Alibaba Cloud E-MapReduce

常见问题

文档版本：20200312

法律声明

阿里云提醒您在阅读或使用本文档之前仔细阅读、充分理解本法律声明各条款的内容。如果您阅读或使用本文档，您的阅读或使用行为将被视为对本声明全部内容的认可。

1. **您应当通过阿里云网站或阿里云提供的其他授权通道下载、获取本文档，且仅能用于自身的合法合规的业务活动。本文档的内容视为阿里云的保密信息，您应当严格遵守保密义务；未经阿里云事先书面同意，您不得向任何第三方披露本手册内容或提供给任何第三方使用。**
2. **未经阿里云事先书面许可，任何单位、公司或个人不得擅自摘抄、翻译、复制本文档内容的部分或全部，不得以任何方式或途径进行传播和宣传。**
3. **由于产品版本升级、调整或其他原因，本文档内容有可能变更。阿里云保留在没有任何通知或者提示下对本文档的内容进行修改的权利，并在阿里云授权通道中不时发布更新后的用户文档。您应当实时关注用户文档的版本变更并通过阿里云授权渠道下载、获取最新版的用户文档。**
4. **本文档仅作为用户使用阿里云产品及服务的参考性指引，阿里云以产品及服务的“现状”、“有缺陷”和“当前功能”的状态提供本文档。阿里云在现有技术的基础上尽最大努力提供相应的介绍及操作指引，但阿里云在此明确声明对本文档内容的准确性、完整性、适用性、可靠性等不作任何明示或暗示的保证。任何单位、公司或个人因为下载、使用或信赖本文档而发生任何差错或经济损失的，阿里云不承担任何法律责任。在任何情况下，阿里云均不对任何间接性、后果性、惩戒性、偶然性、特殊性或刑罚性的损害，包括用户使用或信赖本文档而遭受的利润损失，承担责任（即使阿里云已被告知该等损失的可能性）。**
5. **阿里云文档中所有内容，包括但不限于图片、架构设计、页面布局、文字描述，均由阿里云和/或其关联公司依法拥有其知识产权，包括但不限于商标权、专利权、著作权、商业秘密等。非经阿里云和/或其关联公司书面同意，任何人不得擅自使用、修改、复制、公开传播、改变、散布、发行或公开发表阿里云网站、产品程序或内容。此外，未经阿里云事先书面同意，任何人不得为了任何营销、广告、促销或其他目的使用、公布或复制阿里云的名称（包括但不限于单独为或以组合形式包含“阿里云”、“Aliyun”、“万网”等阿里云和/或其关联公司品牌，上述品牌的附属标志及图案或任何类似公司名称、商号、商标、产品或服务名称、域名、图案标示、标志、标识或通过特定描述使第三方能够识别阿里云和/或其关联公司）。**
6. **如若发现本文档存在任何错误，请与阿里云取得直接联系。**

通用约定

格式	说明	样例
	该类警示信息将导致系统重大变更甚至故障，或者导致人身伤害等结果。	 **禁止:** 重置操作将丢失用户配置数据。
	该类警示信息可能会导致系统重大变更甚至故障，或者导致人身伤害等结果。	 **警告:** 重启操作将导致业务中断，恢复业务时间约十分钟。
	用于警示信息、补充说明等，是用户必须了解的内容。	 **注意:** 权重设置为0，该服务器不会再接受新请求。
	用于补充说明、最佳实践、窍门等，不是用户必须了解的内容。	 **说明:** 您也可以通过按Ctrl + A选中全部文件。
>	多级菜单递进。	单击设置 > 网络 > 设置网络类型。
粗体	表示按键、菜单、页面名称等UI元素。	在结果确认页面，单击**确定**。
`Courier字体`	命令。	执行`cd /d C:/window`命令，进入Windows系统文件夹。
*## *	表示参数、变量。	`bae log list --instanceid` *`Instance_ID`*
[]或者[a|b]	表示可选项，至多选择一个。	`ipconfig [-all|-t]`
{}或者{a|b}	表示必选项，至多选择一个。	`switch {active|stand}`

目录

1 E-MapReduce 基本问题

Q：作业和执行计划的区别

A：在阿里云 E-MapReduce 中，要运行作业，需要有分成两个步骤，分别是：

- **创建作业**

  在 E-MapReduce 产品中，创建一个作业，实际上是创建一个作业运行配置，它并不能被直接运行。即如果在 E-MapReduce 中创建了一个作业，实际上只是创建了一个作业如何运行的配置，这份配置中包括该作业要运行的 jar 包，数据的输入输出地址，以及一些运行参数。这样的一份配置创建好后，给它命一个名，即定义了一个作业。当您需要调试运行作业的时候就需要执行计划了。

- **创建执行计划**

  执行计划，是将作业与集群关联起来的一个纽带。通过它，我们可以把多个作业组合成一个作业序列，通过它我们可以为作业准备一个运行集群（或者自动创建出一个临时集群或者关联一个已存在的集群），通过它我们可以为这个作业序列设置周期执行计划，并在完成任务后自动释放集群。我们也可以在执行记录列表上查看每一次执行的执行成功情况与日志。

Q：如何查看作业日志

A：在 E-MapReduce 系统里，系统已经将作业运行日志按照 JobID 的规划上传到 OSS 中（路径由用户在创建集群时设置），用户可以直接在网页上单击查看作业日志。如果用户是登录到 Master 机器进行作业提交和脚本运行等，则日志根据用户自己的脚本而定，用户可以自行规划。

Q：如何登录 Core 节点

A：按照如下步骤：

1. 首先在 Master 节点上切换到 Hadoop 账号：

```
su hadoop
```

2. 然后即可免密码 SSH 登录到对应的 Core 节点：

```
ssh emr-worker-1
```

3. 通过 sudo 可以获得 root 权限：

```
sudo vi /etc/hosts
```

Q：直接在 OSS 上查看日志

A：用户也可以直接从 OSS 上查找所有的日志文件，并下载。但是因为 OSS 不能直接查看日志，使用起来会比较麻烦一些。如果用户创建集群时打开了运行日志功能，并且指定了一个 OSS 的日志位置，那么作业的日志要如何找到呢？例如对下面这个保存位置*OSS://mybucket/emr/spark*：

1. 首先来到执行计划的页面，找到对应的执行计划，单击运行记录进入运行记录页面。
2. 在运行记录页面找到具体的那一条执行记录，例如最后的一条执行记录。然后单击它对应的执行集群查看这个执行集群的 ID。
3. 然后在*OSS://mybucket/emr/spark*目录下，寻找*OSS://mybucket/emr/spark/*集群ID这个目录
4. 在*OSS://mybucket/emr/spark/clusterID/jobs* 目录下会按照作业的执行 ID 存放多个目录，每一个目录下存放了这个作业的运行日志文件。

Q：集群、执行计划以及运行作业的计时策略

A：三种计时策略如下：

- 集群的计时策略

  在集群列表里可以看到每个集群的运行时间，该运行时间的计算策略为运行时间 = 集群释放时刻 - 集群开始构建时刻。即集群一旦开始构建就开始计时，直到集群的生命周期结束。

- **执行计划的计时策略**

  在执行计划的运行记录列表，可以看到每次执行记录运行的时间，该时间的计时策略总结为两种情况：

  - **如果执行计划是按需执行的，每次执行记录的运行过程涉及到创建集群、提交作业运行、释放集群。所以按需执行计划的运行时间计算策略为，运行时间 = 构建集群的时间 + 执行计划包含所有作业全部运行结束的总耗时 + 集群释放的时间。**
  - **如果执行计划是关联已有集群运行的，整个运行周期不涉及到创建集群和释放集群，所以其运行时间 = 执行计划包含所有作业全部运行结束的总耗时。**

