标签	JupyterHub 版本	安装的组件 JupyterHub
emr-7.0.0	1.5.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.15.0	1.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

亚马逊EMR发布标签	JupyterHub 版本	安装的组件 JupyterHub
emr-6.14.0	1.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.13.0	1.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

亚马逊EMR发布标签	JupyterHub 版本	安装的组件 JupyterHub
emr-6.12.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.11.1	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

亚马逊EMR发布标签	JupyterHub 版本	安装的组件 JupyterHub
emr-6.11.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.10.1	1.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

亚马逊EMR发布标签	JupyterHub 版本	安装的组件 JupyterHub
emr-6.10.0	1.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.9.1	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

亚马逊EMR发布标签	JupyterHub 版本	安装的组件 JupyterHub
emr-6.9.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.8.1	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

亚马逊EMR发布标签	JupyterHub 版本	安装的组件 JupyterHub
emr-6.8.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.7.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

亚马逊EMR发布标签	JupyterHub 版本	安装的组件 JupyterHub
emr-5.36.1	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.36.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub



亚马逊EMR发布标签	JupyterHub 版本	安装的组件 JupyterHub
emr-6.6.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.35.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

亚马逊EMR发布标签	JupyterHub 版本	安装的组件 JupyterHub
emr-6.5.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.4.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

亚马逊EMR发布标签	JupyterHub 版本	安装的组件 JupyterHub
emr-6.3.1	1.2.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.3.0	1.2.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

亚马逊EMR发布标签	JupyterHub 版本	安装的组件 JupyterHub
emr-6.2.1	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.2.0	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

亚马逊EMR发布标签	JupyterHub 版本	安装的组件 JupyterHub
emr-6.1.1	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.1.0	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

亚马逊EMR发布标签	JupyterHub 版本	安装的组件 JupyterHub
emr-6.0.1	1.0.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.0.0	1.0.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

亚马逊EMR发布标签	JupyterHub 版本	安装的组件 JupyterHub
emr-5.34.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.33.1	1.2.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

亚马逊EMR发布标签	JupyterHub 版本	安装的组件 JupyterHub
emr-5.33.0	1.2.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.32.1	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub



亚马逊EMR发布标签	JupyterHub 版本	安装的组件 JupyterHub
emr-5.32.0	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.31.1	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

亚马逊EMR发布标签	JupyterHub 版本	安装的组件 JupyterHub
emr-5.31.0	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.30.2	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

亚马逊EMR发布标签	JupyterHub 版本	安装的组件 JupyterHub
emr-5.30.1	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.30.0	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

亚马逊EMR发布标签	JupyterHub 版本	安装的组件 JupyterHub
emr-5.29.0	1.0.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.28.1	1.0.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

亚马逊EMR发布标签	JupyterHub 版本	安装的组件 JupyterHub
emr-5.28.0	1.0.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.27.1	1.0.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

亚马逊EMR发布标签	JupyterHub 版本	安装的组件 JupyterHub
emr-5.27.0	1.0.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.26.0	0.9.6	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

亚马逊EMR发布标签	JupyterHub 版本	安装的组件 JupyterHub
emr-5.25.0	0.9.6	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.24.1	0.9.6	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

亚马逊EMR发布标签	JupyterHub 版本	安装的组件 JupyterHub
emr-5.24.0	0.9.6	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.23.1	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub



亚马逊EMR发布标签	JupyterHub 版本	安装的组件 JupyterHub
emr-5.23.0	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.22.0	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

亚马逊EMR发布标签	JupyterHub 版本	安装的组件 JupyterHub
emr-5.21.2	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.21.1	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

亚马逊EMR发布标签	JupyterHub 版本	安装的组件 JupyterHub
emr-5.21.0	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.20.1	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

亚马逊EMR发布标签	JupyterHub 版本	安装的组件 JupyterHub
emr-5.20.0	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.19.1	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

亚马逊EMR发布标签	JupyterHub 版本	安装的组件 JupyterHub
emr-5.19.0	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.18.1	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

亚马逊EMR发布标签	JupyterHub 版本	安装的组件 JupyterHub
emr-5.18.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.17.2	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

亚马逊EMR发布标签	JupyterHub 版本	安装的组件 JupyterHub
emr-5.17.1	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.17.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

亚马逊EMR发布标签	JupyterHub 版本	安装的组件 JupyterHub
emr-5.16.1	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.16.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub



亚马逊EMR发布标签	JupyterHub 版本	安装的组件 JupyterHub
emr-5.15.1	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.15.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

亚马逊EMR发布标签	JupyterHub 版本	安装的组件 JupyterHub
emr-5.14.2	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.14.1	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

亚马逊EMR发布标签	JupyterHub 版本	安装的组件 JupyterHub
emr-5.14.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

# Apache Livy

Livy 支持通过REST接口与运行 Spark 的EMR集群进行交互。您可以使用REST界面或RPC客户端库提交 Spark 作业或 Spark 代码片段、同步或异步检索结果以及管理 Spark Context。有关更多信息，请参阅 [Apache Livy 网站](#)。Livy 包含在亚马逊EMR发行版 5.9.0 及更高版本中。

要访问 Livy Web 界面，请设置SSH通往主节点的隧道和代理连接。有关更多信息，请参阅[查看EMR集群上托管的 Web 界面](#)。

下表列出了最新版本的亚马逊 EMR 7.x 系列中包含的 Livy 版本，以及亚马逊与 Livy 一起EMR安装的组件。

有关此版本中与 Livy 一起安装的组件版本，请参阅 [7.2.0 版本组件版本](#)。

emr-7.2.0 的 Livy 版本信息

亚马逊EMR发行标签	Livy 版本	随 Livy 安装的组件
emr-7.2.0	Livy 0.8.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

下表列出了最新版本的亚马逊 EMR 6.x 系列中包含的 Livy 版本，以及亚马逊与 Livy 一起EMR安装的组件。

有关此发行版中随 Livy 安装的组件版本，请参阅[发行版 6.15.0 组件版本](#)。

## emr-6.15.0 的 Livy 版本信息

亚马逊EMR发行标签	Livy 版本	随 Livy 安装的组件
emr-6.15.0	Livy 0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

下表列出了最新版本的亚马逊 EMR 5.x 系列中包含的 Livy 版本，以及亚马逊与 Livy 一起EMR安装的组件。

有关此版本中与 Livy 一起安装的组件的版本，请参阅 [5.36.2 版组件版本](#)。

## emr-5.36.2 的 Livy 版本信息

亚马逊EMR发行标签	Livy 版本	随 Livy 安装的组件
emr-5.36.2	Livy 0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server,

亚马逊EMR发行标签	Livy 版本	随 Livy 安装的组件
		spark-on-yarn, spark-yarn-slave, livy-server, nginx

## 主题

- [HTTPS使用 Apache Livy 启用](#)
- [Livy 发行历史记录](#)

## HTTPS使用 Apache Livy 启用

1. 配置启用传输加密的 Amazon EMR 集群。要了解有关加密的更多信息，请参阅[加密静态数据和传输中的数据](#)。
2. 创建以下内容的名为 `livy_ssl.sh` 的文件。

```
#!/bin/bash

KEYSTORE_FILE=`awk '/ssl.server.keystore.location/{getline; print}' /etc/hadoop/conf/ssl-server.xml | sed -e 's/<[^>]*>/g' | tr -d ' \t\n\r\f'`
KEYSTORE_PASS=`awk '/ssl.server.keystore.password/{getline; print}' /etc/hadoop/conf/ssl-server.xml | sed -e 's/<[^>]*>/g' | tr -d ' \t\n\r\f'`
KEY_PASS=`awk '/ssl.server.keystore.keypassword/{getline; print}' /etc/hadoop/conf/ssl-server.xml | sed -e 's/<[^>]*>/g' | tr -d ' \t\n\r\f'`

echo "livy.keystore $KEYSTORE_FILE
livy.keystore.password $KEYSTORE_PASS
livy.key-password $KEY_PASS" | sudo tee -a /etc/livy/conf/livy.conf >/dev/null

sudo systemctl restart livy-server.service
```

3. 将以下脚本作为 Amazon EMR 步骤运行。此脚本修改 `/etc/livy/conf/livy.conf` 为激活 SSL。

```
--steps '[{"Args":["s3://DOC-EXAMPLE-BUCKET/livy_ssl.sh"],"Type":"CUSTOM_JAR","ActionOnFailure":"CONTINUE","Jar":"s3://us-east-1.elasticmapreduce/libs/script-runner/script-runner.jar","Properties":"","Name":"Custom JAR"}]'
```

- 重新启动 Apache Livy 服务，以使更改生效。要重新启动 Apache Livy，请参阅[停止和重新启动进程](#)。
- 测试客户端现在是否可以使用进行通信HTTPS。例如，要提交任务，请运行以下代码。

```
curl -k -X POST --data '{"file": "local:///usr/lib/spark/examples/jars/spark-examples.jar",
"className": "org.apache.spark.examples.SparkPi"}' \
-H "Content-Type: application/json" \
https://EMR_Master_Node_Host:8998/batches
```

如果您已HTTPS成功启用，Livy 会发送一个响应，表明该命令已被接受并且批处理作业已提交。

```
{"id":1,"name":null,"owner":null,"proxyUser":null,"state":"starting","appId":null,"appInfo":{"driverLogUrl":null,"sparkUiUrl":null},"log":["stdout: ", "\nstderr: ", "\nYARN Diagnostics: "]}
```

## Livy 发行历史记录

下表列出了 Amazon 的每个发行版本中包含的 Livy 版本EMR，以及随应用程序一起安装的组件。有关每个发行版本中的组件版本，请参阅[亚马逊 EMR 7.x 发布版本](#)、[亚马逊 EMR 6.x 发布版本](#) 或 [亚马逊 EMR 5.x 发布版本](#) 中的发行版“组件版本”部分。

### Livy 版本信息

Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-7.2.0	0.8.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server,

Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
		spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.36.2	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-7.1.0	0.8.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx



Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-7.0.0	0.7.1	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.15.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-6.14.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.13.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-6.12.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.11.1	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-6.11.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.10.1	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-6.10.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.9.1	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-6.9.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.8.1	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-6.8.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.7.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-5.36.1	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.36.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx



Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-6.6.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.35.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-6.5.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.4.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-6.3.1	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.3.0	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-6.2.1	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.2.0	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-6.1.1	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.1.0	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-6.0.1	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.0.0	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-5.34.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.33.1	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-5.33.0	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.32.1	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx



Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-5.32.0	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.31.1	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-5.31.0	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.30.2	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-5.30.1	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.30.0	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-notebook-env, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-5.29.0	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.28.1	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-5.28.0	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.27.1	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-5.27.0	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.26.0	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-5.25.0	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.24.1	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-5.24.0	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.23.1	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx



Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-5.23.0	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.22.0	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-5.21.2	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.21.1	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-5.21.0	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.20.1	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-5.20.0	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.19.1	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-5.19.0	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.18.1	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-5.18.0	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.17.2	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-5.17.1	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.17.0	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-5.16.1	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.16.0	0.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server



Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-5.15.1	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.15.0	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-5.14.2	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.14.1	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-5.14.0	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.13.1	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-5.13.0	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.12.3	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-5.12.2	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.12.1	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-5.12.0	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.11.4	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-5.11.3	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.11.2	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-5.11.1	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.11.0	0.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server



Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-5.10.1	0.4.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.10.0	0.4.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.9.1	0.4.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Amazon EMR 发布标签	Livy 版本	随 Livy 安装的组件
emr-5.9.0	0.4.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

# Apache MXNet

Apache MXNet 是一个加速库，专为构建神经网络和其他深度学习应用程序而设计。MXNet自动执行常见工作流程并优化数值计算。MXNet帮助您设计神经网络架构，而不必专注于实现低级计算，例如线性代数运算。MXNet包含在 Amazon EMR 发行版 5.10.0 及更高版本中。

有关更多信息，请参阅 [Apache MXNet 网站](#)。

下表列出了最新版本的 Amazon EMR 7.x 系列中MXNet包含的版本以及亚马逊随之EMR安装的组件。MXNet

有关此版本MXNet中安装的组件的版本，请参阅 [7.2.0 版组件版本](#)。

MXNetemr-7.2.0 的版本信息

亚马逊EMR发行标签	MXNet版本	随 MXNet 安装的组件
emr-7.2.0	MXNet 1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

下表列出了最新版本的 Amazon EMR 6.x 系列中MXNet包含的版本以及亚马逊随之EMR安装的组件。MXNet

有关此版本MXNet中安装的组件的版本，请参阅 [6.15.0 版组件版本](#)。

MXNetemr-6.15.0 的版本信息

亚马逊EMR发行标签	MXNet版本	随 MXNet 安装的组件
emr-6.15.0	MXNet 1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode,

亚马逊EMR发行标签	MXNet版本	随 MXNet 安装的组件
		hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

下表列出了最新版本的 Amazon EMR 5.x 系列中MXNet包含的版本以及亚马逊随之EMR安装的组件。MXNet

有关此版本MXNet中安装的组件的版本，请参阅 [5.36.2 版组件版本](#)。

MXNetemr-5.36.2 的版本信息

亚马逊EMR发行标签	MXNet版本	随 MXNet 安装的组件
emr-5.36.2	MXNet 1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

## MXNet发布历史

下表列出了 Amazon 的每个发行版本中MXNet包含的版本EMR，以及随应用程序一起安装的组件。有关每个发行版本中的组件版本，请参阅 [亚马逊 EMR 7.x 发布版本](#)、[亚马逊 EMR 6.x 发布版本](#) 或 [亚马逊 EMR 5.x 发布版本](#) 中的发行版“组件版本”部分。

## MXNet版本信息

亚马逊EMR发布标签	MXNet版本	安装的组件 MXNet
emr-7.2.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.36.2	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-7.1.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-7.0.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-

亚马逊EMR发布标签	MXNet版本	安装的组件 MXNet
		httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.15.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.14.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.13.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

亚马逊EMR发布标签	MXNet版本	安装的组件 MXNet
emr-6.12.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.11.1	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.11.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

亚马逊EMR发布标签	MXNet版本	安装的组件 MXNet
emr-6.10.1	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.10.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.9.1	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv



亚马逊EMR发布标签	MXNet版本	安装的组件 MXNet
emr-6.9.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.8.1	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.8.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv

亚马逊EMR发布标签	MXNet版本	安装的组件 MXNet
emr-6.7.0	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.36.1	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.36.0	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

亚马逊EMR发布标签	MXNet版本	安装的组件 MXNet
emr-6.6.0	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.35.0	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.5.0	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

亚马逊EMR发布标签	MXNet版本	安装的组件 MXNet
emr-6.4.0	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.3.1	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.3.0	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

亚马逊EMR发布标签	MXNet版本	安装的组件 MXNet
emr-6.2.1	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.2.0	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.1.1	1.6.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv

亚马逊EMR发布标签	MXNet版本	安装的组件 MXNet
emr-6.1.0	1.6.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.0.1	1.5.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.0.0	1.5.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

亚马逊EMR发布标签	MXNet版本	安装的组件 MXNet
emr-5.34.0	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.33.1	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.33.0	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv

亚马逊EMR发布标签	MXNet版本	安装的组件 MXNet
emr-5.32.1	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.32.0	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.31.1	1.6.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv



亚马逊EMR发布标签	MXNet版本	安装的组件 MXNet
emr-5.31.0	1.6.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.30.2	1.5.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.30.1	1.5.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

亚马逊EMR发布标签	MXNet版本	安装的组件 MXNet
emr-5.30.0	1.5.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.29.0	1.5.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.28.1	1.5.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

亚马逊EMR发布标签	MXNet版本	安装的组件 MXNet
emr-5.28.0	1.5.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.27.1	1.4.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.27.0	1.4.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv

亚马逊EMR发布标签	MXNet版本	安装的组件 MXNet
emr-5.26.0	1.4.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.25.0	1.4.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.24.1	1.4.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

亚马逊EMR发布标签	MXNet版本	安装的组件 MXNet
emr-5.24.0	1.4.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.23.1	1.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.23.0	1.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

亚马逊EMR发布标签	MXNet版本	安装的组件 MXNet
emr-5.22.0	1.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.21.2	1.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.21.1	1.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

亚马逊EMR发布标签	MXNet版本	安装的组件 MXNet
emr-5.21.0	1.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.20.1	1.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.20.0	1.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

亚马逊EMR发布标签	MXNet版本	安装的组件 MXNet
emr-5.19.1	1.3.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.19.0	1.3.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.18.1	1.2.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv



亚马逊EMR发布标签	MXNet版本	安装的组件 MXNet
emr-5.18.0	1.2.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.17.2	1.2.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.17.1	1.2.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv

亚马逊EMR发布标签	MXNet版本	安装的组件 MXNet
emr-5.17.0	1.2.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.16.1	1.2.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.16.0	1.2.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

亚马逊EMR发布标签	MXNet版本	安装的组件 MXNet
emr-5.15.1	1.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.15.0	1.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.14.2	1.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

亚马逊EMR发布标签	MXNet版本	安装的组件 MXNet
emr-5.14.1	1.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.14.0	1.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.13.1	1.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet

亚马逊EMR发布标签	MXNet版本	安装的组件 MXNet
emr-5.13.0	1.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet
emr-5.12.3	1.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet
emr-5.12.2	1.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet

亚马逊EMR发布标签	MXNet版本	安装的组件 MXNet
emr-5.12.1	1.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet
emr-5.12.0	1.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet
emr-5.11.4	0.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet

亚马逊EMR发布标签	MXNet版本	安装的组件 MXNet
emr-5.11.3	0.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet
emr-5.11.2	0.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet
emr-5.11.1	0.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet

亚马逊EMR发布标签	MXNet版本	安装的组件 MXNet
emr-5.11.0	0.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet
emr-5.10.1	0.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet
emr-5.10.0	0.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet



# Apache Oozie

使用 Apache Oozie 工作流调度器管理和协调 Hadoop 任务。有关详细信息，请参阅 <http://oozie.apache.org/>。

亚马逊EMR不支持 Oozie 原生网页界面。如要使用 Oozie 的前端界面，请尝试 Hue Oozie 应用程序。有关更多信息，请参阅 [Hue](#)。Oozie 包含在亚马逊EMR发行版 5.0.0 及更高版本中。Oozie 作为沙盒应用程序包含在早期版本中。有关更多信息，请参阅 [亚马逊 EMR 4.x 发布版本](#)。

如果你使用AMI基于亚马逊 Linux 的自定义 Amazon LinuxAMI，创建日期为 2018-08-11，Oozie 服务器将无法启动。如果你使用 Oozie，请AMI根据创建日期不同的亚马逊 Linux AMI ID 创建自定义账号。您可以使用以下 AWS CLI 命令返回所有 HVM Amazon Linux AMIs 版本为 2018.03 的图片IDs列表以及发布日期，以便您可以选择合适的 Amazon Linux AMI 作为基础。MyRegion 替换为您的地区标识符，例如 us-west-2。

```
aws ec2 --region MyRegion describe-images --owner amazon --query 'Images[?Name!=`null`][?starts_with(Name, `amzn-ami-hvm-2018.03`) == `true`]. [CreationDate,ImageId,Name]' --output text | sort -rk1
```

下表列出了最新版本的亚马逊 EMR 7.x 系列中包含的 Oozie 版本，以及亚马逊与 Oozie 一起EMR安装的组件。

有关此版本中与 Oozie 一起安装的组件的版本，请参阅 [7.2.0 版本的组件](#)版本。

emr-7.2.0 的 Oozie 版本信息

亚马逊EMR发行标签	Oozie 版本	随 Oozie 安装的组件
emr-7.2.0	Oozie 5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client,

亚马逊EMR发行标签	Oozie 版本	随 Oozie 安装的组件
		oozie-server, tez-on-yarn, tez-on-worker

下表列出了最新版本的亚马逊 EMR 6.x 系列中包含的 Oozie 版本，以及亚马逊与 Oozie 一起EMR安装的组件。

有关此发行版中随 Oozie 安装的组件版本，请参阅[发行版 6.15.0 组件版本](#)。

#### emr-6.15.0 的 Oozie 版本信息

亚马逊EMR发行标签	Oozie 版本	随 Oozie 安装的组件
emr-6.15.0	Oozie 5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker

下表列出了最新版本的亚马逊 EMR 5.x 系列中包含的 Oozie 版本，以及亚马逊与 Oozie 一起EMR安装的组件。

有关此版本中与 Oozie 一起安装的组件的版本，请参阅[5.36.2 版组件版本](#)。

#### emr-5.36.2 的 Oozie 版本信息

亚马逊EMR发行标签	Oozie 版本	随 Oozie 安装的组件
emr-5.36.2	Oozie 5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-

亚马逊EMR发行标签	Oozie 版本	随 Oozie 安装的组件
		cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

## 主题

- [在亚马逊中将 Oozie 与远程数据库配合使用 RDS](#)
- [为 Oozie 配置 Java 版本](#)
- [Oozie 发行历史记录](#)

## 在亚马逊中将 Oozie 与远程数据库配合使用 RDS

默认情况下，Oozie 用户信息和查询历史记录存储在主节点的本地“我的SQL数据库”中。或者，您可以使用存储在 Amazon S3 中的配置和亚马逊关系数据库服务 (Amazon) 中的“我的SQL数据库”来创建一个或多个支持 Oozie 的集群。RDS 这使您可以保留 Oozie 创建的用户信息和查询历史记录，而无需保持您的 Amazon EMR 集群运行。我们建议使用 Amazon S3 服务器端加密来存储配置文件。

首先为 Oozie 创建远程数据库。

### 创建外部“我的SQL数据库”

1. 打开 Amazon RDS 控制台，网址为<https://console.aws.amazon.com/rds/>。
2. 选择 Launch a DB Instance (启动数据库实例)。
3. 选择“我的”，SQL 然后选择“选择”。
4. 保留默认选择“多可用区部署”和“预配置 IOPS 存储”，然后选择“下一步”。
5. 保留 Instance Specifications (实例规格) 的默认值，指定 Settings (设置)，然后选择 Next (下一步)。

- 在 Configure Advanced Settings (配置高级设置) 页面上，选择相应的安全组和数据库名称。您使用的安全组必须至少允许从集群的主节点对端口 3306 进行入站TCP访问。如果此时您尚未创建集群，则可以允许所有主机连接到端口 3306 并在启动集群之后调整安全组。选择 Launch DB Instance (启动数据库实例)。
- 在RDS控制面板中，选择实例，然后选择您刚刚创建的实例。当您的数据库可用时，记下数据库名称、用户名、密码和RDS实例主机名。您将在创建和配置集群时用到此信息。

要在启动集群时使用 Oozie 指定外部“我的SQL数据库” AWS CLI

要在使用启动集群时为 Oozie 指定外部“我的SQL数据库” AWS CLI，请使用您在创建RDS实例时记下的信息来配置配置oozie-site配置对象。

#### Note

您可以创建使用同一个外部数据库的多个集群，但是每个集群将共享查询历史记录和用户信息。

- 使用 AWS CLI，创建安装了 Oozie 的集群，使用您创建的外部数据库，并引用指定数据库属性的具有 Oozie 配置分类的配置文件。以下示例创建一个安装了 Oozie 的集群，引用了 Amazon S3 中的配置文件 myConfig.json，该文件指定数据库配置。

#### Note

为了便于读取，包含 Linux 行继续符 ( \ )。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 ( ^ )。

```
aws emr create-cluster --release-label emr-7.2.0 --applications Name=Oozie
Name=Spark Name=Hive \
--instance-type m5.xlarge --instance-count 3 \
--configurations https://s3.amazonaws.com/mybucket/myfolder/myConfig.json --use-
default-roles
```

下面显示的是 myConfig.json 文件的内容示例。Replace ( 替换 ) *JDBC URL*, *username*，以及 *password* 使用您的RDS实例的JDBCURL、用户名和密码。

**⚠ Important**

JDBCURL必须包含数据库名称作为后缀。例如，jdbc:mysql://xxxxxxxxx.us-east-1.rds.amazonaws.com:3306/dbname。oozie-external-db

```
[{
  "Classification": "oozie-site",
  "Properties": {
    "oozie.service.JPIService.jdbc.driver": "org.mariadb.jdbc.Driver",
    "oozie.service.JPIService.jdbc.url": "JDBC URL",

    "oozie.service.JPIService.jdbc.username": "username",
    "oozie.service.JPIService.jdbc.password": "password"
  },
  "Configurations": []
}]
```

## 为 Oozie 配置 Java 版本

Oozie 运行多个 Java 虚拟机 (JVM) 进程。本页说明如何为每个流程配置 Java 版本。

- Oozie Server：在 oozie-env 分类中设置 JAVA\_HOME，以更新 EmbeddedOozieServer 的 Java 版本。
- Oozie Launcher AM：Oozie Launcher AM 是一项单映射器 MR 作业，它调用相应的应用程序客户端库，例如 Hadoop 和 Hive。除非另有配置，否则 Oozie Launcher AM 的运行时版本与集群中 Hadoop 的 Java 运行时相同。EMR 要为 Oozie Launcher AM 配置 Java 运行时系统，请在作业的 workflow.xml 中设置以下属性：

```
<property>
  <name>mapred.child.env</name>
  <value>JAVA_HOME=/path/to/JAVA_HOME</value>
</property>
```

此属性可确保 Oozie 作业的 Oozie Launcher AM 在您指定的 Java 版本上运行，而不是在 Hadoop 中设置的 Java 版本上运行。

- 应用程序客户端可执行文件：由于 Oozie Launcher AM 默认调用应用程序客户端，因此客户端可执行文件的 Java 运行时系统与 Oozie Launcher AM 相同。
- 由 Oozie 作业启动的应用程序：除非另有说明，否则由 Oozie 作业启动 JVMs 的实际应用程序的运行时版本与集群中 Hadoop 的 Java 运行时相同。EMR 根据用于在 Oozie 作业中启动应用程序的 Oozie 工作流程操作的类型（Spark 或 Hive 操作），您可以在 Oozie 作业 JVMs 中更新实际应用程序的默认 Java 运行时。workflow.xml

## Oozie 发行历史记录

下表列出了 Amazon 的每个发行版本中包含的 Oozie 版本 EMR，以及随应用程序一起安装的组件。有关每个发行版本中的组件版本，请参阅 [亚马逊 EMR 7.x 发布版本](#)、[亚马逊 EMR 6.x 发布版本](#) 或 [亚马逊 EMR 5.x 发布版本](#) 中的发行版“组件版本”部分。

### Oozie 版本信息

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-7.2.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker
emr-5.36.2	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
		httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-7.1.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker
emr-7.0.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-6.15.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker
emr-6.14.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker



Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-6.13.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker
emr-6.12.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-6.11.1	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker
emr-6.11.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-6.10.1	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker
emr-6.10.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-6.9.1	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.9.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-6.8.1	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.8.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-6.7.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.36.1	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.36.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.6.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.35.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.5.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn



Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-6.4.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.3.1	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-6.3.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.2.1	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-6.2.0	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.1.1	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-6.1.0	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.0.1	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-6.0.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.34.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.33.1	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.33.0	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.32.1	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.32.0	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.31.1	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.31.0	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn



Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.30.2	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.30.1	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.30.0	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.29.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.28.1	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.28.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.27.1	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.27.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.26.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.25.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.24.1	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.24.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.23.1	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.23.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.22.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.21.2	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn



Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.21.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.21.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.20.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.20.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.19.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.19.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.18.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.18.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.17.2	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.17.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.17.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.16.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.16.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.15.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.15.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.14.2	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn



Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.14.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.14.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.13.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.13.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.12.3	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.12.2	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.12.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.12.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.11.4	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.11.3	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.11.2	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.11.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.11.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.10.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.10.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.9.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn



Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.9.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.8.3	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.8.2	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.8.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.8.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.7.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.7.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.6.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.6.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.5.4	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.5.3	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.5.2	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.5.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.5.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.4.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.4.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn



Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.3.2	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.3.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.3.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.2.3	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.2.2	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.2.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.2.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.1.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.1.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.0.3	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.0.2	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.0.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Amazon EMR 发布标签	Oozie 版本	随 Oozie 安装的组件
emr-5.0.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

## Oozie 按版本划分的发行说明

### 主题

- [亚马逊 EMR 7.1.0-Oozie 的变化](#)

### 亚马逊 EMR 7.1.0-Oozie 的变化

类型	描述
Upgrade	将 Netty 升级到 4.1.100.Final。
Upgrade	将 Jetty 升级到 9.4.53.v20231009。

# Apache Phoenix

Apache Phoenix 用于OLTP操作分析，允许你使用标准SQL查询和JDBC APIs使用 Apache HBase 后备存储。有关更多信息，请参阅 [Phoenix in 15 minutes or less](#)。Phoenix 包含在亚马逊EMR发行版 4.7.0 及更高版本中。

如果您从亚马逊的早期版本升级EMR到亚马逊EMR发行版 5.4.0 或更高版本并使用二级索引，请按照 [Apache Phoenix](#) 文档中的说明升级本地索引。Amazon 从hbase-site分类中EMR删除了所需的配置，但需要重新填充索引。支持在线和离线升级索引。在线升级为默认值，这意味着，在从版本 4.8.0 或更高版本的 Phoenix 客户端初始化时重新填充索引。要指定离线升级，请在phoenix-site分类中将`phoenix.client.localIndexUpgrade`配置设置为 `false`，然后SSH将配置设置为要运行的主节点`psql [zookeeper] -1`。

下表列出了最新版本的亚马逊 EMR 7.x 系列中包含的 Phoenix 版本，以及亚马逊与 Phoenix 一起EMR安装的组件。

有关此版本中与 Phoenix 一起安装的组件的版本，请参阅 [7.2.0 版本的组件版本](#)。

emr-7.2.0 的 Phoenix 版本信息

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-7.2.0	Phoenix 5.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-



Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
		server, zookeeper-client, zookeeper-server

下表列出了最新版本的亚马逊 EMR 6.x 系列中包含的 Phoenix 版本，以及亚马逊与 Phoenix 一起EMR 安装的组件。

有关此发行版中随 Phoenix 安装的组件版本，请参阅[发行版 6.15.0 组件版本](#)。

#### emr-6.15.0 的 Phoenix 版本信息

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-6.15.0	Phoenix 5.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

下表列出了最新版本的亚马逊 EMR 5.x 系列中包含的 Phoenix 版本，以及亚马逊与 Phoenix 一起EMR 安装的组件。

有关此版本中与 Phoenix 一起安装的组件的版本，请参阅[5.36.2 版组件版本](#)。

## emr-5.36.2 的 Phoenix 版本信息

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.36.2	Phoenix 4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

### 主题

- [使用 Phoenix 创建集群](#)
- [Phoenix 客户端](#)
- [Phoenix 发行历史记录](#)

## 使用 Phoenix 创建集群

您可以通过在控制台中或使用 AWS CLI 创建集群时选择 Phoenix 来安装此应用程序。以下过程和示例说明如何使用 Phoenix 和创建集群 HBase。有关使用控制台创建集群的更多信息，包括高级选项，请参阅 Amazon EMR 管理指南中的[规划和配置集群](#)。

在控制台中通过使用用来创建集群的 Quick Options 安装的 Phoenix 启动集群

1. 打开亚马逊 EMR 控制台，[网址为 https://console.aws.amazon.com/emr](https://console.aws.amazon.com/emr)。

2. 选择 Create cluster (创建集群) 以使用 Quick Create (快速创建)。
3. 在 Software Configuration (软件配置) 下，选择与您的应用程序对应的最新版本。Phoenix 只有在选择了 Amazon 发行版 emr-4.7.0 或更高版本时才会作为选项显示。
4. 对于“应用程序”，选择第二个选项 HBase : HBase *ver* 和 Ganglia *ver* , Hadoop *ver* , Hive *ver* , 色调 *ver* , 菲尼克斯 *ver* , 以及 ZooKeeper *ver* .
5. 根据需要选择其它选项，然后选择 Create cluster (创建集群)。

### Note

为了便于读取，包含 Linux 行继续符 ( \ )。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 ( ^ )。

以下示例通过使用默认配置设置安装的 Phoenix 启动集群。

使用 Phoenix 启动集群并HBase使用 AWS CLI

- 使用下面的命令创建集群：

```
aws emr create-cluster --name "Cluster with Phoenix" --release-label emr-7.2.0 \  
--applications Name=Phoenix Name=HBase --ec2-attributes KeyName=myKey \  
--instance-type m5.xlarge --instance-count 3 --use-default-roles
```

## 自定义 Phoenix 配

在创建集群时，您可使用 hbase-site.xml 配置分类在 hbase-site 中设置值来配置 Phoenix。

有关更多信息，请参阅 Phoenix 文档中的[配置和优化](#)。

以下示例演示如何使用存储在 Amazon S3 中的JSON文件来指定 phoenix.schema.dropMetaData 属 false 性的值。可以为单个分类指定多个属性。有关更多信息，请参阅 [配置应用程序](#)。然后，该 create-cluster 命令引用该JSON文件作为 --configurations 参数。

保存到 /mybucket/myfolder/myconfig.json JSON 的文件内容如下。

```
[
```

```
{
  "Classification": "hbase-site",
  "Properties": {
    "phoenix.schema.dropMetaData": "false"
  }
}
```

以下示例显示了引用该JSON文件的create cluster命令。

```
aws emr create-cluster --release-label emr-7.2.0 --applications Name=Phoenix \
Name=HBase --instance-type m5.xlarge --instance-count 2 \
--configurations https://s3.amazonaws.com/mybucket/myfolder/myconfig.json
```

### Note

只有亚马逊EMR版本 5.23.0 及更高版本支持对任何 Phoenix 配置分类的重新配置请求，亚马逊版本 5.21.0 或 5.22.0 不支持重新配置请求。EMR有关更多信息，请参阅[为运行的集群中的实例组提供配置](#)

## Phoenix 客户端

您可以使用完全依赖项构建的JDBC客户端，或者使用使用 Phoenix Query Server 且只能在集群的主节点上运行的“瘦客户端”（例如，通过使用SQL客户端、步骤、命令行、SSH端口转发等）连接到 Phoenix。使用“fat”JDBC 客户端时，它仍然需要访问集群的所有节点，因为它直接连接到HBase服务。“瘦”Phoenix 客户端只需要通过默认端口 8765 访问 Phoenix 查询服务器。Phoenix 中有几个[脚本](#)使用这些客户端。

使用 Amazon EMR 步骤通过 Phoenix 进行查询

以下过程从中恢复快照，HBase并使用该数据运行 Phoenix 查询。您可以扩展此示例或创建一个新脚本，以利用 Phoenix 的客户端来满足您的需求。

1. 使用以下命令创建安装了 Phoenix 的集群：

```
aws emr create-cluster --name "Cluster with Phoenix" --log-uri s3://myBucket/
myLogFolder --release-label emr-7.2.0 \
--applications Name=Phoenix Name=HBase --ec2-attributes KeyName=myKey \
```

```
--instance-type m5.xlarge --instance-count 3 --use-default-roles
```

## 2. 创建后，将以下文件上传到 Amazon S3：

### copySnapshot.sh

```
sudo su hbase -s /bin/sh -c 'hbase snapshot export \
  -D hbase.rootdir=s3://us-east-1.elasticmapreduce.samples/hbase-demo-customer-data/
  snapshot/ \
  -snapshot customer_snapshot1 \
  -copy-to hdfs://masterDNSName:8020/user/hbase \
  -mappers 2 -chuser hbase -chmod 700'
```

### runQuery.sh

```
aws s3 cp s3://myBucket/phoenixQuery.sql /home/hadoop/
/usr/lib/phoenix/bin/sqlline-thin.py http://localhost:8765 /home/hadoop/
phoenixQuery.sql
```

### phoenixQuery.sql

#### Note

当您使用 Amazon 5.26.0 及更高EMR版本时，您只需要在以下示例COLUMN\_ENCODED\_BYTES=0中包含即可。

```
CREATE VIEW "customer" (
pk VARCHAR PRIMARY KEY,
"address"."state" VARCHAR,
"address"."street" VARCHAR,
"address"."city" VARCHAR,
"address"."zip" VARCHAR,
"cc"."number" VARCHAR,
"cc"."expire" VARCHAR,
"cc"."type" VARCHAR,
"contact"."phone" VARCHAR)
COLUMN_ENCODED_BYTES=0;

CREATE INDEX my_index ON "customer" ("customer"."state") INCLUDE("PK",
"customer"."city", "customer"."expire", "customer"."type");
```

```
SELECT "customer"."type" AS credit_card_type, count(*) AS num_customers FROM
"customer" WHERE "customer"."state" = 'CA' GROUP BY "customer"."type";
```

使用 AWS CLI 将文件提交到 S3 存储桶：

```
aws s3 cp copySnapshot.sh s3://myBucket/
aws s3 cp runQuery.sh s3://myBucket/
aws s3 cp phoenixQuery.sql s3://myBucket/
```

3. 使用以下步骤创建提交给您在步骤 1 中创建的集群的表：

createTable.json

```
[
  {
    "Name": "Create HBase Table",
    "Args": ["bash", "-c", "echo '$create \"customer\", \"address\", \"cc\", \"contact\" | hbase shell"],
    "Jar": "command-runner.jar",
    "ActionOnFailure": "CONTINUE",
    "Type": "CUSTOM_JAR"
  }
]
```

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \
--steps file:///./createTable.json
```

4. 使用 script-runner.jar 运行您之前上传到 S3 存储桶的 copySnapshot.sh 脚本：

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \
--steps Type=CUSTOM_JAR,Name="HBase Copy Snapshot",ActionOnFailure=CONTINUE,\
Jar=s3://region.elasticmapreduce/libs/script-runner/script-
runner.jar,Args=["s3://myBucket/copySnapshot.sh"]
```

这将运行一项 MapReduce 任务，将您的快照数据复制到集群HDFS。

5. 使用以下步骤还原复制到集群的快照：

restoreSnapshot.json

```
[
```

```
{
  "Name": "restore",
  "Args": ["bash", "-c", "echo '$disable \"customer\"; restore_snapshot \"customer_snapshot1\"; enable \"customer\"' | hbase shell"],
  "Jar": "command-runner.jar",
  "ActionOnFailure": "CONTINUE",
  "Type": "CUSTOM_JAR"
}
]
```

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \
--steps file:///./restoreSnapshot.json
```

6. 使用 `script-runner.jar` 运行您之前上传到 S3 存储桶的 `runQuery.sh` 脚本：

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \
--steps Type=CUSTOM_JAR,Name="Phoenix Run Query",ActionOnFailure=CONTINUE,\
Jar=s3://region.elasticmapreduce/libs/script-runner/script-
runner.jar,Args=["s3://myBucket/runQuery.sh"]
```

查询运行并将结果返回到步骤的 `stdout`。完成此步骤可能需要几分钟时间。

7. `stdout` 在步骤 1 中创建集群时 URI 使用的日志中检查步骤的结果。结果应该类似以下内容：

```
+-----+-----+
|          CREDIT_CARD_TYPE          |          NUM_CUSTOMERS          |
+-----+-----+
| american_express                    | 5728                            |
| dankort                             | 5782                            |
| diners_club                         | 5795                            |
| discover                            | 5715                            |
| forbrugsforeningen                 | 5691                            |
| jcb                                  | 5762                            |
| laser                               | 5769                            |
| maestro                             | 5816                            |
| mastercard                          | 5697                            |
| solo                                | 5586                            |
| switch                              | 5781                            |
| visa                                 | 5659                            |
+-----+-----+
```

## Phoenix 发行历史记录

下表列出了 Amazon 的每个发行版本中包含的 Phoenix 版本EMR，以及随应用程序一起安装的组件。有关每个发行版本中的组件版本，请参阅 [亚马逊 EMR 7.x 发布版本](#)、[亚马逊 EMR 6.x 发布版本](#) 或 [亚马逊 EMR 5.x 发布版本](#) 中的发行版“组件版本”部分。

### Phoenix 版本信息

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-7.2.0	5.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.36.2	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-



Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
		timeline-server, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-7.1.0	5.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-7.0.0	5.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-6.15.0	5.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-6.14.0	5.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-6.13.0	5.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-6.12.0	5.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-6.11.1	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-6.11.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server



Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-6.10.1	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-6.10.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-6.9.1	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-6.9.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-6.8.1	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-hmaster, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-6.8.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-6.7.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.36.1	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.36.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-6.6.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server



Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.35.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-6.5.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-6.4.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-6.3.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-6.3.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-6.2.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-6.2.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-6.1.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-6.1.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-6.0.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-6.0.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.34.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.33.1	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.33.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.32.1	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.32.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server



Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.31.1	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.31.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.30.2	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.30.1	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.30.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.29.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.28.1	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.28.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.27.1	4.14.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.27.0	4.14.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.26.0	4.14.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.25.0	4.14.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.24.1	4.14.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.24.0	4.14.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.23.1	4.14.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.23.0	4.14.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server



Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.22.0	4.14.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.21.2	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.21.1	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.21.0	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.20.1	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.20.0	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.19.1	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.19.0	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.18.1	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.18.0	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.17.2	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.17.1	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.17.0	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.16.1	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.16.0	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.15.1	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server



Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.15.0	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.14.2	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.14.1	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.14.0	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.13.1	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.13.0	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.12.3	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.12.2	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.12.1	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.12.0	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.11.4	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.11.3	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.11.2	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.11.1	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.11.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.10.1	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server



Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.10.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.9.1	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.9.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.8.3	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.8.2	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.8.1	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.8.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.7.1	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.7.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.6.1	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.6.0	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.5.4	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.5.3	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.5.2	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.5.1	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.5.0	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server



Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.4.1	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.4.0	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.3.2	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.3.1	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.3.0	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.2.3	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.2.2	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.2.1	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.2.0	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.1.1	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.1.0	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.0.3	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.0.2	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.0.1	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-5.0.0	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.9.6	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server



Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-4.9.5	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.9.4	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-4.9.3	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.9.2	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-4.9.1	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.8.5	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-4.8.4	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.8.3	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-4.8.2	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.8.1	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-4.8.0	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.7.4	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-4.7.3	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.7.2	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	Phoenix 版本	随 Phoenix 安装的组件
emr-4.7.1	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.7.0	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server



# Apache Pig

Apache Pig 是一个开源 Apache 库，它在 Hadoop 的顶层运行，提供一种脚本语言，您可以用来转换大型数据集，而不必用底层计算机语言 (例如 Java) 编写复杂的代码。该库采用一种名为 Pig Latin 的语言编写的 SQL 类似命令，并根据有向无环图 (DAGs) 或 MapReduce 程序将这些命令转换为 Tez 作业。Pig 可与各种格式的结构化和非结构化数据结合使用。有关 Pig 的详细信息，请参阅 <http://pig.apache.org/>。

您可以通过交互方式或批处理方式执行 Pig 命令。要以交互方式使用 Pig，请创建 SSH 与主节点的连接并使用 `Grunt shell` 提交命令。要以批处理方式使用 Pig，请编写 Pig 脚本，将脚本上传到 Amazon S3，并作为集群步骤提交。有关向集群提交工作的更多信息，请参阅 Amazon EMR 管理指南中的 [向集群提交工作](#)。

当您使用 Pig 将输出写入 Amazon S3 中的 HCatalog 表时，请通过将 `mapred.output.direct.NativeS3FileSystem` 和 `mapred.output.direct.EmrFileSystem` 属性设置为来禁用 Amazon EMR 直接写入 `false`。有关更多信息，请参阅 [使用 HCatalog](#)。在 Pig 脚本中，可使用 `SET mapred.output.direct.NativeS3FileSystem false` 和 `SET mapred.output.direct.EmrFileSystem false` 命令。

下表列出了最新版本的亚马逊 EMR 7.x 系列中包含的 Pig 版本，以及亚马逊与 Pig 一起 EMR 安装的组件。

有关此版本中与 Pig 一起安装的组件的版本，请参阅 [7.2.0 版组件版本](#)。

emr-7.2.0 的 Pig 版本信息

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-7.2.0	Pig 0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
		timeline-server, pig-client, tez-on-yarn, tez-on-worker

下表列出了最新版本的亚马逊 EMR 6.x 系列中包含的 Pig 版本，以及亚马逊与 Pig 一起EMR安装的组件。

有关此发行版中随 Pig 安装的组件版本，请参阅[发行版 6.15.0 组件版本](#)。

#### emr-6.15.0 的 Pig 版本信息

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-6.15.0	Pig 0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker

下表列出了最新版本的亚马逊 EMR 5.x 系列中包含的 Pig 版本，以及亚马逊与 Pig 一起EMR安装的组件。

有关此版本中与 Pig 一起安装的组件的版本，请参阅[5.36.2 版组件版本](#)。

#### emr-5.36.2 的 Pig 版本信息

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.36.2	Pig 0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-ma

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
		pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

## 主题

- [提交 Pig 工作](#)
- [从 Pig 调用由用户定义的函数](#)
- [Pig 发行历史记录](#)

## 提交 Pig 工作

本节演示如何向亚马逊EMR集群提交 Pig 作品。后面的示例会生成一个报告，其中包含已传输的字节总数、前 50 个 IP 地址的列表、前 50 个外部引用站点的列表以及前 50 个使用 Bing 和 Google 的搜索词。Pig 脚本位于 Amazon S3 存储桶 `s3://elasticmapreduce/samples/pig-apache/doreports2.pig` 中。输入数据位于 Amazon S3 存储桶 `s3://elasticmapreduce/samples/pig-apache/input` 中。输出保存到 Amazon S3 存储桶。

## 使用亚马逊EMR控制台提交 Pig 作品

此示例介绍如何使用 Amazon EMR 控制台向集群添加 Pig 步骤。

### 提交 Pig 步骤

1. 打开亚马逊EMR控制台，[网址为 https://console.aws.amazon.com/emr](https://console.aws.amazon.com/emr)。
2. 选择创建集群以创建安装有 Pig 的集群。有关如何创建集群的步骤，请参阅[规划和配置 Amazon EMR 集群](#)。
3. 按照“[使用连接到主节点](#)”中概述的步骤打开终端并SSH进入集群的主节点SSH。完成此操作后，请运行以下步骤。

```
sudo mkdir -p /home/hadoop/lib/pig/  
sudo aws s3 cp s3://elasticmapreduce/libs/pig/0.3/piggybank-0.3-amzn.jar /home/  
hadoop/lib/pig/piggybank.jar
```

4. 在控制台中，点击 Cluster List (集群列表)，并选择您创建的集群。
5. 滚动到 Steps (步骤) 部分并展开它，然后选择 Add step (添加步骤)。
6. 在 Add step (添加步骤) 对话框中：
  - 对于 Step type (步骤类型)，选择 Pig program (Pig 程序)。
  - 对于 Name (名称)，接受默认名称 (Pig program) 或键入新名称。
  - 对于 Script S3 location (脚本 S3 位置)，键入 Pig 脚本的位置。例如：**s3://elasticmapreduce/samples/pig-apache/do-reports2.pig**。
  - 对于 Input S3 location (输入 S3 位置)，键入输入数据的位置。例如：**s3://elasticmapreduce/samples/pig-apache/input**。
  - 对于 Output S3 location (输出 S3 位置)，键入或浏览到您的 Amazon S3 输出存储桶的名称。
  - 对于 Arguments (参数)，将该字段保留为空白。
  - 对于 Action on failure (出现故障时的操作)，接受默认选项 Continue (继续)。
7. 选择添加。步骤会出现在控制台中，其状态为“Pending”。
8. 步骤的状态会随着步骤的运行从“Pending”变为“Running”，再变为“Completed”。要更新状态，请选择 Actions (操作) 列上方的 Refresh (刷新) 图标。步骤完成后，请检查 Amazon S3 存储桶以确认 Pig 步骤的输出文件存在。

## 使用提交 Pig 作品 AWS CLI

要提交 Pig 步骤，请使用 AWS CLI

使用启动集群时 AWS CLI，使用 `--applications` 参数安装 Pig。要提交 Pig 步骤，请使用 `--steps` 参数。

1. 要启动安装了 Pig 的集群，请键入以下命令，替换 *myKey* 以及 *DOC-EXAMPLE-BUCKET/* 使用您的 EC2 密钥对和 Amazon S3 存储桶的名称。

```
aws emr create-cluster \  
--name "Test cluster" \  
--log-uri s3://DOC-EXAMPLE-BUCKET/ \  
--release-label emr-5.36.2 \  

```

```
--applications Name=Pig \  
--use-default-roles \  
--ec2-attributes KeyName=myKey \  
--instance-type m5.xlarge \  
--instance-count 3
```

### Note

为了便于读取，包含 Linux 行继续符 (\)。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (^)。

如果不使用 `--instance-groups` 参数指定实例计数，则将启动单个主节点，其余实例将作为核心节点启动。所有节点都使用该命令中指定的实例类型。

### Note

如果您之前没有创建过默认EMR服务角色和EC2实例配置文件，请在键入 `create-cluster` 子命令之前键入 `aws emr create-default-roles` 以创建它们。

