MySQL安装

- 卸载centos7自带的mariadb

```
#查看是否有进程
rpm -qa| grep mariadb
#如果有卸载
rpm -e 查看到的进程名字 --nodeps
#再次查看
rpm -qa| grep mariadb
```

- 安装mysql

```
#下载mysql
wget https://dev.mysql.com/get/Downloads/MySQL-8.0/mysql-8.0.27-1.el8.x86_64.rpm-bundle.tar
#创建文件夹
mkdir /usr/local/mysql
#进入创建的文件夹 把安装包解压
cd /usr/local/mysql
tar -zxvf 压缩包路径 解压路径
#然后安装
 rpm -ivh mysql-community-common-8.0.27-1.el8.x86_64.rpm --nodeps --force
 rpm -ivh mysql-community-libs-8.0.27-1.el8.x86_64.rpm --nodeps --force
 rpm -ivh mysql-community-client-8.0.27-1.el8.x86_64.rpm --nodeps --force
 rpm -ivh mysql-community-server-8.0.27-1.el8.x86_64.rpm --nodeps --force
 #初始化MySQL
 mysqld --initialize;
 #更改所属组
 chown mysql:mysql /var/lib/mysql -R;
 #启动服务
 systemctl start mysqld.service;
 #设置自动启动
 systemctl enable mysqld;
 #会临时生成一个密码,如果没看到则可以通过
 cat /var/log/mysqld.log | grep password
 #登录MySQL
 mysql -uroot -p
 #修改密码,此时密码是root可以改
  ALTER USER 'root'@'localhost' IDENTIFIED WITH mysql_native_password BY '123456';
  #退出然后重新登录
  #进行远程访问的授权
```

```
create user 'root'@'%' identified with mysql_native_password by 'root';
grant all privileges on *.* to 'root'@'%' with grant option;
flush privileges;
#修改加密规则，MySql8.0 版本 和 5.0 的加密规则不一样，而现在的可视化工具只支持旧的加密方式。
ALTER USER 'root'@'localhost' IDENTIFIED BY 'root' PASSWORD EXPIRE NEVER;
flush privileges;
#安装完成

#修改密码5.0版本
 SET PASSWORD FOR 'root'@'%' = PASSWORD('123456');
```

Hive安装

- 上传安装包解压
- 解决Hive与Hadoop之间huava版本差异

```
cd /export/server/apache-hive-3.1.2-bin/
rm -rf lib/guava-19.0.jar
cp /export/server/hadoop-3.3.0/share/hadoop/common/lib/guava-27.0-jre.jar
/export/server/apache-hive-3.1.2-bin/lib/
```

修改配置文件

- hive-env.sh

```
cd /export/server/apache-hive-3.1.2-bin/conf
mv hive-env.sh.template hive-env.sh

vim hive-env.sh
export HADOOP_HOME=/export/server/hadoop-3.3.0
export HIVE_CONF_DIR=/export/server/apache-hive-3.1.2-bin/conf
export HIVE_AUX_JARS_PATH=/export/server/apache-hive-3.1.2-bin/lib
```

- hive-site.xml

```
vim hive-site.xml
<configuration>
<!-- 存储元数据mysql相关配置 -->
<property>
    <name>javax.jdo.option.ConnectionURL</name>
    <value>jdbc:mysql://node1:3306/hive3?
createDatabaseIfNotExist=true&amp;useSSL=false</value>
```

```
</property>

<property>
    <name>javax.jdo.option.ConnectionDriverName</name>
    <value>com.mysql.jdbc.Driver</value>
</property>

<property>
    <name>javax.jdo.option.ConnectionUserName</name>
    <value>root</value>
</property>

<property>
    <name>javax.jdo.option.ConnectionPassword</name>
    <value>hadoop</value>
</property>

<!-- H2S运行绑定host -->
<property>
    <name>hive.server2.thrift.bind.host</name>
    <value>node1</value>
</property>

<!-- 远程模式部署metastore metastore地址 -->
<property>
    <name>hive.metastore.uris</name>
    <value>thrift://node1:9083</value>
</property>

<!-- 关闭元数据存储授权  -->
<property>
    <name>hive.metastore.event.db.notification.api.auth</name>
    <value>false</value>
</property>
</configuration>
```

- 上传mysql jbdc驱动到hive安装lib下

```
根据MySQL版本驱动版本放入
```

- 在hdfs创建hive存储目录

```
hadoop fs -mkdir /tmp
hadoop fs -mkdir -p /user/hive/warehouse
hadoop fs -chmod g+w /tmp
hadoop fs -chmod g+w /user/hive/warehouse
```