- **作业的计时策略**

  这里的作业指的是被挂载到执行计划里面的作业。在每条执行计划运行记录右侧的查看作业列表单击进去可以查看到该执行作计划下作业列表。这里每个作业的运行时间的计算策略为，运行时间 = 作业运行结束的实际时间 - 作业开始运行的实际时间。作业运行开始（结束）的实际时间指的是作业被 Spark 或 Hadoop 集群实际开始调度运行或运行结束的时间点。

Q：第一次使用执行计划时没有安全组可选

A：因为一些安全的原因，E-MapReduce目前的安全组并不能直接选择用户的已有安全组来使用，所以如果您还没有在E-MapReduce中创建过安全组的话，在执行计划上将无法选择到可用的安全组。我们推荐您先手动创建一个按需集群来进行作业的测试，手动创建集群的时候可以创建一个新的E-MapReduce 安全组，等到测试都通过了以后，再设置您的执行计划来周期调度。这个时候之前创建的安全组也会出现在这里可供选择。

Q：读写 MaxCompute 时，抛出java.lang.RuntimeException.Parse response failed: ‘<!DOCTYPE html>…’

A：检查 MaxCompute tunnel endpoint 是否正确，如果写错会出现这个错误。

Q：多个 ConsumerID 消费同一个 Topic 时出现 TPS 不一致问题

A：有可能这个 Topic 在公测或其他环境创建过，导致某些 Consumer 组消费数据不一致。请在工单系统中将对应的 Topic 和 ConsumerID 提交到 ONS 处理。

Q：E-MapReduce 中能否查看作业的 Worker 上日志

A：可以。前置条件：是创建集群时打开运行日志选项。查看日志位置：执行计划列表 > 更多 > 运行记录 > 运行记录 > 查看作业列表 > 作业列表 > 作业实例。

Q：Hive 创建外部表，没有数据

A：例如：

```
CREATE EXTERNAL TABLE storage_log(content STRING) PARTITIONED BY (ds
STRING)
    ROW FORMAT DELIMITED
    FIELDS TERMINATED BY '\t'
    STORED AS TEXTFILE
    LOCATION 'oss://log-124531712/biz-logs/airtake/pro/storage';
    hive> select * from storage_log;
    OK
    Time taken: 0.3 seconds
    创建完外部表后没有数据
```

实际上 Hive 并不会自动关联指定目录的 partitions 目录，您需要手动操作，例如：

```
alter table storage_log add partition(ds=123);

                                                OK
    Time taken: 0.137 seconds
    hive> select * from storage_log;
    OK
    abcd    123
    efgh    123
```

Q：Spark Streaming 作业运行一段时间后无故结束

A：首先检查 Spark 版本是否是 1.6 之前版本。Spark 1.6 修复了一个内存泄漏的 BUG，这个 BUG 会导致 container 内存超用然后被kill掉（当然，这只是可能的原因之一，不能说明Spark 1.6 不存在任何问题）。此外，检查自己的代码在使用内存上有没有做好优化。

Q：Spark Streaming 作业已经结束，但是 E-MapReduce 控制台显示作业还处于“运行中”状态

A：检查 Spark Streaming 作业的运行模式是否是 yarn-client，若是建议改成 yarn-cluster 模式。E-MapReduce 对 yarn-client 模式的 Spark Streaming 作业的状态监控存在问题，会尽快修复。

Q：“Error: Could not find or load main class”

A：检查作业配置中作业jar包的路径协议头是否是 `ossref`，若不是请改为 `ossref`。

Q：集群机器分工使用说明

A：E-MapReduce 中包含一个 Master 节点和多个 Slave（或者 Worker）节点。其中 Master 节点不参与数据存储和计算任务，Slave 节点用来存储数据和计算任务。例如 3 台 4 核 8 G 机型的集群，其中一台机器用来作为 Master 节点，另外两台用来作为 Slave 节点，也就是集群的可用计算资源为 2 台 4 核 8 G 机器。

Q：如何在 MR 作业中使用本地共享库

A：方法有很多，这里给出一种方式。修改 *mapred-site.xml* **文件，例如：**

```
<property>
    <name>mapred.child.java.opts</name>
    <value>-Xmx1024m -Djava.library.path=/usr/local/share/</value>
  </property>
  <property>
    <name>mapreduce.admin.user.env</name>
    <value>LD_LIBRARY_PATH=$HADOOP_COMMON_HOME/lib/native:/usr/local/
lib</value>
  </property>
```

只要加上您所需的库文件即可。

Q：如何在 MR/Spark 作业中指定 OSS 数据源文件路径

A：如下：`OSS URL： oss://[accessKeyId:accessKeySecret@]bucket[.endpoint]/object/path`

用户在作业中指定输入输出数据源时使用这种 URI，可以类比`hdfs://`**。用户操作 OSS 数据时：**

- **（建议）E-MapReduce提供了 MetaService 服务，支持免 AK 访问 OSS 数据，直接写***oss://bucket/object/path*。
- **（不建议）可以将 AccessKeyId，AccessKeySecret 以及 endpoint 配置到Configuration（Spark 作业是 SparkConf，MR 类作业是 Configuration）中，也可以在 URI中直接指定 AccessKeyId，AccessKeySecret 以及 endpoint。具体请参考开发准备一节。**

Q：Spark SQL抛出“Exception in thread “main” java.sql.SQLException: No suitable driver found for jdbc:mysql:xxx”报错

A：

- **低版本 mysql-connector-java 有可能出现类似问题，更新到最新版本。**
- **作业参数中使用** *–driver-class-path ossref://bucket/…/mysql-connector-java-[version].jar* **来加载** *mysql-connector-java* **包，直接将** *mysql-connector-java* **打进作业 jar 包也会出现上述问题。**

Q：Spark SQL 连 RDS 出现“Invalid authorization specification, message from server: ip not in whitelist”

A：检查 RDS 的白名单设置，将集群机器的内网地址加到 RDS 的白名单中。

Q : 创建低配置机型集群注意事项

A：

- **若 Master 节点选择 2 核 4 G 机型，则 Master 节点内存非常吃紧，很容易造成物理内存不够用，建议调大Master内存。**
- **若 Slave节点选择 2 核 4 G 机型，在运行 MR 作业或者 Hive 作业时，请调节参数。MR 作业添加参数**`-D yarn.app.mapreduce.am.resource.mb=1024`**；Hive 作业设置参数** `set yarn.app.mapreduce.am.resource.mb=1024`**;避免作业挂起。**

Q：Hive/Impala 作业读取 SparkSQL 导入的 Parquet 表报错（表包含 Decimal 格式的列）： Failed with exception java.io.IOException:org.apache.parquet.io.ParquetDecodingException: Can not read value at 0 in block -1 in file hdfs://…/…/part-00000-xxx.snappy.parquet

A： 由于 Hive 和 SparkSQL 在 Decimal 类型上使用了不同的转换方式写入 Parquet，导致 Hive 无法正确读取 SparkSQL 所导入的数据。对于已有的使用 SparkSQL 导入的数据，如果有被 Hive/Impala 使用的需求，建议加上 `spark.sql.parquet.writeLegacyFormat=true` **，重新导入数据。**

Q:beeline 如何访问 Kerberos 安全集群

A:

- **HA 集群（Discovery 模式）**

```
!connect jdbc:hive2://emr-header-1:2181,emr-header-2:2181,emr-
header-3:2181/;serviceDiscoveryMode=zooKeeper;zooKeeperNamespace=
hiveserver2;principal=hive/_HOST@EMR.${clusterId).COM
```

- **HA 集群（直连某台机器）**

  连 emr-header-1

  ```
  !connect jdbc:hive2://emr-header-1:10000/;principal=hive/emr-header-
  1@EMR.${clusterId}.COM
  ```

  连 emr-header-2

  ```
  !connect jdbc:hive2://emr-header-2:10000/;principal=hive/emr-header-
  2@EMR.${clusterId}.COM
  ```

- **非 HA 集群**

  ```
  !connect jdbc:hive2://emr-header-1:10000/;principal=hive/emr-header-
  1@EMR.${clusterId}.COM
  ```

Q：ThriftServer 进程正常，但链接出现异常，报错 “Connection refused telnet emr-header-1 10001” 无法连接

A：

可以查看 *`/mnt/disk1/log/spark`* **日志**。

该问题是由于 thrift server oom，需要调大内存，调大 `spark.driver.memory` **值即可**。

Q：如何查看 E-MapReduce 服务的日志?