- 要提交 Pig 步骤，请输入以下命令，替换 *myClusterId* 以及 *DOC-EXAMPLE-BUCKET* 附上您的集群 ID 和 Amazon S3 存储桶的名称。

```
aws emr add-steps \  
--cluster-id myClusterId \  
--steps Type=PIG,Name="Pig Program",ActionOnFailure=CONTINUE,Args=[-f,s3://  
elasticmapreduce/samples/pig-apache/do-reports2.pig,-p,INPUT=s3://elasticmapreduce/  
samples/pig-apache/input,-p,OUTPUT=s3://DOC-EXAMPLE-BUCKET/pig-apache/output]
```

此命令将返回一个步骤 ID，您可以用它来检查您的步骤的 State。

- 使用带有 `describe-step` 命令的步骤，查询步骤的状态。

```
aws emr describe-step --cluster-id myClusterId --step-id s-1XXXXXXXXXXA
```

随着步骤的运行，步骤的 State 从 PENDING 变为 RUNNING 再变为 COMPLETED。步骤完成后，请检查 Amazon S3 存储桶以确认 Pig 步骤的输出文件存在。

有关在中使用 Amazon EMR 命令的更多信息 AWS CLI，请参阅 [AWS CLI 命令参考](#)。

## 从 Pig 调用由用户定义的函数

Pig 提供了从 Pig 脚本中调用用户定义的函数 (UDFs) 的功能。您可以完成此操作，以便实施自定义处理并在 Pig 脚本中使用。目前支持的语言有 Java、Python/Jython 和 JavaScript（尽管 JavaScript 支持仍处于实验阶段。）

以下部分描述了如何通过 Pig 注册函数，以便从 Pig Shell 或者从 Pig 脚本内部调用它们。有关与 Pig UDFs 一起使用的更多信息，请参阅您的 [Pig 版本的 Pig 文档](#)。

### 从 JAR Pig 那里调用文件

您可以使用 Pig 脚本中的 REGISTER 命令在 Pig 中使用自定义 JAR 文件。该 JAR 文件是本地文件系统或远程文件系统，例如 Amazon S3。当 Pig 脚本运行时，Amazon 会自动将 JAR 文件 EMR 下载到主节点，然后将 JAR 文件上传到 Hadoop 分布式缓存。这样，集群中的所有实例都会根据需要自动使用该 JAR 文件。

将 JAR 文件与 Pig 配合使用

1. 将您的自定义 JAR 文件上传到 Amazon S3。
2. 使用 Pig 脚本中的 REGISTER 命令在 Amazon S3 上指定自定义 JAR 文件中的存储桶。

```
REGISTER s3://mybucket/path/mycustomjar.jar;
```

### 从 Pig 调用 Python/Jython 脚本

您可以通过 Pig 注册 Python 脚本，然后，从 Pig Shell 或者在 Pig 脚本中调用这些脚本中的函数。您可以通过 register 关键字指定该脚本的位置，从而完成此操作。

因为 Pig 是以 Java 编写的，所以它使用 Jython 脚本引擎解析 Python 脚本。有关 Jython 的详细信息，请转到 <http://www.jython.org/>。

从 Pig 调用 Python/Jython 脚本

1. 编写 Python 脚本并将其上传到 Amazon S3 中的位置。它应该是创建该 Pig 集群的同一账户所拥有的存储桶，或者该位置拥有相关权限，使得创建该集群的账户可以进行访问。在此示例中，脚本上传到 s3://mybucket/pig/python。

2. 启动 Pig 集群。如果您正在从 Grunt Shell 访问 Pig，请运行交互式集群。如果您正在从脚本运行 Pig 命令，请启动已编写脚本的 Pig 集群。此示例将启动交互式集群。有关如何创建 Pig 集群的详细信息，请参阅 [提交 Pig 工作](#)。
3. 对于交互式集群，使用 SSH 连接到主节点并运行 Grunt shell。有关更多信息，请参阅 [SSH 主节点](#)。
4. 通过在命令行键入 pig 的方式运行 Pig 的 Grunt Shell：

```
pig
```

5. 在 Grunt 命令提示符中使用 register 关键字，通过 Pig 注册 Jython 库和 Python 脚本（如以下命令所示），您可以其中指定脚本在 Amazon S3 中的位置：

```
grunt> register 'lib/jython.jar';
grunt> register 's3://mybucket/pig/python/myscript.py' using jython as myfunctions;
```

6. 加载输入数据。以下示例从 Amazon S3 位置加载输入：

```
grunt> input = load 's3://mybucket/input/data.txt' using TextLoader as
(line:chararray);
```

7. 现在，您可以通过 myfunctions 引用脚本中的函数的方式，从 Pig 内部调用这些函数：

```
grunt> output=foreach input generate myfunctions.myfunction($1);
```

## Pig 发行历史记录

下表列出了 Amazon 的每个发行版本中包含的 Pig 版本 EMR，以及随应用程序一起安装的组件。有关每个发行版本中的组件版本，请参阅 [亚马逊 EMR 7.x 发布版本](#)、[亚马逊 EMR 6.x 发布版本](#) 或 [亚马逊 EMR 5.x 发布版本](#) 中的发行版“组件版本”部分。

### Pig 版本信息

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-7.2.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode,

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
		hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker
emr-5.36.2	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-7.1.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker



Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-7.0.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker
emr-6.15.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-6.14.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker
emr-6.13.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-6.12.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker
emr-6.11.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-6.11.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker
emr-6.10.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-6.10.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker
emr-6.9.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-6.9.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-6.8.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-6.8.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-6.7.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.36.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.36.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn



Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-6.6.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.35.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-6.5.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-6.4.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-6.3.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-6.3.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-6.2.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-6.2.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-6.1.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-6.1.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.34.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.33.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.33.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.32.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.32.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.31.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn



Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.31.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.30.2	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.30.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.30.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.29.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.28.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.28.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.27.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.27.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.26.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.25.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.24.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.24.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.23.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.23.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.22.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn



Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.21.2	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.21.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.21.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.20.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.20.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.19.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.19.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.18.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.18.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.17.2	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.17.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.17.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.16.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.16.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.15.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.15.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn



Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.14.2	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.14.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.14.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.13.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.13.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.12.3	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.12.2	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.12.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.12.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.11.4	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.11.3	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.11.2	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.11.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.11.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.10.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.10.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn



Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.9.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.9.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.8.3	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.8.2	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.8.1	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.8.0	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.7.1	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.7.0	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.6.1	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.6.0	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.5.4	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.5.3	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.5.2	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.5.1	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.5.0	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.4.1	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn



Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.4.0	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.3.2	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.3.1	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.3.0	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.2.3	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.2.2	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.2.1	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.2.0	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.1.1	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.1.0	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.0.3	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.0.2	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-5.0.1	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.0.0	0.16.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-4.9.6	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.9.5	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.9.4	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client



Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-4.9.3	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.9.2	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.9.1	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-4.8.5	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.8.4	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.8.3	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-4.8.2	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.8.1	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.8.0	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-4.7.4	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.7.3	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.7.2	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-4.7.1	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.7.0	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.6.1	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-4.6.0	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.5.0	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.4.0	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-4.3.0	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.2.0	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.1.0	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client

Amazon EMR 发布标签	Pig 版本	随 Pig 安装的组件
emr-4.0.0	0.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client



# Presto 和 Trino

## Note

Presto SQL 于 2020 年 12 月更名为 Trino。亚马逊 6.4.0 及更高 EMR 版本使用 Trino 这个名字，而早期版本使用 Presto 这个名字。SQL

[Presto](#) 是一款快速 SQL 查询引擎，专为对来自多个来源的大型数据集进行交互式分析查询而设计。有关更多信息，请参阅 [Presto 网站](#)。Presto 包含在亚马逊 5.0.0 及更高 EMR 版本中。早期发行版包含 Presto，将其用作沙盒应用程序。有关更多信息，请参阅 [亚马逊 EMR 4.x 发布版本](#)。除了 Presto 之外，亚马逊 EMR 发布的 6.1.0 及更高版本还支持 [Trino](#) ( PrestoSQL )。有关更多信息，请参阅 [PrestoDB 和 Trino 安装](#)。

下表列出了最新版本的亚马逊 EMR 7.x 系列中包含的 Presto 版本，以及亚马逊与 Presto 一起 EMR 安装的组件。

有关此版本中与 Presto 一起安装的组件的版本，请参阅 [7.2.0 版组件版本](#)。

emr-7.2.0 的 Presto 版本信息

亚马逊 EMR 发行标签	Presto 版本	随 Presto 安装的组件
emr-7.2.0	Presto 0.285	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

下表列出了最新版本的亚马逊 EMR 6.x 系列中包含的 Presto 版本，以及亚马逊与 Presto 一起EMR安装的组件。

有关此发行版中随 Presto 安装的组件版本，请参阅[发行版 6.15.0 组件版本](#)。

#### emr-6.15.0 的 Presto 版本信息

亚马逊EMR发行标签	Presto 版本	随 Presto 安装的组件
emr-6.15.0	Presto 0.283	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

下表列出了最新版本的亚马逊 EMR 5.x 系列中包含的 Presto 版本，以及亚马逊与 Presto 一起EMR安装的组件。

有关此版本中与 Presto 一起安装的组件的版本，请参阅[5.36.2 版组件版本](#)。

#### emr-5.36.2 的 Presto 版本信息

亚马逊EMR发行标签	Presto 版本	随 Presto 安装的组件
emr-5.36.2	Presto 0.267	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-

亚马逊EMR发行标签	Presto 版本	随 Presto 安装的组件
		client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

下表列出了最新版本的亚马逊 EMR 6.x 系列中包含的 Trino ( PrestoSQL ) 版本，以及亚马逊与天合一 ( Presto ) 一起EMR安装的组件。SQL

有关此版本中与 Trino (PrestoSQL) 一起安装的组件的版本，请参阅 [6.15.0 版本](#) 的组件版本。

emr-6.15.0 的 Trino (PrestoSQL) 版本信息

亚马逊EMR发行标签	Trino ( Presto ) 版本 SQL	与天合一一起安装的组件 (Prest SQL o)
emr-6.15.0	Trino (PrestoSQL) 426	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker

## 主题

- [在 AWS Glue 数据目录中使用 Presto](#)
- [使用 S3 Select Pushdown 搭配 Presto 提高性能](#)
- [添加数据库连接器](#)
- [在亚马逊上使用SSL/TLS并LDAPS使用 Presto 进行配置 EMR](#)
- [激活 Presto 严格模式](#)
- [在 Presto 中处理竞价型实例丢失](#)

- [Trino 中的容错执行](#)
- [使用采用 Graceful Decommission 的 Presto 自动扩展配置](#)
- [亚马逊 Presto 的注意事项 EMR](#)
- [Presto 发行历史记录](#)

## 在 AWS Glue 数据目录中使用 Presto

使用亚马逊EMR发布的版本 5.10.0 及更高版本，您可以将 AWS Glue 数据目录指定为 Presto 的默认 Hive 元数据库。当您需要持久性的元数据存储或由不同集群、服务、应用程序和 AWS 账户共享的元数据存储时，我们建议使用此配置。

AWS Glue 是一项完全托管的提取、转换和加载 (ETL) 服务，它使对数据进行分类、清理、丰富数据并在各种数据存储之间可靠地移动数据变得简单且经济高效。AWS Glue 数据目录提供了一个跨各种数据源和数据格式的统一元数据存储库，与亚马逊、亚马逊、亚马逊 Redshift、Redshift Spectrum RDS、Athena EMR 以及任何与 Apache Hive 元数据仓兼容的应用程序集成。AWS Glue 爬虫程序可以自动从 Amazon S3 中的源数据推断出架构，并将相关的元数据存储和数据目录中。有关数据目录的更多信息，请参阅 [AWS Glue 开发者指南中的填充 AWS Glue 数据目录](#)。

Glue 需单独收 AWS 费。在数据目录中存储和访问元数据有月费率，AWS Glue ETL 作业和爬虫运行时按小时费率收费，每个预配置的开发端点按每分钟计费。数据目录让您最多可免费存储一百万个对象。如果您存储的对象超过一百万个，则每超过一百万个 USD 100,000 个对象需要支付 1 美元的费用。数据目录中的对象为表、分区或数据库。有关更多信息，请参阅 [Glue 定价](#)。

### Important

如果您在 2017 年 8 月 14 日之前使用 Amazon Athena 或 Amazon Redshift Spectrum 创建了表，则数据库和表将存储在雅典娜管理的目录中，该目录与 Glue 数据目录是分开的。AWS 要将亚马逊EMR与这些表集成，您必须升级到 AWS Glue 数据目录。有关更多信息，请参阅亚马逊 Athena AWS 用户指南中的升级到 Glue [数据目录](#)。

## 将 AWS Glue 数据目录指定为元数据库

您可以使用 AWS Management Console、AWS CLI 或 Amazon EMR API 将 AWS Glue 数据目录指定为元数据仓。使用 CLI 或 API，您可以使用 Presto 的配置分类来指定数据目录。此外，在 Amazon EMR 5.16.0 及更高版本中，您可以使用配置分类在不同的数据目录中指定数据目录。AWS 账户在使用控制台时，您可以使用 Advanced Options (高级选项) 或 Quick Options (快速选项) 指定数据目录。

## Console

使用新主机 AWS 将 Glue 数据目录指定为 Hive 元数据仓库

1. 登录 AWS Management Console，然后通过 <https://console.aws.amazon.com/emr> 打开亚马逊EMR控制台。
2. EC2在左侧导航窗格的“开”下EMR，选择“集群”，然后选择“创建集群”。
3. 在 Application bundle ( 应用程序包 ) 下，选择 Presto。
4. 在AWS Glue Data Catalog 设置下，选择用于 Presto 表元数据复选框。
5. 选择适用于集群的任何其他选项。
6. 要启动集群，选择 Create cluster ( 创建集群 )。

## CLI

要将 Glue AWS 数据目录指定为默认 Hive 元数据仓库，请使用 AWS CLI

有关如何在创建集群时指定以下配置分类的示例，请参阅[配置应用程序](#)。

亚马逊 EMR 5.16.0 及更高版本

- 将该hive.metastore属性设置glue为，如以下JSON示例所示。

```
[
  {
    "Classification": "presto-connector-hive",
    "Properties": {
      "hive.metastore": "glue"
    }
  }
]
```

要在不同的数据目录中指定数据目录 AWS 账户，请添加hive.metastore.glue.catalogid属性，如以下JSON示例所示。将 *acct-id* 替换为数据目录的 AWS 账户。使用 Amazon 5.15.0 及更早版本时，无法在其他 AWS 账户 EMR版本中使用数据目录。

```
[
  {
    "Classification": "presto-connector-hive",
```

```
"Properties": {
  "hive.metastore": "glue",
  "hive.metastore.glue.catalogid": "acct-id"
}
}
```

### 亚马逊 EMR 5.10.0 到 5.15.0

将该hive.metastore.glue.datacatalog.enabled属性设置为true，如以下JSON示例所示：

```
[
  {
    "Classification": "presto-connector-hive",
    "Properties": {
      "hive.metastore.glue.datacatalog.enabled": "true"
    }
  }
]
```

### 亚马逊 EMR 6.1.0 及更高版本使用 Presto SQL (Trino)

从 6.1.0 EMR 版本开始，Presto SQL 还支持 Glue 作为默认 Hive 元数据仓。使用prestoql-connector-hive配置分类并将hive.metastore属性设置为glue，如以下JSON示例所示。

亚马逊 6.4.0 及更高EMR版本使用新名称 Trino 而不是 Presto。SQL如果您使用 Trino，请在以下配置分类中将 *prestoql-connector-hive* 替换为 trino-connector-hive。

```
[
  {
    "Classification": "prestoql-connector-hive",
    "Properties": {
      "hive.metastore": "glue"
    }
  }
]
```

要在长时间运行的集群上切换元存储，您可以连接到主节点，直接编辑 `/etc/presto/conf/catalog/hive.properties` 文件中的属性值并重新启动 Presto 服务器 ( `sudo restart presto-server` )，以便为您的发行版相应地手动设置这些值。如果您在 Amazon EMR 5.15.0 及更早版本中使用此方法，请确保将其设置 `hive.table-statistics-enabled` 为 `false`。在使用发行版 5.16.0 和更高版本时，不需要使用该设置；但不支持表和分区统计信息。

## IAM权限

集群的EC2实例配置文件必须具有 Glue AWS 操作的IAM权限。此外，如果您为 AWS Glue Data Catalog 对象启用加密，则还必须允许该角色加密、解密和生成 AWS KMS key 用于加密的对象。

### AWS Glue 操作的权限

如果您使用 Amazon 的默认EC2实例配置文件EMR，则无需执行任何操作。附加到 `AmazonElasticMapReduceforEC2Role` 托管策略 `EMR_EC2_DefaultRole` 允许执行所有必要的 AWS Glue 操作。但是，如果您指定自定义EC2实例配置文件和权限，则必须配置相应的 AWS Glue 操作。使用 `AmazonElasticMapReduceforEC2Role` 托管策略作为起点。有关更多信息，请参阅 Amazon EMR 管理指南中的 [集群EC2EC2实例服务角色 \( 实例配置文件 \)](#)。

### 加密和解密 AWS Glue 数据目录的权限

您的实例配置文件需要使用密钥加密和解密数据的权限。如果以下语句适用，您不必配置这些权限：

- 您可以使用 Glue 的 AWS 托管密钥为 Glue 数据目录对象启用加密。 AWS
- 您使用的集群与 Glue 数据目录 AWS 账户 相同。

否则，您必须将以下语句添加到您的EC2实例配置文件所附的权限策略中。

```
[
  {
    "Version": "2012-10-17",
    "Statement": [
      {
        "Effect": "Allow",
        "Action": [
          "kms:Decrypt",
          "kms:Encrypt",
          "kms:GenerateDataKey"
        ]
      }
    ]
  }
]
```

```

    "Resource": "arn:aws:kms:region:acct-id:key/12345678-1234-1234-1234-123456789012"
  }
]
}
]

```

有关 AWS Glue 数据目录加密的更多信息，请参阅 [Glue 开发者指南中的AWS 加密数据目录](#)。

## 基于资源的权限

如果您在亚马逊中将 AWS Glue 与 Hive、Spark 或 Presto 结合使用，EMR 支持基于资源的策略来控制对数据目录资源的访问。这些资源包括数据库、表、连接和用户定义的函数。有关更多信息，请参阅《AWS Glue 开发人员指南》中的 [AWS Glue 资源策略](#)。

使用基于资源的策略限制从 Amazon EMR 内部访问 AWS Glue 时，您在权限策略中指定的委托人必须是与创建集群时指定的 EC2 实例配置文件 ARN 关联的角色。例如，对于附加到目录的基于资源的策略，您可以 ARN 为集群 EC2 实例指定默认服务角色的角色，*EMR\_EC2\_DefaultRole* 正如 Principal，使用以下示例所示的格式：

```
arn:aws:iam::acct-id:role/EMR_EC2_DefaultRole
```

这些区域有：*acct-id* 可以与 AWS Glue 账号 ID 不同。这允许从不同账户的 EMR 集群进行访问。您可以指定多个委托人，且每个委托人都可以来自不同的账户。

## 使用 AWS Glue 数据目录时的注意事项

在 Presto 中使用 AWS Glue 数据目录作为元数据仓库时，请考虑以下事项：

- 不支持在 AWS Glue 中重命名表。
- 当您创建 Hive 表而不指定 LOCATION 时，表数据存储在与通过 `hive.metastore.warehouse.dir` 属性指定的位置。默认情况下，这是 HDFS 中的一个位置。如果另一个集群需要访问该表，则它将失败，除非它有足够的权限访问创建该表的集群。此外，由于 HDFS 存储是临时性的，因此如果集群终止，表数据就会丢失，必须重新创建表。我们建议您在创建 Hive 表时 LOCATION 在 Amazon S3 中指定。此外，也可以使用 `hive-site` 配置分类来为 `hive.metastore.warehouse.dir` 指定 Amazon S3 中的位置，它适用于所有 Hive 表。如果在某个 HDFS 位置创建了表，而创建该表的集群仍在运行，则可以从 Glue 中将表的位置更新为 Amazon S3。有关更多信息，请参阅 [《Glue 开发者指南》中的“在 AWS Glue 控制台上使用表格”](#)。



- 不支持包含引号和撇号的分区值，例如 PARTITION (owner="Doe's")。
- emr-5.31.0 及更高版本支持[列统计数据](#)。
- 不支持使用 [Hive 授权](#)。作为替代方案，考虑使用[基于AWS Glue 资源的策略](#)。有关更多信息，请参阅[使用基于资源的策略让 Amazon EMR 访问 AWS Glue 数据目录](#)。

## 使用 S3 Select Pushdown 搭配 Presto 提高性能

在亚马逊EMR发布版本 5.18.0 及更高版本中，您可以在亚马逊上[将 S3 select Pushdown](#) 与 Presto 一起使用。EMR此功能允许 Presto 将投影操作（例如，SELECT）和谓词操作（例如，WHERE）的计算工作“下推”至 Amazon S3。这允许查询仅从 Amazon S3 中检索所需的数据，这可以提高性能并减少某些应用程序中在 Amazon EMR 和 Amazon S3 之间传输的数据量。

### S3 Select Pushdown 是否适合我的应用程序？

建议您分别在使用和不使用 S3 Select Pushdown 的情况下对您的应用程序进行基准检验，以查看其是否适用于您的应用程序。

使用以下准则来确定您的应用程序是否为使用 S3 Select 的候选项：

- 您的查询将筛选掉原始数据集的一半以上的数据。
- 您的查询筛选谓词使用具有 Presto 和 S3 Select 支持的数据类型的列。S3 Select Pushdown 不支持时间戳、实数和双精度数据类型。建议对数值数据使用十进制数据类型。有关 S3 Select 支持的数据类型的更多信息，请参阅《Amazon Simple Storage Service 用户指南》中的[数据类型](#)。
- 您的 Amazon S3 和 Amazon EMR 集群之间的网络连接具有良好的传输速度和可用带宽。Amazon S3 不压缩响HTTP应，因此压缩后的输入文件的响应大小可能会增加。

### 注意事项和限制

- 仅支持以CSV格式存储的对象。对象可以解压缩，也可以选择使用 gzip 或 bzip2 压缩。
- 不支持 AllowQuotedRecordDelimiters 属性。如果指定该属性，则查询将失败。
- 不支持使用客户提供的加密密钥 (SSE-C) 的 Amazon S3 服务器端加密和客户端加密。
- S3 Select Pushdown 不能替代使用列式或压缩文件格式，例如ORC或 Parquet。

## 启用 S3 Select Pushdown with PrestoDB 或 Trino

要在 Amazon 上 EMR 启用 PrestoDB 的 S3 Select Pushdown，`presto-connector-hive` 请使用配置分类 `hive.s3select-pushdown.enabled` 将其 `true` 设置为，如下例所示。有关更多信息，请参阅 [配置应用程序](#)。还必须设置 `hive.s3select-pushdown.max-connections` 值。对于大部分应用程序，`500` 的默认设置应该能满足需求。有关更多信息，请参阅下面的 [了解和调整 `hive.s3select-pushdown.max-connections`](#)。

对于 EMR 版本为 6.1.0-6.3.0 的 Presto SQL，请在以下示例 `presto-connector-hive` 中替换为 `presto-connector-hive`。

亚马逊 6.4.0 及更高 EMR 版本使用新名称 Trino 而不是 Presto。SQL 如果您使用 Trino，请在以下示例中将 `presto-connector-hive` 替换为 `trino-connector-hive`。

```
[
  {
    "classification": "presto-connector-hive",
    "properties": {
      "hive.s3select-pushdown.enabled": "true",
      "hive.s3select-pushdown.max-connections": "500"
    }
  }
]
```

### 了解和调整 `hive.s3select-pushdown.max-connections`

默认情况下，Presto 使用 EMRFS 作为其文件系统。`emrfs-site` 配置分类 `fs.s3.maxConnections` 中的设置指定了 Presto 允许与 Amazon S3 的最大客户端连接数。EMRFS 默认情况下，此连接数为 500。访问 Amazon S3 进行谓词操作 EMRFS 时，S3 Select Pushdown 会绕过。在此示例中，`hive.s3select-pushdown.max-connections` 的值确定从 Worker 节点执行这些操作所允许的最大客户端连接数。但是，Presto 向 Amazon S3 发起的任何未向下推的请求（例如 GET 操作）将继续受值的约束。`fs.s3.maxConnections`

如果您的应用程序遇到错误“Timeout waiting for connection from pool”，请增加 `hive.s3select-pushdown.max-connections` 和 `fs.s3.maxConnections` 的值。

## 添加数据库连接器

创建集群时，您可以使用配置分类来配置 JDBC 连接器属性。配置分类以 `presto-connector` 开头，例如 `presto-connector-postgresql`。可用的配置分类取决于 Amazon EMR 发布版本。有关最

新版本中可用的配置分类，[the section called “5.36.2 配置分类”](#)请参阅 Amazon EMR 5.36.2。如果您使用的是其他版本的 Amazon EMR，请参阅[亚马逊 EMR 5.x 发布版本](#)以了解配置分类。有关可为每个连接器配置的属性的更多信息，请参见<https://prestodb.io/docs/current/connector.html>。

### Example —使用 Postgre 连接器配置集群 SQL JDBC

要启动安装并配置了 Postgre SQL 连接器的集群，请先使用以下内容创建一个指定配置分类的JSON文件（例如myConfig.json），然后将其保存在本地。

根据您的设置替换连接属性，如 [Presto 文档中的 Postgre SQL 连接器](#)主题所示。

```
[
  {
    "Classification": "presto-connector-postgresql",
    "Properties": {
      "connection-url": "jdbc:postgresql://example.net:5432/database",
      "connection-user": "MYUSER",
      "connection-password": "MYPASS"
    },
    "Configurations": []
  }
]
```

创建集群时，使用--configurations选项引用JSON文件路径，如以下示例所示，其中myConfig.json位于运行命令的同一目录中：

```
aws emr create-cluster --name PrestoConnector --release-label emr-5.36.2 --instance-type m5.xlarge \
--instance-count 2 --applications Name=Hadoop Name=Hive Name=Pig Name=Presto \
--use-default-roles --ec2-attributes KeyName=myKey \
--log-uri s3://my-bucket/logs --enable-debugging \
--configurations file://myConfig.json
```

## 在亚马逊上使用SSL/TLS并LDAPS使用 Presto 进行配置 EMR

在 Amazon EMR 发行版 5.6.0 及更高版本中，您可以启用SSL/TLS以帮助[保护 Presto 节点之间的内部通信](#)。您可以为传输中加密设置安全配置以执行该操作。有关更多信息，请参阅 Amazon EMR 管理指南中的[加密选项](#)和[使用安全配置设置集群安全](#)。

当您使用带有传输中加密的安全配置时，Amazon EMR 会为 Presto 执行以下操作：

- 分发您在整个 Presto 集群中用于传输中加密的加密构件或证书。有关更多信息，请参阅[为传输中的数据加密提供证书](#)。
- 使用 presto-config 配置分类设置以下属性，该分类对应于 Presto 的 config.properties 文件：
  - false在所有节点上设置为http-server.http.enabled，则禁用，取而代HTTP之的是。HTTPS这要求您在为传输中加密设置安全配置DNS时，提供适用于公共和私人的证书。一种方法是使用支持多个域的SAN ( 主题备用名称 ) 证书。
  - 设置 http-server.https.\* 值。有关配置的详细信息，请参阅 Presto 文档中的[LDAP身份验证](#)。
- 对于 6.1.0 及更高EMR版本的 Presto SQL (Trino)，Amazon EMR 会自动配置共享密钥，以实现集群节点之间的安全内部通信。您无需执行任何额外的配置即可启用此安全功能，并且您可以使用自己的私有密钥覆盖配置。有关 Trino 内部身份验证的信息，请参阅 [Trino 353 文档：安全的内部通信](#)。

此外，在 Amazon EMR 发行版 5.10.0 及更高版本中，您可以使用为与 Presto 协调器的客户端连接设置[LDAP身份验证](#)。HTTPS此设置使用安全 LDAP (LDAPS)。TLS必须在您的LDAP服务器上启用，并且 Presto 集群必须使用启用传输中数据加密的安全配置。需要使用额外的配置。根据您使用的Amazon发行版本EMR，配置选项会有所不同。有关更多信息，请参阅 [在亚马逊上使用 Presto 的LDAP身份验证 EMR](#)。

亚马逊上的 Presto 默认EMR使用端口 8446 作为内部端口HTTPS。用于内部通信的端口必须与用于客户端HTTPS访问 Presto 协调器的端口相同。http-server.https.port 配置分类中的 presto-config 属性指定该端口。

## 在亚马逊上使用 Presto 的LDAP身份验证 EMR

按照本节中的步骤进行配置LDAP。请参阅每个步骤以了解示例以及指向更多信息的链接。

### 配置LDAP身份验证的步骤

- [步骤 1：收集有关您的LDAP服务器的信息并将服务器证书复制到 Amazon S3](#)
- [步骤 2：设置安全配置](#)
- [第 3 步：创建具JSON有 Presto 属性的配置 LDAP](#)
- [步骤 4：创建脚本以复制LDAP服务器证书并将其上传到 Amazon S3](#)
- [步骤 5：创建集群](#)

## 步骤 1：收集有关您的LDAP服务器的信息并将服务器证书复制到 Amazon S3

您需要LDAP服务器提供以下部分中的信息和项目来配置LDAP身份验证。

### LDAP服务器的 IP 地址或主机名

Amazon EMR 主节点上的 Presto 协调器必须能够通过指定的 IP 地址或主机名访问LDAP服务器。默认情况下，Presto 使用LDAPS端口 636 与LDAP服务器通信。如果您的LDAP实现需要自定义端口，则可以在 Amazon EMR 5.16.0 或更高版本中使用该`ldap.url`属性进行指定，或者在早期版本中`authentication.ldap.url`使用。将自定义端口替换为 636，如 [第 3 步：创建具有JSON有 Presto 属性的配置 LDAP](#) 中的 `presto-config` 配置分类示例所示。确保任何防火墙和安全组允许端口 636 ( 或自定义端口 ) 以及端口 8446 ( 或自定义端口 ) 上的入站和出站流量，端口 8446 用于内部集群通信。

### LDAP服务器证书

您必须将证书文件上传到 Amazon S3 中的安全位置。有关更多信息，请参阅《Amazon Simple Storage Service 用户指南》中的[如何将文件和文件夹上传到 S3 存储桶](#)。您可以创建一个引导操作，以便在集群启动时将该证书从 Amazon S3 复制到集群中的每个节点。在[步骤 4：创建脚本以复制 LDAP服务器证书并将其上传到 Amazon S3](#)中。示例证书是 `s3://MyBucket/ldap_server.crt`。

### LDAP服务器的匿名绑定设置

如果在 PrestoDB 上禁用了匿名绑定，则需要具有绑定到服务器权限的帐户的用户 ID UID () 和密码，这样 PrestoDB 服务器才能LDAP建立连接。您可以使用`presto-config`配置分类中的`internal-communication.authentication.ldap.user`和`internal-communication.authentication.ldap.password`属性来指定用户名和密码。UID Amazon EMR 5.10.0 不支持这些设置，因此当您使用此发行版时，LDAP服务器上必须支持匿名绑定。

请注意，Trino 不需要匿名绑定配置。

### 获取LDAP服务器上匿名绑定的状态

- 从 Linux 客户端中使用 `ldapwhoami` 命令，如以下示例所示：

```
ldapwhoami -x -H ldaps://LDAPServerHostNameOrIPAddress
```

如果不允许匿名绑定，该命令将返回以下内容：

```
ldap_bind: Inappropriate authentication (48)
```

```
additional info: anonymous bind disallowed
```

## 验证账户是否有权访问使用简单身份验证的LDAP服务器

- 从 Linux 客户端中使用 `ldapwhoami` 命令，如以下示例所示。该示例使用了一个虚构的用户，`presto`，存储在使用虚构主机名的EC2实例上运行的 Open LDAP 服务器中 `ip-xxx-xxx-xxx-xxx.ec2.internal`。用户已与组织单位 (OU) 关联 `admins` 并使用密码 `123456`:

```
ldapwhoami -x -w "123456" -D uid=presto,ou=admins,dc=ec2,dc=internal -H ldaps://ip-xxx-xxx-xxx-xxx.ec2.internal
```

如果该账户有效并具有相应的权限，该命令将返回：

```
dn:uid=presto,ou=admins,dc=ec2,dc=internal
```

为了清楚起见，[第 3 步：创建具JSON有 Presto 属性的配置 LDAP](#) 中的示例配置包含该账户，但 5.10.0 示例除外，该发行版不支持该账户。如果LDAP服务器使用匿名绑定，请移除`internal-communication.authentication.ldap.user`和`internal-communication.authentication.ldap.password`名称/值对。

## Presto 用户LDAP的可分辨名称 (DN)

在指定 Presto 的LDAP配置时，您可以指定一种绑定模式，该模式包括组织单位 (OU) 和其他域组件 (DCs)。\${USER}在密码身份验证期间，Presto 将替换\${USER}为每个用户的实际用户 ID (UID)，以匹配此绑定模式指定的可分辨名称 (DN)。您需要符合条件OUs的用户所属的以及他们DCs的。例如，要允许 `admins` 域上的 `corp.example.com` OU 中的用户在 Presto 中进行身份验证，您可以将 `${USER},ou=admins,dc=corp,dc=example,dc=com` 指定为用户绑定模式。

### Note

使用时 AWS CloudFormation，需要使用 `Fn::Sub` 函数才能替换为实际\${USER}的用户 ID () UID。有关更多信息，请参阅《AWS CloudFormation 用户指南》中的 [Fn::Sub](#) 主题。

使用 Amazon EMR 5.10.0 时，您只能指定一种这样的模式。使用 Amazon EMR 5.11.0 或更高版本，您可以指定用冒号 (:) 分隔的多个模式。尝试在 Presto 中进行身份验证的用户先与第一种模式进行比

较，然后与第二种模式进行比较，依此类推。有关示例，请参阅[第 3 步：创建具JSON有 Presto 属性的配置 LDAP](#)。

## 步骤 2：设置安全配置

创建一个安全配置并启用传输中加密。有关更多信息，请参阅 Amazon EMR 管理指南中的[创建安全配置](#)。在设置传输中加密时提供的加密构件用于加密 Presto 节点之间的内部通信。有关更多信息，请参阅[为传输中的数据加密提供证书](#)。LDAP服务器证书用于验证客户端与 Presto 服务器的连接。

## 第 3 步：创建具JSON有 Presto 属性的配置 LDAP

您可以使用presto-config配置分类来设置 Presto 的属性。LDAP的格式和内容略presto-config有不同，具体取决于亚马逊EMR发布版本和 Presto 安装 ( PrestoDB 或 Trino )。在本节后面提供了配置差异示例。有关更多信息，请参阅[配置应用程序](#)。

以下步骤假设您将JSON数据保存到文件中，*MyPrestoConfig.json*。如果您使用控制台，请将文件上传到 Amazon S3 中的安全位置，以便在创建集群时可以参考该文件。如果使用 AWS CLI，则可以在本地引用该文件。

Example 亚马逊 EMR 6.1.0 及更高版本推出 PrestoSQL ( Trino )

以下示例使用中的LDAP主机名[步骤 1：收集有关您的LDAP服务器的信息并将服务器证书复制到 Amazon S3](#)向LDAP服务器进行身份验证以进行绑定。指定了两种用户绑定模式，这表明 admins OU 和LDAP服务器上的 O datascientists U 中的用户有资格以用户身份向 Trino 服务器进行身份验证。绑定模式由冒号 (:) 分隔。

亚马逊 6.4.0 及更高EMR版本使用新名称 Trino 而不是 Presto。SQL如果你使用 Trino，请更换 *prestoql-config* 在以下配置分类中，使用trino-config和*prestoql-password-authenticator*使用trino-password-authenticator。

```
[
  {
    "Classification": "prestoql-config",
    "Properties": {
      "http-server.authentication.type": "PASSWORD"
    }
  },
  {
    "Classification": "prestoql-password-authenticator",
    "Properties": {
      "password-authenticator.name": "ldap",
```

```

    "ldap.url": "ldaps://ip-xxx-xxx-xxx-xxx.ec2.internal:636",
    "ldap.user-bind-pattern": "uid=${USER},ou=admins,dc=ec2,dc=internal:uid=${USER},ou=datascientists,dc=ec2,dc=internal"
  }
}
]

```

### Example 亚马逊 EMR 5.16.0 及更高版本

以下示例使用LDAP用户 ID 和密码以及中的LDAP主机名[步骤 1：收集有关您的LDAP服务器的信息](#)并将服务器证书复制到 Amazon S3向LDAP服务器进行身份验证以进行绑定。指定了两种用户绑定模式，这表明 admins OU 和LDAP服务器上的 O datascientists U 中的用户有资格以用户身份向 Presto 服务器进行身份验证。绑定模式由冒号 (:) 分隔。

```

[ {
  "Classification": "presto-config",
  "Properties": {
    "http-server.authentication.type": "PASSWORD"
  }
},
{
  "Classification": "presto-password-authenticator",
  "Properties": {
    "password-authenticator.name": "ldap",
    "ldap.url": "ldaps://ip-xxx-xxx-xxx-xxx.ec2.internal:636",
    "ldap.user-bind-pattern": "uid=${USER},ou=admins,dc=ec2,dc=internal:uid=${USER},ou=datascientists,dc=ec2,dc=internal",
    "internal-communication.authentication.ldap.user": "presto",
    "internal-communication.authentication.ldap.password": "123456"
  }
}
]

```

### Example 亚马逊 EMR 5.11.0 到 5.15.0

这些发行版的 presto-config 配置分类的格式略有不同。以下示例指定与上一示例相同的参数。

```

[ {
  "Classification": "presto-config",
  "Properties": {
    "http-server.authentication.type": "LDAP",
    "authentication.ldap.url": "ldaps://ip-xxx-xxx-xxx-xxx.ec2.internal:636",

```



```

        "authentication.ldap.user-bind-pattern": "uid=
        ${USER},ou=admins,dc=ec2,dc=internal:uid=${USER},ou=datascientists,dc=ec2,dc=internal",
        "internal-communication.authentication.ldap.user": "presto",
        "internal-communication.authentication.ldap.password": "123456"
    }
}]]

```

Example 亚马逊 EMR 5.10.0

Amazon EMR 5.10.0 仅支持匿名绑定，因此省略了这些条目。此外，只能指定一种绑定模式。

```

[ {
    "Classification": "presto-config",
    "Properties": {
        "http-server.authentication.type": "LDAP",
        "authentication.ldap.url": "ldaps://ip-xxx-xxx-xxx-xxx.ec2.internal:636",
        "ldap.user-bind-pattern": "uid=
        ${USER},ou=prestousers,dc=ec2,dc=internal"
    }
}]]

```

## 步骤 4：创建脚本以复制LDAP服务器证书并将其上传到 Amazon S3

创建一个脚本以将证书文件复制到集群中的每个节点，然后将其添加到密钥存储中。请使用文本编辑器创建脚本，保存该脚本，然后将其上传到 Amazon S3 中。在中 [步骤 5：创建集群](#)，脚本文件被引用为 *s3://MyBucket/LoadLDAPCert.sh*。

以下示例脚本使用默认的密钥库密码，*changeit*。我们建议您在创建集群后连接到主节点，并使用 `keytool` 命令更改密钥库密码。

```

#!/bin/bash
aws s3 cp s3://MyBucket/ldap_server.crt .
sudo keytool -import -keystore /usr/lib/jvm/jre-1.8.0-openjdk.x86_64/lib/security/cacerts -trustcacerts -alias ldap_server -file ./ldap_server.crt -storepass changeit -noprompt

```

## 步骤 5：创建集群

创建集群时，您需要指定 Presto 以及您希望 Amazon EMR 安装的其他应用程序。以下示例还引用了中的配置分类属性JSON，但您也可以内联指定配置分类。

## 使用亚马逊EMR控制台创建带有LDAP身份验证的 Presto 集群

1. 导航到新的 Amazon EMR 控制台，然后从侧面导航栏中选择“切换到旧主机”。有关切换到旧控制台后预期情况的更多信息，请参阅 [Using the old console](#)。
2. 依次选择 Create cluster (创建集群)、Go to advanced options (转到高级选项)。
3. 选择 Presto 以及其他应用程序供亚马逊EMR安装，然后在“软件配置”下，选择EMR要使用的亚马逊版本。LDAP仅亚马逊 EMR 5.10.0 及更高版本支持身份验证。
4. 在“编辑软件设置”下，选择“JSON从 S3 加载”，输入您在中创建的JSON配置文件在 Amazon S3 中的位置 [第 3 步：创建具JSON有 Presto 属性的配置 LDAP](#)，然后选择下一步。
5. 配置集群硬件和网络，然后选择下一步。
6. 选择 Bootstrap Actions (引导操作)。对于添加引导操作，请选择自定义操作，然后选择配置并添加。
7. 输入引导操作的名称，输入您在中创建的脚本位置 [步骤 4：创建脚本以复制LDAP服务器证书并将其上传到 Amazon S3](#)，例如 s3: MyBucket ///LoadLDAPCert .sh，然后选择添加。
8. 在常规选项、标签和其他选项下面，选择适合您的应用程序的设置，然后选择下一步。
9. 选择身份验证和加密，然后选择您在中创建的安全配置 [步骤 2：设置安全配置](#)。
10. 选择适合您的应用程序的其他安全选项，然后选择创建集群。

要创建带有LDAP身份验证的 Presto 集群，请使用 AWS CLI

- 使用 `aws emr create-cluster` 命令。至少，指定 Presto 应用程序，以及在以前步骤中创建的 Presto 配置分类、引导脚本和安全配置。以下示例将配置文件作为文件引用，该JSON文件保存在您运行命令的同一目录中。另一方面，引导脚本必须保存在 Amazon S3 中。下面的示例使用了 `s3://MyBucket/LoadLDAPCert.sh`。

### Note

为了便于读取，包含 Linux 行继续符 (`\`)。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (^)。

```
aws emr create-cluster --applications Name=presto --release-label emr-5.16.0 \  
--use-default-roles --ec2-attributes KeyName=MyKeyPair,SubnetId=subnet-1234ab5 \  
\ --instance-count 3 --instance-type m5.xlarge --region us-west-2 --name \  
"MyPrestoWithLDAPAuth" \  

```

```
--bootstrap-actions Name="Distribute LDAP server cert",Path="s3://MyBucket/
LoadLDAPCert.sh" \
--security-configuration MyPrestoLDAPSecCfg --configurations file://
MyPrestoConfig.json
```

## 激活 Presto 严格模式

在某些情况下，长时间运行的查询可能会导致高昂的成本，并导致 Amazon EMR 使用更多的集群资源。这会与集群上的其他工作负载竞争资源。在 Amazon 6.8 及更高EMR版本中，您可以使用严格模式功能，拒绝或警告您以下类型的长时间运行的查询：

- 分区列上没有谓词的查询，这会导致对大量数据进行表扫描
- 在两个大表之间使用交叉连接的查询
- 对大量行进行无限制排序的查询

在 Presto 完全优化查询计划后，严格模式将开始运行。要根据您的查询需求使用和自定义严格模式，您可以通过以下方式配置 Presto。

### Presto 严格模式配置

设置	描述	默认
<code>strict-mode-enabled</code>	开启和关闭严格模式。值 <code>true</code> 表示严格模式已开启。	<code>false</code>
<code>strict-mode-fail-query</code>	如果严格模式检测到可能长时间运行的查询，则会拒绝相关查询。如果是 <code>false</code> ，Amazon EMR 只会发出警告。	<code>false</code>
<code>strict-mode-restrictions</code>	指定严格模式开启时将应用的限制。严格模式支持以下限制：MANDATORY_PARTITION_PREDICATE JOIN、DISALLOW_CROSS_JOIN 和 LIMITED_SORT。	MANDATORY_PARTITION_PREDICATE, DISALLOW_CROSS_JOIN, LIMITED_SORT

要试用严格模式，您可以覆盖这些配置，或者在使用 Presto 客户端时将其设置为会话属性。

要在创建集群时使用设置配置 AWS Management Console

1. 选择“创建集群”，然后选择 Amazon EMR 版本 6.8.0，然后选择 Presto 或 Trino。有关更多信息，请参阅 [安装 PrestoDB 和 Trino](#)。
2. 直接为严格模式指定配置属性，或者将JSON文件上传到 Amazon S3。您也可为您的元数据存储选择。指定您的子网VPC、引导操作、key pair 和安全组。选择 Create cluster ( 创建集群 ) 以创建集群。
3. 登录到集群的主节点，然后运行 `presto-cli` 或 `trino-cli`。
4. 提交查询。严格模式会验证每个查询并确定它是否在长时间运行。根据您的 `strict-mode-fail-query` 设置，Amazon EMR 会拒绝查询或发出警告。
5. 完成查询后，终止集群并删除您的资源。

使用在正在运行的集群上设置配置 AWS CLI

1. 使用 AWS CLI 和运行 `presto-cli` 或，登录到集群的主节点 `trino-cli`。
2. 使用所需的值运行以下命令。

```
set session strict_mode_enabled = true;
set session strict_mode_fail_query = false;
set session strict_mode_restrictions = 'DISALLOW_CROSS_JOIN,LIMITED_SORT';
```

## 注意事项

在使用严格模式时，请注意以下几点：

- 在某些情况下，严格模式可能会拒绝并未消耗大量资源的短时查询。例如，对小表的查询未应用动态筛选，也未用交叉连接替内部连接。这可能导致查询使用强制分区谓词或禁止交叉连接。发生这种情况时，严格模式会拒绝该查询。
- 严格模式检查仅适用于 `SELECT`、`INSERT`、`A CREATE TABLE S SELECT` 和 `EXPLAINANALYZE` 查询类型。
- 您只能通过 Hive 连接器使用对强制分区谓词的限制。

## 在 Presto 中处理竞价型实例丢失

借助 Amazon 中的竞价型实例EMR，您可以以更低的成本使用剩余的 Amazon EC2 容量运行大数据工作负载。为了换取更低的成本，Amazon EC2 可以通过两分钟的通知中断竞价型实例。当您终止某个节点时，Presto 最长可能需要 10 分钟才会返回错误。这会导致错误报告出现不必要的延迟，并可能导致重试。快速终止功能可让您控制 Presto 处理已终止节点的方式。

Presto 协调器的作用是通过定期轮询其状态来跟踪所有 Worker 节点。如果不快速终止，协调器就不会查阅每个节点的状态。YARN NodeManager 这可能导致在查询失败之前出现长时间的重试循环。通过快速终止，NodeManager 只要轮询未能到达主机，Presto 协调器就会立即在中查看节点状态。如果 NodeManager 显示该节点处于非活动状态，Presto 将放弃进一步的重试，查询失败并返回错误。NODE\_DECOMMISSIONED

下面的一组配置参数允许您控制和自定义 Presto 在节点终止时的行为。

### Presto 的节点故障处理配置

设置	描述	默认
<code>query.remote-task.max-backoff-duration</code>	协调器会继续尝试从 Worker 节点获取远程任务状态的持续时间。	10 分钟
<code>query.remote-task.quick-terminate-node-failure</code>	<p>如果协调器无法连接到该节点或在该节点上运行的工作线程，则将激活快速节点故障。<code>query.remote-task.terminate-on-connect-exception</code> 的值将决定协调器是必须连接到该节点还是必须连接到工作线程。</p> <p>该节点查询失败，Amazon EMR 将该节点从可用工作线程列表中删除。发生这种情况时，您将无法使用该节点来计划新的查询。</p>	true

设置	描述	默认
	如果您将此值设置为 <code>false</code> ，Presto 会恢复其先前的行为，即 Presto 协调器在将该节点标记为不可用之前再次尝试连接到该节点（对于 <code>query.remote-task.max-backoff-duration</code> ），并且将该节点上正在进行的查询设为失败。	
<code>query.remote-task.terminate-on-connect-exception</code>	指定如果主机可以访问但协调器无法连接到主机的工作进程，Amazon 是否 EMR 应设置节点。将此值设置为 <code>true</code> 时，则将在无法访问主机时激活快速查询失败。	<code>false</code>

## Trino 中的容错执行

容错执行是 Trino 中的一种机制，集群可以使用该机制来减少查询失败。为此，它会在查询失败时重试查询或其组件任务。激活容错执行后，中间交换数据会假脱机，并且如果在查询执行期间发生 Worker 中断或其他故障，可被其他 Worker 重用。

有关 Trino 容错执行的更多信息，请参阅 Trino 博客上的 [Project Tardigrade 以 Trino 的速度向早期用户提供 ETL](#)。

### 配置

默认情况下停用容错执行。若要激活该功能，请根据所需的重试策略将 `trino-config` 分类中的 `retry-policy` 配置属性设置为 `QUERY` 或 `TASK`，如下所示。

```
{
  "classification":
    "trino-config",
  "properties":
    {
      "retry-policy":
```

```

    "QUERY"
  }
}