- **初始化元数据**

```
cd /export/server/apache-hive-3.1.2-bin/

bin/schematool -initSchema -dbType mysql -verbos
#手动创建74张表
```

启动hive

- 1.启动metastore服务

```
#三种方式,一般用第三种
#前台启动  关闭ctrl+c
/export/server/apache-hive-3.1.2-bin/bin/hive --service metastore

#前台启动开启debug日志
/export/server/apache-hive-3.1.2-bin/bin/hive --service metastore --hiveconf hive.root.logger=DEBUG,console

#后台启动 进程挂起  关闭使用jps+ kill -9
nohup /export/server/apache-hive-3.1.2-bin/bin/hive --service metastore &
```

- 2.启动hiveserver2服务

```
  nohup /export/server/apache-hive-3.1.2-bin/bin/hive --service hiveserver2 &

  #注意 启动hiveserver2需要一定的时间  不要启动之后立即beeline连接 可能连接不上
```

- 3.beeline客户端连接
  - 拷贝node1安装包到beeline客户端机器上

```
#如果在本机安装的则忽略
scp -r /export/server/apache-hive-3.1.2-bin/ node3:/export/server/
```

Hive的客户端

- Hive的第一代客户端

- ==bin/hive==
- 直接访问metastore服务
- 配置

```
<configuration>
<property>
        <name>hive.metastore.uris</name>
        <value>thrift://node1:9083</value>
</property>
</configuration>
```

- 弊端:
  - 第一代客户端属于shell脚本客户端 性能友好安全方面存在不足 Hive已经不推荐使用官方建议使用第二代客户端beeline

- Hive的第二代客户端
  - bin/beeline
  - 无法访问metastore服务，只能够访问==Hiveserver2服务==。
  - 使用

```
# 拷贝node1上 hive安装包到beeline客户端机器上（node3）
scp -r /export/server/apache-hive-3.1.2-bin/ node3:/export/server/

#1、在安装hive的服务器上 首先启动metastore服务 再启动hiveserver2服务
nohup /export/server/apache-hive-3.1.2-bin/bin/hive --service metastore &
nohup /export/server/apache-hive-3.1.2-bin/bin/hive --service hiveserver2 &

#2、在任意机器(如node3)上使用beeline客户端访问
[root@node3 ~]# /export/server/apache-hive-3.1.2-bin/bin/beeline
beeline> ! connect jdbc:hive2://node1:10000    #jdbc访问HS2服务
Connecting to jdbc:hive2://node1:10000
Enter username for jdbc:hive2://node1:10000: root  #用户名 要求具备HDFS读写权限
Enter password for jdbc:hive2://node1:10000:       #密码可以没有
```

- 错误

```
Error: Could not open client transport with JDBC Uri: jdbc:hive2://node1:10000:
Failed to open new session: java.lang.RuntimeException:
org.apache.hadoop.ipc.RemoteException(org.apache.hadoop.security.authorize.AuthorizationException):
User: root is not allowed to impersonate root (state=08S01,code=0)
```

- 修改

```
在hadoop的配置文件core-site.xml中添加如下属性:
<property>
        <name>hadoop.proxyuser.root.hosts</name>
      <value>*</value>
</property>
<property>
        <name>hadoop.proxyuser.root.groups</name>
        <value>*</value>
</property>
```

```

```

- 错误解决:Hive3执行insert插入操作 statstask异常
  - 现象

```
在执行insert + values操作的时候  虽然最终执行成功，结果正确。但是在执行日志中会出现如下的错误信息。
```

```
Moving data to directory hdfs://node1:8020/user/hive/warehouse/t_1/.hive-staging_hive_2020-11-09_14-09-06_874204851885-1/-ext-10000
Loading data to table default.t_1
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.StatsTask
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 5.54 sec   HDFS Read: 15118 HDFS Write: 237 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 540 msec
hive> select count(*) from t_1;
```

  - 开启hiveserver2执行日志。查看详细信息

```
2020-11-09 00:37:48,963 WARN  [5ce14c58-6b36-476a-bab8-89cba7dd1706 main]
metastore.RetryingMetaStoreClient: MetaStoreClient lost connection. Attempting to reconnect (1 of 1)
after 1s. setPartitionColumnStatistics

ERROR [5ce14c58-6b36-476a-bab8-89cba7dd1706 main] exec.StatsTask: Failed to run stats task
```