A：

登录 master 节点在 *`/mnt/disk1/log`* **中查看对应服务的日志**。

2 集群创建相关

如果创建E-MapReduce集群失败，您可以根据创建失败的错误提示信息查找对应的解决方法。

错误提示：“The specified zone is not available for purchase.”

出现这个情况一般是您选择创建集群的可用区暂时停止了售卖按量ECS，建议您更换可用区购买。

错误提示：“The request processing has failed due to some unknown error, exception or failure.”

E-MapReduce依赖于阿里云的ECS，这个错误是依赖的ECS管控系统出现一个未知的错误，您可以稍等一会儿，也可以立即提交工单给ECS等待工程师为您排查。

错误提示：“The Node Controller is temporarily unavailable.”

E-MapReduce依赖于阿里云的ECS，这个错误是依赖的ECS管控系统出现了暂时性的问题。请稍等一会儿再重试集群的创建。

错误提示：“Zone或者Cluster的库存不够了”

E-MapReduce集群的创建依赖于阿里云云服务器ECS，出现该错误提示是由于该可用区的ECS实例库存不足。您可以尝试手动选择其他可用区重新创建集群，或者使用随机模式创建集群。

错误提示：“指定的InstanceType未授权使用”

ECS的按量付费高配机型（8核以上的所有机型）需要用户申请开通以后才可以使用，请单击提交工单申请。请申请8核16G、8核32G、16核32G和16核64G这四种目前E-MapReduce使用的机型。

3 产品使用相关

Q： E-MapReduce 按量高配节点问题

按量的 8 核以及 8 核以上的机器默认是不展示的，您可以去 ECS 申请开通 ECS 的高配节点，然后就能在E-MapReduce这里使用了。推荐您使用包年包月集群，不受高配的限制。

Q： 高安全集群

创建集群时的高安全集群，指的是 Kerberos。具体请参见官方文档 *Kerberos* 简介。 如果用户创建了这个类型的集群，低版本还不能支持直接关闭。 3.12 以及后续的版本可以支持关闭，之前的版本目前只能重新创建。

Q： 创建集群失败，构建失败 “The specified instance Type exceeds the maximum limit for the PostPaid instances”

通常是用户的按量节点数量的上限到了。ECS 根据不同用户，按量节点上限是不一样的，需要用户去申请加大。如果确认不是上述的原因，还有一种可能是用户没有创建这个机型的权限，需要去 ECS 开通这个机型的使用权限。

Q：集群续费问题

续费操作请参见集群续费。经常会有用户反馈续费了但是还是会收到没有续费的通知。这是因为E-MapReduce现在有 2 块，一块是 E-MapReduce，一块是 ECS，大部分的用户都只是续费了 ECS 而没有续费 E-MapReduce。您可以打开续费界面查看 ECS 和 E-MapReduce到期时间。

Q：自动续费

E-MapReduce支持自动续费操作，支持 E-MapReduce和 ECS 的自动续费。

Q：已有/现存 ECS 是否可以用到 E-MapReduce集群中

目前还不能支持，用户要创建 E-MapReduce集群，需要在 E-MapReduce控制台上来创建 ECS 。

Q：E-Mapreduce 主节点不允许安装其它软件

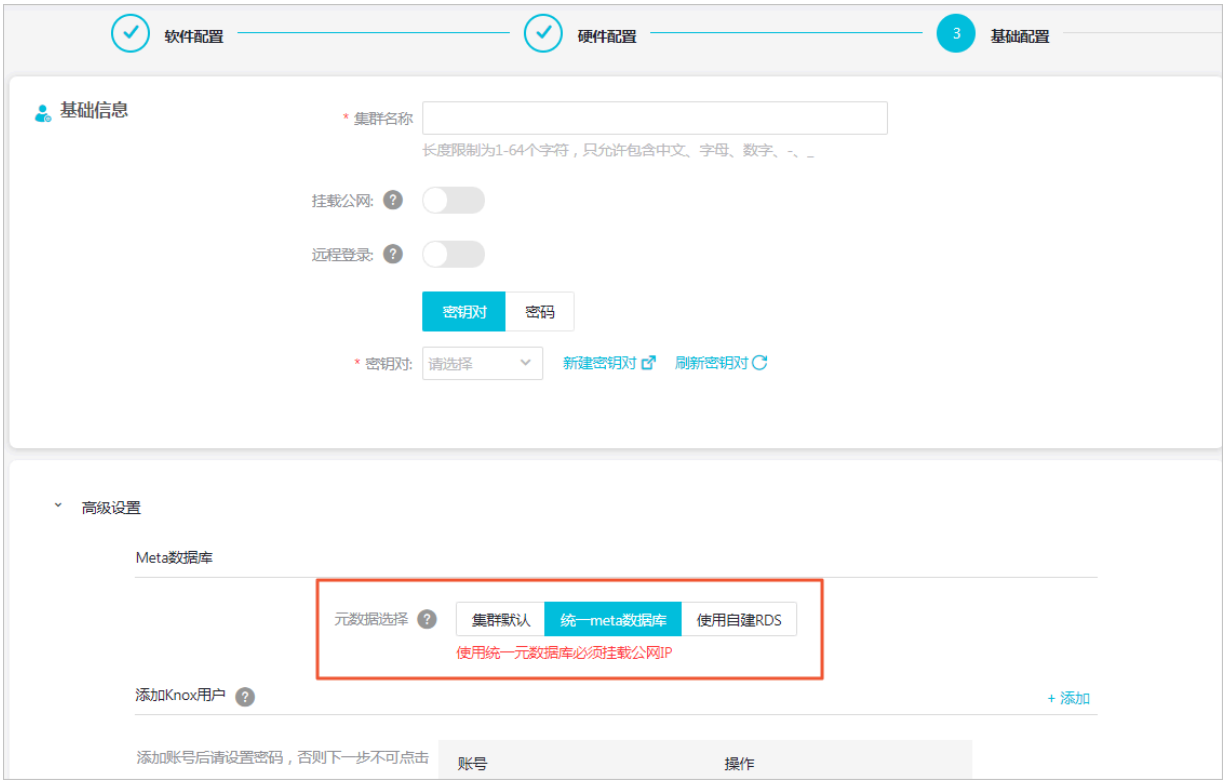
理论上可以在不破坏集群环境的前提下安装。但是这些软件的运行可能会影响到集群的稳定可靠性，不建议进行此类操作。

Q：如何登录 Core 节点，并进行 root 权限操作

请参见文档创建集群中的登录 Core 节点部分。

Q：E-MapReduce 集群 header 节点有公网 IP，存在安全风险，是否可以通过 ECS 控制台关闭公网 IP，关闭公网 IP 是否会对 E-MapReduce 服务产生影响?