```

**QUERY** 重试策略指示 Trino 在 Worker 节点上发生错误时自动重试查询。当 Trino 集群的大部分工作负载包含许多小查询时，建议您使用 QUERY 重试策略。

**TASK** 重试策略指示 Trino 在失败时重试单个查询任务。建议在 Trino 执行大批量查询时使用此策略。集群可以更有效地重试查询中的较小任务，而不是重试整个查询。

## 交换管理器

交换管理器存储和管理假脱机数据，以实现容错执行。它利用外部存储来存储超出内存缓冲区大小的溢出数据。您可以配置基于文件系统的交换管理器，将后台处理后的数据存储到指定位置，例如 Amazon S3、兼容 Amazon S3 的系统或 HDFS。

Amazon 6.9.0 及更高 EMR 版本包含用于配置交易所管理器的 `trino-exchange-manager` 分类。这些版本还支持 HDFS 后台处理。

### 设置交换管理器

使用 `trino-exchange-manager` 配置分类来配置交换管理器。该分类会在协调器和所有 Worker 节点上创建 `etc/exchange-manager.properties` 配置文件。分类还将 `exchange-manager.name` 配置属性设置为 `filesystem`。

默认情况下，Amazon EMR 发布 6.9.0 及更高版本 HDFS 用作交易所管理器。HDFS 在 Amazon EMR EC2 集群中可用，默认情况下，在 `trino-exchange/` 目录中进行后台处理。要使用默认设置，请设置以下配置：

```

{"Classification":
  "trino-exchange-manager"
}

```

如果要提供自定义位置，请在 `trino-exchange-manager` 分类中设置以下属性：

- 将 `exchange.use-local-hdfs` 设置为 `true`。
- 例如 `exchange.base-directories`，设置为 HDFS 自定义目录位置 `exchange.base-directories=/exchange`。如果自定义目录尚未存在 HDFS，Amazon EMR 将创建它。

### HDFS交易所管理器配置

根据内部测试结果，我们建议您使用后台处理到本地HDFS，以便与其他基于云的文件系统相比，查询性能更好。您可以使用为 Exchange 管理器设置以下配置HDFS。

配置	描述	默认设置
<code>exchange.hdfs.block-size</code>	用于HDFS存储的区块大小	4 MB
<code>hdfs.config.resources</code>	要配置的文件路径列表 HDFS	如果 <code>exchange.use-local-hdfs</code> 为 <code>true</code> ，则使用 <code>core-site.xml</code> 、 <code>hdfs-site.xml</code> 文件的路径；否则为 <code>null</code>

有关其他容错执行配置属性及如何设置 Amazon S3 或其他 Amazon S3 兼容系统以进行假脱机的信息，请参阅 Trino 文档的[容错执行](#)页面。

## 注意事项和限制

- 如果启用容错执行，则会在设置 `retry-policy` 时禁用不支持 `write` 的连接器的 `write` 操作。从亚马逊6.9.0 EMR 版本开始，Delta Lake、Hive和Iceberg连接器支持与的操作。`write retry-policy`
- 如果您使用交换管理器并执行高成本的 I/O 操作，当交换管理器将中间数据假脱机到外部存储时，查询性能可能会下降。

## 使用采用 Graceful Decommission 的 Presto 自动扩展配置

Amazon EMR 发布版本 5.30.0 及更高版本包含一项功能，您可以使用该功能为某些扩展操作设置宽限期。宽限期允许 Presto 任务在节点因横向缩减大小调整操作或自动扩展策略请求而终止之前继续运行。有关扩展规则的更多信息，请参阅 Amazon EMR 管理指南中的[了解自动扩展规则](#)。采用 Graceful Decommission 的 Presto 弹性伸缩配置可防止在正在停用的节点上计划新任务，同时允许在达到关机超时之前完成已在运行的任务。正在运行的查询将在节点停用之前完成执行。实例集不支持弹性伸缩。

您可以控制在收到自动扩展关闭请求后必须完成 Presto 任务的时间长度。默认情况下，Amazon 的关闭超时EMR为0分钟，这意味着如果扩容请求需要，Amazon 会EMR立即终止该节点及其上运行的任何 Presto 任务。要在 Amazon 上为 Presto 任务设置更长的超时时间，EMR以允许在缩小集群规模之前



完成正在运行的查询，请使用

```
presto-config
```

配置分类将`graceful-shutdown-timeout`参数设置为大于零的秒或分钟以秒或分钟为单位的值。有关更多信息，请参阅 [配置应用程序](#)。

例如，将 `graceful-shutdown-timeout` 值增大至 "30m" 以指定 30 分钟的超时时段。在关闭超时期结束后，如果标记为停用的节点正在等待查询任务完成，则系统将强制终止该节点，查询失败。如果查询任务在五分钟内完成，则标记为停用的节点将在五分钟后终止，前提是其他YARN应用程序已完成执行。

#### Example 采用 Graceful Decommission 的 Presto 自动扩展配置示例

将 `graceful-shutdown-timeout` 值替换为适合您的设置的分钟数。没有最大值。下面的示例将超时值设置为 1800 秒 ( 30 分钟 )。

```
[
  {
    "classification": "presto-config",
    "properties": {
      "graceful-shutdown-timeout": "1800s"
    }
  }
]
```

#### 限制

PrestoDB 优雅停用不适用于禁用连接EMR的集群HTTP，例如设置为的集群。`http-server.http.enabled false`Trino完全不支持 Graceful Decommission，不论 `http-server.http.enabled` 设置如何。

## 亚马逊 Presto 的注意事项 EMR

在亚马逊EMR上运行 [Presto 时](#)，请考虑以下限制。

### Presto 命令行可执行文件

在亚马逊中EMR，PrestoDB 和 Trino 都使用相同的命令行可执行文件`presto-cli`，如下例所示。

```
presto-cli --catalog hive
```

## 不可配置的 Presto 部署属性

您EMR使用的亚马逊版本决定了可用的 Presto 部署配置。有关这些配置属性的更多信息，请参阅 Presto 文档中的 [部署 Presto](#)。下表显示了 Presto properties 文件的不同配置选项。

文件	可配置
log.properties	<p>PrestoDB：可在EMR亚马逊 4.0.0 及更高版本中进行配置。使用 presto-log 配置分类。</p> <p>Trino ( PrestoSQL )：可在亚马逊 6.1.0 及EMR更高版本中进行配置。使用 prestosql-log 或 trino-log 配置分类。</p>
config.properties	<p>PrestoDB：可在EMR亚马逊 4.0.0 及更高版本中进行配置。使用 presto-config 配置分类。</p> <p>Trino ( PrestoSQL )：可在亚马逊 6.1.0 及EMR更高版本中进行配置。使用 prestosql-config 或 trino-config 配置分类。</p>
hive.properties	<p>PrestoDB：可在EMR亚马逊 4.1.0 及更高版本中进行配置。使用 presto-connector-hive 配置分类。</p> <p>Trino ( PrestoSQL )：可在亚马逊 6.1.0 及EMR更高版本中进行配置。使用 prestosql-connector-hive 或 trino-connector-hive 配置分类。</p>
node.properties	<p>PrestoDB：可在EMR亚马逊 5.6.0 及更高版本中进行配置。使用 presto-node 配置分类。</p> <p>Trino ( PrestoSQL )：可在亚马逊 6.1.0 及EMR更高版本中进行配置。使用 prestosql-node 或 trino-node 配置分类。</p>
jvm.config	不可配置。

## PrestoDB 和 Trino 安装

继续使用应用程序名称 Presto 在集群上安装 PrestoDB。要在集群上安装 Trino，请使用应用程序名称 Trino（或旧版本的亚马逊SQL中的 Presto）。EMR

您可以安装 PrestoDB 或 Trino，但不能在同一个集群上同时安装两者。如果在尝试创建集群时同时指定了 PrestoDB 和 Trino，则会出现验证错误，并且集群创建请求将会失败。

## EMRFS和 Presto FileSystem S3 配置

在亚马逊 5.12.0 及更高EMR版本中，PrestoDB 可以使用。EMRFS这是默认配置。EMRFS也是亚马逊 6.1.0 及更高EMR版本中 Trino (PrestoSQL) 的默认文件系统。有关更多信息，请参阅《Amazon EMR 管理指南》中的[EMR文件系统 \(EMRFS\)](#)。对于早期版本的亚马逊EMR，PrestoS3 FileSystem 是唯一的配置选项。

您可以使用安全配置为 Amazon S3 中的EMRFS数据设置加密。您也可以使用IAM角色向 Amazon S3 EMRFS 发出请求。有关更多信息，请参阅 Amazon EMR管理指南中的[了解加密选项](#)和为向 Amazon S3 发出的EMRFS[请求配置IAM角色](#)。

### Note

如果您使用亚马逊EMR版本 5.12.0 在 Amazon S3 中查询基础数据，则可能会出现 Presto 错误。这是因为 Presto 无法从 `emrfs-site.xml` 提取配置分类值。解决方法是在 `usr/lib/presto/plugin/hive-hadoop2/` 下创建一个 `emrfs` 子目录，并在 `usr/lib/presto/plugin/hive-hadoop2/emrfs` 中创建一个指向现有 `/usr/share/aws/emr/emrfs/conf/emrfs-site.xml` 文件的符号链接。然后重新启动 `presto-server` 进程（首先执行 `sudo presto-server stop`，然后执行 `sudo presto-server start`）。

你可以覆盖EMRFS默认值，改用 PrestoS3 FileSystem。为此，请使用 `presto-connector-hive` 配置分类将 `hive.s3-file-system-type` 设置为 `PRESTO`，如以下示例所示。有关更多信息，请参阅 [配置应用程序](#)。

```
[
  {
    "Classification": "presto-connector-hive",
    "Properties": {
      "hive.s3-file-system-type": "PRESTO"
    }
  }
]
```

```
]
```

如果你使用 PrestoS3FileSystem，请使用

```
presto-connector-hive
```

配置分类或用

```
trino-connector-hive
```

于 Trino 来配置 PrestoS3 属性。FileSystem 有关可用属性的更多信息，请参阅 Presto 文档中“Hive 连接器”部分的 [Amazon S3 配置](#)。这些设置不适用于EMRFS。

## 终端用户模拟的默认设置

默认情况下，Amazon 5.12.0 及更高EMR版本允许通过模拟最终用户进行访问。HDFS有关更多信息，请参阅 Presto 文档中的[终端用户模拟](#)。要使用 `presto-config` 配置分类更改此设置，请将 `hive.hdfs.impersonation.enabled` 属性设置为 `false`。

## Presto Web 界面的默认端口

默认情况下，亚马逊EMR将 Presto 协调器上的 Presto 网页界面配置为使用端口 8889（用于 PrestoDB 和 Trino）。要更改端口，请使用 `presto-config` 配置分类设置 `http-server.http.port` 属性。有关更多信息，请参阅 Presto 文档的部署 Presto 部分中的[配置属性](#)。

## 某些版本中的 Hive 存储桶执行问题

Presto 发行版 152.3 存在一个与 Hive 存储桶执行有关的问题，此问题在某些情况下可能会显著降低 Presto 的查询性能。亚马逊EMR版本 5.0.3、5.1.0 和 5.2.0 包含此版本的 Presto。要解决此问题，请使用 `presto-connector-hive` 配置分类将 `hive.bucket-execution` 属性设置为 `false`，如下示例所示。

```
[
  {
    "Classification": "presto-connector-hive",
    "Properties": {
      "hive.bucket-execution": "false"
    }
  }
]
```

## Presto 发行历史记录

下表列出了 Amazon 的每个发行版本中包含的 Presto 版本EMR，以及随应用程序一起安装的组件。有关每个发行版本中的组件版本，请参阅 [亚马逊 EMR 7.x 发布版本](#)、[亚马逊 EMR 6.x 发布版本](#) 或 [亚马逊 EMR 5.x 发布版本](#) 中的发行版“组件版本”部分。

## Presto 版本信息

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-7.2.0	0.285	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-5.36.2	0.267	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-7.1.0	0.284	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
		client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-7.0.0	0.283	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.15.0	0.283	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-6.14.0	0.281	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.13.0	0.281	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-6.12.0	0.281	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.11.1	0.279	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker



亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-6.11.0	0.279	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.10.1	0.278	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-6.10.0	0.278	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.9.1	0.276	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-6.9.0	0.276	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.8.1	0.273	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-6.8.0	0.273	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.7.0	0.272	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-5.36.1	0.267	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-5.36.0	0.267	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-6.6.0	0.267	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-5.35.0	0.266	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-6.5.0	0.261	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.4.0	0.254.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-6.3.1	0.245.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.3.0	0.245.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker



亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-6.2.1	0.238.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.2.0	0.238.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-6.1.1	0.232	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.1.0	0.232	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-6.0.1	0.230	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.0.0	0.230	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-5.34.0	0.261	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-5.33.1	0.245.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-5.33.0	0.245.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-5.32.1	0.240.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-5.32.0	0.240.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-5.31.1	0.238.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-5.31.0	0.238.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-5.30.2	0.232	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-5.30.1	0.232	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-5.30.0	0.232	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker



亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-5.29.0	0.227	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.28.1	0.227	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.28.0	0.227	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mysql-server, presto-coordinator, presto-worker

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-5.27.1	0.224	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.27.0	0.224	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.26.0	0.220	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-5.25.0	0.220	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.24.1	0.219	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.24.0	0.219	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-5.23.1	0.215	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.23.0	0.215	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.22.0	0.215	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-5.21.2	0.215	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.21.1	0.215	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.21.0	0.215	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-5.20.1	0.214	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.20.0	0.214	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.19.1	0.212	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-5.19.0	0.212	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.18.1	0.210	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.18.0	0.210	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-5.17.2	0.206	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.17.1	0.206	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.17.0	0.206	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker



亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-5.16.1	0.203	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.16.0	0.203	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.15.1	0.194	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-5.15.0	0.194	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.14.2	0.194	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.14.1	0.194	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-5.14.0	0.194	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.13.1	0.194	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.13.0	0.194	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-5.12.3	0.188	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.12.2	0.188	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.12.1	0.188	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-5.12.0	0.188	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.11.4	0.187	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.11.3	0.187	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-5.11.2	0.187	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.11.1	0.187	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.11.0	0.187	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-5.10.1	0.187	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.10.0	0.187	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.9.1	0.184	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-5.9.0	0.184	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.8.3	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.8.2	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker



亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-5.8.1	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.8.0	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.7.1	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-5.7.0	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.6.1	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.6.0	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-5.5.4	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.5.3	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.5.2	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-5.5.1	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.5.0	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.4.1	0.166	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-5.4.0	0.166	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.3.2	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.3.1	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-5.3.0	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.2.3	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.2.2	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-5.2.1	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.2.0	0.152.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.1.1	0.152.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-5.1.0	0.152.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.0.3	0.152.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.0.2	0.150	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker



亚马逊EMR发布标签	Presto 版本	随 Presto 安装的组件
emr-5.0.1	0.150	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.0.0	0.150	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

下表列出了亚马逊每个发行版本中包含的 Trino (PrestoSQL) 版本EMR，以及随应用程序一起安装的组件。Presto 从 351 版本开始SQL更名为 Trino。

#### Trino (PrestoSQL) 版本信息

亚马逊EMR发布标签	Trino ( Presto ) 版本 SQL	与 Trino (Prest SQL o) 一起安装的组件
emr-7.2.0	436	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodema

亚马逊EMR发布标签	Trino ( Presto ) 版本 SQL	与 Trino (Prest SQL o) 一起安装的组件
		nager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-7.1.0	435	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-7.0.0	426	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker

亚马逊EMR发布标签	Trino ( Presto ) 版本 SQL	与 Trino (Prest SQL o) 一起安装的组件
emr-6.15.0	426	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-6.14.0	422	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker

亚马逊EMR发布标签	Trino ( Presto ) 版本 SQL	与 Trino (Prest SQL o) 一起安装的组件
emr-6.13.0	414	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-6.12.0	414	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker

亚马逊EMR发布标签	Trino ( Presto ) 版本 SQL	与 Trino (Prest SQL o) 一起安装的组件
emr-6.11.1	410	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-6.11.0	410	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker

亚马逊EMR发布标签	Trino ( Presto ) 版本 SQL	与 Trino (Prest SQL o) 一起安装的组件
emr-6.10.1	403	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-6.10.0	403	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker

亚马逊EMR发布标签	Trino ( Presto ) 版本 SQL	与 Trino (Prest SQL o) 一起安装的组件
emr-6.9.1	398	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-6.9.0	398	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker

亚马逊EMR发布标签	Trino ( Presto ) 版本 SQL	与 Trino (Prest SQL o) 一起安装的组件
emr-6.8.1	388	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-6.8.0	388	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker



亚马逊EMR发布标签	Trino ( Presto ) 版本 SQL	与 Trino (Prest SQL o) 一起安装的组件
emr-6.7.0	378	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-6.6.0	367	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker

亚马逊EMR发布标签	Trino ( Presto ) 版本 SQL	与 Trino (Prest SQL o) 一起安装的组件
emr-6.5.0	360	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-6.4.0	359	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker

亚马逊EMR发布标签	Trino ( Presto ) 版本 SQL	与 Trino (Prest SQL o) 一起安装的组件
emr-6.3.1	350	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-prestosql, hcatalog-server, mariadb-server, prestosql-coordinator, prestosql-worker
emr-6.3.0	350	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-prestosql, hcatalog-server, mariadb-server, prestosql-coordinator, prestosql-worker

亚马逊EMR发布标签	Trino ( Presto ) 版本 SQL	与 Trino (Prest SQL o) 一起安装的组件
emr-6.2.1	343	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-prestosql, hcatalog-server, mariadb-server, prestosql-coordinator, prestosql-worker
emr-6.2.0	343	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-prestosql, hcatalog-server, mariadb-server, prestosql-coordinator, prestosql-worker

亚马逊EMR发布标签	Trino ( Presto ) 版本 SQL	与 Trino (Prest SQL o) 一起安装的组件
emr-6.1.1	338	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-prestosql, hcatalog-server, mariadb-server, prestosql-coordinator, prestosql-worker
emr-6.1.0	338	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-prestosql, hcatalog-server, mariadb-server, prestosql-coordinator, prestosql-worker

## 各版本的 Trino (PrestoSQL) 发行说明

- [亚马逊 EMR 6.9.0-Trino \(Presto\) 发行说明 SQL](#)

## 亚马逊 EMR 6.9.0-Trino (Presto) 发行说明 SQL

### 亚马逊 EMR 6.9.0-Trino (Presto) 新功能 SQL

- 为了支持长时间运行的查询，Trino 现在包括容错执行机制。容错执行通过重试失败的查询或其组件任务来减少查询失败。有关更多信息，请参阅 [Trino 中的容错执行](#)。

### 亚马逊 EMR 6.9.0-Trino ( Presto ) 的变化 SQL

#### 亚马逊 EMR 6.9.0-PrestoDB 变更

类型	描述
Upgrade	将 PrestoDB 升级到 0.276
Upgrade	对于 Hadoop 3.3.3 的支持
Upgrade	将 Hudi 升级到 0.12.1
功能	Amazon EMR 和 Presto 与 Presto 集成 AWS Lake Formation ，用于交互式GCSCAPI工作负载。
功能	Security Configs for PrestoDB 中添加了与 Kerberos 相关的配置，以启用 Keberos。
错误修复	已恢复OSS拉取请求 #18115，添加该请求是为了减少hdfsConfiguration 副本数量。这会导致在使用EMRFS或 Hudi 表时HDFS配置副本不正确。

### 亚马逊 EMR 6.9.0-Trino 的变化

类型	描述
Upgrade	将 Trino 升级到 398
Upgrade	对于 Hadoop 3.3.3 的支持

类型	描述
功能	Tardigrade 支持：在 Amazon HDFS 上添加对交易所后台处理的支持。有关更多信息，请参阅 <a href="#">Trino 中的容错执行</a>
错误修复	当使用 Trino Iceberg 并启用 Glue 目录时，避免在 <code>iceberg.properties</code> 中添加元存储 URI

### 亚马逊 EMR 6.9.0-Trino (PrestoSQL) 已知问题

- 对于亚马逊EMR版本 6.9.0，Trino 不适用于启用 Apache Ranger 的集群。如果您需要将 Trino 与 Ranger 结合使用，请联系 [AWS Support](#)。

# Apache Spark

[Apache Spark](#) 是一种分布式处理框架和编程模型，可帮助您使用 Amazon EMR 集群进行机器学习、流处理或图形分析。Spark 与 Apache Hadoop 类似，也是一款常用于大数据工作负载的开源、分布式处理系统。但是，Spark 与 Hadoop MapReduce 有几个显著的区别。Spark 具有经过优化的定向非循环图 (DAG) 执行引擎，可以主动在内存中缓存数据，这可以提高性能，特别是对于某些算法和交互式查询。

Spark 内在支持使用 Scala、Python 和 Java 编写的应用程序。它还包括几个紧密集成的库，用于 SQL ([Spark SQL](#))、机器学习 ([MLlib](#))、流处理 ([Spark 流媒体](#)) 和图形处理 ([GraphX](#))。这些工具可让您更轻松地在各种使用案例中充分发挥 Spark 框架的优势。

你可以将 Spark 与其他 Hadoop 应用程序一起安装在亚马逊EMR集群上，它还可以利用亚马逊EMR文件系统 (EMRFS) 直接访问亚马逊 S3 中的数据。Hive 还与 Spark 集成，因此你可以使用 HiveContext 对象通过 Spark 运行 Hive 脚本。Hive 上下文作为 sqlContext 包含在 Spark Shell 中。

有关使用 Spark 设置EMR集群和分析示例数据集的示例教程，请参阅 AWS 新闻博客EMR上的[教程：Amazon 入门](#)。

## Important

[Apache Spark 版本 2.3.1，从亚马逊EMR版本 5.16.0 开始推出，地址CVE为 -2018-8024 和 -2018-1334。CVE建议您将 Spark 的早期版本迁移到 Spark 2.3.1 版本或更高版本。](#)

下表列出了最新版本的亚马逊 EMR 7.x 系列中包含的 Spark 版本，以及亚马逊与 Spark 一起EMR安装的组件。

有关此版本中与 Spark 一起安装的组件的版本，请参阅 [7.2.0 版本的组件版本](#)。

emr-7.2.0 的 Spark 版本信息

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-7.2.0	Spark 3.5.1	delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-librar



Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
		y, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

下表列出了最新版本的亚马逊 EMR 6.x 系列中包含的 Spark 版本，以及亚马逊与 Spark 一起EMR安装的组件。

有关此发行版中随 Spark 安装的组件版本，请参阅[发行版 6.15.0 组件版本](#)。

#### emr-6.15.0 的 Spark 版本信息

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-6.15.0	Spark 3.4.1	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

**Note**

亚马逊EMR版本 6.8.0 附带了 Apache Spark 3.0。此 Spark 发行版使用 Apache Log4j 2 和 `log4j2.properties` 文件，在 Spark 进程中配置 Log4j。如果您在集群中使用 Spark 或使用自定义配置参数创建EMR集群，并且想要升级到 Amazon EMR 版本 6.8.0，则必须迁移到 Apache Log4j 2 的新 `spark-log4j2` 配置分类和密钥格式。有关更多信息，请参阅 [从 Apache Log4j 1.x 迁移到 Log4j 2.x](#)。

下表列出了最新版本的亚马逊 EMR 5.x 系列中包含的 Spark 版本，以及亚马逊与 Spark 一起EMR安装的组件。

有关此版本中与 Spark 一起安装的组件的版本，请参阅 [5.36.2 版组件版本](#)。

## emr-5.36.2 的 Spark 版本信息

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.36.2	Spark 2.4.8	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

## 主题

- [使用 Apache Spark 创建集群](#)
- [在亚马逊 EMR 6.x 上使用 Docker 运行 Spark 应用程序](#)

- [使用 AWS Glue 数据目录作为 Spark 的元数据库 SQL](#)
- [配置 Spark](#)
- [优化 Spark 性能](#)
- [Spark 结果片段缓存](#)
- [使用 Apache Spark RAPIDS 的 Nvidia 加速器](#)
- [访问 Spark Shell](#)
- [使用 Amazon SageMaker Spark 进行机器学习](#)
- [编写 Spark 应用程序](#)
- [使用 Amazon S3 提高 Spark 性能](#)
- [添加 Spark 步骤](#)
- [查看 Spark 应用程序历史记录](#)
- [访问 Spark 网站 UIs](#)
- [使用 Spark 结构化流媒体 Amazon Kinesis Data Streams 连接器](#)
- [使用亚马逊 Redshift 与 Apache Spark 的集成与亚马逊 EMR](#)
- [Spark 发行历史记录](#)

## 使用 Apache Spark 创建集群

以下过程使用亚马逊EMR控制台中的快速选项创建安装了 [Spark](#) 的集群。

作为替代，您可以使用 Advanced Options (高级选项) 进一步自定义您的集群设置，或是提交步骤以编程方式安装应用程序，然后执行自定义应用程序。无论哪个集群创建选项，您都可以选择使用 AWS Glue 作为 Spark SQL 元数据仓。请参阅[使用 AWS Glue 数据目录作为 Spark 的元数据库 SQL](#)了解更多信息。

启动安装了 Spark 的集群

1. 在 <https://console.aws.amazon.com/emr> 上打开亚马逊EMR控制台。
2. 选择 Create cluster (创建集群) 以使用 Quick Options (快速选项)。
3. 输入 Cluster name (集群名称)。您的集群名称不能包含字符 <、>、\$、| 或 ` (反引号)。
4. 在 Software Configuration (软件配置) 中，选择 Release (发行版) 选项。
5. 在 Applications (应用程序) 中，选择 Spark 应用程序捆绑包。

## 6. 根据需要进行其他选项，然后选择 Create cluster (创建集群)。

### Note

要在创建集群时配置 Spark，请参阅[配置 Spark](#)。

要启动安装了 Spark 的集群，请使用 AWS CLI

- 使用下面的命令创建集群。

```
aws emr create-cluster --name "Spark cluster" --release-label emr-7.2.0 --
applications Name=Spark \
--ec2-attributes KeyName=myKey --instance-type m5.xlarge --instance-count 3 --use-
default-roles
```

### Note

为了便于读取，包含 Linux 行继续符 (\)。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (^)。

使用 SDK 适用于 Java 的，启动安装了 Spark 的集群

通过 SupportedProductConfig 中使用的 RunJobFlowRequest 指定 Spark 作为应用程序。

- 下面的实例显示如何通过 Java 使用 Spark 创建集群：

```
import com.amazonaws.AmazonClientException;
import com.amazonaws.auth.AWSCredentials;
import com.amazonaws.auth.AWSStaticCredentialsProvider;
import com.amazonaws.auth.profile.ProfileCredentialsProvider;
import com.amazonaws.services.elasticmapreduce.AmazonElasticMapReduce;
import com.amazonaws.services.elasticmapreduce.AmazonElasticMapReduceClientBuilder;
import com.amazonaws.services.elasticmapreduce.model.*;
import com.amazonaws.services.elasticmapreduce.util.StepFactory;

public class Main {
```

```
public static void main(String[] args) {
    AWSCredentials credentials_profile = null;
    try {
        credentials_profile = new
ProfileCredentialsProvider("default").getCredentials();
    } catch (Exception e) {
        throw new AmazonClientException(
            "Cannot load credentials from .aws/
credentials file. " +
                "Make sure that the
credentials file exists and the profile name is specified within it.",
            e);
    }

    AmazonElasticMapReduce emr =
AmazonElasticMapReduceClientBuilder.standard()
        .withCredentials(new
AWSStaticCredentialsProvider(credentials_profile))
        .withRegion(Regions.US_WEST_1)
        .build();

    // create a step to enable debugging in the AWS Management Console
    StepFactory stepFactory = new StepFactory();
    StepConfig enableddebugging = new StepConfig()
        .withName("Enable debugging")
        .withActionOnFailure("TERMINATE_JOB_FLOW")

.withHadoopJarStep(stepFactory.newEnableDebuggingStep());

    Application spark = new Application().withName("Spark");

    RunJobFlowRequest request = new RunJobFlowRequest()
        .withName("Spark Cluster")
        .withReleaseLabel("emr-5.20.0")
        .withSteps(enableddebugging)
        .withApplications(spark)
        .withLogUri("s3://path/to/my/logs/")
        .withServiceRole("EMR_DefaultRole")
        .withJobFlowRole("EMR_EC2_DefaultRole")
        .withInstances(new JobFlowInstancesConfig()
            .withEc2SubnetId("subnet-12ab3c45")
            .withEc2KeyName("myEc2Key")
            .withInstanceCount(3))
    }
```

```
.withKeepJobFlowAliveWhenNoSteps(true)
                                .withMasterInstanceType("m4.large")

.withSlaveInstanceType("m4.large"));
    RunJobFlowResult result = emr.runJobFlow(request);
    System.out.println("The cluster ID is " + result.toString());
}
}
```

## 在亚马逊 EMR 6.x 上使用 Docker 运行 Spark 应用程序

在 Amazon EMR 6.0.0 中，Spark 应用程序可以使用 Docker 容器来定义其库依赖关系，而不必在集群中的各个亚马逊 EC2 实例上安装依赖关系。要使用 Docker 运行 Spark，您必须首先配置 Docker 注册表，并在提交 Spark 应用程序时定义其他参数。有关更多信息，请参阅[配置 Docker 集成](#)。

提交应用程序后，YARN 调用 Docker 来提取指定的 Docker 镜像，然后在 Docker 容器内运行 Spark 应用程序。这让您可以轻松定义和隔离依赖项。它缩短了 Amazon EMR 集群中引导或准备实例的时间，其中包含任务执行所需的库。

### 利用 Docker 运行 Spark 时的注意事项

使用 Docker 运行 Spark 时，请确保满足以下先决条件：

- docker 软件包和 CLI 仅安装在核心节点和任务节点上。
- 在 Amazon EMR 6.1.0 及更高版本中，您也可以使用以下命令在主节点上安装 Docker。

```
sudo yum install -y docker
sudo systemctl start docker
```

- 该 spark-submit 命令应始终从 Amazon EMR 集群上的主实例运行。
- 用于解析 Docker 镜像的 Docker 注册表必须使用 API 带有分类密钥的分类进行定义，以便在 container-executor 启动集群时定义其他参数：
  - docker.trusted.registries
  - docker.privileged-containers.registries
- 要在 Docker 容器中执行 Spark 应用程序，需要以下配置选项：
  - YARN\_CONTAINER\_RUNTIME\_TYPE=docker
  - YARN\_CONTAINER\_RUNTIME\_DOCKER\_IMAGE={DOCKER\_IMAGE\_NAME}

- 使用 Amazon ECR 检索 Docker 镜像时，您必须将集群配置为对其进行身份验证。为此，您必须使用以下配置选项：
  - `YARN_CONTAINER_RUNTIME__DOCKER_CLIENT_CONFIG = {DOCKER__CLIENT__PATH_ON_CONFIG_HDFS }`
- 在 Amazon EMR 6.1.0 及更高版本中，启用 ECR 自动身份验证功能 `YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG={DOCKER_CLIENT_CONFIG_PATH_ON_HDFS}` 您无需使用列出的命令。
- 与 Spark 一起使用的任何 Docker 映像都必须在 Docker 映像中安装了 Java。

有关先决条件的更多信息，请参阅[配置 Docker 集成](#)。

## 创建 Docker 镜像

Docker 映像使用 Dockerfile 创建的，该文件定义了要包含在映像中的程序包和配置。以下两个 Dockerfile 使用和 PySpark SparkR 的示例。

### PySpark Dockerfile

使用此 Dockerfile 创建的 Docker 镜像包括 Python 3 和 Python 软件包 NumPy。这个 Dockerfile 使用亚马逊 Linux 2 和 Amazon Corretto 8。JDK

```
FROM amazoncorretto:8

RUN yum -y update
RUN yum -y install yum-utils
RUN yum -y groupinstall development

RUN yum list python3*
RUN yum -y install python3 python3-dev python3-pip python3-virtualenv

RUN python -V
RUN python3 -V

ENV PYSPARK_DRIVER_PYTHON python3
ENV PYSPARK_PYTHON python3

RUN pip3 install --upgrade pip
RUN pip3 install numpy pandas

RUN python3 -c "import numpy as np"
```

## SparkR Dockerfile

使用此 Dockerfile 创建的 Docker 镜像包括 R 和软件包。randomForest CRAN这个 Dockerfile 包括亚马逊 Linux 2 和 Amazon Corretto 8。JDK

```
FROM amazoncorretto:8

RUN java -version

RUN yum -y update
RUN amazon-linux-extras install R4

RUN yum -y install curl hostname

#setup R configs
RUN echo "r <- getOption('repos'); r['CRAN'] <- 'http://cran.us.r-project.org';
  options(repos = r);" > ~/.Rprofile

RUN Rscript -e "install.packages('randomForest')"
```

有关 Dockerfile 语法的更多信息，请参阅 [Dockerfile 参考文档](#)。

## 使用来自亚马逊的 Docker 镜像 ECR

Amazon Elastic Container Registry (Amazon ECR) 是一个完全托管的 Docker 容器注册表，可轻松存储、管理和部署 Docker 容器镜像。使用 Amazon 时 ECR，必须将集群配置为信任您的实例 ECR，并且您必须配置身份验证才能让集群使用来自亚马逊 ECR 的 Docker 映像。有关更多信息，请参阅 [配置 YARN 以访问 Amazon ECR](#)。

要确保 Amazon EMR 主机可以访问存储在 Amazon 中的图像 ECR，您的集群必须拥有与实例配置文件关联的 AmazonEC2ContainerRegistryReadOnly 策略的权限。有关更多信息，请参阅 [AmazonEC2ContainerRegistryReadOnly 策略](#)。

在此示例中，必须使用以下附加配置创建集群，以确保 Amazon ECR 注册表是可信的。更换 *123456789123.dkr.ecr.us-east-1.amazonaws.com* 带有您的 Amazon ECR 端点的终端节点。

```
[
  {
    "Classification": "container-executor",
    "Configurations": [
      {
```



```
    "Classification": "docker",
    "Properties": {
      "docker.privileged-containers.registries":
"local,centos,123456789123.dkr.ecr.us-east-1.amazonaws.com",
      "docker.trusted.registries": "local,centos,123456789123.dkr.ecr.us-
east-1.amazonaws.com"
    }
  ],
  "Properties": {}
}
```

## 在 Amazon PySpark 上使用 ECR

以下示例使用 PySpark Dockerfile，该文件将被标记并上传到亚马逊。ECR上传 Dockerfile 后，你可以运行 PySpark 任务并参考来自亚马逊的 Docker 镜像。ECR

启动集群后，使用SSH连接到核心节点并运行以下命令从 Dock PySpark erfile 示例构建本地 Docker 镜像。

首先，创建一个目录和一个 Dockerfile。

```
mkdir pyspark
vi pyspark/Dockerfile
```

粘贴 PySpark Dockerfile 的内容并运行以下命令来构建 Docker 镜像。

```
sudo docker build -t local/pyspark-example pyspark/
```

为示例创建emr-docker-examplesECR存储库。

```
aws ecr create-repository --repository-name emr-docker-examples
```

标记本地构建的映像并将其上传到ECR，替换 *123456789123.dkr.ecr.us-east-1.amazonaws.com* 使用您的ECR终端节点。

```
sudo docker tag local/pyspark-example 123456789123.dkr.ecr.us-east-1.amazonaws.com/emr-
docker-examples:pyspark-example
sudo docker push 123456789123.dkr.ecr.us-east-1.amazonaws.com/emr-docker-
examples:pyspark-example
```

用于SSH连接到主节点并准备一个带有文件名的 Python 脚本main.py。将以下内容粘贴到 main.py 文件中并保存它。

```
from pyspark.sql import SparkSession
spark = SparkSession.builder.appName("docker-numpy").getOrCreate()
sc = spark.sparkContext

import numpy as np
a = np.arange(15).reshape(3, 5)
print(a)
```

在亚马逊 EMR 6.0.0 上，要提交任务，请引用 Docker 镜像的名称。定义其他配置参数，以确保作业执行使用 Docker 作为运行时。使用亚马逊时 ECR，YARN\_CONTAINER\_RUNTIME\_DOCKER\_CLIENT\_CONFIG必须引用包含用于向亚马逊进行身份验证的凭证的config.json文件ECR。

```
DOCKER_IMAGE_NAME=123456789123.dkr.ecr.us-east-1.amazonaws.com/emr-docker-
examples:pyspark-example
DOCKER_CLIENT_CONFIG=hdfs:///user/hadoop/config.json
spark-submit --master yarn \
--deploy-mode cluster \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_TYPE=docker \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_DOCKER_IMAGE=$DOCKER_IMAGE_NAME \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG=
$DOCKER_CLIENT_CONFIG \
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_TYPE=docker \
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_DOCKER_IMAGE=$DOCKER_IMAGE_NAME \
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG=
$DOCKER_CLIENT_CONFIG \
--num-executors 2 \
main.py -v
```

在 Amazon EMR 6.1.0 及更高版本中，要提交任务，请引用 Docker 镜像的名称。启用ECR自动身份验证后，运行以下命令。

```
DOCKER_IMAGE_NAME=123456789123.dkr.ecr.us-east-1.amazonaws.com/emr-docker-
examples:pyspark-example
spark-submit --master yarn \
--deploy-mode cluster \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_TYPE=docker \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_DOCKER_IMAGE=$DOCKER_IMAGE_NAME \
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_TYPE=docker \
```

```
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_DOCKER_IMAGE=$DOCKER_IMAGE_NAME \
--num-executors 2 \
main.py -v
```

作业完成后，记下YARN应用程序 ID，然后使用以下命令获取 PySpark 作业的输出。

```
yarn logs --applicationId application_id | grep -C2 '\[\[\'
LogLength:55
LogContents:
[[ 0  1  2  3  4]
 [ 5  6  7  8  9]
 [10 11 12 13 14]]
```

### 在亚马逊上使用 SparkR ECR

以下示例使用 SparkR Dockerfile，该文件将被标记并上传到。ECR上传 Dockerfile 后，你可以运行 SparkR 作业并参考来自亚马逊的 Docker 镜像。ECR

启动集群后，使用SSH连接到核心节点并运行以下命令从 SparkR Dockerfile 示例中构建本地 Docker 镜像。

首先，创建一个目录和该 Dockerfile。

```
mkdir sparkr
vi sparkr/Dockerfile
```

粘贴 SparkR Dockerfile 的内容并运行以下命令来构建 Docker 映像。

```
sudo docker build -t local/sparkr-example sparkr/
```

标记本地构建的图像并将其上传到 Amazon ECR，替换 *123456789123.dkr.ecr.us-east-1.amazonaws.com* 使用您的亚马逊 ECR 终端节点。

```
sudo docker tag local/sparkr-example 123456789123.dkr.ecr.us-east-1.amazonaws.com/emr-docker-examples:sparkr-example
sudo docker push 123456789123.dkr.ecr.us-east-1.amazonaws.com/emr-docker-examples:sparkr-example
```

用于SSH连接到主节点并准备一个名为的 R 脚本 sparkR.R。将以下内容粘贴到 sparkR.R 文件中。

```
library(SparkR)
```

```
sparkR.session(appName = "R with Spark example", sparkConfig =
  list(spark.some.config.option = "some-value"))

sqlContext <- sparkRSQL.init(spark.sparkContext)
library(randomForest)
# check release notes of randomForest
rfNews()

sparkR.session.stop()
```

在 Amazon EMR 6.0.0 上，要提交任务，请参阅 Docker 镜像的名称。定义其他配置参数，以确保作业执行使用 Docker 作为运行时。使用 Amazon 时 ECR，YARN\_CONTAINER\_RUNTIME\_DOCKER\_CLIENT\_CONFIG 必须参考包含用于进行身份验证的凭证的 config.json 文件 ECR。

```
DOCKER_IMAGE_NAME=123456789123.dkr.ecr.us-east-1.amazonaws.com/emr-docker-
examples:sparkr-example
DOCKER_CLIENT_CONFIG=hdfs:///user/hadoop/config.json
spark-submit --master yarn \
--deploy-mode cluster \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_TYPE=docker \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_DOCKER_IMAGE=$DOCKER_IMAGE_NAME \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG=
$DOCKER_CLIENT_CONFIG \
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_TYPE=docker \
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_DOCKER_IMAGE=$DOCKER_IMAGE_NAME \
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG=
$DOCKER_CLIENT_CONFIG \
sparkR.R
```

在 Amazon EMR 6.1.0 及更高版本中，要提交任务，请引用 Docker 镜像的名称。启用 ECR 自动身份验证后，运行以下命令。

```
DOCKER_IMAGE_NAME=123456789123.dkr.ecr.us-east-1.amazonaws.com/emr-docker-
examples:sparkr-example
spark-submit --master yarn \
--deploy-mode cluster \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_TYPE=docker \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_DOCKER_IMAGE=$DOCKER_IMAGE_NAME \
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_TYPE=docker \
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_DOCKER_IMAGE=$DOCKER_IMAGE_NAME \
sparkR.R
```

作业完成后，记下YARN应用程序 ID，然后使用以下命令获取 SparkR 作业的输出。此示例包括测试以确保 randomForest 库、已安装的版本和发行说明都可用。

```
yarn logs --applicationId application_id | grep -B4 -A10 "Type rfNews"
randomForest 4.6-14
Type rfNews() to see new features/changes/bug fixes.
Wishlist (formerly TODO):

* Implement the new scheme of handling classwt in classification.

* Use more compact storage of proximity matrix.

* Allow case weights by using the weights in sampling?

=====
Changes in 4.6-14:
```

## 使用 AWS Glue 数据目录作为 Spark 的元数据库 SQL

使用亚马逊EMR版本 5.8.0 或更高版本，您可以将 Spark 配置SQL为使用 Glue 数据 AWS 目录作为其元数据库。当您需要持久的元数据仓或由不同集群、服务、应用程序和 AWS 账户共享的元数据仓时，我们建议使用此配置。

AWS Glue 是一项完全托管的提取、转换和加载 (ETL) 服务，它使对数据进行分类、清理、丰富数据并在各种数据存储之间可靠地移动数据变得简单且经济高效。AWS Glue 数据目录提供了一个跨各种数据源和数据格式的统一元数据存储库，与亚马逊、亚马逊、亚马逊 Redshift、Redshift Spectrum RDS、Athena EMR 以及任何与 Apache Hive 元数据仓兼容的应用程序集成。AWS Glue 爬虫程序可以自动从 Amazon S3 中的源数据推断出架构，并将相关的元数据存储入数据目录中。有关数据目录的更多信息，请参阅 [AWS Glue 开发者指南中的填充 G AWS Glue 数据目录](#)。

Glue 需单独收 AWS 费。在数据目录中存储和访问元数据有月费率，Glue ETL 作业和爬虫运行时按小时费率收费，每个预配置的开发端点按每分钟计费。数据目录让您最多可免费存储一百万个对象。如果您存储的对象超过一百万个，则每超过一百万个 USD 100,000 个对象需要支付 1 美元的费用。数据目录中的对象为表、分区或数据库。有关更多信息，请参阅 [Glue 定价](#)。

### Important

如果您在 2017 年 8 月 14 日之前使用 Amazon Athena 或 Amazon Redshift Spectrum 创建了表，则数据库和表将存储在雅典娜管理的目录中，该目录与 Glue 数据目录是分开的。AWS 要

将亚马逊EMR与这些表集成，您必须升级到 AWS Glue 数据目录。有关更多信息，请参阅亚马逊 Athena AWS 用户指南中的升级到 Glue [数据目录](#)。

## 将 AWS Glue 数据目录指定为元数据库

您可以使用 AWS Management Console、AWS CLI 或 Amazon EMR API 将 AWS Glue 数据目录指定为元数据仓库。使用 CLI 或 API，可以使用 Spark 的配置分类来指定数据目录。此外，在 Amazon EMR 5.16.0 及更高版本中，您可以使用配置分类在不同的目录中指定数据目录。AWS 账户在使用控制台时，您可以使用 Advanced Options (高级选项) 或 Quick Options (快速选项) 指定数据目录。

### Note

齐柏林飞艇还提供使用 AWS Glue Data Catalog 的选项，因为齐柏林飞艇安装了 Spark 组件。SQL

## Console

使用新主机 AWS 将 Glue 数据目录指定为 Spark 元数据仓库

1. 登录 AWS Management Console，然后通过 <https://console.aws.amazon.com/emr> 打开亚马逊 EMR 控制台。
2. EC2 在左侧导航窗格的 Amazon on EMR 下，选择集群，然后选择创建集群。
3. 在 Application bundle (应用程序包) 下，选择 Spark 或 Custom (自定义)。如果您自定义集群，请确保选择 Zeppelin 或 Spark 作为应用程序之一。
4. 在 AWS Glue Data Catalog 设置下，选择用于 Spark 表元数据复选框。
5. 选择适用于集群的任何其他选项。
6. 要启动集群，选择 Create cluster (创建集群)。

## AWS CLI

要将 AWS Glue 数据目录指定为 Spark 元数据库，请使用 AWS CLI

有关使用 AWS CLI 和 Amazon 指定配置分类的更多信息 EMR API，请参阅 [配置应用程序](#)。

- 使用 spark-hive-site 分类指定 `hive.metastore.client.factory.class` 的值，如下例所示：

```
[
  {
    "Classification": "spark-hive-site",
    "Properties": {
      "hive.metastore.client.factory.class":
      "com.amazonaws.glue.catalog.metastore.AWSGlueDataCatalogHiveClientFactory"
    }
  }
]
```

要在其他 AWS 账户中指定数据目录，请添加hive.metastore.glue.catalogid属性，如下示例所示。将 *acct-id* 替换为数据目录的 AWS 账户。

```
[
  {
    "Classification": "spark-hive-site",
    "Properties": {
      "hive.metastore.client.factory.class":
      "com.amazonaws.glue.catalog.metastore.AWSGlueDataCatalogHiveClientFactory",
      "hive.metastore.glue.catalogid": "acct-id"
    }
  }
]
```

## IAM权限

集群的EC2实例配置文件必须具有 Glue AWS 操作的IAM权限。此外，如果您为 AWS Glue Data Catalog 对象启用加密，则还必须允许该角色加密、解密和生成 AWS KMS key 用于加密的对象。

### AWS Glue 操作的权限

如果您使用 Amazon 的默认EC2实例配置文件EMR，则无需执行任何操作。附加到的AmazonElasticMapReduceforEC2Role托管策略EMR\_EC2\_DefaultRole允许执行所有必要的AWS Glue 操作。但是，如果您指定自定义EC2实例配置文件和权限，则必须配置相应的 AWS Glue 操作。使用 AmazonElasticMapReduceforEC2Role 托管策略作为起点。有关更多信息，请参阅 Amazon EMR 管理指南中的[集群EC2EC2实例服务角色（实例配置文件）](#)。

## 加密和解密 AWS Glue 数据目录的权限

您的实例配置文件需要使用密钥加密和解密数据的权限。如果以下语句适用，您不必配置这些权限：

- 您可以使用 Glue 的 AWS 托管密钥为 Glue 数据目录对象启用加密。AWS
- 您使用的集群与 Glue 数据目录 AWS 账户 相同。

否则，您必须将以下语句添加到您的EC2实例配置文件所附的权限策略中。

```
[
  {
    "Version": "2012-10-17",
    "Statement": [
      {
        "Effect": "Allow",
        "Action": [
          "kms:Decrypt",
          "kms:Encrypt",
          "kms:GenerateDataKey"
        ],
        "Resource": "arn:aws:kms:region:acct-
id:key/12345678-1234-1234-1234-123456789012"
      }
    ]
  }
]
```

有关 AWS Glue 数据目录加密的更多信息，请参阅 [Glue 开发者指南中的AWS 加密数据目录](#)。

## 基于资源的权限

如果您在亚马逊中将 AWS Glue 与 Hive、Spark 或 Presto 结合使用，EMR Glue 支持基于资源的策略来控制对数据目录资源的访问。这些资源包括数据库、表、连接和用户定义的函数。有关更多信息，请参阅《AWS Glue 开发人员指南》中的 [AWS Glue 资源策略](#)。

使用基于资源的策略限制从 Amazon EMR 内部访问 AWS Glue 时，您在权限策略中指定的委托人必须是与创建集群时指定的EC2实例配置文件ARN关联的角色。例如，对于附加到目录的基于资源的策略，您可以ARN为集群EC2实例指定默认服务角色的角色，*EMR\_EC2\_DefaultRole* 正如Principal，使用以下示例所示的格式：

```
arn:aws:iam::acct-id:role/EMR_EC2_DefaultRole
```



这些区域有：`acct-id` 可以与 AWS Glue 账号 ID 不同。这允许从不同账户的EMR集群进行访问。您可以指定多个委托人，且每个委托人都可以来自不同的账户。