如果您没有使用 E-MapReduce的统一元数据库功能，可以关闭公网 IP。



Q：各个节点之上的服务开机会自动启动吗？服务异常中断也会自动重启吗

会自动启动。服务异常中断会自动恢复，但是如果启动有错误导致无法启动，会重试 3 次。

Q：子账号需要什么权限才能操作E-MapReduce

首先主账号必须有 EMRdefaultRole 和 ECSdefaultRole 权限策略，否则子账号也开通不了E-MapReduce服务。然后子账号必须有EMRfullaccessrole 权限策略才能操作E-MapReduce。

Q：HUE相关问题

HUE 的默认初始用户和密码，参见*Hue* 使用说明 。如果忘记了用户名和密码，也请参见上面的文档进行操作。如果出现Hue无法访问 Web HDFS 的情况是因为 HDFS 的 `dfs.webhdfs.enabled` 的值为 false，在 E-MapReduce控制台上设置这个参数为 true， 重启一下 HDFS 就好了。

Q：E-Mapreduce 服务，Zeppelin 默认可以匿名访问，请问想关闭匿名访问要怎么操作？E-Mapreduce 中的 Zeppelin 配置管理页面没有配置选项，要修改 master 节点上的配置文件吗

需要手动修改配置，并重启 Zeppelin，可以参见*Apache Zeppelin* 设置访问登录。

Q：用 thrift 连接 Hbase，程序中用端口号 9090 连接不上，请问 thrift 的 Hbase 端口号用多少

使用 9099 端口。

Q：在控制台的表管理里面新建了新的数据库，创建成功后刷新页面，库不见了，但是在Zeppelin 中使用 show databases 命令查看可以查看新建的库，请问这是什么原因

目前和这个功能主要是配合统一 meta 数据库使用，才能在页面上看到库和表，如果没有选择，无法使用界面上的查看库和表，直接 Hue 建库建表或者登录集群使用 `hive cli` 操作。 目前已经创建的集群列表页不支持查看 Hive 算数据存储的方式，可以通过以下方式来确认：登录 header 1 查看 *`/etc/ecm/hive-conf/hive-site.xml`* 文件，确认一下 `javax.jdo.option.ConnectionURL` 这个配置的值，如果配置的是 emr-header-1，那说明就是没有选择统一 meta 数据库。

Q：E-MapReduce 是否支持竞价实例

E-MapReduce目前在使用弹性伸缩功能时，支持抢占式实例*#unique_11*。

Q：创建集群时是选择了统一meta库的，不能执行 Hive，错误信息如下：`FAILED: SemanticException org.apache.hadoop.hive.ql.metadata.HiveException: java.lang.RuntimeException: Unable to instantiate org.apache.hadoop.hive.ql.metadata.SessionHiveMetaStoreClient`

通常是集群没有公网 EIP 导致的，统一 meta 是需要公网 IP 的，用户的 EIP 满了，导致创建集群的时候没有创建出来，然后就无法连接上。用户需要手工绑定，然后通知我们给这个 EIP 加到数据库的安全组中。

Q：关于低配节点

低配节点指的是 2c8g 的配置的节点。

> **说明:**
> **可以给部分用户开白名单使用，但是需要注意的是出问题的概率很高，用户需要自己解决。**

Q：Master 节点没有可用的机型

用户在该可用区没有可选的 master 机型，可以切换可用区。

Q：E-MapReduce 如何让其他的 ECS 机器提交任务，并获取任务结果

可以使用 Gateway 功能，请参见文档 *Gateway* 实例，直接在E-MapReduce控制台上购买 Gateway。

Q：E-MapReduce 磁盘如何进行扩容

请参见文档磁盘扩容。

Q：已建集群如果没有配置 OSS 运行日志，后期可以变更配置吗

目前不支持。推荐使用新版工作流上去，就不需要这个设置了。可以直接查看日志。

Q：使用 E-MapReduce 的服务搭建集群使用Zookeeper，Kafka，Storm， 然后购买阿里云 Hbase 服务，双方能数据会连接吗

可以连接的。请使用专有网络，E-MapReduce 集群和 HBase 集群需要置于同一个 VPC 下。且打开了 HBase 的白名单，使E-MapReduce集群可以访问到 HBase。

Q：E-MapReduce 集群如何减少 Core 和 Task 节点？将现有集群 4 个 Core 和 2 个 Task 节点变为只有 2 个Core节点可以实现吗

Core 节点目前不支持减少，如果您需要缩容，可以通过提交 ECS 工单的方式来退掉集群中的 ECS 节点。部分有 ZK 服务的节点，退节点的时候按照 worker 的编号，从后往前退，例如 worker1、worker2、worker3和worker4，那么在退的时候就按照4->3->2 这样的顺序。包月的 Core 和 Task 节点目前都无法支持控制台上缩容。按量的 Task 节点目前可以直接在控制台上进行缩容。

Q：E-MapReduce 退款操作

您如需申请退款，请提交工单联系E-MapReduce产品团队，并提供您退款的原因。

Q：E-MapReduce 和 MaxCompute 的区别

两者的使用场景差异不大，都是大数据处理解决方案。E-MapReduce 是完全在开源的基础上构建的大数据平台，对开源软件的使用方式和实践方式都 100% 兼容的。MaxCompute 是由阿里巴巴开发的，对外不开源，但是封装后使用起来比较方便，而且运维成本也较低。E-MapReduce 是基于开源的 Hadoop 体系做的产品 。如果您的开发人员有比较多的 Hadoop 经验的话，可以直接使用。而用 MaxCompute 的话，需要对代码做一些修改，但修改量并不大 。

Q：E-MapReduce 集群的 mysql 密码是什么

可以直接在集群的配置管理，单击 Hive 服务进入，然后单击配置查看。

```
javax.jdo.option.ConnectionURL
```

```
javax.jdo.option.ConnectionUserName
```

```
javax.jdo.option.ConnectionPassword
```

Q：购买了 E-mapreduce 3.x 版本，在 Zeppelin 中运行 Spark 例子时，提示“Spark 2.2.1 is not supported”，经确认 Zeppelin 0.71 确实不支持 Spark 2.2，请问这种情况要怎么处理

旧版本中会有这个问题。目前在 3.11 以及以后的版本中已经解决了这个问题，Zeppelin 升级到了 0.73。

Q：之前购买了一套 E-MapReduce，版本是3.4.3的，现在要购买另一套，为什么选不到3.4.3版本了

E-MapReduce的主版本会定期更新升级，对于一些老版本会做下线处理。如果您需要已下线的E-MapReduce，首先请您关注现有E-MapReduce各版本中各个软件的版本，如Hive、Spark，看现有E-MapReduce主版本是否满足您的需求。如果还是无法满足您的版本需求，可以通过提交工单联系工程师，帮您以开启白名单的方式使用老版本E-MapReduce。

Q：存储会自动负载均衡还是需要手动均衡？如果需要手动均衡在哪里操作

1. **需要手动均衡，在阿里云 *E-MapReduce* 控制台。**
2. **单击集群管理。**
3. **单击待操作集群所在行的详情。**
4. **单击集群服务 > HDFS。**
5. **单击右上角的操作 > Rebalance。**

Q：E-MapReduce 可以减配置（降低配置）吗，例如 Master 或者 Core、Task 节点由 16CPU/32G 内存减为 8CPU16G 内存

目前还不能支持。

Q： “The specified DataDisk Size beyond the permitted range, or the capacity of snapshot exceeds the size limit of the specified disk category”

通过 SDK/API 的方式创建的集群，磁盘参数设置的太小了，建议您调整到40G以上重试就可以了。

Q： “Your account does not have enough balance”

用户的账号余额不足。

Q： “The maximum number of Pay-As-You-Go instances is exceeded: create ecs vcpu quota per region limited by user quota [xxx]”

您的按量 ECS 使用量达到上限，需要到 ECS 申请提高上限，或者是释放部分按量 ECS 才能继续创建E-MapReduce。

Q：使用 Flume 推送数据到 Hadoop 配置的 OSS 中，但是发现E-MapReduce集群没有 Flume 程序，请问如何集成

目前E-MapReduce还不没有集成 Flume，可以先自己安装使用。

Q：Spark 提交任务是否可以使用 standlone模式

目前在E-MapReduce中默认使用 Spark on Yarn 模式，还不支持 standlone 模式。

Q：软件配置上没有自定义的选项

因为早期的版本还不支持该功能，需要后台来补充升级。如果您需要修改软件配置，且老版本还无法支持的情况下，可以采用如下的方式：

1. **登录到集群的 Master 节点上。**
2. **前往配置模板的目录。**

```
cd /var/lib/ecm-agent/cache/ecm/service/A
```