## 使用 AWS Glue 数据目录时的注意事项

使用 AWS Glue 数据目录作为带有 Spark 的元数据仓时，请考虑以下事项：

- 如果默认数据库没有位置，则在创建表时URI会导致失败。作为解决方法，请在您使用 `LOCATION` 时使用 `s3://EXAMPLE-DOC-BUCKET` 子句指定一个存储桶位置，如 `CREATE TABLE`。或者，在除默认数据库之外的数据库内创建表。
- 不支持在 AWS Glue 中重命名表。
- 当您创建 Hive 表而不指定 `LOCATION` 时，表数据存储在与通过 `hive.metastore.warehouse.dir` 属性指定的位置。默认情况下，这是中的一个位置HDFS。如果另一个集群需要访问该表，则它将失败，除非它有足够的权限访问创建该表的集群。此外，由于HDFS存储是临时性的，因此如果集群终止，表数据就会丢失，必须重新创建表。我们建议您在使用 AWS Glue 创建 Hive 表时`LOCATION`在 Amazon S3 中指定。此外，也可以使用 `hive-site` 配置分类来为 `hive.metastore.warehouse.dir` 指定 Amazon S3 中的位置，它适用于所有 Hive 表。如果在某个HDFS位置创建了表，而创建该表的集群仍在运行，则可以从 Glue 中将该表的位置更新为 Amazon S AWS 3。有关更多信息，请参阅 [《Glue 开发者指南》中的“在 AWS Glue 控制台AWS 上使用表格”](#)。
- 不支持包含引号和撇号的分区值，例如 `PARTITION (owner="Doe's")`。
- `emr-5.31.0` 及更高版本支持[列统计数据](#)。
- 不支持使用 [Hive 授权](#)。作为替代方案，考虑使用[基于AWS Glue 资源的策略](#)。有关更多信息，请参阅[使用基于资源的策略让 Amazon EMR 访问 AWS Glue 数据目录](#)。

## 配置 Spark

您可以使用配置分类在 [Amazon EMR 上配置 Spark](#)。有关使用配置分类的更多信息，请参阅 [配置应用程序](#)。

亚马逊 Spark 的配置分类EMR包括以下内容：

- **spark** – 将 `maximizeResourceAllocation` 属性设置为 `true` 或 `false`。如果为 `true`，Amazon EMR 会根据集群硬件配置自动配置`spark-defaults`属性。有关更多信息，请参阅 [使用 maximizeResourceAllocation](#)。
- **spark-defaults** – 在 `spark-defaults.conf` 文件中设置值。有关更多信息，请参阅 Spark 文档中的 [Spark 配置](#)。

- **spark-env** – 在 `spark-env.sh` 文件中设置值。有关更多信息，请参阅 Spark 文档中的[环境变量](#)。
- **spark-hive-site** – 在 `hive-site.xml` 中为 Spark 设置值。
- **spark-log4j**— ( 亚马逊EMR发布 6.7.x 及更低版本 ) 在文件中`log4j.properties`设置值。有关更多信息，请参阅 Github 上的 [log4j.properties.template](#) 文件。
- **spark-log4j2**— ( Amazon 6.8.0 及更高EMR版本 ) 在文件中`log4j2.properties`设置值。有关更多信息，请参阅 Github 上的 [log4j2.properties.template](#) 文件。
- **spark-metrics** – 在 `metrics.properties` 文件中设置值。有关设置和更多信息，请参阅 Github 上的 [metrics.properties.template](#) 文件和 Spark 文档中的[指标](#)。

#### Note

如果您要将 Spark 工作负载EMR从其他平台迁移到 Amazon，我们建议您在添加自定义配置[亚马逊设置的 Spark 默认值 EMR](#) 之前使用测试您的工作负载。大多数客户都通过我们的默认设置见证了性能有所改善

#### 主题

- [亚马逊设置的 Spark 默认值 EMR](#)
- [在亚马逊 EMR 6.1.0 上配置 Spark 垃圾收集](#)
- [使用 maximizeResourceAllocation](#)
- [配置节点停用行为](#)
- [Spark ThriftServer 环境变量](#)
- [更改 Spark 默认设置](#)
- [从 Apache Log4j 1.x 迁移到 Log4j 2.x](#)

## 亚马逊设置的 Spark 默认值 EMR

下表显示了 Amazon 如何EMR设置影响应用程序`spark-defaults`的默认值。

## 亚马逊设置的 Spark 默认值 EMR

设置	描述	默认值
<code>spark.executor.memory</code>	每个执行程序进程要使用的内存量。例如，1g、2g。	此设置由集群中的核心实例和任务实例类型决定。
<code>spark.executor.cores</code>	要对每个执行程序使用的内核的数量。	此设置由集群中的核心实例和任务实例类型决定。
<code>spark.dynamicAllocation.enabled</code>	如果为 true，则使用动态资源分配，以基于工作负载增大和减小注册到应用程序的执行程序的数目。	true (使用亚马逊 EMR 4.4.0 及更高版本)
<code>spark.sql.hive.advancedPartitionPredicatePushdown.enabled</code>	如果为 true，则启用高级分区谓词下推到 Hive 元数据仓库。	true
<code>spark.sql.hive.stringLikePartitionPredicatePushdown.enabled</code>	将 <code>startsWith</code> 、 <code>contains</code> 和 <code>endsWith</code> 筛选条件向下推送到 Hive 元数据仓库中。	true

 **Note**

Spark 随机播放服务由亚马逊EMR自动配置。

 **Note**

Glue 不支持 `startsWith`、`contains` 或 `endsWith` 谓词下推。如果您使用的是 Glue 元数据仓库，并且由于这些函数的谓词向下推而

设置	描述	默认值
	遇到错误，请将此配置设置为 false。	

## 在亚马逊 EMR 6.1.0 上配置 Spark 垃圾收集

由于与 `spark.driver.extraJavaOptions` Amazon 6.1.0 的垃圾收集配置存在冲突，因此在 Amazon 6.1 中设置自定义垃圾收集配置会导致 `spark.executor.extraJavaOptions` 致 Amazon EMR 6.1 的驱动程序或执行程序启动失败。对于 Amazon EMR 6.1.0，默认的垃圾收集配置是通过 `spark.driver.defaultJavaOptions` 和 `spark.executor.defaultJavaOptions` 设置的。此配置仅适用于亚马逊 EMR 6.1.0。JVM 仍然可以通过设置与垃圾收集无关的选项，例如用于配置日志记录的选项 (`-verbose:class`) `extraJavaOptions`。有关更多信息，请参阅 [Spark 应用程序属性](#)。

## 使用 `maximizeResourceAllocation`

您可以使用 `spark` 配置分类将 `maximizeResourceAllocation` 设置为 `true`，以将执行程序配置为使用集群中的每个节点上尽可能多的资源。特定 `maximizeResourceAllocation` 于 Amazon EMR。启用后 `maximizeResourceAllocation`，Amazon EMR 会计算核心实例组中某个实例上执行者可用的最大计算和内存资源。然后，它将根据计算出的最大值设置相应的 `spark-defaults` 设置。

Amazon EMR 根据核心实例队列中的实例类型计算执行者可用的最大计算和内存资源。由于每个实例队列在队列中可能具有不同的实例类型和大小，因此 Amazon EMR 使用的执行程序配置可能不适合您的集群，因此我们不建议在使用最大资源分配时使用默认设置。为您的实例队列集群配置自定义设置。

### Note

您不应在带有其他分布式应用程序的集群上使用该 `maximizeResourceAllocation` 选项，例如 HBase。Amazon 对分布式应用程序 EMR 使用自定义 YARN 配置，这可能会与 Spark 应用程序冲突 `maximizeResourceAllocation` 并导致 Spark 应用程序失败。

以下是一个 `maximizeResourceAllocation` 设置为 `true` 的 Spark 分类配置。

```
[
```

```

{
  "Classification": "spark",
  "Properties": {
    "maximizeResourceAllocation": "true"
  }
}
]

```

## 启用 `spark-defaults` 时在 `maximizeResourceAllocation` 中配置的设置

设置	描述	值
<code>spark.default.parallelism</code>	join、 <code>parallelize</code> 等转换 RDDs 返回的默认分区数（如果 <code>reduceByKey</code> 用户未设置则为 <code>parallelize</code> ）。	YARN 容器可用 CPU 内核数量的 2 倍。
<code>spark.driver.memory</code>	用于驱动程序进程的内存量， <code>SparkContext</code> 即初始化位置。（例如，1g、2g）。	基于集群中的实例类型配置设置。但是，由于 Spark 驱动程序应用程序可以在主实例或其中一个核心实例上运行（例如，分别在 YARN 客户端和集群模式下），因此它是根据这两个实例组中较小的实例类型进行设置的。
<code>spark.executor.memory</code>	每个执行者进程要使用的内存量。（例如，1g、2g）	基于集群中的核心和任务实例类型配置设置。
<code>spark.executor.cores</code>	要对每个执行程序使用的内核的数量。	基于集群中的核心和任务实例类型配置设置。
<code>spark.executor.instances</code>	执行程序数。	基于集群中的核心和任务实例类型配置设置。除非同时将 <code>spark.dynamicAllocation.enabled</code> 显式设置为 <code>true</code> ，否则将设置。

## 配置节点停用行为

在 Amazon 5.9.0 及更高EMR版本中，Spark on Amazon EMR 包含了一组功能，可帮助确保 Spark 优雅地处理由于手动调整大小或自动扩展策略请求而导致的节点终止。Amazon 在 Spark 中EMR实施了一种建立在YARN停用机制之上的拒绝上架机制。此机制有助于确保不会在即将停用的节点上计划新任务，同时允许正在运行的任务完成。此外，有些功能可以在节点终止导致随机数据块丢失时帮助更快地恢复 Spark 任务。可以更快触发并优化重新计算进程，从而加快重新计算和减少阶段重试，并防止因丢失随机数据块引发的提取失败所导致的任务失败。

### Important

该`spark.decommissioning.timeout.threshold`设置已添加到 Amazon EMR 版本 5.11.0 中，以提高您使用竞价型实例时的 Spark 弹性。在早期发行版中，当节点使用竞价型实例且该实例因出价而终止时，Spark 可能无法正常地处理终止。任务可能失败，而且随机重新计算可能花费大量时间。为此，如果您使用竞价型实例，建议使用发行版 5.11.0 或更高版本。

### Spark 节点停用设置

设置	描述	默认值
<code>spark.blacklist.decommissioning.enabled</code>	设置为 <code>true</code> ，Spark 拒绝会列出 <code>decommissioning</code> 处于状态的节点YARN。Spark 不在于该节点上运行的执行程序上安排新任务。允许已经在运行的任务完成。	<code>true</code>
<code>spark.blacklist.decommissioning.timeout</code>	处于 <code>decommissioning</code> 状态的节点被加入拒绝名单的时间量。默认情况下，此值设置成一小时，这也是 <code>yarn.resourcemanager.decommissioning.timeout</code> 的默认值。要确保节点在其整个退役周期内都处于拒绝名单中，请将此值设置为等于或大于 <code>yarn.resourcemanager</code>	<code>1h</code>

设置	描述	默认值
	<p><code>er.decommissioning.timeout</code>。停用超时到期后，节点将转换为<code>decommissioned</code> 状态，Amazon EMR 可以终止该节点的EC2实例。如果超时过期后有任何任务仍在运行，这些任务会丢失或被终止并在于其他节点上运行的其他执行程序上重新安排。</p>	
<p><code>spark.decommissioning.timeout.threshold</code></p>	<p>在亚马逊 5.11.0 或更高EMR版本中可用。以秒为单位指定。当节点过渡到停用状态时，如果主机将在等于或小于该值的时间段内停用，Amazon EMR 不仅会列出该节点，而且还会在不等待节点过渡到已停用状态的情况下清理主机状态（如所指定<code>spark.resourceManager.cleanupExpiredHost</code> ）。这使 Spark 能够更好地处理竞价型实例终止，因为无论<code>yarn.resourcemanager.decommissioning.timeout</code> 的值如何，Spot 实例都会在 20 秒的超时时间后停用，因此可能没有足够的时间提供其他节点来随机读取文件。</p>	20s

设置	描述	默认值
<code>spark.resourceManager.cleanupExpiredHost</code>	当设置为 <code>true</code> 时，Spark 会注销处于 <code>decommissioned</code> 状态的节点上的执行程序存储的所有已缓存数据和随机数据块。这会加速恢复过程。	<code>true</code>
<code>spark.stage.attempt.ignoreOnDecommissionFetchFailure</code>	当设置为 <code>true</code> 时，有助于防止 Spark 进入失败阶段并因为从退役节点获取失败次数过多而导致任务失败。从处于 <code>decommissioned</code> 状态的节点获取随机数据块失败的次数不会计入连续获取失败的最大数量。	<code>true</code>

## Spark ThriftServer 环境变量

Spark 将 Hive Thrift 服务器端口环境变量 `HIVE_SERVER2_THRIFT_PORT` 设置为 10001。

## 更改 Spark 默认设置

您可以使用 `spark-defaults` 配置分类或 `spark` 配置分类中的 `maximizeResourceAllocation` 设置更改 `spark-defaults.conf` 中的默认值。

以下过程说明如何使用 CLI 或控制台修改设置。

要创建将 `spark.executor.memory` 设置为 2g 的集群，请使用 CLI

- 使用以下命令创建一个安装了 Spark 且 `spark.executor.memory` 设为 2g 的集群，该集群引用存储在 Amazon S3 中的 `myConfig.json` 文件。

```
aws emr create-cluster --release-label emr-7.2.0 --applications Name=Spark \
--instance-type m5.xlarge --instance-count 2 --service-role EMR_DefaultRole_V2
--ec2-attributes InstanceProfile=EMR_EC2_DefaultRole --configurations https://
s3.amazonaws.com/mybucket/myfolder/myConfig.json
```



**Note**

为了便于读取，包含 Linux 行继续符 ( \ )。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 ( ^ )。

myConfig.json:

```
[
  {
    "Classification": "spark-defaults",
    "Properties": {
      "spark.executor.memory": "2G"
    }
  }
]
```

使用控制台创建一个 spark.executor.memory 设为 2g 的集群

1. 导航到新的 Amazon EMR 控制台，然后从侧面导航栏中选择“切换到旧主机”。有关切换到旧控制台后预期情况的更多信息，请参阅 [Using the old console](#)。
2. 依次选择 Create cluster (创建集群)、Go to advanced options (转到高级选项)。
3. 选择 Spark。
4. 在 Edit software settings (编辑软件设置) 下，将 Enter configuration (输入配置) 保留选中状态并输入以下配置：

```
classification=spark-defaults,properties=[spark.executor.memory=2G]
```

5. 选择其他选项，选择 ，然后选择 Create cluster (创建集群)。

要设置 maximizeResourceAllocation

- 使用存储在 Amazon S3 中的文件 AWS CLI，创建安装了 Spark 并将其 maximizeResourceAllocation 设置为 true 的集群。myConfig.json

```
aws emr create-cluster --release-label emr-7.2.0 --applications Name=Spark \
```

```
--instance-type m5.xlarge --instance-count 2 --service-role EMR_DefaultRole_V2
--ec2-attributes InstanceProfile=EMR_EC2_DefaultRole --configurations https://
s3.amazonaws.com/mybucket/myfolder/myConfig.json
```

### Note

为了便于读取，包含 Linux 行继续符 (\)。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (^)。

myConfig.json:

```
[
  {
    "Classification": "spark",
    "Properties": {
      "maximizeResourceAllocation": "true"
    }
  }
]
```

### Note

在 Amazon 5.21.0 及更高EMR版本中，您可以覆盖集群配置，并为正在运行的集群中的每个实例组指定其他配置分类。为此，您可以使用 Amazon EMR 控制台、AWS Command Line Interface (AWS CLI) 或 AWS SDK。有关更多信息，请参阅[为运行的集群中的实例组提供配置](#)。

## 从 Apache Log4j 1.x 迁移到 Log4j 2.x

[Apache Spark](#) 版本 3.2.x 及更早版本使用旧版 Apache Log4j 1.x 和 log4j.properties 文件在 Spark 进程中配置 Log4j。Apache Spark 版本 3.3.0 及更高版本使用 Apache Log4j 2.x 和 log4j2.properties 文件在 Spark 进程中配置 Log4j。

如果您使用低于 6.8.0 的亚马逊EMR版本配置 Apache Spark Log4j，则必须删除旧版spark-log4j配置分类并迁移到spark-log4j2配置分类和密钥格式，然后才能升级到 Amazon 6.8.0 或更高版本。EMR在 Amazon 6.8.0 及更高EMR版本中，旧spark-log4j分类会导致集群创建失败，

并ValidationException出现错误。不会因为与 Log4j 不兼容相关的故障而向您收费，但您必须删除已失效的 spark-log4j 配置分类才能继续。

有关从 Apache Log4j 1.x 迁移到 Log4j 2.x 的更多信息，请参阅 Github 上的《[Apache Log4j 迁移指南](#)》和 [Spark Log4j 2 模板](#)。

#### Note

对于亚马逊EMR，Apache Spark 使用的是log4j2.properties文件，而不是 [Apache Log4j 迁移指南](#)中描述的.xml 文件。此外，我们不建议使用 Log4j 1.x 桥接方法转换为 Log4j 2.x。

## 优化 Spark 性能

亚马逊为 Spark EMR 提供了多种性能优化功能。本主题详细介绍了各个优化功能。

有关如何设置 Spark 配置的更多信息，请参阅 [配置 Spark](#)。

### 自适应查询执行

自适应查询执行是一个根据运行时统计信息重新优化查询计划的框架。从亚马逊 EMR 5.30.0 开始，以下来自 Apache Spark 3 的自适应查询执行优化可在适用于 Spark 2 的 Apache Amazon Runtime 上使用。EMR

- 自适应连接转换
- 随机分区的自适应合并

#### 自适应连接转换

自适应联接转换可根据查询阶段的运行时大小将 broadcast-hash-joins 操作转换为 sort-merge-join 操作，从而提高查询性能。当联接的一侧足够小，可以有效地向所有执行器广播其输出时，B 的性能 broadcast-hash-joins 往往会更好，从而无需对联接的两边进行随机交换和排序。自适应联接转换扩大了 Spark 自动执行 broadcast-hash-joins 的情况范围。

该功能已默认启用。可以通过将 spark.sql.adaptive.enabled 设置为 false 来禁用它，同时会禁用自适应查询执行框架。broadcast-hash-join 当其中一个 sort-merge-join 联接方的运行时大小统计数据不超过（默认为 10,485,760 字节 (10 MiB)）时 spark.sql.autoBroadcastJoinThreshold，Spark 决定将 a 转换为 a。

## 随机分区的自适应合并

随机分区的自适应合并通过合并小的连续随机分区来避免产生太多小任务的开销，从而提高查询性能。这样，您就可以预先配置更多的初始随机分区，然后在运行时将其减少到目标大小，从而提高拥有更均匀分配的随机分区的可能性。

此功能默认情况下已启用，除非 `spark.sql.shuffle.partitions` 采用显式设置。可以通过将 `spark.sql.adaptive.coalescePartitions.enabled` 设置为 `true` 来启用它。初始数量的随机分区和目标分区大小都可以分别使用 `spark.sql.adaptive.coalescePartitions.minPartitionNum` 和 `spark.sql.adaptive.advisoryPartitionSizeInBytes` 属性进行优化。有关此功能的相关 Spark 属性的详细信息，请参阅下表。

### Spark 自适应合并分区属性

属性	默认值	描述
<code>spark.sql.adaptive.coalescePartitions.enabled</code>	<code>true</code> ，除非 <code>spark.sql.shuffle.partitions</code> 为显式设置	如果为 <code>true</code> 且 <code>spark.sql.adaptive.enabled</code> 为 <code>true</code> ，则 Spark 将根据目标大小合并连续的随机分区（通过 <code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code> 指定），以避免过多的小任务。
<code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code>	64MB	合并时，随机分区q的指导大小（按字节计算）。此配置仅在 <code>spark.sql.adaptive.enabled</code> 和 <code>spark.sql.adaptive.coalescePartitions.enabled</code> 两者都为 <code>true</code> 时才有效。
<code>spark.sql.adaptive.coalescePartitions.minPartitionNum</code>	25	合并后的最小随机分区数。此配置仅在 <code>spark.sql.adaptive.enabled</code> 和 <code>spark.sql.adaptive.coalescePartition</code>

属性	默认值	描述
		<code>s.enabled</code> 两者都为 <code>true</code> 时才有效。
<code>spark.sql.adaptive.coalescePartitions.initialPartitionNum</code>	1000	合并前的随机分区的初始数量。此配置仅在 <code>spark.sql.adaptive.enabled</code> 和 <code>spark.sql.adaptive.coalescePartitions.enabled</code> 两者都为 <code>true</code> 时才有效。

## 动态分区修剪

动态分区修剪通过针对特定的查询更准确地选择表中需要读取和处理的特定分区来提高作业性能。通过减少读取和处理的数据量，可节省大量的作业执行时间。在 Amazon EMR 5.26.0 中，此功能默认处于启用状态。在 Amazon EMR 5.24.0 和 5.25.0 中，您可以通过在 Spark 内部或在创建集群时设置 Spark 属性 `spark.sql.dynamicPartitionPruning.enabled` 来启用此功能。

### Spark 动态分区修剪分区属性

属性	默认值	描述
<code>spark.sql.dynamicPartitionPruning.enabled</code>	<code>true</code>	如果为 <code>true</code> ，则启用动态分区修剪。
<code>spark.sql.optimizer.dynamicPartitionPruning.enforceBroadcastReuse</code>	<code>true</code>	为 <code>true</code> 时，Spark 会在查询执行之前执行防御性检查，以确保动态修剪筛选条件中广播交换的重复使用不会被以后的准备规则（如用户定义的列式规则）中断。当重用被中断且此配置是 <code>true</code> 时，Spark 会删除受影响的动态修剪筛选条件，以防止发生性能和正确性问题。当动态修剪筛选条件的

属性	默认值	描述
		广播交换从相应连接操作的广播交换产生不同且不一致的结果时，可能会出现正确性问题。将此配置设置为 false 应谨慎执行；它允许解决如下类似场景：当用户定义的列式规则中断重用。启用“自适应查询执行”后，将始终强制执行广播重用。

这种优化功能在 Spark 2.4.2 的现有功能基础之上进行改进，只支持向下推送可以在计划时解析的静态谓词。

以下是 Spark 2.4.2 中静态谓词向下推送的示例。

```
partition_col = 5

partition_col IN (1,3,5)

partition_col between 1 and 3

partition_col = 1 + 3
```

动态分区修剪允许 Spark 引擎在运行时动态地推断哪些分区需要读取，哪些分区可以安全地消除。例如，以下查询涉及两个表：store\_sales 表，其中包含所有店铺的全部总销售额（按区域分区）；以及 store\_regions 表，其中包含每个国家/地区的区域映射。这些表包含有关分布于全球的存储的数据，但我们只查询北美的数据。

```
select ss.quarter, ss.region, ss.store, ss.total_sales
from store_sales ss, store_regions sr
where ss.region = sr.region and sr.country = 'North America'
```

如果没有动态分区修剪，此查询将读取所有区域，然后过滤出与子查询的结果匹配的区域子集。使用动态分区修剪，此查询将只读取和处理子查询中返回的区域的分区。这样，通过减少数据存储和处理较少的记录，节省了时间和资源。

## 展平标量子查询

这种优化功能通过对同一个表执行标量子查询来提高查询的性能。在 Amazon EMR 5.26.0 中，此功能默认处于启用状态。在 Amazon EMR 5.24.0 和 5.25.0 中，你可以通过在 Spark 中或在创建集群时设置 Spark 属性 `spark.sql.optimizer.flattenScalarSubqueriesWithAggregates.enabled` 来启用它。当此属性设置为 `true` 时，查询优化程序会展平使用相同关系的聚合标量子查询（如果可能）。标量子查询通过以下方法展平：将子查询中存在的任何谓词推送到聚合函数，然后执行一个聚合（针对所有聚合函数，按每个关系）。

以下示例是一个将受益于此优化的查询示例。

```
select (select avg(age) from students          /* Subquery 1 */
        where age between 5 and 10) as group1,
       (select avg(age) from students          /* Subquery 2 */
        where age between 10 and 15) as group2,
       (select avg(age) from students          /* Subquery 3 */
        where age between 15 and 20) as group3
```

此优化将之前的查询重写为：

```
select c1 as group1, c2 as group2, c3 as group3
from (select avg (if(age between 5 and 10, age, null)) as c1,
           avg (if(age between 10 and 15, age, null)) as c2,
           avg (if(age between 15 and 20, age, null)) as c3 from students);
```

请注意，重写的查询将只读取一次 `student` 表，而三个子查询的谓词将推送到 `avg` 函数中。

## DISTINCT之前 INTERSECT

此优化可在使用 `INTERSECT` 时优化联接。在 Amazon EMR 5.26.0 中，此功能默认处于启用状态。在 Amazon EMR 5.24.0 和 5.25.0 中，你可以通过在 Spark 中或在创建集群时设置 Spark 属性 `spark.sql.optimizer.distinctBeforeIntersect.enabled` 来启用它。使用的查询 `INTERSECT` 会自动转换为使用左半联接。当此属性设置为 `true` 时，`INTERSECT` 如果查询优化器检测到 `DISTINCT` 运算符可以将左半联接设为 `a BroadcastHashJoin` 而不是 `a`，则查询优化器会将 `DISTINCT` 运算符推送到的子级。 `SortMergeJoin`

以下示例是一个将受益于此优化的查询示例。

```
(select item.brand brand from store_sales, item
```

```
    where store_sales.item_id = item.item_id)
intersect
(select item.brand cs_brand from catalog_sales, item
    where catalog_sales.item_id = item.item_id)
```

如果没有启用此属性 `spark.sql.optimizer.distinctBeforeIntersect.enabled`，则查询将被重写，如下所示。

```
select distinct brand from
  (select item.brand brand from store_sales, item
    where store_sales.item_id = item.item_id)
left semi join
  (select item.brand cs_brand from catalog_sales, item
    where catalog_sales.item_id = item.item_id)
on brand <=> cs_brand
```

当您启用此属性 `spark.sql.optimizer.distinctBeforeIntersect.enabled` 时，查询将被重写，如下所示。

```
select brand from
  (select distinct item.brand brand from store_sales, item
    where store_sales.item_id = item.item_id)
left semi join
  (select distinct item.brand cs_brand from catalog_sales, item
    where catalog_sales.item_id = item.item_id)
on brand <=> cs_brand
```

## Bloom 筛选条件连接

这种优化可以通过使用从联接另一端的值生成的 [Bloom 筛选条件](#) 对联接的一端进行预筛选，来提高部分联接的性能。在 Amazon EMR 5.26.0 中，此功能默认处于启用状态。在 Amazon EMR 5.25.0 中，您可以通过在 Spark 内部或在创建集群时将 Spark 属性设置 `spark.sql.bloomFilterJoin.enabled` 为 `true` 启用此功能。

下面是一个可以受益于 Bloom 筛选条件的示例查询。

```
select count(*)
from sales, item
where sales.item_id = item.id
and item.category in (1, 10, 16)
```



启用此功能后，Bloom 筛选条件将根据所有类别位于要查询的类别集中的项目ID构建。扫描销售表时，Bloom 筛选条件用于确定哪些销售属于肯定不在 Bloom 筛选条件定义的集中的项目。借此，可以尽早筛选出这些被标识的销售。

## 优化的连接重新排序

这项优化通过将涉及带筛选条件的表的联接进行重新排序来提高查询性能。在 Amazon EMR 5.26.0 中，此功能默认处于启用状态。在 Amazon EMR 5.25.0 中，您可以通过将 Spark 配置参数设置为 `true spark.sql.optimizer.sizeBasedJoinReorder.enabled` 来启用此功能。Spark 的默认行为是从左到右联接表，如查询中所列。此策略可能会错过首先使用筛选条件执行较小联接的机会，以便之后利用更昂贵的联接。

下面的示例查询报告了一个国家/地区所有商店的所有退回商品。如果不经优化的联接重新排序，Spark 首先会联接两个大型表 `store_sales` 和 `store_returns`，然后将其与 `store` 联接，最终再联接 `item`。

```
select ss.item_value, sr.return_date, s.name, i.desc,
from store_sales ss, store_returns sr, store s, item i
where ss.id = sr.id and ss.store_id = s.id and ss.item_id = i.id
and s.country = 'USA'
```

经过优化的联接重新排序，Spark 首先会联接 `store_sales` 与 `store`，因为 `store` 有一个筛选条件并且小于 `store_returns` 和 `broadcastable`。然后，Spark 会联接 `store_returns`，最后联接 `item`。如果 `item` 有一个筛选条件并且可广播，则其也符合重新排序的条件，这会使 `store_sales` 与 `store` 联接，之后联接 `item`，并在最后联接 `store_returns`。

## Spark 结果片段缓存

Amazon EMR 6.0 及更高版本包括可选的 Spark 结果片段缓存功能，该功能可自动缓存结果片段。这些结果片段是查询子树的结果的一部分，其存储在您选择的 Amazon S3 存储桶中。存储的查询结果片段将在后续查询执行时重复使用，从而加快查询速度。

结果片段缓存会分析您的 Spark SQL 查询，并将符合条件的结果片段缓存到您指定的 S3 位置。在后续查询运行中，系统会自动检测并从 S3 中获取可用的查询结果片段。结果片段缓存不同于结果集缓存，其中，后续查询必须与原始查询完全匹配才能从缓存返回结果。当用于重复以静态数据子集为目标的查询时，结果片段缓存可显著提高性能。

请考虑以下查询，它计算 2022 年之前的订单：

```
select
  l_returnflag,
  l_linestatus,
  count(*) as count_order
from
  lineitem
where
  l_shipdate <= current_date
  and year(l_shipdate) == '2022'
group by
  l_returnflag,
  l_linestatus
```

随着时间推移，此查询需要每天运行以报告当年的总销售额。如果没有结果片段缓存，则需要每天重新计算一年中所有日期的结果。随着时间推移，查询速度会变慢，并且在年底最慢，届时将需要重新计算所有 365 天的结果。

当您激活结果片段缓存时，将使用缓存中一年所有以前日期的结果。每天，该功能只能重新计算一天的结果。在该功能计算结果片段后，该功能将缓存片段。因此，启用缓存的查询时间很快，并且每次后续查询都保持不变。

## 启用 Spark 结果片段缓存

要启用 Spark 结果片段缓存，请执行以下步骤：

1. 在 Amazon S3 中创建缓存存储桶并授予读/写访问权限。EMRFS有关更多信息，请参阅 [授权访问 Amazon S3 中的EMRFS数据](#)。
2. 设置 Amazon EMR Spark 配置以启用该功能。

```
spark.subResultCache.enabled = true
spark.subResultCache.fs.root.path = s3://DOC-EXAMPLE-BUCKET/cache_dir/
```

3. 为存储桶启用 S3 生命周期管理以自动清理缓存文件。
4. 或者，配置 `reductionRationThreshold` 和 `maxBufferSize` 属性以进一步调整该功能。

```
spark.sql.subResultCache.reductionRatioThreshold
spark.sql.subResultCache.maxBufferSize
```

## 使用结果片段缓存时的注意事项

当您使用已缓存在 Amazon S3 中的结果而不是重新计算它们时，所节省的成本会随着使用相同缓存结果的次数而增加。对于具有大表扫描后跟筛选条件或散列聚合，并且将结果大小减少至少 8 倍（即输入大小:结果的比率至少为 8:1）的查询将从此功能受益最多。输入和结果之间的缩减率越大，成本效益就越大。只要生成结果的成本高于从 Amazon S3 获取结果的成本，缩减率较小、但在表扫描和筛选条件或聚合之间包含昂贵计算步骤的查询也将受益。默认情况下，结果片段缓存仅在检测到缩减率至少为 8:1 时才生效。

当查询重复使用缓存的结果时，此功能的好处最大。滚动和增量窗口查询就是很好的例子。例如，一个 30 天滚动窗口查询已经运行了 29 天，它只需要从其原始输入源提取 1/30 的目标数据，并将使用前 29 天的缓存结果片段。增量窗口查询将受益更多，因为窗口的开始保持固定：在每次调用查询时，需要从输入源读取的处理比例较小。

以下是使用结果片段缓存时的其他注意事项：

- 如果查询的目标不是具有相同查询片段的相同数据，则缓存命中率较低，因此不会从此功能受益。
- 如果查询的缩减率较低且不包含昂贵的计算步骤，则将导致缓存结果的读取开销与初始处理的开销大致相同。
- 由于写入缓存的成本，第一个查询将始终显示较小的回归。
- 结果片段缓存功能仅适用于 Parquet 文件。不支持其他文件格式。
- 结果片段缓存功能缓冲区将仅尝试缓存文件拆分大小为 128 MB 或更大的扫描。在默认 Spark 配置下，如果扫描大小（正扫描的所有文件的总大小）除以执行程序内核数小于 128 MB，则结果片段缓存将被禁用。如果设置了下面所列的任何 Spark 配置，则文件拆分大小将为：

```
min(maxPartitionBytes, max(openCostInBytes, scan size / minPartitionNum))
```

- spark.sql。leafNodeDefault并行度（默认值为 spark.default.parallelism）
- spark.sql.files。minPartitionNum（默认值为 spark.sql。leafNodeDefault并行度）
- spark.sql.files。openCostIn字节
- spark.sql.files。maxPartitionBytes
- 结果片段缓存功能按RDD分区粒度缓存。先前描述的默认为 8:1 的缩减率是按RDD分区评估的。与每次RDD缩减比率一直低于 8:1 的工作负载相比，每次缩RDD减比率一直低于 8:1 的工作负载获得的性能优势可能较小。
- 默认情况下，结果片段缓存功能对每个正在缓存的RDD分区使用 16MB 的写入缓冲区。如果每个RDD分区的缓存量超过 16MB，则确定不可能写入的成本可能会导致性能下降。

- 虽然默认情况下，Result Fragment Caching 不会尝试缓存缩减比小于 8:1 的RDD分区结果，并且其写入缓冲区上限为 16MB，但这两个值都可以通过以下配置进行调整：

```
spark.sql.subResultCache.reductionRatioThreshold (default: 8.0)
spark.sql.subResultCache.maxBufferSize (default: 16MB, max: 64MB)
```

- 使用同一 Amazon EMR 版本的多个集群可以共享相同的缓存位置。为了确保结果的正确性，结果片段缓存不会使用不同版本的 Amazon 写入的缓存结果。EMR
- 对于 Spark Streaming 用例或使用 Apache Ranger 或时 RecordServer，结果片段缓存将自动禁用。AWS Lake Formation
- 结果片段缓存读/写使用量EMRFS和 Amazon S3 存储桶。CSE/SSSES3/ 支持SSEKMS加密。

## 使用 Apache Spark RAPIDS 的 Nvidia 加速器

在亚马逊EMR版本 6.2.0 及更高版本中，您可以使用 Nvidia 的 [Apache Spark RAPIDS 加速器](#) 插件使用 EC2 图形处理单元 (GPU) 实例类型加速 Spark。RAPIDSGPUAccelerator 无需更改代码即可加速您的 Apache Spark 3.0 数据科学管道，加快数据处理和模型训练，同时大幅降低基础架构成本。

以下各节将指导您配置EMR集群以使用适用于 Spark 的 Spark-RAPIDS 插件。

### 选择实例类型

要使用适用于 Spark 的 Nvidia Spark-RAPIDS 插件，核心实例组和任务EC2GPU实例组必须使用满足 Spark-[硬件要求](#)的实例类型RAPIDS。要查看亚马逊EMR支持的GPU实例类型的完整列表，请参阅亚马逊EMR管理指南中的[支持的实例类型](#)。主实例组的实例类型可以是两种类型，GPU也可以是非GPU类型，但不支持ARM实例类型。

### 为集群设置应用程序配置

1. 允许 Amazon EMR 在您的新集群上安装插件

要安装插件，请在创建集群时提供以下配置：

```
{
  "Classification": "spark",
  "Properties": {
    "enableSparkRapids": "true"
  }
}
```

```
}

```

## 2. 配置YARN为使用 GPU

有关如何在GPU上使用的详细信息YARN，请参阅 Apache Hadoop 文档YARN中的在[GPU上使用](#)。以下示例显示了 Amazon EMR 6.x 和 7.x 版本的示例YARN配置：

### Amazon EMR 7.x

#### 亚马逊 EMR 7. YARN x 的配置示例

```
{
  "Classification": "yarn-site",
  "Properties": {
    "yarn.nodemanager.resource-plugins": "yarn.io/gpu",
    "yarn.resource-types": "yarn.io/gpu",
    "yarn.nodemanager.resource-plugins.gpu.allowed-gpu-devices": "auto",
    "yarn.nodemanager.resource-plugins.gpu.path-to-discovery-executables": "/usr/
bin",
    "yarn.nodemanager.linux-container-executor.cgroups.mount": "true",
    "yarn.nodemanager.linux-container-executor.cgroups.mount-path": "/spark-
rapids-cgroup",
    "yarn.nodemanager.linux-container-executor.cgroups.hierarchy": "yarn",
    "yarn.nodemanager.container-
executor.class": "org.apache.hadoop.yarn.server.nodemanager.LinuxContainerExecutor"
  }
}, {
  "Classification": "container-executor",
  "Properties": {
  },
  "Configurations": [
    {
      "Classification": "gpu",
      "Properties": {
        "module.enabled": "true"
      }
    },
    {
      "Classification": "cgroups",
      "Properties": {
        "root": "/spark-rapids-cgroup",
        "yarn-hierarchy": "yarn"
      }
    }
  ]
}
```

```

    }
  }
]
}

```

## Amazon EMR 6.x

### 亚马逊 EMR 6. YARN x 的配置示例

```

{
  "Classification": "yarn-site",
  "Properties": {
    "yarn.nodemanager.resource-plugins": "yarn.io/gpu",
    "yarn.resource-types": "yarn.io/gpu",
    "yarn.nodemanager.resource-plugins.gpu.allowed-gpu-devices": "auto",
    "yarn.nodemanager.resource-plugins.gpu.path-to-discovery-executables": "/usr/
bin",
    "yarn.nodemanager.linux-container-executor.cgroups.mount": "true",
    "yarn.nodemanager.linux-container-executor.cgroups.mount-path": "/sys/fs/
cgroup",
    "yarn.nodemanager.linux-container-executor.cgroups.hierarchy": "yarn",
    "yarn.nodemanager.container-
executor.class": "org.apache.hadoop.yarn.server.nodemanager.LinuxContainerExecutor"
  }
}, {
  "Classification": "container-executor",
  "Properties": {
  },
  "Configurations": [
    {
      "Classification": "gpu",
      "Properties": {
        "module.enabled": "true"
      }
    },
    {
      "Classification": "cgroups",
      "Properties": {
        "root": "/sys/fs/cgroup",
        "yarn-hierarchy": "yarn"
      }
    }
  ]
}

```

```
}
```

### 3. 配置 Spark 以供使用 RAPIDS

以下是让 Spark 能够使用RAPIDS插件所需的配置：

```
{
  "Classification": "spark-defaults",
  "Properties": {
    "spark.plugins": "com.nvidia.spark.SQLPlugin",
    "spark.executor.resource.gpu.discoveryScript": "/usr/lib/spark/scripts/gpu/getGpusResources.sh",
    "spark.executor.extraLibraryPath": "/usr/local/cuda/targets/x86_64-linux/lib:/usr/local/cuda/extras/CUPTI/lib64:/usr/local/cuda/compat/lib:/usr/local/cuda/lib:/usr/local/cuda/lib64:/usr/lib/hadoop/lib/native:/usr/lib/hadoop-lzo/lib/native:/docker/usr/lib/hadoop/lib/native:/docker/usr/lib/hadoop-lzo/lib/native"
  }
}
```

[XGBoost4J-当集群上启用 Spark RAPIDS 插件时](#)，[XGBoost文档中的 Spark 库](#)也可用。您可以使用以下配置XGBoost与您的 Spark 作业集成：

```
{
  "Classification": "spark-defaults",
  "Properties": {
    "spark.submit.pyFiles": "/usr/lib/spark/jars/xgboost4j-spark_3.0-1.4.2-0.3.0.jar"
  }
}
```

有关可用于调整GPU加速EMR集群的其他 Spark 配置，请参阅 [nvidia.github.io](https://nvidia.github.io) 文档中的[适用于 Apache Spark 的 Rapids Accelerator 调优指南](#)。

### 4. 配置YARN容量调度器

DominantResourceCalculator必须配置为启用GPU调度和隔离。有关更多信息，请参阅 Apache Hadoop YARN 文档中的 [GPU“使用on”](#)。

```
{
  "Classification": "capacity-scheduler",
  "Properties": {
```

```
"yarn.scheduler.capacity.resource-calculator":"org.apache.hadoop.yarn.util.resource.DominantResourceCalculator"
}
}
```

## 5. 创建包含您的配置的JSON文件

您可以创建一个包含您的配置的JSON文件，以便在 Spark 集群中使用该RAPIDS插件。您稍后在启动集群时需提供该文件。

您可以将文件存储在本地或 S3 上。有关如何为集群提供应用程序配置的详细信息，请参阅[配置应用程序](#)。

使用以下示例文件作为模板来构建自己的配置。

### Amazon EMR 7.x

#### 亚马逊 EMR 7.x 的示例my-configurations.json文件

```
[
  {
    "Classification":"spark",
    "Properties":{
      "enableSparkRapids":"true"
    }
  },
  {
    "Classification":"yarn-site",
    "Properties":{
      "yarn.nodemanager.resource-plugins":"yarn.io/gpu",
      "yarn.resource-types":"yarn.io/gpu",
      "yarn.nodemanager.resource-plugins.gpu.allowed-gpu-devices":"auto",
      "yarn.nodemanager.resource-plugins.gpu.path-to-discovery-executables":"/usr/bin",
      "yarn.nodemanager.linux-container-executor.cgroups.mount":"true",
      "yarn.nodemanager.linux-container-executor.cgroups.mount-path":"/spark-rapids-cgroup",
      "yarn.nodemanager.linux-container-executor.cgroups.hierarchy":"yarn",
      "yarn.nodemanager.container-executor.class":"org.apache.hadoop.yarn.server.nodemanager.LinuxContainerExecutor"
    }
  },
  {
    "Classification":"container-executor",
```



```

    "Properties":{
    },
    "Configurations":[
      {
        "Classification":"gpu",
        "Properties":{
          "module.enabled":"true"
        }
      },
      {
        "Classification":"cgroups",
        "Properties":{
          "root":"/spark-rapids-cgroup",
          "yarn-hierarchy":"yarn"
        }
      }
    ]
  },
  {
    "Classification":"spark-defaults",
    "Properties":{
      "spark.plugins":"com.nvidia.spark.SQLPlugin",
      "spark.executor.resource.gpu.discoveryScript":"/usr/lib/spark/scripts/
gpu/getGpusResources.sh",
      "spark.executor.extraLibraryPath":"/usr/local/cuda/targets/x86_64-
linux/lib:/usr/local/cuda/extras/CUPTI/lib64:/usr/local/cuda/compat/lib:/usr/local/
cuda/lib:/usr/local/cuda/lib64:/usr/lib/hadoop/lib/native:/usr/lib/hadoop-lzo/lib/
native:/docker/usr/lib/hadoop/lib/native:/docker/usr/lib/hadoop-lzo/lib/native",
      "spark.submit.pyFiles":"/usr/lib/spark/jars/xgboost4j-
spark_3.0-1.4.2-0.3.0.jar",
      "spark.rapids.sql.concurrentGpuTasks":"1",
      "spark.executor.resource.gpu.amount":"1",
      "spark.executor.cores":"2",
      "spark.task.cpus":"1",
      "spark.task.resource.gpu.amount":"0.5",
      "spark.rapids.memory.pinnedPool.size":"0",
      "spark.executor.memoryOverhead":"2G",
      "spark.locality.wait":"0s",
      "spark.sql.shuffle.partitions":"200",
      "spark.sql.files.maxPartitionBytes":"512m"
    }
  },
  {

```

```

    "Classification": "capacity-scheduler",
    "Properties": {
      "yarn.scheduler.capacity.resource-
calculator": "org.apache.hadoop.yarn.util.resource.DominantResourceCalculator"
    }
  }
]

```

## Amazon EMR 6.x

### 亚马逊 EMR 6.x 的示例 `my-configurations.json` 文件

```

[
  {
    "Classification": "spark",
    "Properties": {
      "enableSparkRapids": "true"
    }
  },
  {
    "Classification": "yarn-site",
    "Properties": {
      "yarn.nodemanager.resource-plugins": "yarn.io/gpu",
      "yarn.resource-types": "yarn.io/gpu",
      "yarn.nodemanager.resource-plugins.gpu.allowed-gpu-devices": "auto",
      "yarn.nodemanager.resource-plugins.gpu.path-to-discovery-executables": "/usr/bin",
      "yarn.nodemanager.linux-container-executor.cgroups.mount": "true",
      "yarn.nodemanager.linux-container-executor.cgroups.mount-path": "/sys/fs/cgroup",
      "yarn.nodemanager.linux-container-executor.cgroups.hierarchy": "yarn",
      "yarn.nodemanager.container-executor.class": "org.apache.hadoop.yarn.server.nodemanager.LinuxContainerExecutor"
    }
  },
  {
    "Classification": "container-executor",
    "Properties": {
    },
    "Configurations": [
      {
        "Classification": "gpu",
        "Properties": {

```

```

        "module.enabled":"true"
      }
    },
    {
      "Classification":"cgroups",
      "Properties":{
        "root":"/sys/fs/cgroup",
        "yarn-hierarchy":"yarn"
      }
    }
  ]
},
{
  "Classification":"spark-defaults",
  "Properties":{
    "spark.plugins":"com.nvidia.spark.SQLPlugin",
    "spark.executor.resource.gpu.discoveryScript":"/usr/lib/spark/scripts/
gpu/getGpusResources.sh",
    "spark.executor.extraLibraryPath":"/usr/local/cuda/targets/x86_64-
linux/lib:/usr/local/cuda/extras/CUPTI/lib64:/usr/local/cuda/compat/lib:/usr/local/
cuda/lib:/usr/local/cuda/lib64:/usr/lib/hadoop/lib/native:/usr/lib/hadoop-lzo/lib/
native:/docker/usr/lib/hadoop/lib/native:/docker/usr/lib/hadoop-lzo/lib/native",
    "spark.submit.pyFiles":"/usr/lib/spark/jars/xgboost4j-
spark_3.0-1.4.2-0.3.0.jar",
    "spark.rapids.sql.concurrentGpuTasks":"1",
    "spark.executor.resource.gpu.amount":"1",
    "spark.executor.cores":"2",
    "spark.task.cpus":"1",
    "spark.task.resource.gpu.amount":"0.5",
    "spark.rapids.memory.pinnedPool.size":"0",
    "spark.executor.memoryOverhead":"2G",
    "spark.locality.wait":"0s",
    "spark.sql.shuffle.partitions":"200",
    "spark.sql.files.maxPartitionBytes":"512m"
  }
},
{
  "Classification":"capacity-scheduler",
  "Properties":{
    "yarn.scheduler.capacity.resource-
calculator":"org.apache.hadoop.yarn.util.resource.DominantResourceCalculator"
  }
}

```

]

## 为您的集群添加引导操作

有关如何在创建集群时提供引导操作脚本的更多信息，请参阅 Amazon EMR 管理指南中的[引导操作基础知识](#)。

以下示例脚本展示了如何为 Amazon EMR 6.x 和 7.x 创建引导操作文件：

### Amazon EMR 7.x

亚马逊 EMR 7.x 的示例 **my-bootstrap-action.sh** 文件

YARN 要使用管理具有 Amazon EMR 7.x 版本的 GPU 资源，您必须在集群 CGroup v1 上手动挂载。您可以使用引导操作脚本来执行此操作，如本示例所示。

```
#!/bin/bash
set -ex

sudo mkdir -p /spark-rapids-cgroup/devices
sudo mount -t cgroup -o devices cgroupv1-devices /spark-rapids-cgroup/devices
sudo chmod a+rx -R /spark-rapids-cgroup
```

### Amazon EMR 6.x

亚马逊 EMR 6.x 的示例 **my-bootstrap-action.sh** 文件

对于 Amazon EMR 6.x 版本，您必须在集群 YARN 上打开 CGroup 权限。您可以使用引导操作脚本来执行此操作，如本示例所示。

```
#!/bin/bash
set -ex

sudo chmod a+rx -R /sys/fs/cgroup/cpu,cpuacct
sudo chmod a+rx -R /sys/fs/cgroup/devices
```

## 启动您的集群。

最后一步是使用上述集群配置启动您的集群。以下是从 Amazon 启动集群的命令示例 EMRCLI：

```
aws emr create-cluster \  
--release-label emr-7.2.0 \  
--applications Name=Hadoop Name=Spark \  
--service-role EMR_DefaultRole_V2 \  
--ec2-attributes KeyName=my-key-pair,InstanceProfile=EMR_EC2_DefaultRole \  
--instance-groups InstanceGroupType=MASTER,InstanceCount=1,InstanceType=m4.4xlarge \  
  
                InstanceGroupType=CORE,InstanceCount=1,InstanceType=g4dn.2xlarge \  
  
                InstanceGroupType=TASK,InstanceCount=1,InstanceType=g4dn.2xlarge \  
--configurations file:///my-configurations.json \  
--bootstrap-actions Name='My Spark Rapids Bootstrap action',Path=s3://my-bucket/my-  
bootstrap-action.sh
```

## 访问 Spark Shell

Spark Shell 基于 Scala REPL (Read-Eval-Print-Loop)。它让您能够以交互方式创建 Spark 程序并将工作提交到框架。您可以通过连接主节点SSH并调用`spark-shell`来访问 Spark 外壳。有关连接到主节点的更多信息，请参阅 Amazon EMR 管理指南SSH中的[使用连接到主节点](#)。以下示例使用存储在 Amazon S3 中的 Apache HTTP 服务器访问日志。

### Note

这些示例中的存储桶对能够访问美国东部（弗吉尼亚州北部）的客户端可用。

默认情况下，Spark 外壳会创建自己的名为的[SparkContext](#)对象`sc`。如果需要在中使用此上下文，则可以使用该上下文REPL。 `sqlContext`也有外壳可供选择，它是 [HiveContext](#)。

Example 使用 Spark Shell 统计存储在 Amazon S3 中的某个文件中的某个字符串的出现次数

本示例使用 `sc` 读取存储在 Amazon S3 中的文本文件。

```
scala> sc  
res0: org.apache.spark.SparkContext = org.apache.spark.SparkContext@404721db  
  
scala> val textFile = sc.textFile("s3://elasticmapreduce/samples/hive-ads/tables/  
impressions/dt=2009-04-13-08-05/ec2-0-51-75-39.amazon.com-2009-04-13-08-05.log")
```

Spark 创建 `textFile` 和关联的[数据结构](#)。然后，示例会统计此日志文件中包含字符串“cartoonnetwork.com”的行数：

```
scala> val linesWithCartoonNetwork = textFile.filter(line =>
  line.contains("cartoonnetwork.com")).count()
linesWithCartoonNetwork: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[2] at
  filter at <console>:23
<snip>
<Spark program runs>
scala> linesWithCartoonNetwork
res2: Long = 9
```

Example 使用基于 Python 的 Spark Shell 统计存储在 Amazon S3 中的某个文件中的某个字符串的出现次数

Spark 还包含一个基于 Python 的 Shell `pyspark`，您可以用它来设计以 Python 编写的 Spark 程序的原型。就像在主节点 `pyspark` 上调用一样；它也有相同的[SparkContext](#)对象。 `spark-shell`

```
>>> sc
<pyspark.context.SparkContext object at 0x7fe7e659fa50>
>>> textfile = sc.textFile("s3://elasticmapreduce/samples/hive-ads/tables/impressions/
dt=2009-04-13-08-05/ec2-0-51-75-39.amazon.com-2009-04-13-08-05.log")
```

Spark 创建 `textFile` 和关联的[数据结构](#)。然后，示例会统计此日志文件中包含字符串“cartoonnetwork.com”的行数。

```
>>> linesWithCartoonNetwork = textfile.filter(lambda line: "cartoonnetwork.com" in
  line).count()
15/06/04 17:12:22 INFO lzo.GPLNativeCodeLoader: Loaded native gpl library from the
  embedded binaries
15/06/04 17:12:22 INFO lzo.LzoCodec: Successfully loaded & initialized native-lzo
  library [hadoop-lzo rev EXAMPLE]
15/06/04 17:12:23 INFO fs.EmrFileSystem: Consistency disabled, using
  com.amazon.ws.emr.hadoop.fs.s3n.S3NativeFileSystem as filesystem implementation
<snip>
<Spark program continues>
>>> linesWithCartoonNetwork
9
```

## 使用 Amazon SageMaker Spark 进行机器学习

使用亚马逊EMR版本 5.11.0 及更高版本时，该aws-sagemaker-spark-sdk组件将与 Spark 一起安装。此组件安装 Amazon SageMaker Spark 和相关依赖项，以便将 Spark 与[亚马逊](#)集成 SageMaker。您可以使用 Amazon SageMaker Spark 通过亚马逊 SageMaker 阶段构建 Spark 机器学习 (ML) 管道。有关更多信息，请参阅《[亚马逊 SageMaker 开发者指南](#)》中的 [Amaz README on SageMaker Spark GitHub](#) 和 SageMaker在[亚马逊上使用 Apache Spark](#)。

## 编写 Spark 应用程序

可使用 [Scala](#)、[Java](#) 或 [Python](#) 来编写 Spark 应用程序。Apache Spark 文档的 [Spark 示例](#)主题包含多个 Spark 应用程序示例。下面所示为三个内在支持的应用程序中的 Estimating Pi 示例。您还可以在\$SPARK\_HOME/examples和中查看完整的示例[GitHub](#)。有关如何为 Spark 编译JARs的更多信息，请参阅 Apache Spark 文档中的[快速入门](#)主题。

## Scala

为避免 Scala 兼容性问题，我们建议您在为亚马逊EMR集群编译 Spark 应用程序时使用正确的 Scala 版本的 Spark 依赖项。您应该使用的 Scala 版本取决于您的集群上安装的 Spark 版本。例如，亚马逊EMR版本5.30.1使用的是使用Scala 2.11构建的Spark 2.4.5。如果您的集群使用亚马逊EMR版本 5.30.1，请使用适用于 Scala 2.11 的 Spark 依赖项。有关 Spark 使用的 Scala 版本的更多信息，请参阅 [Apache Spark 文档](#)

```
package org.apache.spark.examples
import scala.math.random
import org.apache.spark._

/** Computes an approximation to pi */
object SparkPi {
  def main(args: Array[String]) {
    val conf = new SparkConf().setAppName("Spark Pi")
    val spark = new SparkContext(conf)
    val slices = if (args.length > 0) args(0).toInt else 2
    val n = math.min(100000L * slices, Int.MaxValue).toInt // avoid overflow
    val count = spark.parallelize(1 until n, slices).map { i =>
      val x = random * 2 - 1
      val y = random * 2 - 1
      if (x*x + y*y < 1) 1 else 0
    }.reduce(_ + _)
  }
}
```

```
println("Pi is roughly " + 4.0 * count / n)
spark.stop()
}
}
```

## Java

```
package org.apache.spark.examples;

import org.apache.spark.SparkConf;
import org.apache.spark.api.java.JavaRDD;
import org.apache.spark.api.java.JavaSparkContext;
import org.apache.spark.api.java.function.Function;
import org.apache.spark.api.java.function.Function2;

import java.util.ArrayList;
import java.util.List;

/**
 * Computes an approximation to pi
 * Usage: JavaSparkPi [slices]
 */
public final class JavaSparkPi {

    public static void main(String[] args) throws Exception {
        SparkConf sparkConf = new SparkConf().setAppName("JavaSparkPi");
        JavaSparkContext jsc = new JavaSparkContext(sparkConf);

        int slices = (args.length == 1) ? Integer.parseInt(args[0]) : 2;
        int n = 100000 * slices;
        List<Integer> l = new ArrayList<Integer>(n);
        for (int i = 0; i < n; i++) {
            l.add(i);
        }

        JavaRDD<Integer> dataSet = jsc.parallelize(l, slices);

        int count = dataSet.map(new Function<Integer, Integer>() {
            @Override
            public Integer call(Integer integer) {
                double x = Math.random() * 2 - 1;
                double y = Math.random() * 2 - 1;
                return (x * x + y * y < 1) ? 1 : 0;
            }
        }).count();
    }
}
```



```
    }
  }).reduce(new Function2<Integer, Integer, Integer>() {
    @Override
    public Integer call(Integer integer, Integer integer2) {
      return integer + integer2;
    }
  });

  System.out.println("Pi is roughly " + 4.0 * count / n);

  jsc.stop();
}
}
```

## Python

```
import argparse
import logging
from operator import add
from random import random

from pyspark.sql import SparkSession

logger = logging.getLogger(__name__)
logging.basicConfig(level=logging.INFO, format="%(levelname)s: %(message)s")

def calculate_pi(partitions, output_uri):
    """
    Calculates pi by testing a large number of random numbers against a unit circle
    inscribed inside a square. The trials are partitioned so they can be run in
    parallel on cluster instances.

    :param partitions: The number of partitions to use for the calculation.
    :param output_uri: The URI where the output is written, typically an Amazon S3
        bucket, such as 's3://example-bucket/pi-calc'.
    """

    def calculate_hit(_):
        x = random() * 2 - 1
        y = random() * 2 - 1
        return 1 if x**2 + y**2 < 1 else 0
```

```
tries = 100000 * partitions
logger.info(
    "Calculating pi with a total of %s tries in %s partitions.", tries, partitions
)
with SparkSession.builder.appName("My PyPi").getOrCreate() as spark:
    hits = (
        spark.sparkContext.parallelize(range(tries), partitions)
        .map(calculate_hit)
        .reduce(add)
    )
    pi = 4.0 * hits / tries
    logger.info("%s tries and %s hits gives pi estimate of %s.", tries, hits, pi)
    if output_uri is not None:
        df = spark.createDataFrame([(tries, hits, pi)], ["tries", "hits", "pi"])
        df.write.mode("overwrite").json(output_uri)

if __name__ == "__main__":
    parser = argparse.ArgumentParser()
    parser.add_argument(
        "--partitions",
        default=2,
        type=int,
        help="The number of parallel partitions to use when calculating pi.",
    )
    parser.add_argument(
        "--output_uri", help="The URI where output is saved, typically an S3 bucket."
    )
    args = parser.parse_args()

    calculate_pi(args.partitions, args.output_uri)
```

## 使用 Amazon S3 提高 Spark 性能

当使用 Spark 查询、读取和写入保存在亚马逊 S3 中的数据时，亚马逊提供的功能可帮助优化性能。EMR

[S3 Select](#) 可以通过将处理“下推”到 Amazon S3 来提高某些应用程序中 CSV 和 JSON 文件的查询性能。

EMRFSS3 优化的提交者是该[OutputCommitter](#)类的替代方案，在使用 Spark SQL、和数据集将 Parquet 文件写入 Amazon S3 时，该类使用分段上传功能来提高性能。EMRFS DataFrames

## 主题

- [将 S3 Select 与 Spark 结合使用以提高查询性能](#)
- [使用 EMRFS S3 优化的提交器](#)
- [使用 EMRFS S3 优化的提交协议](#)
- [使用重试 Amazon S3 请求 EMRFS](#)

## 将 S3 Select 与 Spark 结合使用以提高查询性能

在亚马逊EMR版本 5.17.0 及更高版本中，你可以在亚马逊上使用带有 Spark 的 [S3 Select](#)。EMRS3 Select 可让应用程序仅从对象检索数据子集。对于亚马逊来说EMR，筛选大型数据集进行处理的计算工作从集群“下推”到Amazon S3，这可以提高某些应用程序的性能并减少亚马逊EMR和Amazon S3之间传输的数据量。

S3 Select 支持使用CSVs3selectCSV和s3selectJSON值来指定数据格式的JSON文件。有关更多信息以及示例，请参阅 [在代码中指定 S3 Select](#)。

## S3 Select 是否适合我的应用程序？

建议您分别在使用和不使用 S3 Select 的情况下测试您的应用程序，以查看 S3 Select 是否适用于您的应用程序。

使用以下准则来确定您的应用程序是否为使用 S3 Select 的候选项：

- 您的查询将筛选掉原始数据集的一半以上的数据。
- 您的 Amazon S3 和 Amazon EMR 集群之间的网络连接具有良好的传输速度和可用带宽。Amazon S3 不压缩响HTTP应，因此压缩后的输入文件的响应大小可能会增加。

## 注意事项和限制

- 不支持使用客户提供的加密密钥 (SSE-C) 的 Amazon S3 服务器端加密和客户端加密。
- 不支持 AllowQuotedRecordDelimiters 属性。如果指定该属性，则查询将失败。
- 仅支持CSV和 UTF -8 格式的JSON文件。不支持CSVs多行。
- 仅支持未压缩文件或 gzip 文件。

- 不支持 Spark CSV 和JSON选项，例如nanValue、positiveInfnegativeInf、，以及与损坏的记录相关的选项（例如，failfast 和 dropformalform 模式）。
- 不支持在十进制数中使用逗号(,)。例如，不支持 10,000，支持 10000。
- 不支持最后一行中的注释字符。
- 文件末尾的空行不会被处理。
- 以下筛选条件不会向下推送到 Amazon S3：
  - 聚合函数（如 COUNT() 和 SUM()）。
  - 对属性进行 CAST() 的筛选条件。例如，CAST(stringColumn as INT) = 1。
  - 具有作为对象或很复杂的属性的筛选条件。例如，intArray[1] = 1, objectColumn.objectNumber = 1。
  - 值不是文本值的筛选条件。例如，intColumn1 = intColumn2
  - 仅支持 [S3 Select 支持的数据类型](#)，但存在记录的限制。

## 在代码中指定 S3 Select

以下示例演示如何指定 S3 Select CSV 以使用 Scala SQL、R 和 PySpark。您可以使用同样的方式将 S3 Select JSON 用于。有关选项、默认值和限制的列表，请参阅[Options](#)。

### PySpark

```
spark
  .read
  .format("s3selectCSV") // "s3selectJson" for Json
  .schema(...) // optional, but recommended
  .options(...) // optional
  .load("s3://path/to/my/datafiles")
```

### R

```
read.df("s3://path/to/my/datafiles", "s3selectCSV", schema, header = "true",
        delimiter = "\t")
```

### Scala

```
spark
  .read
  .format("s3selectCSV") // "s3selectJson" for Json
```

```
.schema(...) // optional, but recommended
.options(...) // optional. Examples:
// .options(Map("quote" -> "\"", "header" -> "true")) or
// .option("quote", "\"").option("header", "true")
.load("s3://path/to/my/datafiles")
```

## SQL

```
CREATE TEMPORARY VIEW MyView (number INT, name STRING) USING s3selectCSV OPTIONS
(path "s3://path/to/my/datafiles", header "true", delimiter "\t")
```

## Options

使用 `s3selectCSV` 和 `s3selectJSON` 时，有以下选项可用。如果未指定，将使用默认值。

### 使用 `s3Select` 的选项 CSV

选项	默认	使用量
<code>compression</code>	"none"	指示是否使用了压缩。"gzip" 是除 "none" 之外唯一受支持的设置。
<code>delimiter</code>	","	指定字段分隔符。
<code>quote</code>	"\""	指定引号字符。不支持指定空字符串，这会导致格式XML错误的错误。
<code>escape</code>	"\\"	指定转义字符。
<code>header</code>	"false"	"false" 指定不存在标头。"true" 指定第一行中存在标头。仅支持第一行中的标头，不支持标头前面的空行。
<code>comment</code>	"#"	指定注释字符。无法禁用注释标记。换句话说，不支持值 <code>\u0000</code> 。

选项	默认	使用量
nullValue	""	

## 使用 s3Select 的选项 JSON

选项	默认	使用量
compression	"none"	指示是否使用了压缩。"gzip" 是除 "none" 之外唯一受支持的设置。
multiline	"false"	"false" 指定 JSON is 采用 S3 Select LINES t 格式，这意味着输入数据中的每一行都包含一个JSON对象。 "true" 指定采用JSON用 S3 Select DOCUMENT 格式，这意味着JSON对象可以在输入数据中跨越多行。

## 使用 EMRFS S3 优化的提交器

EMRFS S3 优化的提交器是一种替代 [OutputCommitter](#) 实现，针对使用时向 Amazon S3 写入文件进行了优化。EMRFS S3 优化的提交器通过避免在任务和任务提交阶段在 Amazon S3 中执行列出和重命名操作来提高应用程序性能。提交器适用于亚马逊 5.19.0 及更高 EMR 版本，亚马逊 EMR 5.20.0 及更高版本默认启用。提交器用于使用 Spark SQL DataFrames、或数据集的 Spark 作业。从 Amazon EMR 6.4.0 开始，此提交器可用于所有常见格式，ORC 包括 parquet 和基于文本的格式（包括 CSV 和）。JSON 对于亚马逊 EMR 6.4.0 之前的版本，仅支持 Parquet 格式。在某些情况下，不使用提交程序。有关更多信息，请参阅 [对 EMRFS S3 优化的提交者的要求](#)。

### 主题

- [对 EMRFS S3 优化的提交者的要求](#)
- [EMRFS S3 优化的提交者和分段上传](#)
- [任务优化注意事项](#)

- [为亚马逊 5.19.0 启用 EMRFS S3 优化的提交器 EMR](#)

## 对 EMRFS S3 优化的提交者的要求

满足以下条件时使用 EMRFS S3 优化的提交器：

- 你运行 Spark 作业，这些任务使用 Spark SQL DataFrames、或数据集将文件写入 Amazon S3。从 Amazon EMR 6.4.0 开始，此提交器可用于所有常见格式，ORC包括 parquet 和基于文本的格式（包括CSV和）。JSON对于亚马逊 EMR 6.4.0 之前的版本，仅支持 Parquet 格式。
- 在 Amazon 中，分段上传已启用。EMR这是默认模式。有关更多信息，请参阅 [EMRFS S3 优化的提交者和分段上传](#)。
- 使用 Spark 的内置文件格式支持。内置文件格式支持用于以下情况：
  - 对于 Hive metastore 表，当 `spark.sql.hive.convertMetastoreParquet` Parquet 表设置为 `true`对于亚马逊 EMR 6.4.0 或更高版本 `spark.sql.hive.convertMetastoreOrc` 的 Orc 表，则设置为 `true`。这些是默认设置。
  - 当任务写入文件格式数据来源或表时，例如，使用 `USING parquet` 子句创建目标表。
  - 当作业写入未分区的 Hive 元存储 Parquet 表时。Spark 的内置 Parquet 支持不支持分区的 Hive 表，这是一个已知限制。有关更多信息，请参阅《[Apache Spark SQL](#)》[DataFrames](#) 和《[数据集指南](#)》中的 [Hive metastore Parquet 表转换](#)。
- 写入默认分区位置的 Spark 任务操作，例如 `${table_location}/k1=v1/k2=v2/`，使用提交程序。如果任务操作写入自定义分区位置，则不使用提交程序，例如，如果使用 `ALTER TABLE SQL` 命令设置自定义分区位置。
- 必须使用 Spark 的以下值：
  - `spark.sql.parquet.fs.optimized.committer.optimization-enabled` 属性必须设置为 `true`。这是亚马逊 EMR 5.20.0 及更高版本的默认设置。在亚马逊 EMR 5.19.0 中，默认值为 `false`。有关配置此值的信息，请参阅[为亚马逊 5.19.0 启用 EMRFS S3 优化的提交器 EMR](#)。
  - 如果写入未分区的 Hive 元数据仓库表，则仅支持 Parquet 和 Orc 文件格式。  
`spark.sql.hive.convertMetastoreParquet``true`如果写入未分区 Parquet Hive 元数据仓库表，则必须设置为 `true`。  
`spark.sql.hive.convertMetastoreOrc``true`如果写入未分区的 Orc Hive 元数据仓库表，则必须设置为 `true`。这些是默认设置。
  - `spark.sql.parquet.output.committer.class` 必须设置为 `com.amazon.emr.committer.EmrOptimizedSparkSqlParquetOutputCommitter`。这是默认设置。

- `spark.sql.sources.commitProtocolClass` 必须设置为 `org.apache.spark.sql.execution.datasources.SQLEmrOptimizedCommitProtocol` 或 `org.apache.spark.sql.execution.datasources.SQLEmrOptimizedCommitProtocol` 是亚马逊 EMR 5.x 系列版本 5.30.0 及更高版本以及亚马逊 6. EMR x 系列版本 6.2.0 及更高版本的默认设置。  
`org.apache.spark.sql.execution.datasources.SQLHadoopMapReduceCommitProtocol` 是先前 Amazon EMR 版本的默认设置。
- 如果 Spark 作业用动态分区覆盖分区的 Parquet 数据集，则 `partitionOverwriteMode` 写入选项和 `spark.sql.sources.partitionOverwriteMode` 必须设置为 `static`。这是默认设置。

#### Note

Spark 2.4.0 中引入了 `partitionOverwriteMode` 写入选项。对于亚马逊 EMR 发行版 5.19.0 中包含的 Spark 版本 2.3.2，请设置该属性。 `spark.sql.sources.partitionOverwriteMode`

不使用 EMRFS S3 优化的提交者的场合

通常，在 EMRFS 以下情况下不使用 S3 优化的提交器。

情况	为什么不使用提交程序
当你写信给 HDFS	提交者仅支持使用 EMRFS 写入 Amazon S3。
当您使用 S3A 文件系统时	提交者只支持 EMRFS。
当你使用 MapReduce 或 Spark 时 RDD API	提交者仅支持使用 Spark SQL DataFrame、或数据集 APIs。

以下 Scala 示例演示了一些其他情况，这些情况会阻止 EMRFS S3 优化的提交器全部使用（第一个示例）和部分使用（第二个示例）。



## Example – 动态分区覆盖模式

以下 Scala 示例指示 Spark 使用不同的提交算法，这样可以完全阻止使用 EMRFS S3 优化的提交器。该代码将 `partitionOverwriteMode` 属性设置为 `dynamic`，以仅覆盖您的数据所写入到的分区。然后，通过 `partitionBy` 指定动态分区列，并将写入模式设置为 `overwrite`。

```
val dataset = spark.range(0, 10)
  .withColumn("dt", expr("date_sub(current_date(), id)"))

dataset.write.mode("overwrite")
  .option("partitionOverwriteMode", "dynamic")
  .partitionBy("dt")
  .parquet("s3://EXAMPLE-DOC-BUCKET/output")
```

您必须配置所有三个设置，以避免使用 EMRFS S3 优化的提交器。当您执行此操作时，Spark 会执行在 Spark 的提交协议中指定的不同提交算法。对于 EMR 5.30.0 之前的 Amazon 5.x 版本和 EMR 6.2.0 之前的 Amazon 6.x 版本，提交协议使用 Spark 的暂存目录，该目录是在以开头的输出位置下创建的临时目录。`.spark-staging` 该算法按顺序对分区目录进行重命名，这可能会对性能产生负面影响。有关 Amazon 5.30.0 及更高 EMR 版本以及 6.2.0 及更高版本的更多信息，请参阅 [使用 EMRFS S3 优化的提交协议](#)

Spark 2.4.0 中的算法遵循以下步骤：

1. 任务尝试将其输出写入 Spark 的暂存目录下的分区目录，例如 `${outputLocation}/spark-staging-${jobID}/k1=v1/k2=v2/`。
2. 对于写入的每个分区，任务尝试都跟踪相对分区路径，例如 `k1=v1/k2=v2`。
3. 任务成功完成后，它会为驱动程序提供它跟踪的所有相对分区路径。
4. 完成所有任务后，作业提交阶段将收集成功任务尝试在 Spark 的暂存目录下写入的所有分区目录。Spark 使用目录树重命名操作按顺序将这些目录的每一个都重命名为其最终输出位置。
5. 暂存目录会在作业提交阶段完成之前删除。

## Example – 自定义分区位置

在此示例中，Scala 代码插入到两个分区中。一个分区具有自定义分区位置。另一个分区使用默认分区位置。EMRFS S3 优化的提交器仅用于将任务输出写入使用默认分区位置的分区。

```
val table = "dataset"
```

```
val location = "s3://bucket/table"

spark.sql(s"""
  CREATE TABLE $table (id bigint, dt date)
  USING PARQUET PARTITIONED BY (dt)
  LOCATION '$location'
  """)

// Add a partition using a custom location
val customPartitionLocation = "s3://bucket/custom"
spark.sql(s"""
  ALTER TABLE $table ADD PARTITION (dt='2019-01-28')
  LOCATION '$customPartitionLocation'
  """)

// Add another partition using default location
spark.sql(s"ALTER TABLE $table ADD PARTITION (dt='2019-01-29')")

def asDate(text: String) = lit(text).cast("date")

spark.range(0, 10)
  .withColumn("dt",
    when($"id" > 4, asDate("2019-01-28")).otherwise(asDate("2019-01-29")))
  .write.insertInto(table)
```

Scala 代码创建以下 Amazon S3 对象：

```
custom/part-00001-035a2a9c-4a09-4917-8819-e77134342402.c000.snappy.parquet
custom_${folder$}
table/_SUCCESS
table/dt=2019-01-29/part-00000-035a2a9c-4a09-4917-8819-e77134342402.c000.snappy.parquet
table/dt=2019-01-29_${folder$}
table_${folder$}
```

当写入到自定义位置的分区时，Spark 会使用类似于上一个示例的提交算法，如下所述。与前面的示例一样，该算法会导致顺序重命名，这可能会影响性能。

1. 在将输出写入自定义位置的分区时，任务会写入到 Spark 的暂存目录下的文件中，该目录是在最终输出位置下创建的。文件名中包含一个随机名称 UUID，以防文件冲突。任务尝试跟踪每个文件以及最终所需的输出路径。
2. 在任务成功完成后，它会为驱动程序提供这些文件及其最终所需的输出路径。

3. 所有任务都完成后，作业提交阶段会按顺序将为自定义位置的分区写入的所有文件重命名为其最终输出路径。
4. 暂存目录会在作业提交阶段完成之前删除。

## EMRFSS3 优化的提交者和分段上传

要使用 EMRFS S3 优化的提交器，您必须为 Amazon 启用分段上传。EMR 默认启用分段上传。需要时，您可以重新启用它。有关更多信息，请参阅《亚马逊EMR管理指南》中的[为 Amazon S3 配置分段上传](#)。

EMRFSS3 优化的提交者使用分段上传的类似事务的特性来确保任务尝试写入的文件仅在任务提交后出现在作业的输出位置。通过以这种方式使用分段上传，提交者比默认 FileOutputCommitter 算法版本 2 提高了任务提交性能。在使用 EMRFS S3 优化的提交器时，需要考虑与传统分段上传行为的一些主要区别：

- 无论文件大小如何，分段上传都会执行。这与您使用的默认行为不同EMRFS，后者的 `fs.s3n.multipart.uploads.split.size` 属性控制触发分段上传的文件大小。
- 在任务提交或中止之前，分段上传在较长时间内都保持在未完成状态。这与任务完成写入给定文件后分段上传完成的默认行为不同。EMRFS

由于这些差异，如果 Spark Executor 在任务运行和向 Amazon S3 写入数据时 JVM 崩溃或被终止，则不完整的分段上传更有可能被遗忘。因此，在使用 EMRFS S3 优化的提交器时，请务必遵循管理失败的分段上传的最佳实践。有关更多信息，请参阅《亚马逊EMR管理指南》中的[使用 Amazon S3 存储桶的最佳实践](#)。

## 任务优化注意事项

EMRFSS3 优化的提交者会为任务尝试写入的每个文件消耗少量内存，直到任务被提交或中止。在大多数作业中，占用的内存量可以忽略不计。对于包含写入大量文件的长期任务的作业，提交程序占用的内存可能很大，需要调整为 Spark 执行程序分配的内存。您可以使用 `spark.executor.memory` 属性调整执行程序内存。作为指导，编写 100000 个文件的单个任务通常需要额外的 100 MB 内存。有关更多信息，请参阅 Apache Spark 配置文档中的[应用程序属性](#)。

## 为亚马逊 5.19.0 启用 EMRFS S3 优化的提交器 EMR

如果您使用的是 Amazon EMR 5.19.0，则可以在创建集群 `true` 时手动将该 `spark.sql.parquet.fs.optimized.committer.optimization-enabled` 属性设置为 `true`，如果您使用的是亚马逊，则可以从 Spark 中手动设置为 `EMRFS_S3`。

## 创建集群时启用 EMRFS S3 优化的提交器

使用 spark-defaults 配置分类将

spark.sql.parquet.fs.optimized.committer.optimization-enabled 属性设置为 true。有关更多信息，请参阅 [配置应用程序](#)。

启用来自 Sp EMRFS ark 的 S3 优化提交者

您可以将 spark.sql.parquet.fs.optimized.committer.optimization-enabled 设置为 true，方法是在 SparkConf 中对其进行硬编码，并在 Spark Shell 或 spark-submit 和 spark-sql 工具或 conf/spark-defaults.conf 中将其作为 --conf 参数进行传递。有关更多信息，请参阅 Apache Spark 文档中的 [Spark 配置](#)。

以下示例显示了如何在运行 spark-sql 命令的同时启用提交程序。

```
spark-sql \  
  --conf spark.sql.parquet.fs.optimized.committer.optimization-enabled=true \  
  -e "INSERT OVERWRITE TABLE target_table SELECT * FROM source_table;"
```

## 使用 EMRFS S3 优化的提交协议

EMRFS S3 优化的提交协议是一种替代 [FileCommitProtocol](#) 实现方案，它针对在使用时将 Spark 动态分区覆盖的文件写入 Amazon S3 进行了优化。EMRFS 该协议可在 Spark 动态分区覆盖任务提交阶段避免 Amazon S3 中的重命名操作，从而提高应用程序性能。

请注意，[使用 EMRFS S3 优化的提交器](#) 也可以通过避免重命名操作来提高性能。但是，它不适用于动态分区覆盖情况，同时提交协议的改进仅针对动态分区覆盖情况。

提交协议在 Amazon EMR 版本 5.30.0 及更高版本以及 6.2.0 及更高版本中可用，并且默认处于启用状态。从 5.31.0 版本开始，亚马逊 EMR 增加了并行度改进。该协议用于使用 Spark SQL DataFrames、或数据集的 Spark 作业。在某些情况下，不使用提交协议。有关更多信息，请参阅 [对 EMRFS S3 优化的提交协议的要求](#)。

主题

- [对 EMRFS S3 优化的提交协议的要求](#)
- [EMRFS S3 优化的提交协议和分段上传](#)
- [任务优化注意事项](#)

## 对 EMRFS S3 优化的提交协议的要求

满足以下条件时使用 EMRFS S3 优化的提交协议：

- 你运行使用 Spark SQL DataFrames、或数据集来覆盖分区表的 Spark 作业。
- 您可以运行分区覆盖模式为 `dynamic` 的 Spark 任务。
- 在 Amazon 中，分段上传已启用。EMR这是默认模式。有关更多信息，请参阅 [EMRFS S3 优化的提交协议和分段上传](#)。
- 的文件系统缓存已启EMRFS用。这是默认模式。检查设置 `fs.s3.impl.disable.cache` 是否设置为 `false`。
- 使用 Spark 的内置数据来源支持。内置数据来源支持用于以下情况：
  - 当任务写入内置数据来源或表时。
  - 当任务写入 Hive 元存储 Parquet 表时。当 `spark.sql.hive.convertInsertingPartitionedTable` 和 `spark.sql.hive.convertMetastoreParquet` 都设置为 `true` 时，就会发生这种情况。这些是默认设置。
  - 当作业写入 Hive 元数据仓库表ORC时。当 `spark.sql.hive.convertInsertingPartitionedTable` 和 `spark.sql.hive.convertMetastoreOrc` 都设置为 `true` 时，就会发生这种情况。这些是默认设置。
- 写入默认分区位置的 Spark 任务操作，例如 `/${table_location}/k1=v1/k2=v2/`，使用提交协议。如果任务操作写入自定义分区位置，则不使用协议，例如，如果使用 `ALTER TABLE SQL` 命令设置自定义分区位置。
- 必须使用 Spark 的以下值：
  - `spark.sql.sources.commitProtocolClass` 必须设置为 `org.apache.spark.sql.execution.datasources.SQLEmrOptimizedCommitProtocol`。这是亚马逊EMR版本 5.30.0 及更高版本以及 6.2.0 及更高版本的默认设置。
  - `partitionOverwriteMode` 写入选项或 `spark.sql.sources.partitionOverwriteMode` 必须设置为 `dynamic`。默认设置为 `static`。

**Note**

Spark 2.4.0 中引入了 `partitionOverwriteMode` 写入选项。对于亚马逊EMR发行版 5.19.0 中包含的 Spark 版本 2.3.2，请设置该属性。`spark.sql.sources.partitionOverwriteMode`

- 如果 Spark 任务覆盖了 Hive 元存储 Parquet 表，则 `spark.sql.hive.convertMetastoreParquet`、`spark.sql.hive.convertInsertingPartitions` 和 `spark.sql.hive.convertMetastore.partitionOverwriteMode` 必须设置为 `true`。这些是默认设置。
- 如果 Spark 作业覆盖到 Hive 元数据仓ORC表，则`spark.sql.hive.convertMetastore.partitionOverwriteMode`必须将和`spark.sql.hive.convertMetastoreOrcspark.sql.hive.convertInsertingPartitions`和设置为。`true`这些是默认设置。

**Example – 动态分区覆盖模式**

在此 Scala 示例中已触发优化。首先，将 `partitionOverwriteMode` 属性设置为 `dynamic`。这只会覆盖您正在写入数据的那些分区。然后，通过 `partitionBy` 指定动态分区列，并将写入模式设置为 `overwrite`。

```
val dataset = spark.range(0, 10)
  .withColumn("dt", expr("date_sub(current_date(), id)"))

dataset.write.mode("overwrite")           // "overwrite" instead of "insert"
  .option("partitionOverwriteMode", "dynamic") // "dynamic" instead of "static"
  .partitionBy("dt")                       // partitioned data instead of
  unpartitioned data
  .parquet("s3://EXAMPLE-DOC-BUCKET/output") // "s3://" to use Amazon EMR file
  system, instead of "s3a://" or "hdfs://"
```

**不使用 EMRFS S3 优化的提交协议时**

通常，EMRFS S3 优化的提交协议与开源默认 Spark SQL 提交协议的工作原理相同。`org.apache.spark.sql.execution.datasources.SQLHadoopMapReduceCommitProtocol`以下情况下不会进行优化。

情况	为什么不使用提交协议
当你写信给 HDFS	提交协议仅支持使用写入 Amazon S3 EMRFS。
当您使用 S3A 文件系统时	仅支持提交协议EMRFS。
当你使用 MapReduce 或 Spark 时 RDD API	提交协议仅支持使用 Spark SQL DataFrame、或数据集APIs。
当没有触发动态分区覆盖时	提交协议仅优化动态分区覆盖情况。有关其他情况，请参阅 <a href="#">使用 EMRFS S3 优化的提交器</a> 。

以下 Scala 示例演示了 EMRFS S3 优化的提交协议委托的其他一些情况。SQLHadoopMapReduceCommitProtocol

#### Example – 具有自定义分区位置的动态分区覆盖模式

在此示例中，Scala 程序以动态分区覆盖模式覆盖两个分区。一个分区具有自定义分区位置。另一个分区使用默认分区位置。EMRFS S3 优化的提交协议只会改进使用默认分区位置的分区。

```
val table = "dataset"
val inputView = "tempView"
val location = "s3://bucket/table"

spark.sql(s"""
  CREATE TABLE $table (id bigint, dt date)
  USING PARQUET PARTITIONED BY (dt)
  LOCATION '$location'
  """)

// Add a partition using a custom location
val customPartitionLocation = "s3://bucket/custom"
spark.sql(s"""
  ALTER TABLE $table ADD PARTITION (dt='2019-01-28')
  LOCATION '$customPartitionLocation'
  """)

// Add another partition using default location
spark.sql(s"ALTER TABLE $table ADD PARTITION (dt='2019-01-29')")
```

```
def asDate(text: String) = lit(text).cast("date")

spark.range(0, 10)
  .withColumn("dt",
    when($"id" > 4, asDate("2019-01-28")).otherwise(asDate("2019-01-29")))
  .createTempView(inputView)

// Set partition overwrite mode to 'dynamic'
spark.sql(s"SET spark.sql.sources.partitionOverwriteMode=dynamic")

spark.sql(s"INSERT OVERWRITE TABLE $table SELECT * FROM $inputView")
```

Scala 代码创建以下 Amazon S3 对象：

```
custom/part-00001-035a2a9c-4a09-4917-8819-e77134342402.c000.snappy.parquet
custom_${folder}$
table/_SUCCESS
table/dt=2019-01-29/part-00000-035a2a9c-4a09-4917-8819-e77134342402.c000.snappy.parquet
table/dt=2019-01-29_${folder}$
table_${folder}$
```

#### Note

在早期 Spark 版本中写入自定义分区位置可能会导致数据丢失。在此示例中，分区 `dt='2019-01-28'` 将丢失。有关更多详细信息，请参阅 [SPARK-35106](#)。此问题已在亚马逊 EMR 版本 5.33.0 及更高版本中修复，不包括 6.0.x 和 6.1.x。

当写入到自定义位置的分区时，Spark 会使用类似于上一个示例的提交算法，如下所述。与前面的示例一样，该算法会导致顺序重命名，这可能会影响性能。

Spark 2.4.0 中的算法遵循以下步骤：

1. 在将输出写入自定义位置的分区时，任务会写入到 Spark 的暂存目录下的文件中，该目录是在最终输出位置下创建的。文件名中包含一个随机名称 UUID，以防文件冲突。任务尝试跟踪每个文件以及最终所需的输出路径。
2. 在任务成功完成后，它会为驱动程序提供这些文件及其最终所需的输出路径。
3. 所有任务都完成后，作业提交阶段会按顺序将为自定义位置的分区写入的所有文件重命名为其最终输出路径。
4. 暂存目录会在作业提交阶段完成之前删除。



## EMRFSS3 优化的提交协议和分段上传

要使用优化 EMRFS S3 的提交协议中的动态分区覆盖优化，必须在 Amazon 中启用分段上传。EMR 默认启用分段上传。需要时，您可以重新启用它。有关更多信息，请参阅《亚马逊EMR管理指南》中的[为 Amazon S3 配置分段上传](#)。

在动态分区覆盖期间，EMRFSS3 优化的提交协议使用分段上传的类似事务的特性来确保任务尝试写入的文件仅在作业提交时出现在作业的输出位置。通过以这种方式使用分段上传，提交协议提高了默认 SQLHadoopMapReduceCommitProtocol 的任务提交性能。在使用 EMRFS S3 优化的提交协议时，需要考虑与传统分段上传行为的一些关键区别：

- 无论文件大小如何，分段上传都会执行。这不同于默认行为不同EMRFS，后者的 `fs.s3n.multipart.uploads.split.size` 属性控制触发分段上传的文件大小。
- 在任务提交或中止之前，分段上传在较长时间内都保持在未完成状态。这与任务完成写入给定文件后分段上传完成的默认行为不同。EMRFS

由于这些差异，如果 Spark Executor 在任务运行并将数据写入 Amazon S3 时 JVM 崩溃或被终止，或者 Spark Driver 在任务运行时 JVM 崩溃或被终止，则不完整的分段上传更有可能被遗忘。因此，当您使用 EMRFS S3 优化的提交协议时，请务必遵循管理失败的分段上传的最佳实践。有关更多信息，请参阅《亚马逊EMR管理指南》中的使用 Amazon S3 存储桶的[最佳实践](#)。

### 任务优化注意事项

在 Spark 执行器上，EMRFSS3 优化的提交协议会为任务尝试写入的每个文件消耗少量内存，直到任务被提交或中止。在大多数作业中，占用的内存量可以忽略不计。

在 Spark 驱动程序上，EMRFSS3 优化的提交协议需要内存来存储每个已提交文件的元数据信息，直到任务提交或中止。在大多数任务中，默认的 Spark 驱动程序内存设置可以忽略不计。

对于包含写入大量文件的长期任务的作业，提交协议占用的内存可能很大，需要调整分配给 Spark 的内存，尤其是 Spark 执行程序。您可以使用 Spark 驱动程序的 `spark.driver.memory` 属性和 `spark.executor.memory` 属性来优化内存。作为指导，编写 100000 个文件的单个任务通常需要额外的 100 MB 内存。有关更多信息，请参阅 Apache Spark 配置文档中的[应用程序属性](#)。

## 使用重试 Amazon S3 请求 EMRFS

本主题提供有关向 Amazon S3 发出请求时可以使用的重试策略的信息。EMRFS 当您的请求速率提高时，S3 会尝试扩展以支持新的速率。在此过程中，S3 可以限制请求并返回 503 Slow Down 错误。为了提高 S3 请求的成功率，您可以通过在 `emrfs-site` 配置中配置属性以调整重试策略。

您可以通过以下方法调整重试策略。

- 提高默认指数退避重试策略的最大重试限制。
- 启用并配置加法-增加/乘法-减少 () 重试策略。AIMDAIMD支持 Amazon EMR 版本 6.4.0 及更高版本。

## 使用默认的指数退避策略

默认情况下，EMRFS使用指数退避策略重试 Amazon S3 请求。默认EMRFS重试限制为 15。为避免 S3 503 Slow Down 错误，您可以在创建新集群时、在正在运行的群集上或应用程序运行时提高重试限制。

要提高重试限制，您必须在您的 `emrfs-site` 配置中更改 `fs.s3.maxRetries` 的值。以下示例配置将 `fs.s3.maxRetries` 设置为自定义值 30。

```
[
  {
    "Classification": "emrfs-site",
    "Properties": {
      "fs.s3.maxRetries": "30"
    }
  }
]
```

有关使用配置对象的更多信息，请参阅 [配置应用程序](#)。

## 使用AIMD重试策略

在 Amazon 6.4.0 及更高EMR版本中，EMRFS支持基于加法-增加/乘法-减少 () 模型的替代重试策略。AIMD当您使用大型 Amazon EMR 集群时，AIMD重试策略特别有用。

AIMD使用有关最近成功请求的数据计算自定义请求率。此策略减少了受限请求的数量和每个请求所需的总尝试次数。

要启用AIMD重试策略，您必须在`emrfs-site`配置`true`中将该`fs.s3.aimd.enabled`属性设置为，如下例所示。

```
[
  {
```

```

    "Classification": "emrfs-site",
    "Properties": {
      "fs.s3.aimd.enabled": "true"
    }
  }
]

```

有关使用配置对象的更多信息，请参阅 [配置应用程序](#)。

## 高级AIMD重试设置

在使用重试策略时，您可以配置下表中列出的属性以优化AIMD重试行为。对于大多数使用案例，我们建议您使用默认值。

### 高级AIMD重试策略属性

属性	默认值	描述
<code>fs.s3.aimd.increaseIncrement</code>	0.1	控制连续请求成功时请求速率的增长速度。
<code>fs.s3.aimd.reductionFactor</code>	2	控制 Amazon S3 返回 503 响应时请求速率降低的速度。默认因子 2 将请求率降低一半。
<code>fs.s3.aimd.minRate</code>	0.1	设置请求经历 S3 持续限制时的请求速率的下限。
<code>fs.s3.aimd.initialRate</code>	5500	设置初始请求速率，然后该速率将根据您为 <code>fs.s3.aimd.increaseIncrement</code> 和 <code>fs.s3.aimd.reductionFactor</code> 指定的值变化。  初始速率也用于GET请求，并按比例缩放 (3500/5500) 来处理请求。PUT
<code>fs.s3.aimd.adjustWindow</code>	2	控制调整请求速率的频率，以响应数量衡量。

属性	默认值	描述
fs.s3.aimd.maxAttempts	100	设置尝试请求的最大尝试次数。

## 添加 Spark 步骤

您可以使用 Amazon EMR 步骤将工作提交到安装在EMR集群上的 Spark 框架。有关更多信息，请参阅《Amazon EMR 管理指南》中的[步骤](#)。在控制台和中CLI，您可以使用 Spark 应用程序步骤执行此操作，该步骤将spark-submit脚本作为步骤代表您运行。使用时API，您可以使用步骤来调spark-submit用command-runner.jar。

有关向 Spark 提交应用程序的更多信息，请参阅 Apache Spark 文档中的[提交应用程序](#)主题。

### 使用控制台提交 Spark 步骤

1. 在 <https://console.aws.amazon.com/emr> 上打开亚马逊EMR控制台。
2. 在 Cluster List (集群列表) 中，选择您的集群的名称。
3. 滚动到 Steps (步骤) 部分并展开它，然后选择 Add step (添加步骤)。
4. 在 Add Step (添加步骤) 对话框中：
  - 对于 Step type (步骤类型)，选择 Spark application (Spark 应用程序)。
  - 对于 Name (名称)，接受原定设置名称 (Spark application) 或键入一个新名称。
  - 对于 Deploy mode (部署模式)，选择 Client (客户端) 或 Cluster (集群) 模式。客户端模式在集群的主实例上启动驱动程序，而集群模式在集群上启动驱动程序。对于客户端模式，驱动程序的日志输出显示在步骤日志中，而对于集群模式，驱动程序的日志输出显示在第一个YARN容器的日志中。有关更多信息，请参阅 Apache Spark 文档中的[集群模式概览](#)。
  - 指定所需的 Spark-submit options。有关 spark-submit 选项的更多信息，请参阅[使用 spark-submit 启动应用程序](#)。
  - 在“应用程序位置”中，指定应用程序的本地或 S3 URI 路径。
  - 对于 Arguments (参数)，将该字段保留为空白。
  - 对于 Action on failure (出现故障时的操作)，接受默认选项 Continue (继续)。
5. 选择添加。步骤会出现在控制台中，其状态为“Pending”。
6. 步骤的状态会随着步骤的运行从“Pending”变为“Running”，再变为“Completed”。要更新状态，请选择 Actions (操作) 列上方的 Refresh (刷新) 图标。

7. 如果您配置了日志记录，则该步骤的结果将位于 Amazon EMR 控制台集群详情页面，该页面位于日志文件下的步骤旁边。在启动集群时，您可以选择在配置的日志存储桶中查找步骤信息。

要使用 Spark 将作品提交给 Spark AWS CLI

在创建集群时提交步骤，或使用 `aws emr add-steps` 子命令在现有集群中提交步骤。

1. 使用 `create-cluster`，如以下示例所示。

#### Note

为了便于读取，包含 Linux 行继续符 ( \ )。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 ( ^ )。

```
aws emr create-cluster --name "Add Spark Step Cluster" --release-label emr-7.2.0 --
applications Name=Spark \
--ec2-attributes KeyName=myKey --instance-type m5.xlarge --instance-count 3 \
--steps Type=Spark,Name="Spark Program",ActionOnFailure=CONTINUE,Args=[--
class,org.apache.spark.examples.SparkPi,/usr/lib/spark/examples/jars/spark-
examples.jar,10] --use-default-roles
```

作为替代方法，您可使用 `command-runner.jar`，如以下示例所示。

```
aws emr create-cluster --name "Add Spark Step Cluster" --release-label emr-7.2.0 \
--applications Name=Spark --ec2-attributes KeyName=myKey --instance-type m5.xlarge
--instance-count 3 \
--steps Type=CUSTOM_JAR,Name="Spark Program",Jar="command-
runner.jar",ActionOnFailure=CONTINUE,Args=[spark-example,SparkPi,10] --use-default-
roles
```

#### Note

为了便于读取，包含 Linux 行继续符 ( \ )。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 ( ^ )。

2. 或者，向正在运行的集群添加步骤。使用 `add-steps`。

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF --steps
  Type=Spark,Name="Spark Program",ActionOnFailure=CONTINUE,Args=[--
class,org.apache.spark.examples.SparkPi,/usr/lib/spark/examples/jars/spark-
examples.jar,10]
```

作为替代方法，您可使用 `command-runner.jar`，如以下示例所示。

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF --steps Type=CUSTOM_JAR,Name="Spark
  Program",Jar="command-runner.jar",ActionOnFailure=CONTINUE,Args=[spark-
example,SparkPi,10]
```

使用SDK适用于 Java 的，向 Spark 提交作品

1. 下面的示例显示如何通过 Java 添加步骤至带有 Spark 的集群：

```
AWSCredentials credentials = new BasicAWSCredentials(accessKey, secretKey);
AmazonElasticMapReduce emr = new AmazonElasticMapReduceClient(credentials);

StepFactory stepFactory = new StepFactory();
AmazonElasticMapReduceClient emr = new AmazonElasticMapReduceClient(credentials);
AddJobFlowStepsRequest req = new AddJobFlowStepsRequest();
req.withJobFlowId("j-1K48XXXXXXHCB");

List<StepConfig> stepConfigs = new ArrayList<StepConfig>();

HadoopJarStepConfig sparkStepConf = new HadoopJarStepConfig()
    .withJar("command-runner.jar")
    .withArgs("spark-submit", "--executor-memory", "1g", "--
class", "org.apache.spark.examples.SparkPi", "/usr/lib/spark/examples/jars/spark-
examples.jar", "10");

StepConfig sparkStep = new StepConfig()
    .withName("Spark Step")
    .withActionOnFailure("CONTINUE")
    .withHadoopJarStep(sparkStepConf);

stepConfigs.add(sparkStep);
req.withSteps(stepConfigs);
AddJobFlowStepsResult result = emr.addJobFlowSteps(req);
```

- 您可以通过查看该步骤的日志来了解该步骤的结果。AWS Management Console 如果您启用了日志记录，则可以在中执行此操作，方法是选择“步骤”，选择您的步骤，然后在“日志文件”中，选择stdout或stderr。要查看可用日志，请选择 View Logs (查看日志)。