3. **找到对应服务的目录，例如这里是 Hue，那就进入到 Hue 的目录。**
4. **选择对应的版本，例如** */var/lib/ecm-agent/cache/ecm/service/HUE/4.1.0.1.3*。
5. **进入** */package/templates/* **目录会看到对应的配置文件。**
6. **添加您需要的配置就可以了，添加配置分为两种情况：**
   - **配置是新增，那么需要按照格式来添加，注意不要写错换行，空格等。**
   - **配置本身存在，直接打开注释，或者是修改参数值即可。**
7. **修改完成以后，在E-MapReduce控制台界面上，选择重启服务生效。**
8. **最后可以确认，在实际的配置目录** */etc/ecm/ ### -conf/* **对应的文件内容已经被修改。**

4 作业异常诊断

Q: Spark 作业报错 "Container killed by YARN for exceeding memory limits." 或者 MapReduce 作业报错 "Container is running beyond physical memory limits."

> **A：App 提交时申请的内存量较低，但 JVM 启动占用了更多的内存，超过了自身的申请量，导致被 NodeManager异常终止；特别是 Spark 类型作业，可能会占用较多的堆外内存，很容易被异常终止。对于 Spark 作业，尝试提高** `spark.yarn.driver.memoryOverhead` **或** `spark.yarn.executor.memoryOverhead`**；对于MapReduce 作业，尝试提高** `mapreduce.map.memory.mb` **和** `mapreduce.reduce.memory.mb`。

Q: "Error: Java heap space"

> **A: 作业 Task 处理的数据量较大，但同时 Task JVM 申请的内存量不足，JVM 内存不足从而抛出 OutOfMemoryError。对于 Tez 作业，尝试提高 Hive 参数** `hive.tez.java.opts`。**对于 Spark 作业，尝试提高** `spark.executor.memory` **或** `spark.driver.memory`。**对于 MapReduce 作业，尝试提高** `mapreduce.map.java.opts`**或**`mapreduce.reduce.java.opts`。

Q: "No space left on device"

> **A: Master 或 Worker 节点空间不足，导致作业失败。同时磁盘空间满也会导致本地 Hive 元数据库（MySQL Server）异常，Hive Metastore 连接报错。建议清理 Master 节点磁盘空间，特别是系统盘的空间，清理 HDFS 空间。**

Q: 访问 OSS 或 LogService 报错 ConnectTimeoutException 或者 ConnectionException

> **A: OSS endpoint 配置为公网地址，但 EMR worker 节点并无公网 IP，所以无法访问。一个典型的场景是 Hive SQL：** `select * from tbl limit 10` **可以正常运行，但是**`Hive SQL: select count(1) from tbl` **报错。**
>
> **将OSS endpoint 地址修改为内网地址，比如** *oss-cn-hangzhou-internal.aliyuncs.com* **等。或者使用 EMR metaservice 功能，不用指定 endpoint。**

```
alter table tbl set location "oss://bucket.oss-cn-hangzhou-internal.aliyuncs.com/xxx"
```

```
alter table tbl partition (pt = 'xxxx-xx-xx') set location "oss://
bucket.oss-cn-hangzhou-internal.aliyuncs.com/xxx"
```

Q: 读取 Snappy 文件时报错 OutOfMemoryError

A: LogService 等服务写入的标准 Snappy 文件和 Hadoop 的 Snappy 文件格式不同，EMR 默认处理的是 Hadoop 修改过的 Snappy 格式，处理标准格式时会抛出 OutOfMemoryError。对于 Hive 配置`set io.compression.codec.snappy.native=true`。对于 MapReduce 配置`-Dio.compression.codec.snappy.native=true`。对于 Spark 作业配置`spark.hadoop.io.compression.codec.snappy.native=true`。

Q: 访问 RDS 时报错 Invalid authorization specification, message from server: “ip not in whitelist or in blacklist, client ip is xxx”

A: EMR 集群访问 RDS 需要设置白名单，如果没有配置集群各节点的 IP 到白名单，特别是扩容后忘记添加新增加节点的 IP 到白名单，会出现访问 RDS 出错。

Q："Exception in thread main java.lang.RuntimeException: java.lang.ClassNotFoundException: Class com.aliyun.fs.oss.nat.NativeOssFileSystem not found”

A：在 Spark 作业中读写 OSS 数据时，需要将 E-MapReduce 提供的 SDK 打进作业 jar 包中，详情请参见准备工作。

Q：Spark 接 Flume 时出现内存超用问题

A：检查是否是以 Push-based 方式接收数据，若不是，请尝试改成 Push-based 方式接收数据。详情请参见这里。

Q："Caused by: java.io.IOException: Input stream cannot be reset as 5242880 bytes have been written, exceeding the available buffer size of 524288"

A：（OSS）网络连接重试时缓存不足的 BUG，请使用 1.1.0 版本以上的 emr-sdk。

Q："Failed to access metastore. This class should not accessed in runtime.org.apache.hadoop.hive.ql.metadata.HiveException: java.lang.RuntimeException: Unable to instantiate org.apache.hadoop.hive.ql.metadata.SessionHiveMetaStoreClient"

A：首先，Spark 处理 Hive 数据需要作业的执行模式为 yarn-client（local 也行），不能为 yarn-cluster，否则会报上述异常。其次，作业 jar 中引入一些第三方的包也有可能导致 Spark 运行期间报上述异常。

Q：Spark 程序中使用 OSS SDK 出现 "java.lang.NoSuchMethodError:org.apache.http.conn.ssl.SSLConnetionSocketFactory.init(Ljavax/net/ssl/SSLContext;Ljavax/net/ssl/HostnameVerifier)"

A：OSS SDK依赖的 http-core 和 http-client 包与 Spark 和 Hadoop 的运行环境存在版本依赖冲突，不建议在代码中使用 OSS SDK，否则需要手动解决依赖冲突问题，比较麻烦。如果有需要对 OSS 中的文件做一些基础操作例如 list 等等，可以参见这里的用法进行操作。

Q："java.lang.IllegalArgumentException: Wrong FS: oss://xxxxx, expected: hdfs://ip:9000"

A： 因为在操作 OSS 数据的时候，使用 HDFS 的默认的 fs，所以在初始化的时候，要使用 OSS 的路径来初始化fs，这样在后续的操作中，才能使用这个 fs 来操作 OSS 源上的数据。

```
Path outputPath = new Path(EMapReduceOSSUtil.buildOSSCompleteUri("oss
://bucket/path", conf));        org.apache.hadoop.fs.FileSystem fs =
org.apache.hadoop.fs.FileSystem.get(outputPath.toUri(), conf);
if (fs.exists(outputPath)) {
  fs.delete(outputPath, true);
}
```

Q: 作业长时间 GC 导致作业运行较慢

A: 作业的 JVM heap size 设置过小，可能会引起长时间的 GC，影响作业性能。建议提升 `Java Heap Size`，对于 Tez，尝试提高 Hive 参数 `hive.tez.java.opts`。对于 Spark，尝试提高 `spark.executor.memory` 或 `spark.driver.memory`。对于 MapReduce，尝试提高 `mapreduce.map.java.opts`或`mapreduce.reduce.java.opts`。

Q: AppMaster 调度启动 Task 的时间过长

A: 作业 Task 数目过多（或 Spark Executor 数目过多），AppMaster 调度启动 Task 的时间过长，单个 Task 运行时间较短，作业调度的 overhead 较大。建议减少 Task 数目，使用 CombinedInputFormat。或者提高前序作业产出数据的 block size（`dfs.blocksize`）。或者提高 `mapreduce.input.fileinputformat.split.maxsize`。对于 Spark 作业，减少 executor 数目（`spark.executor.instances`）或者降低并发数（`spark.default.parallelism`）。