## 覆盖 Spark 默认配置设置

建议您为不同的应用程序覆盖 Spark 默认配置值。您可以在提交应用程序时使用步骤完成此操作 (实质上是向 spark-submit 传递选项)。例如，您可能需要通过更改 spark.executor.memory 来更改为执行者进程分配的内存。您可以为 --executor-memory 开关提供类似下面的参数：

```
spark-submit --executor-memory 1g --class org.apache.spark.examples.SparkPi /usr/lib/spark/examples/jars/spark-examples.jar 10
```

同样地，您也可以调节 --executor-cores 和 --driver-memory。在步骤中，您可以向步骤提供以下参数：

```
--executor-memory 1g --class org.apache.spark.examples.SparkPi /usr/lib/spark/examples/jars/spark-examples.jar 10
```

您还可以使用 --conf 选项调节没有内置开关的设置。有关可调节的其他设置的更多信息，请参阅 Apache Spark 文档中的[动态加载 Spark 属性](#)主题。

## 查看 Spark 应用程序历史记录

您可以使用控制台中集群详细信息页面的YARN应用程序用户界面选项卡查看 Spark、应用程序和 Tez UI 的详细信息。Amazon EMR 应用程序用户界面 (UI) 使您可以更轻松地对活动任务和任务历史进行故障排除和分析。

有关更多信息，请参阅《Amazon EMR 管理指南》中的[查看应用程序历史记录](#)。

## 访问 Spark 网站 UIs

您可以UIs按照《Amazon EMR 管理指南》中名为“[Connect to to the Clust er](#)”一节中的创建SSH隧道或创建代理的过程查看 Spark 网页，然后导航到您的集群YARN ResourceManager 对应的。在 Tracking UI (跟踪 UI) 下选择适合您的应用程序的链接。如果您的应用程序正在运行，则会看到ApplicationMaster。这会将您带到主应用程序的 Web UI，在端口 20888 上，无论驱动程序位于何处都是如此。如果您在YARN客户端模式下运行，则驱动程序可能位于群集的主节点上。如果您在YARN集群模式下运行应用程序，则驱动程序位于集群上应用程序 ApplicationMaster 的驱动程序

中。如果您的应用程序已完成，则会看到历史记录，它会将您带到EMR集群主节点 18080 处的 Spark HistoryServer UI 端口号。这适用于已经完成的应用程序。你也可以直接通过 `http://` 导航到 Spark HistoryServer 用户界面 `master-public-dns-name:18080/`。

在 Amazon 5.25.0 及更高EMR版本中，您可以从控制台访问 Spark 历史服务器用户界面，而无需通过连接设置网络代理。SSH有关更多信息，请参阅[查看持久性应用程序用户界面](#)。

## 使用 Spark 结构化流媒体 Amazon Kinesis Data Streams 连接器

亚马逊EMR发布的7.1.0及更高版本在发布图片中包括火花结构化的流媒体Amazon Kinesis Data Streams连接器。使用此连接器，您可以使用亚马逊上的 Spark EMR 来处理存储在 Amazon Kinesis Data Streams 中的数据。该连接器支持两种使用者类型 `GetRecords` ( 共享吞吐量 ) 和 `SubscribeToShard` ( 增强扇出 )。此集成基于 [spark-sql-kinesis-connector](#)。有关如何开始使用连接器的详细信息，请参阅[README](#)。

以下示例演示如何使用连接器在 Amazon 上启动 Spark 应用程序 EMR

```
spark-submit my_kinesis_streaming_script.py
```

## 使用亚马逊 Redshift 与 Apache Spark 的集成与亚马逊 EMR

在亚马逊6.4.0及更高EMR版本中，每张发布的图片都包含A [pache Spark和Amazon Redshift](#) 之间的连接器。使用此连接器，您可以使用亚马逊上的 Spark EMR 来处理存储在 Amazon Redshift 中的数据。[对于亚马逊EMR版本 6.4.0 到 6.8.0，集成基于开源连接器。spark-redshift](#)对于亚马逊EMR版本 6.9.0 及更高版本，A [pache Spark 的 Amazon Redshift 集成](#)已从社区版本迁移到原生集成。

### 主题

- [使用适用于 Apache Spark 的 Amazon Redshift 集成启动 Spark 应用程序](#)
- [使用适用于 Apache Spark 的 Amazon Redshift 集成进行身份验证](#)
- [在 Amazon Redshift 中进行读取和写入](#)
- [使用 Spark 连接器时的注意事项和限制](#)

## 使用适用于 Apache Spark 的 Amazon Redshift 集成启动 Spark 应用程序

对于 Amazon 6.4 到 6.9 EMR 版本，您必须使用 `--jars` 或 `--packages` 选项来指定要使用以下哪个 JAR 文件。该 `--jars` 选项指定存储在本地 HDFS、中或使用 HTTP /S 存储的依赖关系。要查看该 `--`



`jars`选项支持的其他文件位置，请参阅 Spark 文档中的[高级依赖关系管理](#)。 `--packages` 选项指定存储在公共 Maven 存储库中的依赖项。

- `spark-redshift.jar`
- `spark-avro.jar`
- `RedshiftJDBC.jar`
- `minimal-json.jar`

Amazon 6.10.0 及更高EMR版本不需要`minimal-json.jar`依赖关系，并且默认情况下会自动将其他依赖项安装到每个集群。以下示例显示了如何使用适用于 Apache Spark 的 Amazon Redshift 集成启动 Spark 应用程序。

#### Amazon EMR 6.10.0 +

以下示例说明如何使用亚马逊EMR版本 6.10 及更高版本的`spark-redshift`连接器启动 Spark 应用程序。

```
spark-submit my_script.py
```

#### Amazon EMR 6.4.0 - 6.9.x

要在 Amazon 6.4 到 6.9 EMR 版本上启动带有`spark-redshift`连接器的 Spark 应用程序，必须使用`--jars`或`--packages`选项，如以下示例所示。请注意，`--jars`选项中列出的路径是JAR文件的默认路径。

```
spark-submit \  
  --jars /usr/share/aws/redshift/jdbc/RedshiftJDBC.jar,/usr/share/aws/redshift/  
spark-redshift/lib/spark-redshift.jar,/usr/share/aws/redshift/spark-redshift/lib/  
spark-avro.jar,/usr/share/aws/redshift/spark-redshift/lib/minimal-json.jar \  
  my_script.py
```

## 使用适用于 Apache Spark 的 Amazon Redshift 集成进行身份验证

### AWS Secrets Manager 用于检索凭证并连接亚马逊 Redshift

以下代码示例显示了如何使用检索凭证，AWS Secrets Manager 以便通过 Python 中的 Apache Spark PySpark 接口连接到 Amazon Redshift 集群。

```

from pyspark.sql import SQLContext
import boto3

sc = # existing SparkContext
sql_context = SQLContext(sc)

secretsmanager_client = boto3.client('secretsmanager')
secret_manager_response = secretsmanager_client.get_secret_value(
    SecretId='string',
    VersionId='string',
    VersionStage='string'
)
username = # get username from secret_manager_response
password = # get password from secret_manager_response
url = "jdbc:redshift://redshifthost:5439/database?user=" + username + "&password=" +
password

# Read data from a table
df = sql_context.read \
    .format("io.github.spark_redshift_community.spark.redshift") \
    .option("url", url) \
    .option("dbtable", "my_table") \
    .option("tempdir", "s3://path/for/temp/data") \
    .load()

```

## IAM用于检索凭证并连接亚马逊 Redshift

您可以使用 Amazon Redshift 提供的第 2 JDBC 版驱动程序通过 Spark 连接器连接到亚马逊 Redshift。要使用 AWS Identity and Access Management (IAM)，[请将您的配置JDBCURL为使用IAM身份验证](#)。要从亚马逊连接到 Redshift 集群EMR，您必须向您的IAM角色授予检索临时IAM证书的权限。为您的IAM角色分配以下权限，使其可以检索证书并运行 Amazon S3 操作。

- [Redshift: GetClusterCredentials](#) ( 适用于预配置的亚马逊 Redshift 集群 )
- [Redshift: DescribeClusters](#) ( 适用于预配置的亚马逊 Redshift 集群 )
- [Redshift : GetWorkgroup](#) ( 适用于亚马逊 Redshift 无服务器工作组 )
- [Redshift : GetCredentials](#) ( 适用于亚马逊 Redshift 无服务器工作组 )
- [s3 : GetBucket](#)
- [s3 : GetBucketLocation](#)
- [s3 : GetObject](#)

- [s3 : PutObject](#)
- [s3 : GetBucketLifecycleConfiguration](#)

有关 `GetClusterCredentials` 的更多信息，请参阅 [GetClusterCredentials 的资源策略](#)。

您还必须确保 Amazon Redshift 可以在 COPY 和 UNLOAD 操作期间担任该 IAM 角色。

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": "redshift.amazonaws.com"
      },
      "Action": "sts:AssumeRole"
    }
  ]
}
```

以下示例使用了 Spark 和 Amazon Redshift 之间的 IAM 身份验证：

```
from pyspark.sql import SQLContext
import boto3

sc = # existing SparkContext
sql_context = SQLContext(sc)

url = "jdbc:redshift:iam//redshift-host:redshift-port/db-name"
iam_role_arn = "arn:aws:iam::account-id:role/role-name"

# Read data from a table
df = sql_context.read \
    .format("io.github.spark_redshift_community.spark.redshift") \
    .option("url", url) \
    .option("aws_iam_role", iam_role_arn) \
    .option("dbtable", "my_table") \
    .option("tempdir", "s3a://path/for/temp/data") \
    .mode("error") \
    .load()
```

## 在 Amazon Redshift 中进行读取和写入

以下代码示例用于使用数据源和使用 PySpark Spark 从包含数据源的 Amazon Redshift 数据库中读取 API 和写入示例数据。SQL

### Data source API

PySpark 用于从带有数据源的 Amazon Redshift 数据库中读取和写入示例数据。API

```
import boto3
from pyspark.sql import SQLContext

sc = # existing SparkContext
sql_context = SQLContext(sc)

url = "jdbc:redshift:iam://redshifthost:5439/database"
aws_iam_role_arn = "arn:aws:iam::accountID:role/roleName"

df = sql_context.read \
    .format("io.github.spark_redshift_community.spark.redshift") \
    .option("url", url) \
    .option("dbtable", "tableName") \
    .option("tempdir", "s3://path/for/temp/data") \
    .option("aws_iam_role", "aws_iam_role_arn") \
    .load()

df.write \
    .format("io.github.spark_redshift_community.spark.redshift") \
    .option("url", url) \
    .option("dbtable", "tableName_copy") \
    .option("tempdir", "s3://path/for/temp/data") \
    .option("aws_iam_role", "aws_iam_role_arn") \
    .mode("error") \
    .save()
```

### SparkSQL

PySpark 用于通过 Spark 读取和写入亚马逊 Redshift 数据库的示例数据。SQL

```
import boto3
import json
import sys
import os
```

```
from pyspark.sql import SparkSession

spark = SparkSession \
    .builder \
    .enableHiveSupport() \
    .getOrCreate()

url = "jdbc:redshift:iam://redshifthost:5439/database"
aws_iam_role_arn = "arn:aws:iam::accountID:role/roleName"

bucket = "s3://path/for/temp/data"
tableName = "tableName" # Redshift table name

s = f"""CREATE TABLE IF NOT EXISTS {tableName} (country string, data string)
    USING io.github.spark_redshift_community.spark.redshift
    OPTIONS (dbtable '{tableName}', tempdir '{bucket}', url '{url}', aws_iam_role
    '{aws_iam_role_arn'} '); """

spark.sql(s)

columns = ["country" ,"data"]
data = [("test-country","test-data")]
df = spark.sparkContext.parallelize(data).toDF(columns)

# Insert data into table
df.write.insertInto(tableName, overwrite=False)
df = spark.sql(f"SELECT * FROM {tableName}")
df.show()
```

## 使用 Spark 连接器时的注意事项和限制

- 我们建议你开启从亚马逊上SSL的 Spark EMR 到亚马逊 Redshift 的JDBC连接。
- 作为最佳实践，建议在 AWS Secrets Manager 中管理 Amazon Redshift 集群的凭证。有关示例 [AWS Secrets Manager](#)，请参阅[使用检索连接至亚马逊 Redshift 的凭证](#)。
- 我们建议你传递一个带有 Amazon Redshift 身份验证参数参数参数aws\_iam\_role的IAM角色。
- 它们tempdirURI指向 Amazon S3 的一个地点。此临时目录不会自动清理，因此可能会增加额外成本。
- 请考虑以下针对 Amazon Redshift 的建议：
  - 建议阻止对 Amazon Redshift 集群的公有访问。

- 建议启用 [Amazon Redshift 审计日志记录](#)。
- 建议启用 [Amazon Redshift 静态加密](#)。
- 请考虑以下针对 Amazon S3 的建议：
  - 建议[阻止对 Amazon S3 存储桶的公有访问](#)。
  - 建议使用 [Amazon S3 服务器端加密](#)以加密使用的 Amazon S3 存储桶。
  - 建议使用 [Amazon S3 生命周期策略](#)定义 Amazon S3 存储桶的保留规则。
  - Amazon EMR 始终会验证从开源导入镜像的代码。出于安全原因，我们不支持从 Spark 到 Amazon S3 的以下身份验证方法：
    - 在hadoop-env配置分类中设置 AWS 访问密钥
    - 在中 AWS 对访问密钥进行编码 tempdir URI

有关使用连接器及其支持参数的更多信息，请参阅以下资源：

- Amazon Redshift Management Guide (《Amazon Redshift 管理指南》) 中的 [Amazon Redshift integration for Apache Spark](#) (适用于 Apache Spark 的 Amazon Redshift 集成)
- Github 上的 [spark-redshift 社区存储库](#)

## Spark 发行历史记录

下表列出了 Amazon 的每个发行版本中包含的 Spark 版本EMR，以及随应用程序一起安装的组件。有关每个发行版本中的组件版本，请参阅 [亚马逊 EMR 7.x 发布版本](#)、[亚马逊 EMR 6.x 发布版本](#) 或 [亚马逊 EMR 5.x 发布版本](#) 中的发行版“组件版本”部分。

### Important

[Apache Spark 版本 2.3.1](#)，从亚马逊EMR版本 5.16.0 开始推出，地址CVE为 [-2018-8024](#) 和 [-2018-1334](#)。CVE建议您将 Spark 的早期版本迁移到 Spark 2.3.1 版本或更高版本。

### Spark 版本信息

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-7.2.0	3.5.1	delta, emrfs, emr-goodies, emr-ddb, emr-s3-select,

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
		hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.36.2	2.4.8	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-7.1.0	3.5.0	delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-7.0.0	3.5.0	delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave



Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-6.15.0	3.4.1	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.14.0	3.4.1	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-6.13.0	3.4.1	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.12.0	3.4.0	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-6.11.1	3.3.2	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.11.0	3.3.2	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-6.10.1	3.3.1	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.10.0	3.3.1	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-6.9.1	3.3.0	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.9.0	3.3.0	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-6.8.1	3.3.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.8.0	3.3.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-6.7.0	3.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.36.1	2.4.8	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.36.0	2.4.8	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.6.0	3.2.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave



Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.35.0	2.4.8	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.5.0	3.1.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-6.4.0	3.1.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.3.1	3.1.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-6.3.0	3.1.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.2.1	3.0.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-6.2.0	3.0.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.1.1	3.0.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-6.1.0	3.0.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.0.1	2.4.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-6.0.0	2.4.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.34.0	2.4.8	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.33.1	2.4.7	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.33.0	2.4.7	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.32.1	2.4.7	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.32.0	2.4.7	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave



Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.31.1	2.4.6	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.31.0	2.4.6	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.30.2	2.4.5	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.30.1	2.4.5	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.30.0	2.4.5	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-notebook-env, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.29.0	2.4.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.28.1	2.4.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.28.0	2.4.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.27.1	2.4.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.27.0	2.4.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.26.0	2.4.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.25.0	2.4.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.24.1	2.4.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.24.0	2.4.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.23.1	2.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.23.0	2.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave



Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.22.0	2.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.21.2	2.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.21.1	2.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.21.0	2.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.20.1	2.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.20.0	2.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.19.1	2.3.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.19.0	2.3.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.18.1	2.3.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.18.0	2.3.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.17.2	2.3.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.17.1	2.3.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.17.0	2.3.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.16.1	2.3.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.16.0	2.3.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.15.1	2.3.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave



Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.15.0	2.3.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.14.2	2.3.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.14.1	2.3.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.14.0	2.3.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.13.1	2.3.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.13.0	2.3.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.12.3	2.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.12.2	2.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.12.1	2.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.12.0	2.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.11.4	2.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.11.3	2.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.11.2	2.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.11.1	2.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.11.0	2.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.10.1	2.2.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave



Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.10.0	2.2.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.9.1	2.2.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.9.0	2.2.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.8.3	2.2.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.8.2	2.2.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.8.1	2.2.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.8.0	2.2.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.7.1	2.1.1	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.7.0	2.1.1	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.6.1	2.1.1	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.6.0	2.1.1	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.5.4	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.5.3	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.5.2	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.5.1	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.5.0	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.4.1	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.4.0	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.3.2	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave



Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.3.1	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.3.0	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.2.3	2.0.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.2.2	2.0.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.2.1	2.0.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.2.0	2.0.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.1.1	2.0.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.1.0	2.0.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.0.3	2.0.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-5.0.2	2.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.0.1	2.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.0.0	2.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-4.9.6	1.6.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.9.5	1.6.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.9.4	1.6.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-4.9.3	1.6.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.9.2	1.6.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.9.1	1.6.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-4.8.5	1.6.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.8.4	1.6.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.8.3	1.6.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-4.8.2	1.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.8.1	1.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.8.0	1.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave



Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-4.7.4	1.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.7.3	1.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.7.2	1.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-4.7.1	1.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.7.0	1.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.6.1	1.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-4.6.0	1.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resource manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.5.0	1.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resource manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.4.0	1.6.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resource manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-4.3.0	1.6.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.2.0	1.5.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.1.0	1.5.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Amazon EMR 发布标签	Spark 版本	随 Spark 安装的组件
emr-4.0.0	1.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datano de, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop- yarn-nodemanager, hadoop- yarn-resourcemanager, spark- client, spark-history-server, spark-on-yarn, spark-yarn- slave

# Apache Sqoop

Apache Sqoop 是一款在亚马逊 S3、Hadoop 和数据库之间传输数据的工具。HDFS RDBMS有关更多信息，请参阅 [Apache Sqoop 网站](#)。Sqoop 包含在亚马逊 5.0.0 及更高EMR版本中。早期发行版包含包含 Sqoop，将其用作沙盒应用程序。有关更多信息，请参阅 [亚马逊 EMR 4.x 发布版本](#)。

## 主题

- [Sqoop 版本信息](#)
- [在亚马逊上使用 Sqoop 的注意事项 EMR](#)
- [Sqoop 发行历史记录](#)

## Sqoop 版本信息

### 适用于 7.2.0 的 Sqoop 版本

下表列出了最新版本的亚马逊 EMR 7.x 系列中包含的 Sqoop 版本，以及亚马逊与 Sqoop 一起EMR安装的组件。

有关此版本中与 Sqoop 一起安装的组件的版本，请参阅 [7.2.0 版本](#)的组件版本。

### emr-7.2.0 的 Sqoop 版本信息

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-7.2.0	Sqoop 1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

### Sqoop 的 6.15.0 版本

下表列出了最新版本的亚马逊 EMR 6.x 系列中包含的 Sqoop 版本，以及亚马逊与 Sqoop 一起EMR安装的组件。

有关此发行版中随 Sqoop 安装的组件版本，请参阅[发行版 6.15.0 组件版本](#)。

#### emr-6.15.0 的 Sqoop 版本信息

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-6.15.0	Sqoop 1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

#### 适用于 5.36.2 的 Sqoop 版本

下表列出了最新版本的亚马逊 EMR 5.x 系列中包含的 Sqoop 版本，以及亚马逊与 Sqoop 一起EMR安装的组件。

有关此版本中与 Sqoop 一起安装的组件的版本，请参阅[发行版 5.36.2 组件版本](#)。

#### emr-5.36.2 的 Sqoop 版本信息

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-5.36.2	Sqoop 1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resour

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
		cemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

## 在亚马逊上使用 Sqoop 的注意事项 EMR

在亚马逊上运行 Sqoop 时，请考虑以下几点。EMR

### 将 Sqoop 与集成一起使用 HCatalog

亚马逊上的 Sqoop EMR 支持 [Sqoop](#) 集成。HCatalog 当您使用 Sqoop 将输出写入 Amazon S3 中的 HCatalog 表时，请通过

将 `mapred.output.direct.NativeS3FileSystem` 和 `mapred.output.direct.EmrFileSystem` 属性设置为来禁用亚马逊 EMR 直接写入。false 有关更多信息，请参阅 [使用 HCatalog](#)。您可使用 `Hadoop -D mapred.output.direct.NativeS3FileSystem=false` 和 `-D mapred.output.direct.EmrFileSystem=false` 命令。如果您未禁用直接写入，则不会发生错误，但在 Amazon S3 中创建表时不会写入任何数据。

### Sqoop JDBC 和数据库支持

默认情况下，Sqoop 安装了 MariaDB 和 Postgre 驱动程序。SQL 为 Sqoop 安装的 Postgre SQL 驱动程序仅适用于 Postgre 8.4。SQL 要为 Sqoop 安装一组备用 JDBC 连接器，请连接到群集主节点并将其安装在中。/usr/lib/sqoop/lib 以下是各种 JDBC 连接器的链接：

- MariaDB：[关于 MariaDB Connector/J](#)。
- PostgreSQL：[Postgre 驱动程序 SQL JDBC](#)。
- SQL Server：[下载适用于 SQL 服务器的微软 JDBC 驱动程序](#)。
- 我的 SQL：[下载 Connector/J](#)
- 甲骨文：[UCP 从 Oracle Maven 存储库中获取甲骨文 JDBC 驱动程序](#)

Sqoop 支持的数据库列在以下网址上 [http://sqoop.apache.org/docs/\*version\*/SqoopUserGuide.html#\\_supported\\_databases](http://sqoop.apache.org/docs/version/SqoopUserGuide.html#_supported_databases)，哪里 *version* 是你正在使用的 Sqoop 版本，例如 1.4.6。如果 JDBC 连接字符串与此列表中的字符串不匹配，则必须指定驱动程序。

例如，您可以使用以下命令导出到 Amazon Redshift 数据库表（对于 JDBC 4.1）：



```
sqoop export --connect jdbc:redshift://$MYREDSHIFTHOST:5439/mydb --table mysqoopexport
--export-dir s3://mybucket/myinputfiles/ --driver com.amazon.redshift.jdbc41.Driver --
username master --password Mymasterpass1
```

你可以同时使用 MariaDB 和 My 的 SQL 连接字符串，但是如果你指定了 MariaDB 连接字符串，则需要指定驱动程序：

```
sqoop export --connect jdbc:mariadb://$HOSTNAME:3306/mydb --table mysqoopexport
--export-dir s3://mybucket/myinputfiles/ --driver org.mariadb.jdbc.Driver --
username master --password Mymasterpass1
```

如果您使用安全套接字层加密来访问数据库，则需要使用以下 Sqoop 导出示例中的 JDBCURI 类似内容：

```
sqoop export --connect jdbc:mariadb://$HOSTNAME:3306/mydb?
verifyServerCertificate=false&useSSL=true&requireSSL=true --table mysqoopexport
--export-dir s3://mybucket/myinputfiles/ --driver org.mariadb.jdbc.Driver --
username master --password Mymasterpass1
```

有关 SSL 加密的更多信息 RDS，请参阅 Amazon RDS 用户指南中的[使用 SSL 加密与数据库实例的连接](#)。

有关更多信息，请参阅 [Apache Sqoop](#) 文档。

## 保护密码

您可以选择以下几种方法来安全地传递密码：

### Java KeyStore

首选方法使用 Java KeyStore (JKS) 对密码进行加密，无需以可读格式存储密码。

1. 创建密码别名。当出现提示时，输入您用于访问数据库的密码。

```
hadoop credential create mydb.password.alias -provider jceks://hdfs/user/root/
mysql.password.jceks
```

2. 使用密码别名启动 Sqoop 作业：

```
sqoop export -Dhadoop.security.credential.provider.path=jceks://hdfs/user/
root/mysql.password.jceks --connect jdbc:mariadb://$HOSTNAME:3306/mydb
```

```
--table mysqoopexport --export-dir s3://mybucket/myinputfiles/ --driver
org.mariadb.jdbc.Driver --username master --password-alias mydb.password.alias
```

## --password-file

您可以使用 `--password-file` 命令通过文件传递密码，如以下示例所示：

1. 创建包含密码的新文件：

```
echo -n 'Mymasterpass1' > /home/hadoop/mysql-pass.password
```

2. 使用该文件启动 Sqoop 作业：

```
sqoop export --connect jdbc:mariadb://$HOSTNAME:3306/mydb --table mysqoopexport
--export-dir s3://mybucket/myinputfiles/ --driver org.mariadb.jdbc.Driver --
username master --password-file /home/hadoop/mysql-pass.password
```

## -P

您可以使用 `-P` 命令通过提示符输入密码，如以下示例所示：

```
sqoop export --connect jdbc:mariadb://$HOSTNAME:3306/mydb --table mysqoopexport --
export-dir s3://mybucket/myinputfiles/ --driver org.mariadb.jdbc.Driver --username
master -P
```

## Sqoop 发行历史记录

下表列出了 Amazon 的每个发行版本中包含的 Sqoop 版本EMR，以及随应用程序一起安装的组件。有关每个发行版本中的组件版本，请参阅 [亚马逊 EMR 7.x 发布版本](#)、[亚马逊 EMR 6.x 发布版本](#) 或 [亚马逊 EMR 5.x 发布版本](#) 中的发行版“组件版本”部分。

### Sqoop 版本信息

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-7.2.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode,

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
		hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-5.36.2	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-7.1.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-7.0.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.15.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.14.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-6.13.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.12.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.11.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-6.11.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.10.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.10.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-6.9.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.9.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.8.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-6.8.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.7.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-5.36.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client



Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-5.36.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.6.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-5.35.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-6.5.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.4.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.3.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-6.3.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.2.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.2.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-6.1.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.1.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-5.34.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-5.33.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-5.33.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-5.32.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-5.32.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-5.31.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-5.31.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-5.30.2	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-5.30.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-5.30.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-5.29.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.28.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.28.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client



Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-5.27.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.27.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.26.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-5.25.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.24.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.24.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-5.23.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.23.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.22.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-5.21.2	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.21.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.21.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-5.20.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.20.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.19.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-5.19.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.18.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.18.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-5.17.2	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.17.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.17.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-5.16.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.16.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.15.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client



Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-5.15.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.14.2	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.14.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-5.14.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.13.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.13.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-5.12.3	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.12.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.12.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-5.12.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.11.4	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.11.3	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-5.11.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.11.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.11.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-5.10.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.10.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.9.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-5.9.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.8.3	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.8.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-5.8.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.8.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.7.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client



Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-5.7.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.6.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.6.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-5.5.4	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.5.3	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.5.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-5.5.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, mysql-server, sqoop-client
emr-5.5.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, mysql-server, sqoop-client
emr-5.4.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, mysql-server, sqoop-client

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-5.4.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.3.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.3.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-5.3.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.2.3	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.2.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-5.2.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.2.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.1.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-5.1.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.0.3	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.0.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Amazon EMR 发布标签	Sqoop 版本	随 Sqoop 安装的组件
emr-5.0.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.0.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client



# TensorFlow

TensorFlow 是一个用于机器智能和深度学习应用程序的开源符号数学库。欲了解更多信息，请[TensorFlow 访问网站](#)。TensorFlow 在 Amazon EMR 发行版 5.17.0 及更高版本中可用。

下表列出了最新版本的 Amazon EMR 7.x 系列中 TensorFlow 包含的版本以及亚马逊随之EMR安装的组件。TensorFlow

有关此版本 TensorFlow 中安装的组件的版本，请参阅 [7.2.0 版组件版本](#)。

TensorFlow emr-7.2.0 的版本信息

Amazon EMR 发布标签	TensorFlow 版本	安装在一起的组件 TensorFlow
emr-7.2.0	TensorFlow 2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

下表列出了最新版本的 Amazon EMR 6.x 系列中 TensorFlow 包含的版本以及亚马逊随之EMR安装的组件。TensorFlow

有关此版本 TensorFlow 中安装的组件的版本，请参阅 [6.15.0 版组件版本](#)。

TensorFlow emr-6.15.0 的版本信息

Amazon EMR 发布标签	TensorFlow 版本	安装在一起的组件 TensorFlow
emr-6.15.0	TensorFlow 2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager

Amazon EMR 发布标签	TensorFlow 版本	安装在一起的组件 TensorFlow
		nager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

下表列出了最新版本的 Amazon EMR 5.x 系列中 TensorFlow 包含的版本以及亚马逊随之EMR安装的组件。 TensorFlow

有关此版本 TensorFlow 中安装的组件的版本，请参阅 [5.36.2 版组件版本](#)。

TensorFlow emr-5.36.2 的版本信息

Amazon EMR 发布标签	TensorFlow 版本	安装在一起的组件 TensorFlow
emr-5.36.2	TensorFlow 2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

## TensorFlow 按 Amazon EC2 实例类型构建

根据您为集群选择的实例类型，Amazon EMR 使用不同的 TensorFlow 库版本。下表按实例类型列出版本。

EC2实例类型	TensorFlow 建造
M5 和 C5	经过英特尔优化的 Tensorflow 1.9.0 MKL
P2	Tensorflow 1.9.0 带有 CUDA 9.2，cu 7.1 DNN
P3	Tensorflow 1.9.0 带有 CUDA 9.2、cu 7.1、2.2.13 DNN NCCL

EC2实例类型	TensorFlow 建造 <a href="#">Nvid NCCL ia</a> 仅在 P3 实例上可用。最终用户许可协议 (EULA)：在亚马逊上使用 Nvidia 组件 EMR，即表示您同意 <a href="#">产品</a> 中概述的条款和条件 EULA。
所有其他地址	Tensorflow 1.9.0

## 安全性

除了遵循 [TensorFlow 安全使用](#) 中的指导外，我们还建议您在私有子网中启动集群，以帮助限制对可信来源的访问。有关更多信息，请参阅《[亚马逊EMR管理指南](#)》中的[亚马逊VPC选项](#)。

## 使用 TensorBoard

TensorBoard 是一套用于 TensorFlow 程序的可视化工具。有关更多信息，请参阅 [TensorBoard : Tensorflow 网站上的可视化学习](#)。

要在 TensorBoard Amazon 上使用EMR，您必须从集群主节点 TensorBoard 上启动。

在亚马逊上将 tensorboard 与 Tensorflow 配合使用 EMR

1. 使用连接到集群的主节点SSH。有关更多信息，请参阅 [《Amazon EMR 管理指南》SSH中的“使用连接到主节点”](#)。
2. 键入以下命令，在主节点上启动 TensorFboard。将 `/my/log/directory` 替换为您使用摘要写入器生成和存储摘要数据的主节点上的目录。

Amazon EMR 5.19.0 and later

```
python3 -m tensorboard.main --logdir=/home/hadoop/tensor --bind_all
```

Amazon EMR 5.18.1 and earlier

```
python3 -m tensorboard.main --logdir=/my/log/dir
```

默认情况下，主节点 TensorBoard 使用端口 6006 和主节点公共DNS名称进行主机。启动后 TensorBoard，命令行输出URL将显示可用于连接的 TensorBoard，如以下示例所示：

```
TensorBoard 1.9.0 at http://master-public-dns-name:6006 (Press CTRL+C to quit)
```

3. 设置来自受信任客户端对主节点上 Web 界面的访问权限。有关更多信息，请参阅《亚马逊EMR管理指南》中的[查看亚马逊EMR集群上托管的 Web 界面](#)。
4. 开 TensorBoard 业于http://*master-public-dns-name*:6006。

## TensorFlow 发布历史

下表列出了 Amazon 的每个发行版本中 TensorFlow 包含的版本EMR，以及随应用程序一起安装的组件。有关每个发行版本中的组件版本，请参阅 [亚马逊 EMR 7.x 发布版本](#)、[亚马逊 EMR 6.x 发布版本](#) 或 [亚马逊 EMR 5.x 发布版本](#) 中的发行版“组件版本”部分。

### TensorFlow 版本信息

Amazon EMR 发布标签	TensorFlow 版本	安装在一起的组件 TensorFlow
emr-7.2.0	2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.36.2	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resour

Amazon EMR 发布标签	TensorFlow 版本	安装在一起的组件 TensorFlow
		cemanager, hadoop-yarn-timeline-server, tensorflow
emr-7.1.0	2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow
emr-7.0.0	2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow
emr-6.15.0	2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow

Amazon EMR 发布标签	TensorFlow 版本	安装在一起的组件 TensorFlow
emr-6.14.0	2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.13.0	2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.12.0	2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Amazon EMR 发布标签	TensorFlow 版本	安装在一起的组件 TensorFlow
emr-6.11.1	2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.11.0	2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.10.1	2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Amazon EMR 发布标签	TensorFlow 版本	安装在一起的组件 TensorFlow
emr-6.10.0	2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.9.1	2.10.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.9.0	2.10.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow



Amazon EMR 发布标签	TensorFlow 版本	安装在一起的组件 TensorFlow
emr-6.8.1	2.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.8.0	2.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.7.0	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Amazon EMR 发布标签	TensorFlow 版本	安装在一起的组件 TensorFlow
emr-5.36.1	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.36.0	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.6.0	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Amazon EMR 发布标签	TensorFlow 版本	安装在一起的组件 TensorFlow
emr-5.35.0	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.5.0	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.4.0	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Amazon EMR 发布标签	TensorFlow 版本	安装在一起的组件 TensorFlow
emr-6.3.1	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.3.0	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.2.1	2.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Amazon EMR 发布标签	TensorFlow 版本	安装在一起的组件 TensorFlow
emr-6.2.0	2.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.1.1	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.1.0	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Amazon EMR 发布标签	TensorFlow 版本	安装在一起的组件 TensorFlow
emr-6.0.1	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.0.0	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.34.0	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Amazon EMR 发布标签	TensorFlow 版本	安装在一起的组件 TensorFlow
emr-5.33.1	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.33.0	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.32.1	2.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Amazon EMR 发布标签	TensorFlow 版本	安装在一起的组件 TensorFlow
emr-5.32.0	2.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.31.1	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.31.0	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow



Amazon EMR 发布标签	TensorFlow 版本	安装在一起的组件 TensorFlow
emr-5.30.2	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.30.1	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.30.0	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Amazon EMR 发布标签	TensorFlow 版本	安装在一起的组件 TensorFlow
emr-5.29.0	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.28.1	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.28.0	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Amazon EMR 发布标签	TensorFlow 版本	安装在一起的组件 TensorFlow
emr-5.27.1	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.27.0	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.26.0	1.13.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Amazon EMR 发布标签	TensorFlow 版本	安装在一起的组件 TensorFlow
emr-5.25.0	1.13.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.24.1	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.24.0	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Amazon EMR 发布标签	TensorFlow 版本	安装在一起的组件 TensorFlow
emr-5.23.1	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.23.0	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.22.0	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Amazon EMR 发布标签	TensorFlow 版本	安装在一起的组件 TensorFlow
emr-5.21.2	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.21.1	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.21.0	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Amazon EMR 发布标签	TensorFlow 版本	安装在一起的组件 TensorFlow
emr-5.20.1	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.20.0	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.19.1	1.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Amazon EMR 发布标签	TensorFlow 版本	安装在一起的组件 TensorFlow
emr-5.19.0	1.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.18.1	1.9.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.18.0	1.9.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow



Amazon EMR 发布标签	TensorFlow 版本	安装在一起的组件 TensorFlow
emr-5.17.2	1.9.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.17.1	1.9.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.17.0	1.9.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

# Apache Tez

Apache Tez 是一个框架，用于创建用于处理数据的任务的复杂有向无环图 (DAG)。在某些用例中，你可以将其用作 Hadoop MapReduce 的替代方案。例如，你可以使用 Hadoop 运行 Pig 和 Hive 工作流程，MapReduce 也可以使用 Tez 作为执行引擎。有关更多信息，请参阅 <https://tez.apache.org/>。亚马逊EMR发布的4.7.0及更高版本包括Tez。

下表列出了最新版本的亚马逊 EMR 7.x 系列中包含的 Tez 版本，以及亚马逊随着 Tez EMR 安装的组件。

有关此版本中与 Tez 一起安装的组件的版本，请参阅 [7.2.0 版组件版本](#)。

## emr-7.2.0 的 Tez 版本信息

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-7.2.0	Tez 0.10.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker

下表列出了最新版本的亚马逊 EMR 6.x 系列中包含的 Tez 版本，以及亚马逊随着 Tez EMR 安装的组件。

有关此发行版中随 Tez 安装的组件版本，请参阅[发行版 6.15.0 组件版本](#)。

## emr-6.15.0 的 Tez 版本信息

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-6.15.0	Tez 0.10.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
		hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker

下表列出了最新版本的亚马逊 EMR 5.x 系列中包含的 Tez 版本，以及亚马逊随着 Tez EMR 安装的组件。

有关此版本中与 Tez 一起安装的组件的版本，请参阅 [5.36.2 版组件版本](#)。

emr-5.36.2 的 Tez 版本信息

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-5.36.2	Tez 0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

## 主题

- [使用 Tez 创建集群](#)
- [配置 Tez](#)
- [Tez Web UI](#)
- [时间线服务器](#)
- [Tez 发行历史记录](#)

## 使用 Tez 创建集群

要安装 Tez，请在创建集群时选择 Apache Tez 作为应用程序。

使用控制台创建安装了 Tez 的集群

1. 导航到新的 Amazon EMR 控制台，然后从侧面导航栏中选择“切换到旧主机”。有关切换到旧控制台后预期情况的更多信息，请参阅 [Using the old console](#)。
2. 依次选择 Create cluster (创建集群)、Go to advanced options (转到高级选项)。
3. 在软件配置下，选择 emr-4.7.0 或更高的版本。
4. 选择 Tez 以及您想让 Amazon EMR 安装的其他应用程序。
5. 根据需要选择其它选项，然后选择 Create cluster (创建集群)。

要使用 Tez 创建集群，请使用 AWS CLI

- 使用 create-cluster 命令以及 -- applications 选项指定 Tez。以下示例将创建一个安装了 Tez 的集群。

### Note

为了便于读取，包含 Linux 行继续符 (\)。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (^)。

```
aws emr create-cluster --name "Cluster with Tez" --release-label emr-7.2.0 \  
--applications Name=Tez --ec2-attributes KeyName=myKey \  
--instance-type m5.xlarge --instance-count 3 --use-default-roles
```

## 配置 Tez

您可以使用 tez-site 配置分类设置值来自定义 Tez，该配置分类将配置 tez-site.xml 配置文件中的设置。有关更多信息，请参阅 Apache Tez 文档 [TezConfiguration](#) 中的。要将 Hive 或 Pig 更改为使用 Tez 执行引擎，请根据需要使用 hive-site 和 pig-properties 配置分类。示例如下。

## 示例配置

Example 示例：自定义 Tez 根日志记录级别，并将 Tez 设置为 Hive 和 Pig 的执行引擎

下面显示的 `create-cluster` 命令将创建一个安装了 Tez、Hive 和 Pig 的集群。该命令引用存储在 Amazon S3 中的文件 `myConfig.json`，该文件为将 `tez.am.log.level` 设置为 `DEBUG` 的 `tez-site` 分类指定属性，以及将执行引擎设置为 Tez，以便 Hive 和 Pig 使用 `hive-site` 和 `pig-properties` 配置分类。

### Note

为了便于读取，包含 Linux 行继续符 (`\`)。它们可以通过 Linux 命令删除或使用。对于 Windows，请将它们删除或替换为脱字号 (`^`)。

```
aws emr create-cluster --release-label emr-7.2.0 \  
--applications Name=Tez Name=Hive Name=Pig --ec2-attributes KeyName=myKey \  
--instance-type m5.xlarge --instance-count 3 \  
--configurations https://s3.amazonaws.com/mybucket/myfolder/myConfig.json --use-  
default-roles
```

下面显示的是 `myConfig.json` 的内容示例。

```
[  
  {  
    "Classification": "tez-site",  
    "Properties": {  
      "tez.am.log.level": "DEBUG"  
    }  
  },  
  {  
    "Classification": "hive-site",  
    "Properties": {  
      "hive.execution.engine": "tez"  
    }  
  },  
  {  
    "Classification": "pig-properties",  
    "Properties": {  
      "exectype": "tez"  
    }  
  }  
]
```

```

    }
  }
]
```

### Note

在 Amazon 5.21.0 及更高EMR版本中，您可以覆盖集群配置，并为正在运行的集群中的每个实例组指定其他配置分类。为此，您可以使用 Amazon EMR 控制台、AWS Command Line Interface (AWS CLI) 或 AWS SDK。有关更多信息，请参阅[为运行的集群中的实例组提供配置](#)。

## Tez 异步拆分打开

当表路径中有大量小文件并且查询尝试读取所有文件时，与每个单独拆分相对应的每个小文件将合并并在一个 Tez 分组拆分下。然后，单个映射器将处理单个 Tez 分组拆分。由于执行是同步的，因此分组拆分下的每个单独拆分都将逐一得到处理。这就要求 RecordReader 对象同步处理各个拆分。

名称	分类	描述
<code>tez.grouping.split.init.threads</code>	tez-site	指定 Tez 用于预启动 RecordReaders 和打开拆分的进程守护程序线程数。对于 ACID 表， <code>tez.grouping.split.init.threads</code> 支持的最大值为 1。
<code>tez.grouping.split.init.recordreaders</code>	tez-site	指定要由进程守护程序线程保持预初始化 RecordReaders 的数量。当 Tez 分组拆分包含大量 InputSplits 时，这会有所帮助。处理这些输入拆分的 RecordReaders 初始化可以用进程守护线程异步完成，而不是顺序处理。

## Tez 异步拆分打开的基准测试

我们已使用以下环境和配置对 Tez 异步拆分打开功能进行了基准测试：

- 基准测试环境 — Amazon EMR 集群有 1 个主节点在使用 m5.16xlarge，16 个核心节点在使用 m5.16xlarge。
- 基准配置 – 要模拟单个 Tez 分组拆分中存在大量输入拆分的基准测试场景，tez.grouping.split-count 设置为 1。
- 用于基准测试的表 – 该表包含 200 个分区，其中每个分区包含一个文件。基准测试在该表包含 CSV 文件时以及该表包含 parquet 文件时进行。Hive 查询基准测试：从表中 SELECT COUNT(\*) 十次，取平均运行时间。
- 启用 Tez 异步拆分打开的配置 – 如下所示：
  - tez.grouping.split.init.threads = 4
  - tez.grouping.split.init.recordreaders = 10

数据集	功能已禁用 ( 基线 )	功能已启用	改进
CSV 数据集	90.26 秒	79.20 秒	12.25%
Parquet 数据集	54.67 秒	42.23 秒	22.75%

## Tez Web UI

Tez 拥有自己的 Web 用户界面。要查看 Web 用户界面，请参阅以下内容 URL。

```
http://masterDNS:8080/tez-ui
```

要在 Tez Web UI 上启用 Hive 查询选项卡，请设置以下配置。

```
[
  {
    "Classification": "hive-site",
    "Properties": {
      "hive.exec.pre.hooks": "org.apache.hadoop.hive.q1.hooks.ATSHook",
      "hive.exec.post.hooks": "org.apache.hadoop.hive.q1.hooks.ATSHook",
      "hive.exec.failure.hooks": "org.apache.hadoop.hive.q1.hooks.ATSHook"
    }
  }
]
```