Q: 资源申请时间较长，导致作业等待

A: 作业提交之后，AppMaster 需要申请资源启动 Task，在这个过程中如果集群比较繁忙，资源申请时间较长，导致作业等待。建议检查资源组配置是否合理，当前资源组是否繁忙。如果集群整体资源还有富余，可以适当提高关键资源组的资源占比，或者扩容集群。

Q: 一小部分 Task 执行时间超长，作业整体运行时间变长（数据倾斜）

A: 作业的某个 stage 中 Task 数据分布不均衡，导致大部分 Task 很快执行完成，一小部分 Task 因为数据量过大，执行时间超长，作业整体运行时间变长。建议使用 Hive 的 mapjoin 并设置`set hive.optimize.skewjoin = true`。

Q: 失败的 Task attempt 导致作业运行时间变长

A: 一个作业有失败的 Task attempt 或失败的 Job attempt，虽然作业可能正常结束，但失败的 attempt 延长了作业运行时间。建议寻找 Task 失败的原因，做针对性的优化，减少作业运行时间。失败原因可以参考本页其他常见问题。

Q: Spark作业报错 "java.lang.IllegalArgumentException: Size exceeds Integer.MAX_VALUE"

A: 在 shuffle 时，partition 数量过少使得一个 block size 过大，超过最大值 Integer.MAX_VALUE(2 G)。您可以尝试增大 partition 数目，增大 `spark.default.parallelism` **和** `spark.sql.shuffle.partitions`**，或者在 shuffle 前执行repartition。**

5 执行计划使用

本节介绍执行计划相关的使用及其常见问题。

高配机器申请

如果您尚未申请开通高配，那么在使用高配机型创建集群的时候会失败，并在状态右侧的红色提示上出现类似如下的错误。

这个时候您需要提交工单提交工单，开通高配机型。

安全组使用

目前 E-MapReduce 创建集群的时候需要使用在 E-MapReduce 中创建的安全组。主要是因为 E-MapReduce 创建的集群只开放了 22 端口。我们推荐的做法是，用户的现有 ECS 实例也按照功能划分，位于不同的用户安全组中。例如，E-MapReduce 的安全组为“EMR-安全组”，而用户已有的安全组为“用户-安全组”,每个安全组按照不同的需要开启不同的访问控制。 如果需要和已有的集群进行联动请参考如下的做法。

- E-MapReduce 集群加入现有安全组

  在集群上单击详情，会展示集群所有 ECS 实例所在的安全组。前往 ECS 的管理控制台，单击左下方的安全组 页签，找到列表中对应的安全组，如上面提到的“EMR-安全组”。单击该安全组操作中的管理实例，会看到很多 emr-xxx 开头名字的 ECS 实例，这些就是对应的 E-MapReduce 集群中的 ECS。全选这些实例并单击右上的移入安全组选择想要加入的新的安全组即可。
- 现有集群加入 E-MapReduce 安全组

  和上面的操作一样，先找到现有集群的所在安全组，重复如上的操作，移入 E-MapReduce 的安全组即可。 如果是一些零散的机器，也可以直接在 ECS 的控制台界面选择机器然后通过下方的批量操作，移入 E-MapReduce 的安全组。
- 安全组的规则

  一个 ECS 实例在多个不同的安全组的时候，安全组的规则是或的关系。 例如，E-MapReduce 的安全组只开放了 22 端口。“用户-安全组”开放了所有的端口的。那么当 E-MapReduce 的集群加入了“用户-安全组”以后，E-MapReduce 中的机器也会开放所有的端口。所以在使用上请特别注意。

  注意:
  设置安全组规则时，一定要限制访问 IP 范围，不要设置 0.0.0.0，避免被攻击。

执行计划使用常见问题

- **执行计划的编辑。**

  执行计划只有在非运行且非调度中的状态下，才可以被编辑。如果编辑按钮灰色，请确认以上状态。

- **执行计划的执行。**

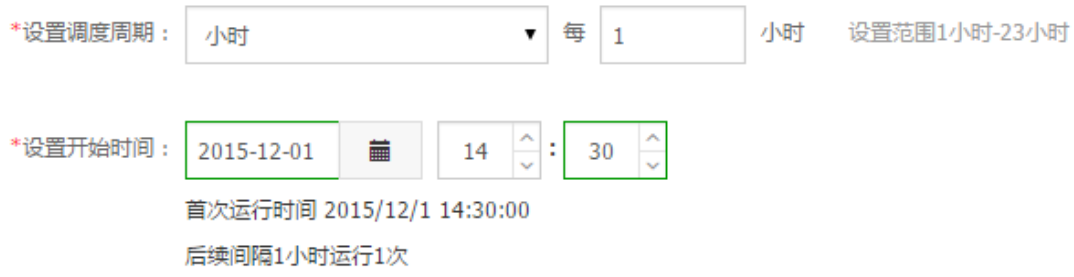
  新创建的执行计划，如果选择的是立即执行，那么在创建完成后，会自动执行。而如果是一个已经存在的执行计划，在创建完成以后并不会立即执行，需要手动执行。

- **周期执行时间。**

  周期执行集群的开始时间表示这个执行计划开始执行的时间，具体到分钟。而调度周期表明从这个开始时间开始之后每一次执行的间隔。

如上图所示。 第一次运行是 2015/12/1 14:30:00，然后第二次执行是 2015/12/2 14:30:00，每一天执行一次。

如果当前时间已经大于这个时间了，那么符合条件的最近一次时间就是第一次的运行时间。



如果现在是 2015/12/2 号的上午 9:30，那么按照调度规则，最近的一次调度是在 2015/12/2 10:00:00。那么首次调度就会在这个时间启动。

6 RDS数据导入时，时间字段显示延时8小时

这个问题发生在当您使用Sqoop将RDS数据导入EMR的时候。

问题症状

当您遇到如下场景的时候：

- **您有一个云数据库RDS数据源，数据表名字为**Test_Table**，表中包含时间戳（timestamp）字段，结构和数据如下。**

- **您通过以下脚本将**Test_Table**里的数据导入到HDFS。**

```
sqoop import \
--connect jdbc:mysql://rm-2ze****341.mysql.rds.aliyuncs.com:3306/s
***o_sqoopp_db \
--username s***o \
--password ****** \
--table play_evolutions \
--target-dir /user/hadoop/output \
--delete-target-dir \
--direct \
--split-by id \
--fields-terminated-by '|' \
-m 1
```

- **您在阿里云E-MapReduce上查询导入结果。**

查询结果显示，源数据的时间字段显示延迟8小时。

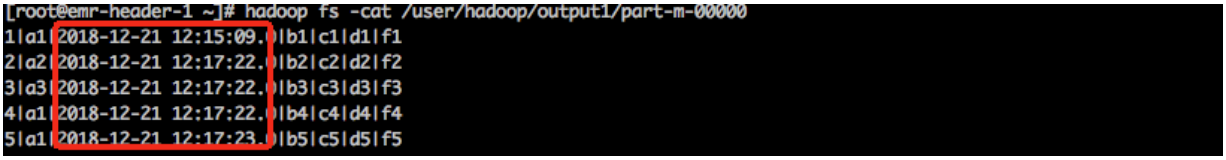
解决办法

经过大量脚本测试，在使用timestamp字段时，导入命令中去掉--direct**参数执行。**

```
sqoop import \
--connect jdbc:mysql://rm-2ze****341.mysql.rds.aliyuncs.com:3306/s***
o_sqoopp_db \
--username s***o \
--password ****** \
--table play_evolutions \
--target-dir /user/hadoop/output \
--delete-target-dir \
--split-by id \
--fields-terminated-by '|' \
```

```
-m 1
```