```
}
]
```

您还可以使用控制台中集群详细信息页面的“YARN应用程序用户界面”选项卡上的链接查看 Tez、Spark 和应用程序 UI 的详细信息。Amazon EMR 应用程序用户界面 (UI) 托管在集群外并在集群终止后可用。它们不需要您设置SSH连接或 Web 代理，因此您可以更轻松地对活跃的作业和作业历史进行故障排除和分析。

有关更多信息，请参阅《Amazon EMR 管理指南》中的[查看应用程序历史记录](#)。

## 时间线服务器

YARN时间轴服务器配置为在安装 Tez 时运行。要查看通过 Tez 或使用时间轴服务器 MapReduce 执行引擎提交的作业，请使用查看 Web 用户界面。URL <http://master-public-DNS:8188>有关更多信息，请参阅《亚马逊EMR管理指南》中的[查看亚马逊EMR集群上托管的 Web 界面](#)。

## Tez 发行历史记录

下表列出了 Amazon 的每个发行版本中包含的 Tez 版本EMR，以及随应用程序一起安装的组件。有关每个发行版本中的组件版本，请参阅 [亚马逊 EMR 7.x 发布版本](#)、[亚马逊 EMR 6.x 发布版本](#) 或 [亚马逊 EMR 5.x 发布版本](#) 中的发行版“组件版本”部分。

### Tez 版本信息

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-7.2.0	0.10.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker
emr-5.36.2	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-



Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
		hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-7.1.0	0.10.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker
emr-7.0.0	0.10.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-6.15.0	0.10.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker
emr-6.14.0	0.10.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker
emr-6.13.0	0.10.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-6.12.0	0.10.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker
emr-6.11.1	0.10.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker
emr-6.11.0	0.10.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-6.10.1	0.10.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker
emr-6.10.0	0.10.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker
emr-6.9.1	0.10.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-6.9.0	0.10.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.8.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.8.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-6.7.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.36.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.36.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-6.6.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.35.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.5.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-6.4.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.3.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.3.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn



Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-6.2.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.2.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.1.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-6.1.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.0.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.0.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-5.34.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.33.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.33.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-5.32.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.32.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.31.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-5.31.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.30.2	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.30.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-5.30.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.29.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.28.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-5.28.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.27.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.27.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-5.26.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.25.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.24.1	0.9.1	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn



Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-5.24.0	0.9.1	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.23.1	0.9.1	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.23.0	0.9.1	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-5.22.0	0.9.1	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.21.2	0.9.1	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.21.1	0.9.1	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-5.21.0	0.9.1	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.20.1	0.9.1	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.20.0	0.9.1	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-5.19.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.19.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.18.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-5.18.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.17.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.17.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-5.17.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.16.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.16.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-5.15.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.15.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.14.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-5.14.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.14.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.13.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn



Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-5.13.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.12.3	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.12.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-5.12.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.12.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.11.4	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-5.11.3	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.11.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.11.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-5.11.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.10.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.10.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-5.9.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.9.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.8.3	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-5.8.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.8.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.8.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-5.7.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.7.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.6.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-5.6.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.5.4	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.5.3	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn



Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-5.5.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.5.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.5.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-5.4.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.4.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.3.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-5.3.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.3.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.2.3	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-5.2.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.2.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.2.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-5.1.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.1.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.0.3	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-5.0.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.0.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.0.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-4.9.6	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-4.9.5	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-4.9.4	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-4.9.3	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-4.9.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-4.9.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn



Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-4.8.5	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-4.8.4	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-4.8.3	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-4.8.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-4.8.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-4.8.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-4.7.4	0.8.3	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-4.7.3	0.8.3	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-4.7.2	0.8.3	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Amazon EMR 发布标签	Tez 版本	随 Tez 安装的组件
emr-4.7.1	0.8.3	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-4.7.0	0.8.3	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

## Tez 发布说明 ( 按版本分类 )

### 主题

- [亚马逊 EMR 7.1.0-Tez 发行说明](#)
- [亚马逊 EMR 6.15.0-Tez 发行说明](#)
- [亚马逊 EMR 6.14.0-Tez 发行说明](#)
- [亚马逊 EMR 6.13.0-Tez 发行说明](#)
- [亚马逊 EMR 6.12.0-Tez 发行说明](#)
- [亚马逊 EMR 6.11.0-Tez 发行说明](#)
- [亚马逊 EMR 6.10.0-Tez 发行说明](#)
- [亚马逊 EMR 6.9.0-Tez 发行说明](#)
- [亚马逊 EMR 6.8.0-Tez 发行说明](#)

- [亚马逊 EMR 6.7.0-Tez 发行说明](#)
- [亚马逊 EMR 6.6.0-Tez 发行说明](#)

## 亚马逊 EMR 7.1.0-Tez 发行说明

### 亚马逊 EMR 7.1.0-Tez 变更

类型	描述
错误修复	<a href="#">TEZ-4394</a> — Netty4 ShuffleHandler : 应该使用 1 个 boss 话题 (#256)
Upgrade	将 Netty 升级到 4.1.100.Final。
Upgrade	将 Jetty 升级到 9.4.53.v20231009。

## 亚马逊 EMR 6.15.0-Tez 发行说明

### 亚马逊 EMR 6.15.0-Tez 变更

类型	描述
功能	<a href="#">TEZ-4397</a> : 打开 Tez 输入异步拆分
Upgrade	<a href="#">TEZ-4493</a> : 将 Apache Hadoop 升级到 3.3.6

### 亚马逊 EMR 6.15.0-Tez 功能

- [Tez 异步拆分打开](#) — Amazon EMR 6.15.0 引入了配置，您可以指定这些配置来异步打开 Tez 分组拆分中的输入拆分。该功能由 [TEZ-4397](#) 启动，但在 Hive 中出现了回归。OSSAmazon EMR Hive 修复了 Hiv ACID e 表中的回归和其他错误。当单个 Tez 分组拆分中有大量输入拆分时，此改进可以提高读取查询的性能。有关更多信息，请参阅 [Tez 异步拆分打开](#)。

## 亚马逊 EMR 6.14.0-Tez 发行说明

### 亚马逊 EMR 6.14.0-Tez 变更

类型	描述
改进	将 Tez 中的 TLS 版本升级到 1.2

## 亚马逊 EMR 6.13.0-Tez 发行说明

### 亚马逊 EMR 6.13.0-Tez 变更

类型	描述
错误修复	REVERT <a href="#">TEZ-4295</a> : 无法解压缩数据。缓冲区长度太小。
错误修复	REVERT <a href="#">TEZ-4302</a> : NullPointerException 在 with 中。CodecUtils GzipCodec
错误修复	REVERT <a href="#">TEZ-4234</a> : 压缩机可能导致 IllegalArgumentException 极限超过 Buffer.limit 容量。
错误修复	REVERT <a href="#">TEZ-4135</a> : 在执行内存中读取时改善内存分配。

## 亚马逊 EMR 6.12.0-Tez 发行说明

### 亚马逊 EMR 6.12.0-Tez 变更

类型	描述
改进	添加了对 JDK 11 和 JDK 17 运行时的 Support

类型	描述
错误修复	<a href="#">TEZ-4492</a> : 更新 Bowerrc 以使用 bower.herokuapp 镜像避免 Bower Registry _ 问题 (-2608) CERT EXPIRE BOWER
Upgrade	将 Surefire 升级到 3.0.0-M7

## 亚马逊 EMR 6.11.0-Tez 发行说明

### 亚马逊 EMR 6.11.0-Tez 变更

类型	描述
错误	修复了在清理随机数据的顶点等级时无效的顶点状态转换
错误	修复 shuffle 数据的清理DAG或顶级别清理不起作用
改进	默认启用 <code>tez.am.dag.cleanup.on.completion</code> 以清除已完成的随机播放数据 DAGs

## 亚马逊 EMR 6.10.0-Tez 发行说明

### 亚马逊 EMR 6.10.0-Tez 变更

类型	描述
功能	默认情况下启用 <code>tez.runtime.transfer.data-via-events.enabled</code>
逆向移植	<a href="#">TEZ-4450</a> : 修复通过数据移动事件传输随机数据时随机数据获取失败的问题
逆向移植	<a href="#">TEZ-4460</a> : 修复从 Tez Shuffle Handler 获取随机播放数据时出现的读取超时错误

类型	描述
逆向移植	<a href="#">TEZ-4455</a> : 添加 ShuffleHandler管道 LoggingHandler 以获得更好的可调试性
错误	修复启用了抢占任务后 Tez 任务间歇性卡住的问题

## 亚马逊 EMR 6.9.0-Tez 发行说明

### 亚马逊 EMR 6.9.0-Tez 变更

类型	描述
Upgrade	Tez 已升级到 0.10.2。有关更多信息，请参阅 <a href="#">change log for Apache Tez 0.10.2</a> ( Apache Tez 0.10.2 的更改日志 )。
Upgrade	将 Hadoop 升级到 3.3.3。
错误	<code>tez.runtime.transfer.data-via-events.enabled</code> 由于 <a href="#">TEZ-4450</a> ，默认情况下处于禁用状态。

## 亚马逊 EMR 6.8.0-Tez 发行说明

### 亚马逊 EMR 6.8.0-Tez 变更

类型	描述
逆向移植	<a href="#">TEZ-3363</a> : 删除 Shuffle Handler 的顶级别别的中间数据
逆向移植	<a href="#">TEZ-4129</a> : 删除 Shuffle Handler 尝试失败的中间尝试数据



类型	描述
逆向移植	<a href="#">TEZ-4430</a> : 修复了 <code>tez.task.launch.cmd-opts</code> 属性不起作用的问题

## 亚马逊 EMR 6.7.0-Tez 发行说明

### 亚马逊 EMR 6.7.0-Tez 变更

类型	描述
逆向移植	<a href="#">TEZ-4403</a> : 将版本升级到 <a href="#">SLF4J 1.7.3 6</a>
逆向移植	<a href="#">TEZ-4405</a> : 用 <code>reload4j</code> 替换 <code>log4j 1.x</code>
逆向移植	<a href="#">TEZ-4411</a> : Tez 构建失败 : <code>FileSaver</code> 找不到.js

## 亚马逊 EMR 6.6.0-Tez 发行说明

### 亚马逊 EMR 6.6.0-Tez 变更

类型	描述
逆向移植	<a href="#">TEZ-3918</a> : 修复了 <code>tez.task.log.level</code> 属性不起作用的问题。
逆向移植	<a href="#">TEZ-4353</a> : 将 <code>commons-io</code> 更新到 2.8.0。
逆向移植	<a href="#">TEZ-4114</a> : 从 <code>tez</code> 中移除直接的码头依赖关系。
逆向移植	<a href="#">TEZ-4323</a> : <code>Jetty</code> 罐子已从 <code>dist</code> 包裹中移除 , 包含 -4114。TEZ

# Apache Zeppelin

使用 Apache Zeppelin 作为用于交互式数据探索的笔记本。有关 Zeppelin 的更多信息，请参阅 <https://zeppelin.apache.org/>。齐柏林飞艇包含在 Amazon 5.0.0 及更高EMR版本中。早期发行版包含 Zeppelin，将其用作沙盒应用程序。有关更多信息，请参阅 [亚马逊 EMR 4.x 发布版本](#)。

要访问齐柏林飞艇网页界面，请设置通往主节点的SSH隧道和代理连接。有关更多信息，请参阅[查看 EMR集群上托管的 Web 界面](#)。

下表列出了最新版本的亚马逊 EMR 7.x 系列中包含的齐柏林飞艇版本，以及亚马逊随齐柏林飞艇一起安装的组件。EMR

有关此版本中与齐柏林飞艇一起安装的组件版本，请参阅 [7.2.0 版本](#) 组件版本。

emr-7.2.0 的齐柏林飞艇版本信息

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-7.2.0	Zeppelin 0.10.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

下表列出了最新版本的亚马逊 EMR 6.x 系列中包含的齐柏林飞艇版本，以及亚马逊随齐柏林飞艇一起安装的组件。EMR

有关此发行版中随 Zeppelin 安装的组件版本，请参阅[发行版 6.15.0 组件版本](#)。

## emr-6.15.0 的 Zeppelin 版本信息

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-6.15.0	Zeppelin 0.10.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

下表列出了最新版本的亚马逊 EMR 5.x 系列中包含的齐柏林飞艇版本，以及亚马逊随齐柏林飞艇一起安装的组件。EMR

有关此版本中与齐柏林飞艇一起安装的组件版本，请参阅 [5.36.2 版本](#) 组件版本。

## emr-5.36.2 的齐柏林飞艇版本信息

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.36.2	Zeppelin 0.10.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
		spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

## 主题

- [在 Amazon 上使用齐柏林飞艇时的注意事项 EMR](#)
- [Zeppelin 发行历史记录](#)

## 在 Amazon 上使用齐柏林飞艇时的注意事项 EMR

- 使用相同的[SSH隧道方法](#)连接齐柏林飞艇，以连接到主节点上的其他 Web 服务器。Zeppelin 服务器可以在端口 8890 上找到。
- [亚马逊上的齐柏林飞艇EMR发布了 5.0.0 及更高版本，支持 Shiro 身份验证。](#)
- 亚马逊上的齐柏林飞艇EMR发布版本5.8.0及更高版本，支持使用Glue Data Catalog作为 AWS Spark的元数据库。SQL有关更多信息，请参阅[使用 AWS Glue 数据目录作为 Spark SQL 的元数据库](#)。
- 齐柏林飞艇不使用集群spark-defaults.conf配置文件中定义的某些设置，尽管它会指示YARN动态分配执行器（如果您已设置为）。spark.dynamicAllocation.enabled true您必须使用Zeppelin Interpreter (解释器) 选项卡设置执行者设置 (如内存和内核)，然后为要使用的设置重新启动解释器。
- 亚马逊EMR发布的 6.10.0 及更高版本支持 Apache Zeppelin 与 Apache Flink 集成。请参阅在[Amazon EMR 中通过 Zeppelin 使用 Flink 作业](#)了解更多信息。
- 亚马逊上的齐柏林飞艇EMR不支持 SparkR 解释器。

## Zeppelin 发行历史记录

下表列出了 Amazon 的每个发行版本中包含的齐柏林飞艇版本EMR，以及随应用程序一起安装的组件。有关每个发行版本中的组件版本，请参阅[亚马逊 EMR 7.x 发布版本](#)、[亚马逊 EMR 6.x 发布版本](#)或[亚马逊 EMR 5.x 发布版本](#)中的发行版“组件版本”部分。

## Zeppelin 版本信息

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-7.2.0	0.10.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.36.2	0.10.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-7.1.0	0.10.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
		server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-7.0.0	0.10.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-6.15.0	0.10.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.14.0	0.10.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-6.13.0	0.10.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.12.0	0.10.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server



Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-6.11.1	0.10.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.11.0	0.10.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-6.10.1	0.10.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.10.0	0.10.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-6.9.1	0.10.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.9.0	0.10.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-6.8.1	0.10.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.8.0	0.10.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-6.7.0	0.10.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.36.1	0.10.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.36.0	0.10.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.6.0	0.10.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.35.0	0.10.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.5.0	0.10.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-6.4.0	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.3.1	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server



Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-6.3.0	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.2.1	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-6.2.0	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.1.1	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-6.1.0	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.0.1	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-6.0.0	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.34.0	0.10.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.33.1	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.33.0	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.32.1	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.32.0	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.31.1	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.31.0	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.30.2	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.30.1	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server



Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.30.0	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.29.0	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.28.1	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.28.0	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.27.1	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.27.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.26.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.25.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.24.1	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.24.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.23.1	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.23.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.22.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.21.2	0.8.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.21.1	0.8.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.21.0	0.8.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server



Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.20.1	0.8.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.20.0	0.8.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.19.1	0.8.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.19.0	0.8.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.18.1	0.8.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.18.0	0.8.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.17.2	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.17.1	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.17.0	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.16.1	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.16.0	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.15.1	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.15.0	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.14.2	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.14.1	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.14.0	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server



Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.13.1	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.13.0	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.12.3	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.12.2	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.12.1	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.12.0	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.11.4	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.11.3	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.11.2	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.11.1	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.11.0	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.10.1	0.7.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.10.0	0.7.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.9.1	0.7.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.9.0	0.7.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.8.3	0.7.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server



Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.8.2	0.7.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.8.1	0.7.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.8.0	0.7.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.7.1	0.7.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.7.0	0.7.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.6.1	0.7.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.6.0	0.7.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.5.4	0.7.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.5.3	0.7.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.5.2	0.7.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.5.1	0.7.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.5.0	0.7.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.4.1	0.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.4.0	0.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.3.2	0.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.3.1	0.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.3.0	0.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.2.3	0.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.2.2	0.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.2.1	0.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server



Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.2.0	0.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.1.1	0.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.1.0	0.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.0.3	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.0.2	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.0.1	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Amazon EMR 发布标签	Zeppelin 版本	随 Zeppelin 安装的组件
emr-5.0.0	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

# Apache ZooKeeper

Apache ZooKeeper 是一项集中式服务，用于维护配置信息、命名、提供分布式同步和提供组服务。有关的更多信息 ZooKeeper，请参阅 <http://zookeeper.apache.org/>。

下表列出了最新版本的 Amazon EMR 7.x 系列中 ZooKeeper 包含的版本以及亚马逊随之EMR安装的组件。 ZooKeeper

有关此版本 ZooKeeper 中安装的组件的版本，请参阅 [7.2.0 版组件版本](#)。

ZooKeeper emr-7.2.0 的版本信息

Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-7.2.0	ZooKeeper 3.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

下表列出了最新版本的 Amazon EMR 6.x 系列中 ZooKeeper 包含的版本以及亚马逊随之EMR安装的组件。 ZooKeeper

有关此版本 ZooKeeper 中安装的组件的版本，请参阅 [6.15.0 版组件版本](#)。

ZooKeeper emr-6.15.0 的版本信息

Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-6.15.0	ZooKeeper 3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-

Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
		server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

下表列出了最新版本的 Amazon EMR 5.x 系列中 ZooKeeper 包含的版本以及亚马逊随之EMR安装的组件。 ZooKeeper

有关此版本 ZooKeeper 中安装的组件的版本，请参阅 [5.36.2 版组件版本](#)。

ZooKeeper emr-5.36.2 的版本信息

Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-5.36.2	ZooKeeper 3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

主题

- [ZooKeeper 发布历史](#)

## ZooKeeper 发布历史

下表列出了 Amazon 的每个发行版本中 ZooKeeper 包含的版本EMR，以及随应用程序一起安装的组件。有关每个发行版本中的组件版本，请参阅 [亚马逊 EMR 7.x 发布版本](#)、[亚马逊 EMR 6.x 发布版本](#) 或 [亚马逊 EMR 5.x 发布版本](#) 中的发行版“组件版本”部分。

## ZooKeeper 版本信息

Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-7.2.0	3.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.36.2	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-7.1.0	3.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-7.0.0	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.15.0	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.14.0	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-6.13.0	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.12.0	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.11.1	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server



Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-6.11.0	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.10.1	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.10.0	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-6.9.1	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.9.0	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.8.1	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-6.8.0	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.7.0	3.5.7	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.36.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-5.36.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.6.0	3.5.7	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.35.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-6.5.0	3.5.7	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.4.0	3.5.7	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.3.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-6.3.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.2.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.2.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-6.1.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.1.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.0.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-6.0.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.34.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.33.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server



Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-5.33.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.32.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.32.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-5.31.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.31.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.30.2	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-5.30.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.30.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.29.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-5.28.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.28.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.27.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-5.27.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.26.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.25.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-5.24.1	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.24.0	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.23.1	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-5.23.0	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.22.0	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.21.2	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-5.21.1	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.21.0	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.20.1	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server



Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-5.20.0	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.19.1	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.19.0	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-5.18.1	3.4.12	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.18.0	3.4.12	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.17.2	3.4.12	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-5.17.1	3.4.12	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.17.0	3.4.12	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.16.1	3.4.12	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-5.16.0	3.4.12	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.15.1	3.4.12	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.15.0	3.4.12	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-5.14.2	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.14.1	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.14.0	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-5.13.1	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.13.0	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.12.3	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-5.12.2	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.12.1	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.12.0	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-5.11.4	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.11.3	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.11.2	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server



Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-5.11.1	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.11.0	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.10.1	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-5.10.0	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.9.1	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.9.0	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-5.8.3	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.8.2	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.8.1	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-5.8.0	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.7.1	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.7.0	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-5.6.1	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.6.0	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.5.4	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-5.5.3	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server
emr-5.5.2	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server
emr-5.5.1	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-5.5.0	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server
emr-5.4.1	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server
emr-5.4.0	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-5.3.2	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-5.3.1	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-5.3.0	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server



Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-5.2.3	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server
emr-5.2.2	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server
emr-5.2.1	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-5.2.0	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-5.1.1	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-5.1.0	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-5.0.3	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-5.0.2	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-5.0.1	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server

Amazon EMR 发布标签	ZooKeeper 版本	安装在一起的组件 ZooKeeper
emr-5.0.0	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server

## 连接器和实用工具

Amazon EMR 提供了多个连接器和实用工具，用于将其他 AWS 服务作为数据源进行访问。您通常可在一个程序内访问这些服务中的数据。例如，您可以在 Hive 查询、Pig 脚本或 MapReduce 应用程序中指定 Kinesis 流，然后对该数据进行操作。

### 主题

- [使用 Amazon 在 DynamoDB 中导出、导入、查询和联接表 EMR](#)
- [Kinesis](#)
- [S3 DistCp \(s3-dist-cp\)](#)
- [在 S3 DistCp 任务失败后进行清理](#)

## 使用 Amazon 在 DynamoDB 中导出、导入、查询和联接表 EMR

### Note

Amazon EMR-DynamoDB Connector 已在上开源。GitHub 有关更多信息，请参阅<https://github.com/aws-labs/emr-dynamodb-connector>。

DynamoDB 是一项完全托管的 SQL 无数据库服务，可提供快速且可预测的性能以及无缝的可扩展性。开发人员可以创建数据库表，并可以不受限制地增加请求流量或存储空间。DynamoDB 可自动将表的数据和流量分布到足够多的服务器中，以便处理客户指定的容量请求和数据存储量，同时还能保持性能一致、访问高效。使用 Amazon EMR 和 Hive，您可以快速高效地处理大量数据，例如存储在 DynamoDB 中的数据。有关 DynamoDB 的更多信息，请参阅 [Amazon DynamoDB 开发人员指南](#)。

Apache Hive 是一个软件层，您可以使用一种名为 HiveQL 的简化 SQL 类似查询语言来查询地图减少集群。它在 Hadoop 架构的顶层运行。有关 Hive 和 HiveQL 的更多信息，请转至 [HiveQL 语言手册](#)。有关 Hive 和 Amazon 的更多信息 EMR，请参阅 [Apache Hive](#)。

您可以将 Amazon EMR 与自定义版本的 Hive 配合使用，该版本包括与 DynamoDB 的连接，以便对存储在 DynamoDB 中的数据执行操作：

- 将 DynamoDB 数据加载到 Hadoop 分布式文件系统 HDFS () 中，并将其用作亚马逊集群的输入。EMR
- SQL 使用类似 ( HiveQL ) 的语句查询实时的 DynamoDB 数据。

- 连接 DynamoDB 中存储的数据并导出这些数据或查询连接的数据。
- 将存储在 DynamoDB 中的数据导出到 Amazon S3。
- 将存储在 Amazon S3 中的数据导入到 DynamoDB。

### Note

[Amazon EMR-DynamoDB 连接器不支持配置为使用 Kerberos 身份验证的集群。](#)

要执行以下每项任务，您需要启动一个 Amazon EMR 集群，在 DynamoDB 中指定数据的位置，然后发出 Hive 命令在 DynamoDB 中操作数据。

有多种启动亚马逊EMR集群的方法：您可以使用亚马逊EMR控制台、命令行界面 (CLI)，或者您可以使用或亚马逊对集群进行 AWS SDK编程EMRAPI。您还可以选择交互运行 Hive 集群还是从脚本运行。在本节中，我们将向您展示如何从 Amazon EMR 控制台启动交互式 Hive 集群，CLI以及。

通过交互的方式使用 Hive 是测试查询性能和调试应用程序的良好方式。在您建立了一组定期运行的 Hive 命令之后，可以考虑创建一个 Amazon EMR 可以为您运行的 Hive 脚本。

### Warning

Amazon 对 DynamoDB 表的EMR读取或写入操作计入您确定的预配置吞吐量，这可能会增加预配置吞吐量异常的频率。对于大型请求，Amazon 会使用指数退避EMR实现重试，以管理 DynamoDB 表上的请求负载。与其他流量同时运行 Amazon EMR 作业可能会导致您超出分配的预配置吞吐量级别。您可以通过查看 Amazon 中的ThrottleRequests指标来监控这一点 CloudWatch。如果请求负载过高，您可以重新启动集群并将[读取百分比设置](#)或设置为较低的值[写入百分比设置](#)以限制 Ama EMR zon 的操作。有关 DynamoDB 吞吐量设置的详细信息，请参阅[预置吞吐量](#)。

如果表配置为[按需模式](#)，则应先将表更改回预配置模式，再运行导出或导入操作。管道需要一个吞吐量比才能从 D 计算要使用的资源ynamoDBtable。按需模式删除预置的吞吐量。要预配置吞吐量，您可以使用 Amazon Ev CloudWatch ents 指标来评估表使用的总吞吐量。

## 主题

- [设置 Hive 表来运行 Hive 命令](#)
- [用于在 DynamoDB 中导出、导入和查询数据的 Hive 命令示例](#)
- [在 DynamoDB 中优化亚马逊EMR运营的性能](#)

## 设置 Hive 表来运行 Hive 命令

Apache Hive 是一款数据仓库应用程序，您可以使用SQL类似语言查询亚马逊EMR集群中包含的数据。有关 Hive 的更多信息，请参阅 <http://hive.apache.org/>。

以下过程假设您已经创建了集群并指定了 Amazon EC2 密钥对。要了解如何开始创建集群，请参阅《[亚马逊EMR管理指南](#)》EMR中的 [Amazon 入门](#)。

### 配置 Hive 以供使用 MapReduce

当你使用亚马逊EMR上的 Hive 查询 DynamoDB 表时，如果 Hive 使用默认的执行引擎 Tez，则可能会出现错误。因此，当您按本节所述使用 Hive 创建与 DynamoDB 集成的集群时，我们建议您使用将 Hive 设置为使用的配置分类。MapReduce有关更多信息，请参阅 [配置应用程序](#)。

以下代码段显示了用于设置 MapReduce 为 Hive 执行引擎的配置分类和属性：

```
[
    {
        "Classification": "hive-site",
        "Properties": {
            "hive.execution.engine": "mr"
        }
    }
]
```

### 以交互方式运行 Hive 命令

1. 连接到主节点。有关更多信息，请参阅 [《Amazon EMR 管理指南》SSH中的“使用连接到主节点”](#)。
2. 当命令提示输入当前主节点时，键入 hive。

您应看到 Hive 提示符：hive>

3. 输入用于将 Hive 应用程序中的表映射到 DynamoDB 中的数据的数据的 Hive 命令。该表充当对 Amazon DynamoDB 中存储的数据的引用；数据未存储在本地的 Hive 中，每次运行命令时，使用此表的任何查询将针对 DynamoDB 中的实时数据运行，从而占用此表的读取或写入容量。如果您需要对同一数据集运行多个 Hive 命令，请考虑先将其导出。

下面说明将 Hive 表映射到 DynamoDB 表的语法。

```
CREATE EXTERNAL TABLE hive_tablename
(hive_column1_name column1_datatype, hive_column2_name column2_datatype...)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodb_tablename",
"dynamodb.column.mapping" =
"hive_column1_name:dynamodb_attribute1_name, hive_column2_name:dynamodb_attribute2_name...")
```

当您在 Hive 中从 DynamoDB 创建表时，必须使用关键字 EXTERNAL 将该表创建为外部表。外部表与内部表之间的区别是：删除内部表时，将删除内部表中的数据。当连接 Amazon DynamoDB 时，这不是所需行为，因此仅支持外部表。

例如，以下 Hive 命令在 Hive 中创建名为 `hivetable1` 的表，该表引用名为 `dynamodhtable1` 的 DynamoDB 表。DynamoDB 表 `dynamodhtable1` 具有主键架构。hash-and-range 哈希键元素是 `name` (字符串类型)。范围键元素是 `year` (数字类型)。每个项目都有 `holidays` (字符串集类型) 的属性值。

```
CREATE EXTERNAL TABLE hivetable1 (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodhtable1",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");
```

第 1 行使用 HiveQL CREATE EXTERNAL TABLE 语句。对于 `hivetable1`，您需要为 DynamoDB 表中的每个属性名称/值对建立一列，并提供数据类型。这些值不区分大小写，并且您可以为列提供任何名称 (保留字除外)。

第 2 行使用 STORED BY 语句。STORED BY 的值是用于处理 Hive 与 DynamoDB 之间连接的类的名称。该值应设置为

```
'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'。
```

第 3 行使用 TBLPROPERTIES 语句将“`hivetable1`”与 DynamoDB 中相应的表和架构相关联。为 TBLPROPERTIES 提供 `dynamodb.table.name` 参数和 `dynamodb.column.mapping` 参数的值。这些值是区分大小写的。

#### Note

此表的所有 DynamoDB 属性名称必须在 Hive 表中有对应的列。根据您的 Amazon EMR 版本，如果 one-to-one 映射不存在，则会出现以下情况：



- 在 Amazon 5.27.0 及更高EMR版本中，连接器经过验证，可确保 DynamoDB 属性名称与 Hive 表中的列 one-to-one 之间存在映射。如果 one-to-one 映射不存在，则会发生错误。
- 在亚马逊EMR版本 5.26.0 及更早版本中，Hive 表将不包含来自 DynamoDB 的名称/值对。如果您未映射 DynamoDB 主键属性，则 Hive 将生成错误。如果您未映射非主键属性，则不会生成错误，但您将无法查看 Hive 表中的数据。如果数据类型不匹配，则值为空。

然后，您可以开始对 hivetable1 运行 Hive 操作。根据 hivetable1 运行的查询也根据 DynamoDB 账户的 DynamoDB 表 dynamodhtable1 在内部运行，在每次执行运行时消耗读取或写入单位。

对 DynamoDB 表运行 Hive 查询时，您需要确保已预置足量的读取容量单位。

例如，假设您为 DynamoDB 表预配置了 100 个读取容量单位。这将允许您每秒执行 100 次读取或读取 409600 字节。如果该表包含 20GB 的数据 ( 21474836480 字节 ) 并且您的 Hive 查询执行全表扫描，则可以估算执行查询将花费多长时间：

$$21474836480/409600 = 52429 \text{ 秒} = 14.56 \text{ 小时}$$

减少所需时间的唯一方法是调整源 DynamoDB 表的读取容量单位。添加更多 Amazon EMR 节点无济于事。

在 Hive 输出中，当一个或多个映射器进程已完成时，将更新完成百分比。对于预配置的读取容量设置较低的大型 DynamoDB 表，完成百分比输出可能会很长时间不更新；在上面的示例中，作业将在几个小时内显示为完成 0%。有关任务进度的更多详细状态，请访问 Amazon EMR 控制台；您将能够查看单个映射器任务状态和数据读取统计信息。您还可以登录主节点的 Hadoop 界面，查看 Hadoop 统计数据。该界面将向您显示单个映射任务状态和一些数据读取统计数据。有关更多信息，请参阅以下主题：

- [托管在主节点 \(master node\) 上的 Web 界面](#)
- [查看 Hadoop Web 界面](#)

有关用于执行从 DynamoDB 导出或导入数据和联接表等任务的示例 HiveQL 语句的更多信息，请参阅[用于在 DynamoDB 中导出、导入和查询数据的 Hive 命令示例](#)。

## 取消 Hive 请求

执行 Hive 查询时，来自服务器的初始响应包含用于取消请求的命令。要在此过程中随时取消请求，请使用服务器响应中的 Kill 命令。

1. 输入 Ctrl+C 可退出命令行客户端。
2. 在 Shell 提示符下，输入服务器对您的请求的初始响应中的 Kill 命令。

或者，您可以从主节点的命令行运行以下命令来终止 Hadoop 作业，其中 *job-id* 是 Hadoop 作业的标识符，可以从 Hadoop 用户界面中检索。

```
hadoop job -kill job-id
```

## Hive 和 DynamoDB 的数据类型

下表显示了可用的 Hive 数据类型、它们对应的默认 DynamoDB 类型以及它们也可以映射到的备用 DynamoDB 类型。

Hive 类型	默认 DynamoDB 类型	备用 DynamoDB 类型
字符串	字符串	
bigint 或 double	数字 (N)	
binary	二进制 (B)	
布尔值	布尔值 (BOOL)	
数组	list (L)	数字集 (NS)、字符串集 (SS) 或二进制集 (BS)
map<string, string>	项目	map (M)
map<string, ?>	map (M)	

Hive 类型	默认 DynamoDB 类型	备用 DynamoDB 类型
	空 (NULL)	

如果要将 Hive 数据作为对应的备用 DynamoDB 类型写入，或者您的 DynamoDB 数据包含备用 DynamoDB 类型的属性值，则可以使用 `dynamodb.type.mapping` 参数指定列和 DynamoDB 类型。以下示例显示了用于指定备用类型映射的语法。

```
CREATE EXTERNAL TABLE hive_tablename (hive_column1_name column1_datatype,
  hive_column2_name column2_datatype...)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodb_tablename",
"dynamodb.column.mapping" =
  "hive_column1_name:dynamodb_attribute1_name,hive_column2_name:dynamodb_attribute2_name...",
"dynamodb.type.mapping" = "hive_column1_name:dynamodb_attribute1_datatype");
```

类型映射参数是可选的，仅必须为使用备用类型的列指定它。

例如，以下 Hive 命令在 Hive 中创建名为 `hivetable2` 的表，该表引用 DynamoDB 表 `dynamodatabase2`。它与 `hivetable1` 相似，不同之处在于它将 `col3` 列映射到字符串集 (SS) 类型。

```
CREATE EXTERNAL TABLE hivetable2 (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodatabase2",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays",
"dynamodb.type.mapping" = "col3:SS");
```

在 Hive 中，`hivetable1` 和 `hivetable2` 是相同的。但是，将这些表中的数据写入其对应的 DynamoDB 表时，`dynamodatabase1` 将包含列表，而 `dynamodatabase2` 将包含字符串集。

如果要将 Hive `null` 值作为 DynamoDB `null` 类型的属性写入，则您可以使用 `dynamodb.null.serialization` 参数来写入。以下示例显示了用于指定 `null` 序列化的语法。

```
CREATE EXTERNAL TABLE hive_tablename (hive_column1_name column1_datatype,
  hive_column2_name column2_datatype...)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodb_tablename",
"dynamodb.column.mapping" =
  "hive_column1_name:dynamodb_attribute1_name,hive_column2_name:dynamodb_attribute2_name...",
"dynamodb.null.serialization" = "true");
```

```
"dynamodb.null.serialization" = "true");
```

空序列化参数是可选的，如果未指定，则设置为 `false`。请注意，无论参数设置如何，DynamoDB `null` 属性都将作为 Hive 中的 `null` 值进行读取。仅当将空序列化参数指定为 `true` 时，才能将具有 `null` 值的 Hive 集合写入 DynamoDB。否则，将出现 Hive 错误。

就精度而言，Hive 中的 `bigint` 类型与 Java `long` 类型相同，而 Hive `double` 类型与 Java `double` 类型相同。这意味着，如果您有精度高于 Hive 数据类型所提供精度的数值数据存储在 DynamoDB 中，则使用 Hive 导出、导入或引用 DynamoDB 数据会导致精度损失或 Hive 查询失败。

将二进制类型从 DynamoDB 导出到亚马逊简单存储服务 (Amazon HDFS S3) Service，或者存储为 Base64 编码的字符串。如果您要从 Amazon S3 或 DynamoDB 二进制类型导入 HDFS 数据，则应将其编码为 Base64 字符串。

## Hive 选项

您可以设置以下 Hive 选项来管理从 Amazon DynamoDB 的数据传出。这些选项只针对当前 Hive 会话保留。如果您在集群上关闭 Hive 命令提示符并稍后重新打开，则这些设置将恢复为默认值。

Hive 选项	描述
<code>dynamodb.throughput.read.percent</code>	<p>设置读取操作的速率，在为您的表分配的范围内保持 DynamoDB 预配置的吞吐速率。该值介于 0.1 到 1.5 之间（包含端点）。</p> <p>值 0.5 是默认读取速率，这意味着，Hive 将在表的整个资源中尝试占用一半的预配读取量。增加此值使之高于 0.5 将提高读取请求速率。减少此值使之低于 0.5 将降低读取请求速率。此读取速率是近似值。实际读取速率取决于 DynamoDB 中是否存在统一分配的键等因素。</p> <p>如果您发现 Hive 操作经常超出您预配的吞吐量，或者如果过多限制了实时读取流量，则可以减少此值使之低于 0.5。如果您有足够的容量并希望 Hive 操作的速度更快，请将此值设置为高于 0.5。如果您认为有可用的输入/输出操作未使用，则还可以通过将此值设置到最高 1.5 来进行超额预订。</p>

Hive 选项	描述
<code>dynamodb.throughput.write.percent</code>	<p>设置写入操作的速率，在为您的表分配的范围保持 DynamoDB 预配置的吞吐速率。该值介于 0.1 到 1.5 之间（包含端点）。</p> <p>值 0.5 是默认写入速率，这意味着，Hive 将在表的整个资源中尝试占用一半的预配写入量。增加此值使之高于 0.5 将提高写入请求速率。减少此值使之低于 0.5 将降低写入请求速率。此写入速率是近似值。实际写入速率取决于 DynamoDB 中是否存在统一分配的键等因素</p> <p>如果您发现 Hive 操作经常超出您预配的吞吐量，或者如果过多限制了实时写入流量，则可以减少此值使之低于 0.5。如果您有足够的容量并希望 Hive 操作的速度更快，请将此值设置为高于 0.5。如果您认为有可用的输入/输出操作未使用或者这是到表的初始数据上载，还没有实时流量，则还可以通过将此值设置到最高 1.5 来进行超额预订。</p>
<code>dynamodb.endpoint</code>	<p>为 DynamoDB 服务指定终端节点。有关可用 DynamoDB 终端节点的更多信息，请参阅 <a href="#">区域和终端节点</a>。</p>
<code>dynamodb.max.map.tasks</code>	<p>指定在从 DynamoDB 读取数据时，映射任务的最大数量。此值必须等于或大于 1。</p>
<code>dynamodb.retry.duration</code>	<p>指定要用作重试 Hive 命令的超时时间的分钟数。此值必须是大于或等于 0 的整数。默认超时持续时间为 2 分钟。</p>

这些选项是使用 SET 命令设置的，如以下示例所示。

```
SET dynamodb.throughput.read.percent=1.0;

INSERT OVERWRITE TABLE s3_export SELECT *
FROM hiveTableName;
```

## 用于在 DynamoDB 中导出、导入和查询数据的 Hive 命令示例

以下示例使用 Hive 命令执行诸如将数据导出到 Amazon S3 或HDFS将数据导入到 DynamoDB、联接表、查询表等操作。

对 Hive 表执行的操作将引用 DynamoDB 中存储的数据。Hive 命令受到 DynamoDB 表预置的吞吐量设置约束，并且检索的数据包括 DynamoDB 处理 Hive 操作请求时写入到 DynamoDB 表的数据。如果数据检索过程需要很长一段时间，则自 Hive 命令开始执行以来，Hive 命令返回的某些数据可能已在 DynamoDB 中更新。

Hive 命令 DROP TABLE 和 CREATE TABLE 仅对 Hive 中的本地表进行操作，而不会在 DynamoDB 中创建或删除表。如果 Hive 查询引用 DynamoDB 中的表，则在您运行查询之前，该表必须已存在。有关在 DynamoDB 中创建和删除表的更多信息，请参阅 Amazon DynamoDB 开发人员指南中的[在 DynamoDB 中处理表](#)。

### Note

当您将 Hive 表映射到 Amazon S3 中的位置时，请勿将其映射到存储桶的根路径 `s3://mybucket`，否则这会在 Hive 将数据写入到 Amazon S3 时导致错误。而是应将表映射到存储桶的子路径 `s3://mybucket/mypath`。

## 从 DynamoDB 中导出数据

可以使用 Hive 从 DynamoDB 中导出数据。

将 DynamoDB 表导出到 Amazon S3 存储桶

- 创建一个引用 DynamoDB 中存储的数据的 Hive 表。然后，您可以调用INSERTOVERWRITE命令将数据写入外部目录。在以下示例中，`s3://bucketname/path/subpath/` 是 Amazon S3 中的有效路径。调整CREATE命令中的列和数据类型，使其与 DynamoDB 中的值相匹配。可以使用此命令在 Amazon S3 中创建 DynamoDB 数据的存档。

```
CREATE EXTERNAL TABLE hiveTableName (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodbtable1",
```

```
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");
```

```
INSERT OVERWRITE DIRECTORY 's3://bucketname/path/subpath/' SELECT *
FROM hiveTableName;
```

## 使用格式设置将 DynamoDB 表导出到 Amazon S3 存储桶

- 创建引用 Amazon S3 中的位置的外部表。此表在下面显示为 `s3_export`。在 CREATE 调用期间，为表指定行格式。然后，当您使用 INSERT OVERWRITE 将数据从 DynamoDB 导出到 `s3_export` 时，数据将以指定的格式写出。在以下示例中，数据以逗号分隔值 () CSV 的形式写出。

```
CREATE EXTERNAL TABLE hiveTableName (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodhtable1",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");
```

```
CREATE EXTERNAL TABLE s3_export(a_col string, b_col bigint, c_col array<string>)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LOCATION 's3://bucketname/path/subpath/';
```

```
INSERT OVERWRITE TABLE s3_export SELECT *
FROM hiveTableName;
```

## 在不指定列映射的情况下将 DynamoDB 表导出到 Amazon S3 存储桶

- 创建一个引用 DynamoDB 中存储的数据的 Hive 表。此例与前面的示例类似，只是不指定列映射。该表必须正好具有类型为 `map<string, string>` 的一个列。如果您随后在 Amazon S3 中创建 EXTERNAL 表，可以调用 INSERT OVERWRITE 命令将数据从 DynamoDB 写入到 Amazon S3。可以使用此命令在 Amazon S3 中创建 DynamoDB 数据的存档。由于没有列映射，因此您无法查询以此方式导出的表。Hive 0.8.1.5 或更高版本支持在不指定列映射的情况下导出数据，Amazon 2.2 支持该版本。EMR AMI x 及更高版本。

```
CREATE EXTERNAL TABLE hiveTableName (item map<string,string>)
```

```

STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodbtable1");

CREATE EXTERNAL TABLE s3TableName (item map<string, string>)
ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' LINES TERMINATED BY '\n'
LOCATION 's3://bucketname/path/subpath/';

INSERT OVERWRITE TABLE s3TableName SELECT *
FROM hiveTableName;

```

## 使用数据压缩将 DynamoDB 表导出到 Amazon S3 存储桶

- Hive 提供多个可以在 Hive 会话期间设置的压缩编解码器。这样做会导致导出的数据以指定的格式进行压缩。以下示例使用 Lempel-Ziv-Oberhumer () 算法压缩导出的文件。LZO

```

SET hive.exec.compress.output=true;
SET io.seqfile.compression.type=BLOCK;
SET mapred.output.compression.codec = com.hadoop.compression.lzo.LzopCodec;

CREATE EXTERNAL TABLE hiveTableName (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodbtable1",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");

CREATE EXTERNAL TABLE lzo_compression_table (line STRING)
ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' LINES TERMINATED BY '\n'
LOCATION 's3://bucketname/path/subpath/';

INSERT OVERWRITE TABLE lzo_compression_table SELECT *
FROM hiveTableName;

```

可用的压缩编解码器包括：

- org.apache.hadoop.io.compress。GzipCodec
- org.apache.hadoop.io.compress。DefaultCodec
- com.hadoop.compression.lzo。LzoCodec



- `com.hadoop.compression.lzo`。 `LzopCodec`
- `org.apache.hadoop.io.compress`。 `BZip2Codec`
- `org.apache.hadoop.io.compress`。 `SnappyCodec`

## 将 DynamoDB 表导出到 HDFS

- 使用以下 Hive 命令，其中 `hdfs:///directoryName` 是有效的HDFS路径并且 `hiveTableName` 是 Hive 中引用 DynamoDB 的表。此导出操作比将 DynamoDB 表导出到 Amazon S3 要快，因为 Hive 0.7.1.1 在将数据导出到亚马逊 S3 时HDFS用作中间步骤。以下示例还显示了如何将 `dynamodb.throughput.read.percent` 设置为 1.0 以提高读取请求速率。

```
CREATE EXTERNAL TABLE hiveTableName (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodhtable1",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");

SET dynamodb.throughput.read.percent=1.0;

INSERT OVERWRITE DIRECTORY 'hdfs:///directoryName' SELECT * FROM hiveTableName;
```

您也可以HDFS使用格式化和压缩将数据导出到，如上所示，以便导出到 Amazon S3。为此，只需将上面示例中的 Amazon S3 目录替换为一个HDFS目录即可。

## 在 Hive 中读取不可打印的 UTF -8 个字符数据

- 通过在创建表时使用 `STORED AS SEQUENCEFILE` 子句，可以使用 Hive 读取和写入不可打印的 UTF -8 个字符的数据。A `SequenceFile` 是 Hadoop 二进制文件格式；你需要使用 Hadoop 来读取这个文件。以下示例显示了如何将数据从 DynamoDB 导出到 Amazon S3 中。您可以使用此功能来处理不可打印的 UTF -8 编码字符。

```
CREATE EXTERNAL TABLE hiveTableName (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodhtable1",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");
```

```
CREATE EXTERNAL TABLE s3_export(a_col string, b_col bigint, c_col array<string>)
STORED AS SEQUENCEFILE
LOCATION 's3://bucketname/path/subpath/';

INSERT OVERWRITE TABLE s3_export SELECT *
FROM hiveTableName;
```

## 将数据导入到 DynamoDB

使用 Hive 将数据写入到 DynamoDB 中时，应确保写入容量单位数大于集群中的映射器数。例如，在 m1.xlarge 实例上运行的集群每个 EC2 实例生成 8 个映射器。对于具有 10 个实例的集群，这意味着生成 80 个映射器。如果写入容量单位数不大于集群中的映射器数，则 Hive 写入操作可能会占用所有写入吞吐量，或者尝试占用超过预配置值的吞吐量。有关每种 EC2 实例类型生成的映射器数量的更多信息，请参阅[配置 Hadoop](#)。

Hadoop 中的映射器数由输入的拆分数控制。如果拆分数过小，写入命令可能无法占用所有可用的写入吞吐量。

如果具有相同键的项目在目标 DynamoDB 表中存在，则将覆盖该项目。如果目标 DynamoDB 表中不存在具有该键的项目，则将插入该项目。

### 要将数据从 Amazon S3 导入 DynamoDB

- 您可以使用亚马逊 EMR ( 亚马逊 EMR ) 和 Hive 将数据从亚马逊 S3 写入 DynamoDB。

```
CREATE EXTERNAL TABLE s3_import(a_col string, b_col bigint, c_col array<string>)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LOCATION 's3://bucketname/path/subpath/';

CREATE EXTERNAL TABLE hiveTableName (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodhtable1",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");

INSERT OVERWRITE TABLE hiveTableName SELECT * FROM s3_import;
```

## 在不指定列映射时将表从 Amazon S3 存储桶导入到 DynamoDB 中

- 创建一个引用 Amazon S3 中存储数据的 EXTERNAL 表，该数据是以前从 DynamoDB 中导出的。在导入之前，请确保该表存在于 DynamoDB 中，并且该表与以前导出的 DynamoDB 表具有相同的键架构。此外，该表还必须正好具有类型为 `map<string, string>` 的一个列。如果您随后创建一个链接到 DynamoDB 的 Hive 表，则可以调用 `INSERT OVERWRITE` 命令将数据从 Amazon S3 写入到 DynamoDB 中。由于没有列映射，因此您无法查询以此方式导入的表。Hive 0.8.1.5 或更高版本支持在不指定列映射的情况下导入数据，Amazon EMR AMI 2.2.3 及更高版本支持该功能。

```
CREATE EXTERNAL TABLE s3TableName (item map<string, string>)
ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' LINES TERMINATED BY '\n'
LOCATION 's3://bucketname/path/subpath/';

CREATE EXTERNAL TABLE hiveTableName (item map<string, string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodhtable1");

INSERT OVERWRITE TABLE hiveTableName SELECT *
FROM s3TableName;
```

## 将表从导入 Dynam HDFS oDB

- 您可以使用 Amazon EMR 和 Hive 将数据从写入 DynamoD HDFS B。

```
CREATE EXTERNAL TABLE hdfs_import(a_col string, b_col bigint, c_col array<string>)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LOCATION 'hdfs:///directoryName';

CREATE EXTERNAL TABLE hiveTableName (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodhtable1",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");

INSERT OVERWRITE TABLE hiveTableName SELECT * FROM hdfs_import;
```

## 查询 DynamoDB 中的数据

以下示例显示了您可以使用 Amazon EMR 查询存储在 DynamoDB 中的数据的各种方式。

### 查找映射列的最大值 (max)

- 使用如下 Hive 命令。在第一个命令中，该CREATE语句创建一个 Hive 表，该表引用存储在 DynamoDB 中的数据。然后，该SELECT语句使用该表来查询存储在 DynamoDB 中的数据。以下示例查找给定客户提交的最大订单。

```
CREATE EXTERNAL TABLE hive_purchases(customerId bigint, total_cost double,  
items_purchased array<String>)  
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'  
TBLPROPERTIES ("dynamodb.table.name" = "Purchases",  
"dynamodb.column.mapping" =  
"customerId:CustomerId,total_cost:Cost,items_purchased:Items");  
  
SELECT max(total_cost) from hive_purchases where customerId = 717;
```

### 使用 GROUP BY 子句聚合数据

- 可以使用 GROUP BY 子句收集多条记录的数据。此子句通常与聚合函数 (如 sum、count、min 或 max) 一起使用。以下示例返回提交了三个以上订单的客户的最小订单列表。

```
CREATE EXTERNAL TABLE hive_purchases(customerId bigint, total_cost double,  
items_purchased array<String>)  
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'  
TBLPROPERTIES ("dynamodb.table.name" = "Purchases",  
"dynamodb.column.mapping" =  
"customerId:CustomerId,total_cost:Cost,items_purchased:Items");  
  
SELECT customerId, max(total_cost) from hive_purchases GROUP BY customerId HAVING  
count(*) > 3;
```

## 连接两个 DynamoDB 表

- 以下示例将两个 Hive 表映射到 DynamoDB 中存储的数据。然后，它对这两个表调用联接。连接在集群上计算并返回。连接不在 DynamoDB 中进行。此示例返回提交了两个以上订单的客户及其购买物的列表。

```
CREATE EXTERNAL TABLE hive_purchases(customerId bigint, total_cost double,
items_purchased array<String>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "Purchases",
"dynamodb.column.mapping" =
"customerId:CustomerId,total_cost:Cost,items_purchased:Items");

CREATE EXTERNAL TABLE hive_customers(customerId bigint, customerName string,
customerAddress array<String>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "Customers",
"dynamodb.column.mapping" =
"customerId:CustomerId,customerName:Name,customerAddress:Address");

Select c.customerId, c.customerName, count(*) as count from hive_customers c
JOIN hive_purchases p ON c.customerId=p.customerId
GROUP BY c.customerId, c.customerName HAVING count > 2;
```

## 联接来自不同源的两个表

- 在以下示例中，Customer\_S3 是一个 Hive 表，用于加载存储在 Amazon S3 中的 CSV 文件，而 *hive\_purchases* 是一个引用 DynamoDB 中数据的表。以下示例将以 CSV 文件形式存储在 Amazon S3 中的客户数据与存储在 DynamoDB 中的订单数据合并在一起，返回一组数据，这些数据表示名字中有“Miller”的客户下达的订单。

```
CREATE EXTERNAL TABLE hive_purchases(customerId bigint, total_cost double,
items_purchased array<String>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "Purchases",
"dynamodb.column.mapping" =
"customerId:CustomerId,total_cost:Cost,items_purchased:Items");
```

```
CREATE EXTERNAL TABLE Customer_S3(customerId bigint, customerName string,  
  customerAddress array<String>)  
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','  
LOCATION 's3://bucketname/path/subpath/';  
  