查询结果，这时结果是正常的。

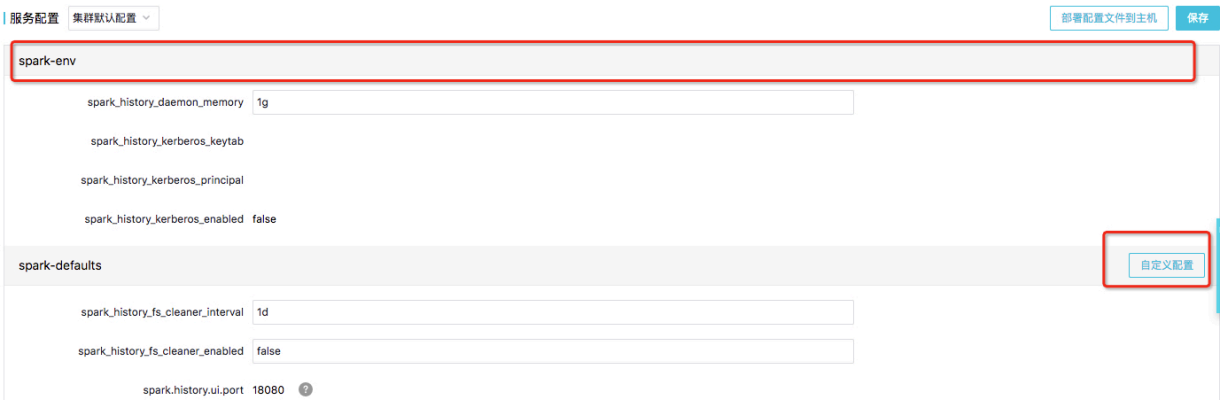


7 如何修改Spark服务的spark-env配置

本文档介绍如何在E-MapReduce的Spark集群中修改spark-env配置。

问题症状

传统Spark集群配置中的*spark-env.sh*配置文件中提供了服务的一些环境变量配置，例如，配置pyspark的Python运行时。当您使用EMR的Spark集群时，在控制台上只能支持修改*spark-defaults*，目前还不支持修改*spark-env*配置。

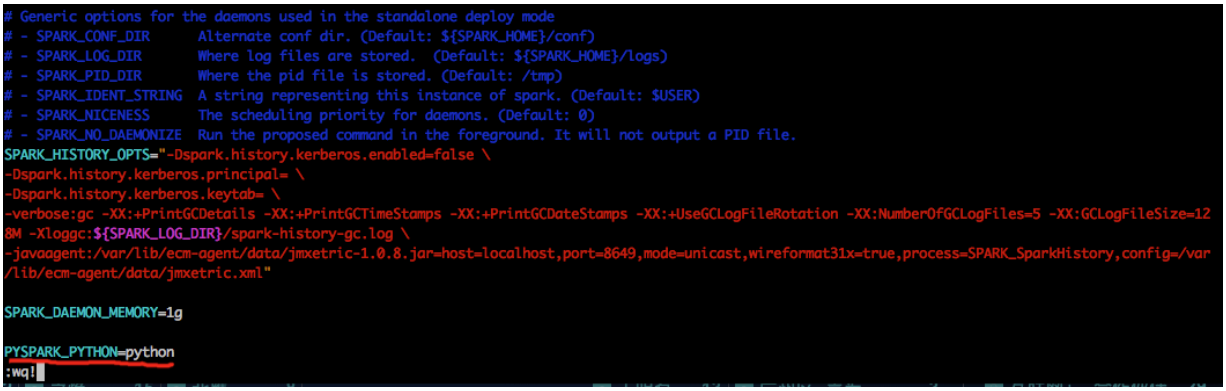


解决办法

解决这个问题，您需要登录到header节点，同时修改*/etc/ecm/spark-conf/spark-env.sh*及*/var/lib/ecm-agent/cache/ecm/service/SPARK/<###>/package/templates/spark-env.sh*两个文件中的配置。

说明:

- **如果您在worker节点提交任务，则需要同步修改worker节点相关配置。**
- ***/etc/ecm/<##>-conf/*是标准的配置文件，但若只更改该目录下的配置，在服务重启后，修改的配置会被还原导致不生效。因此，要同时修改对应服务的模板中配置文件，目录为：*/var/lib/ecm-agent/cache/ecm/service/<###>/<###>/package/templates/*。同时修改两个文件，可以保证手动修改的配置文件不会因服务重启而还原配置。**


# Generic options for the daemons used in the standalone deploy mode
# - SPARK_CONF_DIR Alternate conf dir. (Default: ${SPARK_HOME}/conf)
# - SPARK_LOG_DIR Where log files are stored. (Default: ${SPARK_HOME}/logs)
# - SPARK_PID_DIR Where the pid file is stored. (Default: /tmp)
# - SPARK_IDENT_STRING A string representing this instance of spark. (Default: $USER)
# - SPARK_NICENESS The scheduling priority for daemons. (Default: 0)
# - SPARK_NO_DAEMONIZE Run the proposed command in the foreground. It will not output a PID file.
SPARK_HISTORY_OPTS="-Dspark.history.kerberos.enabled=false \
-Dspark.history.kerberos.principal= \
-Dspark.history.kerberos.keytab= \
-verbose:gc -XX:+PrintGCDetails -XX:+PrintGCTimeStamps -XX:+PrintGCDateStamps -XX:+UseGCLogFileRotation -XX:NumberOfGCLogFiles=5 -XX:GCLogFileSize=12
8M -Xloggc:${SPARK_LOG_DIR}/spark-history-gc.log \
-javaagent:/var/lib/ecm-agent/data/jmxetric-1.0.8.jar=host=localhost,port=8649,mode=unicast,wireformat31x=true,process=SPARK_SparkHistory,config=/var
/lib/ecm-agent/data/jmxetric.xml"
SPARK_DAEMON_MEMORY=1g
PYSPARK_PYTHON=python
:wq!

8 如何将HiveServer2的认证方式设置为LDAP

本节介绍如何将HiveServer2的认证方式设置为LDAP。

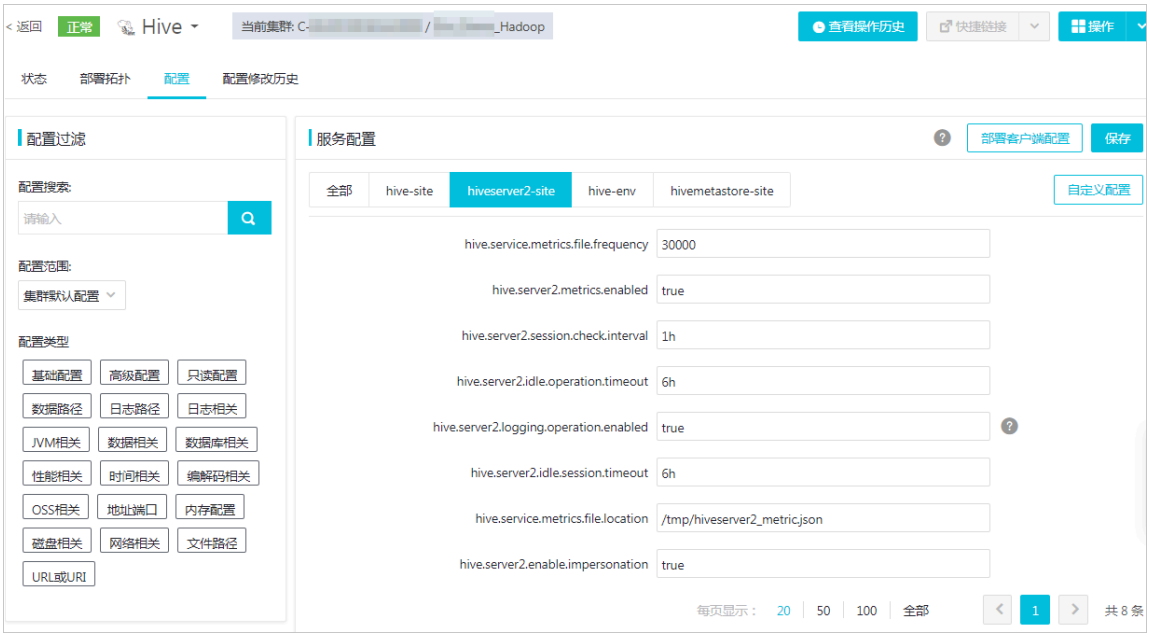
问题

如何将HiveServer2的认证方式设置为LDAP?