Select c.customerId, c.customerName, c.customerAddress from  
Customer_S3 c  
JOIN hive_purchases p  
ON c.customerid=p.customerid  
where c.customerName like '%Miller%';
```

### Note

在前面的示例中，为了清晰和完整，在每个示例中都包含了CREATETABLE语句。针对给定 Hive 表运行多个查询或执行导出操作时，只需在 Hive 会话的开始创建表一次即可。

## 在 DynamoDB 中优化亚马逊EMR运营的性能

Amazon 对 DynamoDB 表的EMR操作算作读取操作，并受该表的预配置吞吐量设置的约束。Amazon EMR 实施自己的逻辑来尝试平衡您的 DynamoDB 表上的负载，以最大限度地减少超出预配置吞吐量的可能性。在每个 Hive 查询结束时，Amazon 都会EMR返回有关用于处理查询的集群的信息，包括超过预配置吞吐量的次数。您可以使用这些信息以及有关您的 DynamoDB 吞吐量的 CloudWatch 指标，在后续请求中更好地管理 DynamoDB 表上的负载。

以下因素会影响 Hive 在处理 DynamoDB 表时的查询性能。

### 预置的读取容量单位

对 DynamoDB 表运行 Hive 查询时，您需要确保已预置足量的读取容量单位。

例如，假设您为 DynamoDB 表预配置了 100 个读取容量单位。这将允许您每秒执行 100 次读取或读取 409600 字节。如果该表包含 20GB 的数据 ( 21474836480 字节 ) 并且您的 Hive 查询执行全表扫描，则可以估算执行查询将花费多长时间：

$$21474836480/409600 = 52429 \text{ 秒} = 14.56 \text{ 小时}$$

减少所需时间的唯一方法是调整源 DynamoDB 表的读取容量单位。向 Amazon EMR 集群添加更多节点无济于事。

在 Hive 输出中，当一个或多个映射器进程已完成时，将更新完成百分比。对于预配置的读取容量设置较低的大型 DynamoDB 表，完成百分比输出可能会很长时间不更新；在上面的示例中，作业将在几个小时内显示为完成 0%。有关任务进度的更多详细状态，请访问 Amazon EMR 控制台；您将能够查看单个映射器任务状态和数据读取统计信息。

您还可以登录主节点的 Hadoop 界面，查看 Hadoop 统计数据。该界面将向您显示单个映射任务状态和一些数据读取统计数据。有关更多信息，请参阅 Amazon EMR 管理指南中的[主节点上托管的 Web 界面](#)。

## 读取百分比设置

默认情况下，Amazon 会根据您当前的预配置吞吐量EMR管理您的 DynamoDB 表的请求负载。但是，当 Amazon EMR 返回的任务信息中包含大量预配置吞吐量超过响应时，您可以在设置 Hive 表时使用 `dynamodb.throughput.read.percent` 参数调整默认读取速率。有关设置读取百分比参数的更多信息，请参阅 [Hive 选项](#)。

## 写入百分比设置

默认情况下，Amazon 会根据您当前的预配置吞吐量EMR管理您的 DynamoDB 表的请求负载。但是，当 Amazon EMR 返回的任务信息中包含大量预配置吞吐量超过响应时，您可以在设置 Hive 表时使用 `dynamodb.throughput.write.percent` 参数调整默认写入速率。有关设置写入百分比参数的更多信息，请参阅[Hive 选项](#)。

## 重试持续时间设置

默认情况下，如果在两分钟（默认EMR重试间隔）内未返回结果，Amazon 会重新运行 Hive 查询。在运行 Hive 查询时，您可以通过设置 `dynamodb.retry.duration` 参数来调整此时间间隔。有关设置写入百分比参数的更多信息，请参阅[Hive 选项](#)。

## 映射任务数

Hadoop 为了处理导出和查询 DynamoDB 中所存储数据的请求而启动的映射器守护进程的读取速率控制在每秒最多 1 MiB 之内，以限制所用的读取容量。如果在 DynamoDB 上有更多预置的吞吐量可用，则可以通过增加映射器守护进程数来改善 Hive 导出和查询操作的性能。为此，您可以增加集群中的 EC2实例数量，也可以增加在每个EC2实例上运行的映射器守护程序的数量。

您可以通过停止当前集群并使用更多EC2实例重新启动集群来增加集群中的EC2实例数量。如果您是从 Amazon EMR 控制台启动集群，则可以在“配置EC2实例”对话框中指定实例的数量；如果要从启动集群，则 `--num-instances` 可以选择此选项CLI。EC2

在实例上运行的地图任务数量取决于EC2实例类型。有关支持的EC2实例类型以及每种实例类型提供的映射器数量的更多信息，请参阅[任务配置](#)。其中，每个受支持的配置都有一个“任务配置”部分。

增加映射器守护程序数的另一个方法是，将 Hadoop 的 `mapreduce.tasktracker.map.tasks.maximum` 配置参数更改为更大的值。这样做的好处是，在不增加EC2实例数量或大小的情况下，为您提供更多的映射器，从而为您节省资金。缺点是，将此值设置得过高可能会导致集群中的EC2实例内存不足。要设置 `mapreduce.tasktracker.map.tasks.maximum`，请启动集群并为 `mapreduce.tasktracker.map.tasks.maximum` 指定一个值，作为 `mapred-site` 配置分类的属性。如以下示例所示。有关更多信息，请参阅[配置应用程序](#)。

```
{
  "configurations": [
    {
      "classification": "mapred-site",
      "properties": {
        "mapred.tasktracker.map.tasks.maximum": "10"
      }
    }
  ]
}
```

## 并行数据请求

从多个用户或多个应用程序向单个表发出的多个数据请求可能会耗尽预配置的读取吞吐量并降低性能。

## 处理持续时间

DynamoDB 中的数据一致性取决于在每个节点上执行读取和写入操作的顺序。当正在进行 Hive 查询时，其它应用程序可能会将新数据加载到 DynamoDB 表，或者修改或删除现有数据。在这种情况下，Hive 查询的结果可能无法反映查询运行时对数据所做的更改。

## 避免超出吞吐量

针对 DynamoDB 运行 Hive 查询时，请注意不要超出您的预置吞吐量，因为这会用尽应用程序调用 `DynamoDB::Get` 时所需的容量。为确保不会发生这种情况，您应 `DynamoDB::Get` 通过查看 Amazon 中的日志和监控指标，定期监控读取量和应用程序调用的限制。CloudWatch



## 请求时间

调度 Hive 查询以便在对 DynamoDB 表的需求较低时访问 DynamoDB 表，可以改善性能。举例来说，如果应用程序的大多数用户住在旧金山，您可以选择在太平洋标准时间凌晨 4 点导出每日数据。PST，此时大多数用户都处于睡眠状态，并且没有更新您的 DynamoDB 数据库中的记录。

## 基于时间的表

如果将数据组织为一系列基于时间的 DynamoDB 表（例如，每天一个表），您可以在该表不再处于活动状态时导出数据。您可以利用此方法将数据持续备份到 Amazon S3 中。

## 已存档数据

如果您计划对存储在 DynamoDB 中的数据运行多个 Hive 查询，并且您的应用程序可以容忍存档数据，则可能需要将数据导出到或 HDFS Amazon S3，然后对数据副本而不是 DynamoDB 运行 Hive 查询。这将节省读取操作和预配置的吞吐量。

## Kinesis

亚马逊EMR集群可以使用 Hadoop 生态系统中熟悉的工具（例如 Hive、Pig、Hadoop Streaming 和 Cascading）直接读取和处理 Amazon Kinesis 流。MapReduce API您还可以将来自 Amazon Kinesis 的实时数据与亚马逊 S3、亚马逊 DynamoDB 和正在运行的集群中的现有数据合并。HDFS您可以直接将数据从亚马逊加载到 Amazon S3 或 DynamoDB EMR 以进行后期处理活动。有关 Amazon Kinesis 服务亮点和定价的信息，请参阅 [Amazon Kinesis](#) 页面。

## 我能用亚马逊EMR和亚马逊 Kinesis 的集成做什么？

亚马逊EMR和亚马逊 Kinesis 之间的集成使某些场景变得更加容易；例如：

- 流式处理日志分析 – 您可以分析流式处理 Web 日志，以便每隔几分钟按区域、浏览器和访问域生成前 10 个错误类型的列表。
- 客户参与 – 您可以编写查询将 Amazon Kinesis 中的点击流数据与存储在 DynamoDB 表中的广告活动信息进行连接，以确定显示在特定网站上的最有效广告类别。
- 临时交互式查询 — 您可以定期将数据从 Amazon Kinesis 流加载到本地 Impala 表中，HDFS并以本地 Impala 表的形式提供，以进行快速、交互式的分析查询。

## 对 Amazon Kinesis 流进行检查点分析

用户可以定期对 Amazon Kinesis 流进行批量分析，这些分析称为迭代。由于 Amazon Kinesis 流数据记录是使用序列号检索的，因此迭代边界由亚马逊EMR存储在 DynamoDB 表中的起始和结束序列号来定义。例如，当 `iteration0` 结束时，它在 DynamoDB 中存储结束序列号，这样在 `iteration1` 作业开始时，它可以检索流的后续数据。迭代在流数据中的这种映射称为检查点操作。有关更多信息，请参阅 [Kinesis 连接器](#)。

如果对迭代进行了检查但任务未能处理迭代，则 Amazon EMR 会尝试重新处理该迭代中的记录。

通过检查点功能，您可以：

- 从运行于相同的流和逻辑名称之上的前一个查询处理的序列号之后，开始数据处理
- 重新处理 Kinesis 中由之前的查询处理的同一批数据

要启用检查点操作，请在脚本中将 `kinesis.checkpoint.enabled` 参数设置为 `true`。此外，请配置以下参数：

配置设置	描述
<code>kinesis.checkpoint.metastore.table.name</code>	用于存储检查点信息的 DynamoDB 表名称
<code>kinesis.checkpoint.metastore.hash.key.name</code>	DynamoDB 表的哈希键名称
<code>kinesis.checkpoint.metastore.hash.range.name</code>	DynamoDB 表的范围键名称
<code>kinesis.checkpoint.logical.name</code>	当前处理的逻辑名称
<code>kinesis.checkpoint.iteration.no</code>	与逻辑名称关联的处理的迭代编号
<code>kinesis.rerun.iteration.without.wait</code>	用来指示是否可以重新运行失败的迭代而不等待超时的布尔值；默认值为 <code>false</code>

### 针对 Amazon D IOPS ynamoDB 表的预配置建议

适用于 Amazon Kinesis 的亚马逊EMR连接器使用 DynamoDB 数据库作为检查点元数据的支持。您必须在 DynamoDB 中创建表，然后才能按检查点间隔使用亚马逊集群的 Amazon Kinesis 流中的数

据。EMR该表必须与您的 Amazon EMR 集群位于同一区域。以下是IOPS您应为 DynamoDB 表预配置的数量的一般建议；j假设可以同时运行的 Hadoop 任务的最大数量（具有不同的逻辑名称+迭代号组合），s并且是任何任务将处理的最大分片数：

对于 Read Capacity Units： $j*s/5$

对于 Write Capacity Units： $j*s$

## 性能注意事项

Amazon Kinesis 分片吞吐量与亚马逊EMR集群中节点的实例大小和流中记录的大小成正比。建议在主节点和核心节点上使用 m5.xlarge 或更大的实例。

## 与亚马逊一起安排亚马逊 Kinesis 分析 EMR

如果要对活动 Amazon Kinesis 流分析数据，由于任何迭代都受超时和最长持续时间限制，您应经常运行分析，以便从流定期收集详细信息，这十分重要。可以通过多种方式定期执行该类脚本和查询；但建议针对此类周期性任务使用 AWS Data Pipeline。有关更多信息，请参阅AWS Data Pipeline 开发人员指南[AWS Data Pipeline HiveActivity](#)中的[AWS Data Pipeline PigActivity](#)和。

## 将适用于亚马逊 7.0 的 Spark Kinesis 连接器迁移到 SDK 2.x EMR

AWS SDK提供了一组丰富的库来与 AWS 云计算服务进行交互，例如管理凭据、连接到 S3 和 Kinesis 服务。APIsSpark Kinesis 连接器用于使用来自 Kinesis Data Streams 的数据，且接收到的数据将在 Spark 的执行引擎中进行转换和处理。目前，此连接器建立在 1.x of AWS SDK 和 Kinesis-client-library (KCL) 之上。

作为 AWS SDK 2.x 迁移的一部分，Spark Kinesis 连接器也相应进行了更新，使其可以在 2.x 上 SDK 运行。在亚马逊 EMR 7.0 版本中，Spark 包含 SDK 2.x 升级，但在 Apache Spark 的社区版本中尚不可用。如果您使用低于 7.0 版本的 Spark Kinesis 连接器，则必须先将应用程序代码迁移到在 SDK 2.x 上运行，然后才能迁移到亚马逊 7.0。EMR

## 迁移指南

本部分介绍将应用程序迁移到升级后的 Spark Kinesis 连接器的步骤。它包括迁移到 Kinesis 客户端库 (KCL) 2.x、AWS 证书提供者和在 2.x 中迁移到 AWS 服务客户端的指南。AWS SDK作为参考，它还包括一个使用 Kinesis 连接器的示例[WordCount](#)程序。

### 主题

- [KCL从 1.x 迁移到 2.x](#)

- [将 AWS 凭证提供程序从 AWS SDK 1.x 迁移到 2.x](#)
- [将 AWS 服务客户端从 AWS SDK 1.x 迁移到 2.x](#)
- [流式传输应用程序的代码示例](#)
- [使用升级后的 Spark Kinesis 连接器时的注意事项](#)

## KCL从 1.x 迁移到 2.x

- **KinesisInputDStream** 中的指标级别和维度

当您实例化 `KinesisInputDStream` 时，您可以控制流的指标级别和维度。以下示例演示了如何使用 KCL 1.x 自定义这些参数：

```
import
  com.amazonaws.services.kinesis.clientlibrary.lib.worker.KinesisClientLibConfiguration
import com.amazonaws.services.kinesis.metrics.interfaces.MetricsLevel

val kinesisStream = KinesisInputDStream.builder
  .streamingContext(ssc)
  .streamName(streamName)
  .endpointUrl(endpointUrl)
  .regionName(regionName)
  .initialPosition(new Latest())
  .checkpointAppName(appName)
  .checkpointInterval(kinesisCheckpointInterval)
  .storageLevel(StorageLevel.MEMORY_AND_DISK_2)
  .metricsLevel(MetricsLevel.DETAILED)

  .metricsEnabledDimensions(KinesisClientLibConfiguration.DEFAULT_METRICS_ENABLED_DIMENSIONS.a
  .build()
```

在 KCL 2.x 中，这些配置设置具有不同的软件包名称。要迁移到 2.x：

### 1. 分别将

`com.amazonaws.services.kinesis.clientlibrary.lib.worker.KinesisClientLibCon`  
和 `com.amazonaws.services.kinesis.metrics.interfaces.MetricsLevel`  
的导入语句更改为 `software.amazon.kinesis.metrics.MetricsLevel` 和  
`software.amazon.kinesis.metrics.MetricsUtil`。

```
// import com.amazonaws.services.kinesis.metrics.interfaces.MetricsLevel
```

```
import software.amazon.kinesis.metrics.MetricsLevel

// import
com.amazonaws.services.kinesis.clientlibrary.lib.worker.KinesisClientLibConfiguration
import software.amazon.kinesis.metrics.MetricsUtil
```

## 2. 将行

`metricsEnabledDimensionsKinesisClientLibConfiguration.DEFAULT_METRICS_ENABLED`  
 替换为  
`metricsEnabledDimensionsSet(MetricsUtil.OPERATION_DIMENSION_NAME,  
 MetricsUtil.SHARD_ID_DIMENSION_NAME)`

以下是包含自定义指标级别和指标维度的 `KinesisInputDStream` 的更新版本：

```
import software.amazon.kinesis.metrics.MetricsLevel
import software.amazon.kinesis.metrics.MetricsUtil

val kinesisStream = KinesisInputDStream.builder
  .streamingContext(ssc)
  .streamName(streamName)
  .endpointUrl(endpointUrl)
  .regionName(regionName)
  .initialPosition(new Latest())
  .checkpointAppName(appName)
  .checkpointInterval(kinesisCheckpointInterval)
  .storageLevel(StorageLevel.MEMORY_AND_DISK_2)
  .metricsLevel(MetricsLevel.DETAILED)
  .metricsEnabledDimensions(Set(MetricsUtil.OPERATION_DIMENSION_NAME,
  MetricsUtil.SHARD_ID_DIMENSION_NAME))
  .build()
```

- `KinesisInputDStream` 中的消息处理程序函数

在实例化 `KinesisInputDStream` 时，您还可以提供一个“消息处理程序函数”，该函数接收 `Kinesis` 记录并返回通用对象 `T`，以备您想使用记录中包含的其他数据（例如分区键）。

在 KCL 1.x 中，消息处理函数签名为 `:Record => T`，其中 `Record` 是 `com.amazonaws.services.kinesis.model.Record`。在 KCL 2.x 中，处理程序的签名更改为 `:KinesisClientRecord => T`，其中 `KinesisClientRecord` 是 `software.amazon.kinesis.retrieval.KinesisClientRecord`。

以下是在 KCL 1.x 中提供消息处理程序的示例：

```
import com.amazonaws.services.kinesis.model.Record

def addFive(r: Record): Int = JavaUtils.bytesToString(r.getData).toInt + 5
val stream = KinesisInputDStream.builder
  .streamingContext(ssc)
  .streamName(streamName)
  .endpointUrl(endpointUrl)
  .regionName(regionName)
  .initialPosition(new Latest())
  .checkpointAppName(appName)
  .checkpointInterval(Seconds(10))
  .storageLevel(StorageLevel.MEMORY_ONLY)
  .buildWithMessageHandler(addFive)
```

要迁移消息处理程序：

1. 将 `com.amazonaws.services.kinesis.model.Record` 的导入语句更改为 `software.amazon.kinesis.retrieval.KinesisClientRecord`。

```
// import com.amazonaws.services.kinesis.model.Record
import software.amazon.kinesis.retrieval.KinesisClientRecord
```

2. 更新消息处理程序的方法签名。

```
//def addFive(r: Record): Int = JavaUtils.bytesToString(r.getData).toInt + 5
def addFive = (r: KinesisClientRecord) => JavaUtils.bytesToString(r.data()).toInt
+ 5
```

以下是在 KCL 2.x 中提供消息处理程序的更新示例：

```
import software.amazon.kinesis.retrieval.KinesisClientRecord

def addFive = (r: KinesisClientRecord) => JavaUtils.bytesToString(r.data()).toInt + 5
val stream = KinesisInputDStream.builder
  .streamingContext(ssc)
  .streamName(streamName)
```

```
.endpointUrl(endpointUrl)
.regionName(regionName)
.initialPosition(new Latest())
.checkpointAppName(appName)
.checkpointInterval(Seconds(10))
.storageLevel(StorageLevel.MEMORY_ONLY)
.buildWithMessageHandler(addFive)
```

有关从 KCL 1.x 迁移到 2.x 的更多信息，请参阅将使用者[从 KCL 1.x 迁移到 2.x](#)。KCL

## 将 AWS 凭证提供程序从 AWS SDK 1.x 迁移到 2.x

凭证提供者用于获取与之交互的 AWS 凭证 AWS。在 2.x 中，有几项与 SDK 2.x 中的凭证提供程序相关的接口和类更改，可以在[在这里](#)找到。Spark Kinesis 连接器定义了一个接口 (`org.apache.spark.streaming.kinesis.SparkAWSCredentials`) 和实现类，用于返回 1.x 版本的 AWS 凭据提供程序。初始化 Kinesis 客户端时需要这些凭证提供程序。例如，如果您在应用程序 `SparkAWSCredentials.provider` 中使用该方法，则需要更新代码以使用 2.x 版本的 AWS 凭据提供程序。

以下是在 AWS SDK 1.x 中使用凭证提供程序的示例：

```
import org.apache.spark.streaming.kinesis.SparkAWSCredentials
import com.amazonaws.auth.AWSCredentialsProvider

val basicSparkCredentials = SparkAWSCredentials.builder
    .basicCredentials("accessKey", "secretKey")
    .build()

val credentialProvider = basicSparkCredentials.provider
assert(credentialProvider.isInstanceOf[AWSCredentialsProvider], "Type should be
AWSCredentialsProvider")
```

要迁移到 SDK 2.x，请执行以下操作：

1. 将 `com.amazonaws.auth.AWSCredentialsProvider` 的导入语句更改为 `software.amazon.awssdk.auth.credentials.AwsCredentialsProvider`

```
//import com.amazonaws.auth.AWSCredentialsProvider
import software.amazon.awssdk.auth.credentials.AwsCredentialsProvider
```

2. 更新使用此类的其余代码。

```
import org.apache.spark.streaming.kinesis.SparkAWSCredentials
import software.amazon.awssdk.auth.credentials.AwsCredentialsProvider

val basicSparkCredentials = SparkAWSCredentials.builder
    .basicCredentials("accessKey", "secretKey")
    .build()

val credentialProvider = basicSparkCredentials.provider
assert (credentialProvider.isInstanceOf[AwsCredentialsProvider], "Type should be
    AwsCredentialsProvider")
```

## 将 AWS 服务客户端从 AWS SDK 1.x 迁移到 2.x

AWS 服务客户端在 2.x 中具有不同的软件包名称 (即 `software.amazon.awssdk`)。而 SDK 1.x 使用 `com.amazonaws` 有关客户端更改的更多信息, 请参阅[此处](#)。如果您在代码中使用这些服务客户端, 则需要相应地迁移客户端。

以下是在 SDK 1.x 中创建客户端的示例:

```
import com.amazonaws.services.dynamodbv2.AmazonDynamoDBClient
import com.amazonaws.services.dynamodbv2.document.DynamoDB

AmazonDynamoDB ddbClient = AmazonDynamoDBClientBuilder.defaultClient();
AmazonDynamoDBClient ddbClient = new AmazonDynamoDBClient();
```

要迁移到 2.x:

1. 请更改服务客户端的导入语句。以 DynamoDB 客户端为例。您需要将 `com.amazonaws.services.dynamodbv2.AmazonDynamoDBClient` 或 `com.amazonaws.services.dynamodbv2.document.DynamoDB` 更改为 `software.amazon.awssdk.services.dynamodb.DynamoDbClient`。

```
// import com.amazonaws.services.dynamodbv2.AmazonDynamoDBClient
// import com.amazonaws.services.dynamodbv2.document.DynamoDB
import software.amazon.awssdk.services.dynamodb.DynamoDbClient
```

2. 更新用初始化客户端的代码

```
// AmazonDynamoDB ddbClient = AmazonDynamoDBClientBuilder.defaultClient();
```



```
// AmazonDynamoDBClient ddbClient = new AmazonDynamoDBClient();

DynamoDbClient ddbClient = DynamoDbClient.create();
DynamoDbClient ddbClient = DynamoDbClient.builder().build();
```

有关 AWS SDK从 1.x 迁移到 2.x 的更多信息，请参阅[AWS SDK适用于 Java 1.x 和 2.x 的区别](#)

## 流式传输应用程序的代码示例

```
import java.net.URI
import software.amazon.awssdk.auth.credentials.DefaultCredentialsProvider
import software.amazon.awssdk.http.apache.ApacheHttpClient
import software.amazon.awssdk.services.kinesis.KinesisClient
import software.amazon.awssdk.services.kinesis.model.DescribeStreamRequest
import software.amazon.awssdk.regions.Region
import software.amazon.kinesis.metrics.{MetricsLevel, MetricsUtil}

import org.apache.spark.SparkConf
import org.apache.spark.storage.StorageLevel
import org.apache.spark.streaming.{Milliseconds, StreamingContext}
import org.apache.spark.streaming.dstream.DStream.toPairDStreamFunctions
import org.apache.spark.streaming.kinesis.KinesisInitialPositions.Latest
import org.apache.spark.streaming.kinesis.KinesisInputDStream

object KinesisWordCountASLSDKV2 {

  def main(args: Array[String]): Unit = {
    val appName = "demo-app"
    val streamName = "demo-kinesis-test"
    val endpointUrl = "https://kinesis.us-west-2.amazonaws.com"
    val regionName = "us-west-2"

    // Determine the number of shards from the stream using the low-level Kinesis
    Client
    // from the AWS Java SDK.
    val credentialsProvider = DefaultCredentialsProvider.create
    require(credentialsProvider.resolveCredentials() != null,
      "No AWS credentials found. Please specify credentials using one of the methods
    specified " +
      "in https://docs.aws.amazon.com/sdk-for-java/latest/developer-guide/
    credentials.html")
```

```
val kinesisClient = KinesisClient.builder()
    .credentialsProvider(credentialsProvider)
    .region(Region.US_WEST_2)
    .endpointOverride(URI.create(endpointUrl))
    .httpClientBuilder(ApacheHttpClient.builder())
    .build()
val describeStreamRequest = DescribeStreamRequest.builder()
    .streamName(streamName)
    .build()
val numShards = kinesisClient.describeStream(describeStreamRequest)
    .streamDescription
    .shards
    .size

// In this example, we are going to create 1 Kinesis Receiver/input DStream for
each shard.
// This is not a necessity; if there are less receivers/DStreams than the number of
shards,
// then the shards will be automatically distributed among the receivers and each
receiver
// will receive data from multiple shards.
val numStreams = numShards

// Spark Streaming batch interval
val batchInterval = Milliseconds(2000)

// Kinesis checkpoint interval is the interval at which the DynamoDB is updated
with information
// on sequence number of records that have been received. Same as batchInterval for
this
// example.
val kinesisCheckpointInterval = batchInterval

// Setup the SparkConfig and StreamingContext
val sparkConfig = new SparkConf().setAppName("KinesisWordCountASLSDKV2")
val ssc = new StreamingContext(sparkConfig, batchInterval)

// Create the Kinesis DStreams
val kinesisStreams = (0 until numStreams).map { i =>
    KinesisInputDStream.builder
        .streamingContext(ssc)
        .streamName(streamName)
        .endpointUrl(endpointUrl)
```

```

        .regionName(regionName)
        .initialPosition(new Latest())
        .checkpointAppName(appName)
        .checkpointInterval(kinesisCheckpointInterval)
        .storageLevel(StorageLevel.MEMORY_AND_DISK_2)
        .metricsLevel(MetricsLevel.DETAILED)
        .metricsEnabledDimensions(Set(MetricsUtil.OPERATION_DIMENSION_NAME,
MetricsUtil.SHARD_ID_DIMENSION_NAME))
        .build()
    }

    // Union all the streams
    val unionStreams = ssc.union(kinesisStreams)

    // Convert each line of Array[Byte] to String, and split into words
    val words = unionStreams.flatMap(byteArray => new String(byteArray).split(" "))

    // Map each word to a (word, 1) tuple so we can reduce by key to count the words
    val wordCounts = words.map(word => (word, 1)).reduceByKey(_ + _)

    // Print the first 10 wordCounts
    wordCounts.print()

    // Start the streaming context and await termination
    ssc.start()
    ssc.awaitTermination()
}
}

```

## 使用升级后的 Spark Kinesis 连接器时的注意事项

- 如果您的应用程序Kinesis-producer-library使用JDK版本低于 11 的，则可能会遇到异常，例如java.lang.NoClassDefFoundError: javax/xml/bind/DatatypeConverter。之所以发生这种情况，是因EMR为 7.0 默认JDK包含 17 个，而自 Java 11+ 版本以来，J2EE 模块已从标准库中删除。此问题可以通过在 pom 文件中添加以下依赖项来解决。将库版本替换为您认为合适的版本。

```

<dependency>
  <groupId>javax.xml.bind</groupId>
  <artifactId>jaxb-api</artifactId>
  <version>${jaxb-api.version}</version>
</dependency>

```

- 创建EMR集群后，可以在此路径下找到 Spark Kinesis 连接器 jar：`/usr/lib/spark/connector/lib/`

## S3 DistCp (s3-dist-cp)

Apache DistCp 是一个开源工具，你可以用来复制大量数据。S3 与 DistCpAmazon S3 类似 DistCp，但经过了优化 AWS，尤其是 Amazon S3。DistCp 在 Amazon 4.0 及更高EMR版本中s3-dist-cp，S3 的命令是，您可以将其作为集群中的一个步骤或在命令行中添加。使用 S3DistCp，您可以高效地将大量数据从 Amazon S3 复制到可通过亚马逊EMR集群中的后续步骤进行处理HDFS的位置。您也可以使用 S3 DistCp 在亚马逊 S3 存储桶之间复制数据，或者将数据从中复制HDFS到 Amazon S3。S3 DistCp 在跨存储桶和跨 AWS 账户并行复制大量对象方面更具可扩展性和效率。

有关在现实场景中演示 S3DistCp 灵活性的特定命令，请参阅 AWS 大数据博客DistCp上的[S3 使用七个技巧](#)。

比如 DistCp，S3 DistCp MapReduce 使用分布式方式进行复制。它在几个服务器之间共享复制、错误处理、恢复和报告任务。有关 Apache DistCp 开源项目的更多信息，请参阅 Apache Hadoop 文档中的[DistCp指南](#)。

如果 DistCp S3 无法复制部分或全部指定文件，则集群步骤将失败并返回非零错误代码。如果发生这种情况，S3 DistCp 不会清理部分复制的文件。

### Important

S3 DistCp 不支持包含下划线字符的 Amazon S3 存储桶名称。

S3 DistCp 不支持 Parquet 文件的串联。PySpark 改用。有关更多信息，请参阅[在 Amazon 中串联 parquet 文件](#)。EMR

为了避免在使用 S3DistCp 将单个文件（而不是目录）从 S3 复制到 S3 时出现复制错误，HDFS请使用亚马逊EMR版本 5.33.0 或更高版本，或者亚马逊EMR 6.3.0 或更高版本。

## S3 DistCp 选项

尽管与之类似 DistCp，但 S3 DistCp 支持一组不同的选项来更改其复制和压缩数据的方式。

当您调用 S3 时DistCp，您可以指定下表中描述的选项。选项是用参数列表添加到步骤的。下表显示了 S3 DistCp 参数的示例。

选项	描述	必填
<code>--src=LOCATION</code>	<p>待复制数据的位置。这可以是一个HDFS或一个Amazon S3 地点。</p> <p>例如：<code>--src=s3://amzn-s3-demo-bucket1/logs/j-3GYXXXXXX9I0J/node</code></p> <div style="border: 1px solid #f08080; border-radius: 10px; padding: 10px; margin-top: 10px;"> <p><b>⚠ Important</b></p> <p>S3 DistCp 不支持包含下划线字符的 Amazon S3 存储桶名称。</p> </div>	是
<code>--dest=LOCATION</code>	<p>数据的目标位置。这可以是一个HDFS或一个Amazon S3 地点。</p> <p>例如：<code>--dest=hdfs:///output</code></p> <div style="border: 1px solid #f08080; border-radius: 10px; padding: 10px; margin-top: 10px;"> <p><b>⚠ Important</b></p> <p>S3 DistCp 不支持包含下划线字符的 Amazon S3 存储桶名称。</p> </div>	是
<code>--srcPattern=PATTERN</code>	<p>该<a href="#">正则表达式</a>会筛选 <code>--src</code> 的部分数据的复制操作。如果指定的既不是 <code>--srcPattern</code> 也不是 <code>--groupBy</code> ，那么会将 <code>--src</code> 的所有数据复制到 <code>--dest</code>。</p> <p>如果正则表达式参数包含特殊字符，如星号 ( * ) ，那么必须将正则表达式或整个 <code>--args</code> 字符串括在单引号 ( ' ) 中。</p> <p>例如：<code>--srcPattern=.*daemons.*-hadoop-.*</code></p>	否
<code>--groupBy=PATTERN</code>	<p>一种<a href="#">正则表达式</a>，它使 S3 DistCp 连接与该表达式匹配的文件。例如，您可以使用此选项把一个小时内写</p>	否

选项	描述	必填
	<p>入的所有日志文件组合成为单个文件。已连接文件的文件名是与该分组的正则表达式相匹配的值。</p> <p>圆括号表明文件应使用的分组方式，与圆括号内的语句匹配的所有项目都组合成单个输出文件。如果正则表达式不包含带括号的语句，则集群将在 S3 DistCp 步骤上失败并返回错误。</p> <p>如果正则表达式参数包含特殊字符，如星号 ( * ) ，那么必须将正则表达式或整个 --args 字符串括在单引号 ( ' ) 中。</p> <p>如果指定 --groupBy ，则仅复制与指定的模式匹配的文件。您不需要同时指定 --groupBy 和 --srcPattern 。</p> <p>例如：--groupBy=.*subnetid.*([0-9]+-[0-9]+-[0-9]+).*</p>	
--targetSize=SIZE	<p>要根据 --groupBy 选项创建的文件的大小，以兆字节 (MiB) 为单位。此值必须是整数。设置后 --targetSize ，S3 DistCp 会尝试匹配此大小；复制文件的实际大小可能大于或小于此值。基于数据文件大小聚合任务，因此目标文件大小将可能匹配源数据文件大小。</p> <p>如果由 --groupBy 连接的文件大于 --targetSize 的值，则将这些文件分解成部分文件，并在末尾附加一个数值按顺序命名。例如，连接组成 myfile.gz 的一个文件将被分解成部分文件，如：myfile0.gz 、 myfile1.gz 等。</p> <p>例如：--targetSize=2</p>	否

选项	描述	必填
<code>--appendToLastFile</code>	指定从 Amazon S3 复制到已存在的文件 DistCp 时 S3 HDFS 的行为。它向现有文件附加新文件数据。如果将 <code>--appendToLastFile</code> 与 <code>--groupBy</code> 结合使用，则将新数据附加到与相同的组匹配的文件。与 <code>--targetSize</code> 结合使用时，此选项也具有 <code>--groupBy</code> 行为	否
<code>--outputCodec=CODEC</code>	指定用于所复制文件的压缩编解码器。取值可以是 <code>:gzip</code> 、 <code>gz</code> 、 <code>lzo</code> 、 <code>snappy</code> 或 <code>none</code> 。例如，您可以使用此选项将使用 Gzip 压缩的输入文件转换为经过 LZO 压缩的输出文件，或者在复制操作中解压缩这些文件。如果选择输出编解码器，则系统会在文件名末尾追加适当的扩展名 (例如，对于 <code>gz</code> 和 <code>gzip</code> ，将追加扩展名 <code>.gz</code> )。如果不为 <code>--outputCodec</code> 指定值，则复制的文件不会出现压缩方面的变化。  例如： <code>--outputCodec=lzo</code>	否
<code>--s3ServerSideEncryption</code>	确保目标数据使用 AWS 服务端密钥在 Amazon S3 中传输 SSL 并自动加密。使用 S3 检索数据时 DistCp，对象会自动解密。如果尝试将未加密的对象复制到需要加密的 Amazon S3 存储桶，则操作将失败。有关更多信息，请参阅 <a href="#">使用数据加密</a> 。  例如： <code>--s3ServerSideEncryption</code>	否
<code>--deleteOnSuccess</code>	如果复制操作成功，则此选项会导致 S3 DistCp 从源位置删除复制的文件。如果您以计划任务的形式将输出文件 (如日志文件) 从一个位置复制到另一个位置，又不想复制两次相同的文件，那么这个选项会非常有用。  例如： <code>--deleteOnSuccess</code>	否
<code>--disableMultipartUpload</code>	禁用分段上载。  例如： <code>--disableMultipartUpload</code>	否

选项	描述	必填
<code>--multipartUploadChunkSize=SIZE</code>	<p>Amazon S3 分段上传中每个分段的大小，以 MiB 为单位。当 S3 复制的数据大于时，它会 DistCp 使用分段上传。 <code>multipartUploadChunkSize</code> 要提高作业性能，可以增加每个分段的大小。默认大小为 128 MiB。</p> <p>例如：<code>--multipartUploadChunkSize=1000</code></p>	否
<code>--numberOfFiles</code>	<p>在输出文件之前加上序号。计数从 0 开始，除非 <code>--startingIndex</code> 指定一个不同的值。</p> <p>例如：<code>--numberOfFiles</code></p>	否
<code>--startingIndex=INDEX</code>	<p>使用 <code>--numberOfFiles</code> 指定序列中的第一个数字。</p> <p>例如：<code>--startingIndex=1</code></p>	否
<code>--outputManifest=FILENAME</code>	<p>创建使用 Gzip 压缩的文本文件，其中包含 S3 DistCp 复制的所有文件的列表。</p> <p>例如：<code>--outputManifest=manifest-1.gz</code></p>	否
<code>--previousManifest=PATH</code>	<p>DistCp 使用 <code>--outputManifest</code> 标志读取上次调用 S3 期间创建的清单文件。设置该 <code>--previousManifest</code> 标志后，S3 会将清单中列出的文件 DistCp 排除在复制操作之外。如果同时指定 <code>--outputManifest</code> 和 <code>--previousManifest</code>，则之前清单中列出的文件也会出现在新的清单文件中，但不会复制这些文件。</p> <p>例如：<code>--previousManifest=/usr/bin/manifest-1.gz</code></p>	否
<code>--requirePreviousManifest</code>	<p>需要先前在调用 S3 时创建的清单 DistCp。如果将它设置为 <code>false</code>，则不指定之前的清单也不会生成错误。默认值为 <code>true</code>。</p>	否



选项	描述	必填
<code>--copyFromManifest</code>	<p>反转的行为<code>--previousManifest</code>，使 S3 DistCp 使用指定的清单文件作为要复制的文件列表，而不是要从复制中排除的文件列表。</p> <p>例如：<code>--copyFromManifest --previousManifest=/usr/bin/manifest-1.gz</code></p>	否
<code>--s3Endpoint=ENDPOINT</code>	<p>指定上载文件时要使用的 Amazon S3 终端节点。此选项会同时设置源位置和目标位置的终端节点。如果未设置，则默认终端节点是 <code>s3.amazonaws.com</code>。有关 Amazon S3 终端节点的列表，请参阅<a href="#">区域和终端节点</a>。</p> <p>例如：<code>--s3Endpoint=s3.eu-west-1.amazonaws.com</code></p>	否
<code>--storageClass=CLASS</code>	<p>目标为 Amazon S3 时要使用的存储类。有效值为 STANDARD 和 REDUCED_REDUNDANCY。如果未指定此选项，S3 将 DistCp 尝试保留存储类别。</p> <p>例如：<code>--storageClass=STANDARD</code></p>	否
<code>--srcPrefixesFile=PATH</code>	<p>Amazon S3 (<code>s3://</code>)、HDFS (<code>hdfs://</code>) 或本地文件系统 (<code>file://</code>) 中的文本文件，其中包含前缀列表，每行 <code>src</code> 一个前缀。</p> <p>如果提供 <code>srcPrefixesFile</code>，则 S3 DistCp 将不会列出 <code>src</code> 路径。相反，它将生成一个源列表，以作为列出在此文件中指定的所有前缀的组合结果。与 <code>src</code> 相比较的相对路径 (而不是这些前缀) 将用于生成目标路径。如果还指定了 <code>srcPattern</code>，则会将它应用于源前缀的组合列表结果以进一步筛选输入。如果使用了 <code>copyFromManifest</code>，则会复制清单中的对象并忽略 <code>srcPrefixesFile</code>。</p> <p>例如：<code>--srcPrefixesFile=PATH</code></p>	否

除了上述选项外，S3 还DistCp 实现了[工具接口](#)，这意味着它支持通用选项。

## 将 S3 DistCp 作为集群中的一个步骤添加

您可以DistCp 通过将 S3 作为步骤添加到集群中来调用 S3。可以在启动时向集群添加步骤，也可以使用控制台向正在运行的集群添加步骤CLI、或API。以下示例演示如何向正在运行的集群添加 S3 DistCp 步骤。有关向集群添加步骤的更多信息，请参阅 Amazon EMR 管理指南中的向[集群提交工作](#)。

要向正在运行的集群添加 S3 DistCp 步骤，请使用 AWS CLI

有关在中使用 Amazon EMR 命令的更多信息 AWS CLI，请参阅[AWS CLI 命令参考](#)。

- 要向调用 S3 的集群添加步骤DistCp，请将指定 S3 DistCp 应如何执行复制操作的参数作为参数传递。

以下示例将守护进程日志从 Amazon S3 复制到 `hdfs:///output`。在以下命令中：

- `--cluster-id` 指定集群
- `Jar`是 S3 DistCp JAR 文件的位置。有关如何使用 `command-runner.jar` 在集群上运行命令的示例，请参阅[提交自定义JAR步骤以运行脚本或命令](#)。
- `Args`是要传入 S3 的选项名称/值对的逗号分隔列表。DistCp有关可用选项的完整列表，请参阅[S3 DistCp 选项](#)。

要向正在运行的集群添加 S3 DistCp 复制步骤，请将以下内容放入保存在 A JSON mazon S3 或本地文件系统中的文件中，如`myStep.json`本示例所示。Replace（替换）`j-3GYXXXXXX9I0K` 使用您的集群 ID 并替换 `mybucket` 使用您的 Amazon S3 存储桶名称。

```
[
  {
    "Name": "S3DistCp step",
    "Args": ["s3-dist-cp", "--s3Endpoint=s3.amazonaws.com", "--src=s3://mybucket/logs/j-3GYXXXXXX9I0J/node/", "--dest=hdfs:///output", "--srcPattern=.*[a-zA-Z,]+"],
    "ActionOnFailure": "CONTINUE",
    "Type": "CUSTOM_JAR",
    "Jar": "command-runner.jar"
  }
]
```

```
aws emr add-steps --cluster-id j-3GYXXXXXX9I0K --steps file://./myStep.json
```

## Example 将日志文件从 Amazon S3 复制到 HDFS

此示例还说明了如何HDFS通过向正在运行的集群添加步骤，将存储在 Amazon S3 存储桶中的日志文件复制到。在此示例中，`--srcPattern` 选项用于限制复制到守护程序日志的数据。

要HDFS使用该`--srcPattern`选项将日志文件从 Amazon S3 复制到，请将以下内容放入保存在 A JSON mazon S3 或本地文件系统中的文件中，如`myStep.json`本示例所示。Replace ( 替换 ) `j-3GYXXXXXX9I0K` 使用您的集群 ID 并替换 `mybucket` 使用您的 Amazon S3 存储桶名称。

```
[
  {
    "Name": "S3DistCp step",
    "Args": ["s3-dist-cp", "--s3Endpoint=s3.amazonaws.com", "--src=s3://mybucket/logs/j-3GYXXXXXX9I0J/node/", "--dest=hdfs:///output", "--srcPattern=.*daemons.*-hadoop-.*"],
    "ActionOnFailure": "CONTINUE",
    "Type": "CUSTOM_JAR",
    "Jar": "command-runner.jar"
  }
]
```

## 在 S3 DistCp 任务失败后进行清理

如果 S3 DistCp 无法复制部分或全部指定文件，则命令或集群步骤将失败并返回非零错误代码。如果发生这种情况，S3 DistCp 不会清理部分复制的文件。您必须手动删除它们。

部分复制的文件将以 S3 DistCp 作业的唯一标识符保存到子目录中。HDFS tmp您可以在作业的标准输出中找到此 ID。

例如，对于 ID 为 S3 的 DistCp 作业 `4b1c37bb-91af-4391-aaf8-46a6067085a6`，您可以连接到集群的主节点并运行以下命令来查看与该任务关联的输出文件。

```
hdfs dfs -ls /tmp/4b1c37bb-91af-4391-aaf8-46a6067085a6/output
```

该命令将返回与以下类似的文件列表：

```
Found 8 items
-rw-r--r-- 1 hadoop hadoop 0 2018-12-10 06:03 /tmp/4b1c37bb-91af-4391-aaf8-46a6067085a6/output/_SUCCESS
-rw-r--r-- 1 hadoop hadoop 0 2018-12-10 06:02 /tmp/4b1c37bb-91af-4391-aaf8-46a6067085a6/output/part-r-00000
```

```
-rw-r--r--    1 hadoop hadoop          0 2018-12-10 06:02 /tmp/4b1c37bb-91af-4391-  
aaf8-46a6067085a6/output/part-r-00001  
-rw-r--r--    1 hadoop hadoop          0 2018-12-10 06:02 /tmp/4b1c37bb-91af-4391-  
aaf8-46a6067085a6/output/part-r-00002  
-rw-r--r--    1 hadoop hadoop          0 2018-12-10 06:03 /tmp/4b1c37bb-91af-4391-  
aaf8-46a6067085a6/output/part-r-00003  
-rw-r--r--    1 hadoop hadoop          0 2018-12-10 06:03 /tmp/4b1c37bb-91af-4391-  
aaf8-46a6067085a6/output/part-r-00004  
-rw-r--r--    1 hadoop hadoop          0 2018-12-10 06:03 /tmp/4b1c37bb-91af-4391-  
aaf8-46a6067085a6/output/part-r-00005  
-rw-r--r--    1 hadoop hadoop          0 2018-12-10 06:03 /tmp/4b1c37bb-91af-4391-  
aaf8-46a6067085a6/output/part-r-00006
```

然后，您可以运行以下命令来删除目录和所有内容。

```
hdfs dfs rm -rf /tmp/4b1c37bb-91af-4391-aaf8-46a6067085a6
```

## 在 Amazon EMR 集群上运行命令和脚本

本主题介绍如何作为步骤在集群上运行命令或脚本。作为步骤运行命令或脚本是[将工作提交到集群](#)的其中一种方式，在以下情况下非常有用：

- 当您未允许 SSH 访问您的 Amazon EMR 集群时
- 当你想运行 bash 或 shell 命令来排除集群故障时

您可以在创建集群时或者在集群处于 WAITING 状态时运行脚本。要在步骤处理开始前运行脚本，您可使用引导操作。有关引导操作更多信息，请参阅《Amazon EMR 管理指南》中的[创建引导操作以安装其它软件](#)。

Amazon EMR 提供以下工具来帮助您运行脚本、命令和其他集群上的程序。您可以使用 Amazon EMR 管理控制台或 AWS CLI 调用所有工具。

### command-runner.jar

位于集群的 Amazon EMR AMI 上。您可以使用 `command-runner.jar` 在集群上运行命令。你没有使用其完整路径指定 `command-runner.jar`。

### script-runner.jar

Amazon S3 托管在 `s3://<region>.elasticmapreduce/libs/script-runner/script-runner.jar`，其中 `<region>` 是 Amazon EMR 集群所在的区域。您可以使用 `script-runner.jar` 以在集群上运行本地或 Amazon S3 上保存的脚本。当您提交步骤时，您必须指定 `script-runner.jar` 的完整 URI。

## 提交自定义 JAR 步骤以运行脚本或命令

以下 AWS CLI 示例说明了 Amazon EMR 的 `command-runner.jar` 和 `script-runner.jar` 一些常见使用场景。

Example：使用 `command-runner.jar` 在集群上运行命令

当您使用 `command-runner.jar` 时，您可以在步骤的参数列表中指定命令、选项和值。

以下 AWS CLI 示例将步骤提交至调用 `command-runner.jar` 的运行中的集群。Args 列表指定的命令下载名为 `my-script.sh` 的脚本，下载方式是从 Amazon S3 进入 hadoop 用户主目录。然后，该命令修改脚本权限并运行 `my-script.sh`。

当您使用 AWS CLI 时，Args 列表中的项目应该用逗号分隔开来，列表元素之间没有空格。例如，使用 `Args=[example-command,example-option,"example option value"]` 而不是 `Args=[example-command, example-option, "example option value"]`。

```
aws emr add-steps \  
--cluster-id j-2AXXXXXXGAPLF \  
--steps Type=CUSTOM_JAR,Name="Download a script from S3, change its permissions, and  
run it",ActionOnFailure=CONTINUE,Jar=command-runner.jar,Args=[bash,-c,"aws s3 cp s3://  
EXAMPLE-DOC-BUCKET/my-script.sh /home/hadoop; chmod u+x /home/hadoop/my-script.sh; cd /  
home/hadoop; ./my-script.sh"]
```

Example : 使用 **script-runner.jar** 在集群上运行脚本

当您使用 `script-runner.jar` 时，在步骤的参数列表中指定想要运行的脚本。

以下 AWS CLI 示例将步骤提交至调用 `script-runner.jar` 的运行中的集群。在此情况下，称为 *my-script.sh* 的脚本存储在 Amazon S3。您还可以指定存储在集群主节点 (master node) 的本地脚本。

```
aws emr add-steps \  
--cluster-id j-2AXXXXXXGAPLF \  
--steps Type=CUSTOM_JAR,Name="Run a script from S3 with script-  
runner.jar",ActionOnFailure=CONTINUE,Jar=s3://us-west-2.elasticmapreduce/libs/script-  
runner/script-runner.jar,Args=[s3://EXAMPLE-DOC-BUCKET/my-script.sh]
```

## 其他使用 **command-runner.jar** 的方法

您还可以借助类似于 `spark-submit` 或 `hadoop-streaming` 的工具，使用 `command-runner.jar` 将工作提交至集群。当您使用 `command-runner.jar` 启动应用程序时，您指定 `CUSTOM_JAR` 作为步骤类型而不是使用类似于 `SPARK`、`STREAMING` 或者 `PIG` 的值。工具可用性取决于您在集群上安装的应用程序。

以下示例命令借助 `spark-submit` 使用 `command-runner.jar` 提交步骤。Args 列表指定 `spark-submit` 作为命令，接着是附加参数和值的 Spark 应用程序 *my-app.py* Amazon S3 URI。

```
aws emr add-steps \  
--cluster-id j-2AXXXXXXGAPLF \  
--steps Type=CUSTOM_JAR,Name="Run spark-submit using command-  
runner.jar",ActionOnFailure=CONTINUE,Jar=command-runner.jar,Args=[spark-submit,S3://  
DOC-EXAMPLE-BUCKET/my-app.py,ArgName1,ArgValue1,ArgName2,ArgValue2]
```

下表列出了通过 `command-runner.jar` 您可以使用的其他工具。

工具名称	描述
<code>hadoop-streaming</code>	提交 Hadoop 流式处理程序。在控制台和一些开发工具包中，这是流步骤。
<code>hive-script</code>	运行 Hive 脚本。在控制台和开发工具包中，这是 Hive 步骤。
<code>pig-script</code>	运行 Pig 脚本。在控制台和开发工具包中，这是 Pig 步骤。
<code>spark-submit</code>	运行 Spark 应用程序。在控制台中，这是 Spark 步骤。
<code>hadoop-lzo</code>	在目录上运行 <a href="#">Hadoop LZO 索引器</a> 。
<code>s3-dist-cp</code>	将大量数据从 Amazon S3 分布式复制到 HDFS。有关更多信息，请参阅 <a href="#">S3 DistCp (s3-dist-cp)</a> 。

# AWS 术语表

有关最新的 AWS 术语，请参阅《AWS 词汇表参考》中的 [AWS 词汇表](#)。



本文属于机器翻译版本。若本译文内容与英语原文存在差异，则一律以英文原文为准。