解决方法

在E-MapRduce集群中，HiveServer2支持多种认证方式，包括NOSASL、None、LDAP、Kerberos、PAM和Custom，这些认证方式均可通过hive.server2.authentication参数来设置。

1. **登录阿里云 *E-MapReduce* 控制台。**
2. **单击上方的集群管理页签。**
3. **单击待配置集群所在行的详情。**
4. **单击左侧导航栏的集群服务 > Hive。**
5. **新增LDAP认证配置项并重启HiveServer2。**
   a. **进入Hive组件的配置 > hiveserver2-site页面。**



   b. **单击自定义配置，新增配置项。**

LDAP认证方式需要新增如下三个配置项。

配置项	值	说明
hive.server2.authentication	LDAP	认证方式。

配置项	值	说明
hive.server2.authentication.ldap.url	格式为ldap://${emr-header-1-hostname}:10389	${emr-header-1-hostname}需要以实际主机名称为准，您可在集群的emr-header-1上执行hostname命令获取（#unique_20）。
hive.server2.authentication.ldap.baseDN	ou=people,dc=emr	-

c. 完成上述参数配置后，单击右上方的保存。

d. 在弹出的对话框中填写变更描述，然后单击确定，系统提示操作成功。

e. 在右上方单击操作 > 重启HiveServer2，重启HiveServer2。

6. 在LDAP中添加账号。

   在E-MapReduce集群中，OpenLDAP组件是一个LDAP的服务，默认用于管理Knox的用户账号，HiveServer2的LDAP认证方式可以复用Knox的账号体系。添加账号的方法请参见用户管理。本例新增账号为emr-test。

7. 测试新增账号是否可正常登录HiveServer2。

   通过*/usr/lib/hive-current/bin/beeline*登录HiveServer2，正常登录情况如下。

```
beeline> !connect jdbc:hive2://emr-header-1:10000/
Enter username for jdbc:hive2://emr-header-1:10000/: emr-guest
Enter password for jdbc:hive2://emr-header-1:10000/: emr-guest-pwd
Transaction isolation: TRANSACTION_REPEATABLE_READ
```

   如果账号密码不正确，则会显示如下异常。

```
Error: Could not open client transport with JDBC Uri: jdbc:hive2://
emr-header-1:10000/: Peer indicated failure: Error validating the
login (state=08S01,code=0)
```

9 如何使用阿里云E-MapReduce HDFS的Balancer功能

本节介绍如何调优阿里云E-MapReduce HDFS的Balancer参数。

问题

如何使用E-MapReduce HDFS的Balancer功能以及参数调优?

解决方法

HDFS Balancer主要用来防止节点数据分布过度不均衡，平摊DataNode压力，同时计算也可以利用HDFS的Locality特性达到更高的性能，尤其在节点扩容之后，新增了大量的空节点，及时进行平衡操作，可以更好地发挥Hadoop集群性能。

1. **登录到待配置集群的任意节点。**
2. **执行以下操作，切换到hdfs用户并执行Balancer参数。**

```
su hdfs
/usr/lib/hadoop-current/sbin/start-balancer.sh -threshold 10
```

3. **执行以下操作，查看balancer否在运行。**

```
less /var/log/hadoop-hdfs/hadoop-hdfs-balancer-emr-header-xx.cluster
-xxx.log
```

或者

```
tailf /var/log/hadoop-hdfs/hadoop-hdfs-balancer-emr-header-xx.
cluster-xxx.log
```

说明:

当提示信息包含Successfully字样时，表示执行成功。

Balancer的主要参数：

参数	说明
Threshold	默认值为10%，意思是上下浮动10%，即相差20%认为是平均的。 当集群总使用率较高时，需要调小Threshold，避免阈值过高。如果总体利用率不高，较低的阈值会产生更多调整，最终效果也不一定有特别明显的提升。当集群新增较多节点时，可以适当增加Threshold，让数据从高使用率节点移向低使用率的节点，随着Balancer的运行，需要逐步减少，保证数据移出节点的数量是足够的，有足够的并发。
dfs.datanode.balance.max.concurrent.moves	默认值为5。 表示一个DataNode节点并发移动的最大个数，一般考虑和磁盘数匹配，推荐在DataNode端设置为（4 * 磁盘数）作为上限，可用Balancer的值进行调节。 例如：一个DataNode有28块盘，在balancer端设置为28，DataNode端设置为28*4。具体使用时根据集群负载，观察相关指标，进行适当调整。在负载较低时，增加concurrent数，在负载较高时，减少concurrent数。这个设置在Datanode上和Balancer上都有，如果不一致时以较小的值生效。 说明: DataNode端需要重启来刷新配置。

参数	说明
dfs.balancer.dispatcherThreads	Balancer会在移动Block之前，每次迭代会查询出一个Block列表，分发给Mover线程使用。  说明: dispatcherThreads是这个分发线程的个数，默认为200。
dfs.balancer.rpc.per.sec	默认值为20，即每秒发送的rpc数量为20。 因为分发线程会调用大量的getBlocks的rpc查询，为了避免分发线程对NameNode造成过大压力，针对分发线程rpc发送速度进行控制，可在负载高的集群进行调整，例如减小10或者5，对整体移动进度不会产生特别大的影响。
dfs.balancer.getBlocks.size	Balancer会在移动Block之前，每次迭代时查询出一个Block列表，给Mover线程使用，默认Block列表中Block的大小为2GB。因为getBlocks这个过程会对RPC进行加锁，可以根据NameNode压力进行调整。
dfs.balancer.moverThreads	默认值为1000。 表示Balancer处理移动Block的线程数，每个Block移动时会使用一个线程。
dfs.namenode.balancer.request.standby	默认值为false。 表示Balancer是否在 standby NameNode上查询要移动的Blocks，因为此类查询会对NameNode加锁，导致写文件时间较长，HA集群开启后只会在 Standby NameNode上进行查询。

参数	说明
dfs.balancer.getBlocks.min-block-size	Balancer查询需要移动的参数时，对于较小 block（默认10MB），移动起来效率较低，通过此参数过滤掉较小的数据块，增加查询效率。
dfs.balancer.max-iteration-time	默认值为1200000，单位毫秒。 Balancer 在一次迭代的最长时间，超过后将进入下一次迭代。
dfs.balancer.block-move.timeout	默认值为0，默认为不开启，单位毫秒。 Balancer在移动 Block 时，会出现由于个别数据块没有完成而导致迭代较长情况，您可以通过此参数对移动长尾进行控制。

DataNode的主要参数：

参数	说明
dfs.datanode.balance.bandwidthPerSec	默认值为1MB/s。 表示DataNode用于Balancer的带宽，一般推荐使用100MB/s 以上进行快速平衡，也可以观察一下集群负载，适当进行调整，可通过设置dfsadmin -setBalancerBandwidth 参数，不需要重启DataNode。 例如白天负载很低，就增加Balancer的带宽，在负载高的时候减少一些。
dfs.datanode.balance.max.concurrent.moves	表示DataNode上同时用于Balancer待移动block的最大线程个数。