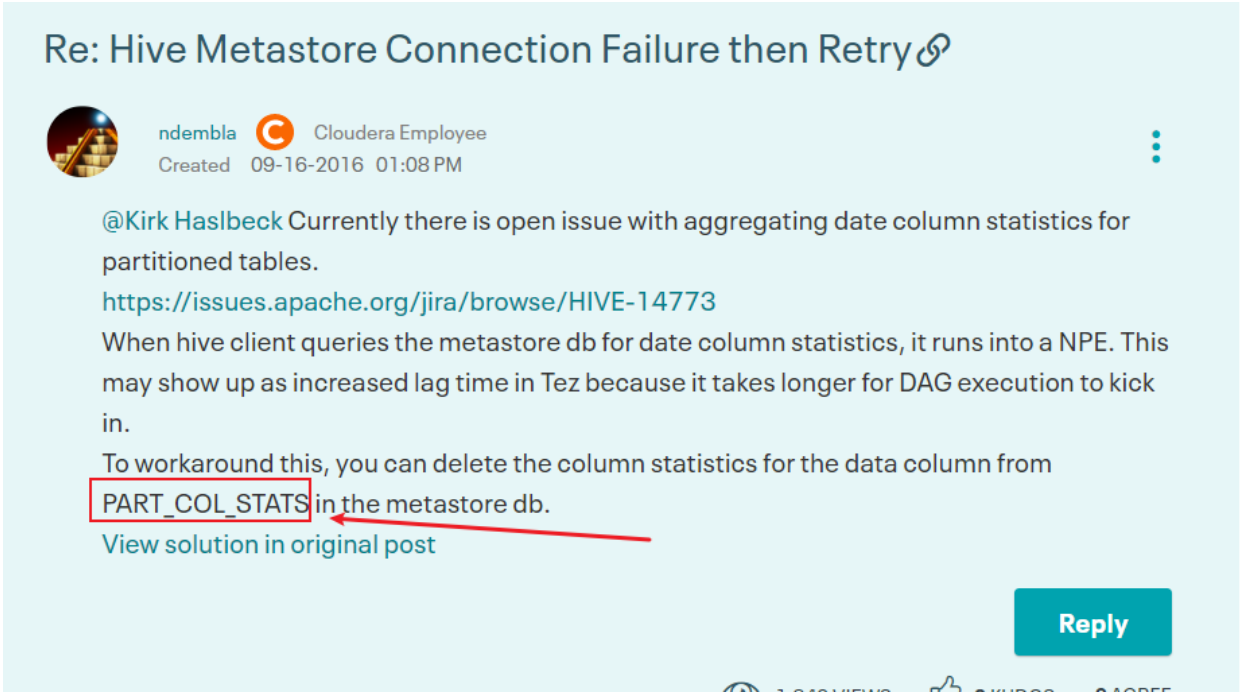
```
7 INFO  [5ce14c58-6b36-476a-bab8-89cba7dd1706 main] metastore.HiveMetaStoreClient: Opened a connection to metasto
8 INFO  [5ce14c58-6b36-476a-bab8-89cba7dd1706 main] metastore.HiveMetaStoreClient: Connected to metastore.
0 ERROR [5ce14c58-6b36-476a-bab8-89cba7dd1706 main] exec.StatsTask: Failed to run stats task
.ql.metadata.HiveException: org.apache.thrift.transport.TTransportException
hadoop.hive.ql.metadata.Hive.setPartitionColumnStatistics(Hive.java:4423)
hadoop.hive.ql.stats.ColStatsProcessor.persistColumnStats(ColStatsProcessor.java:179)
hadoop.hive.ql.stats.ColStatsProcessor.process(ColStatsProcessor.java:83)
```

- 是 ==此错误并不影响最终的插入语句执行成功==。
- 分析原因和解决

- statstask是一个hive中用于统计插入等操作的状态任务 其返回结果如下

```
INFO  : Stage-3 is filtered out by condition resolver.
INFO  : Stage-5 is filtered out by condition resolver.
INFO  : Starting task [Stage-4:MOVE] in serial mode
INFO  : Moving data to: hdfs://node-1:8020/user/hive/warehouse/itcast.db/t_mysql/.hive-staging_hi
_1345143957608079693-1/-ext-10000 from hdfs://node-1:8020/user/hive/warehouse/itcast.db/t_mysql/.
09_15-02-16_954_1345143957608079693-1/-ext-10002
INFO  : Starting task [Stage-0:MOVE] in serial mode
INFO  : Loading data to table itcast.t_mysql from hdfs://node-1:8020/user/hive/warehouse/itcast.d
ve_2020-08-09_15-02-16_954_1345143957608079693-1/-ext-10000
INFO  : Starting task [Stage-2:STATS] in serial mode
INFO  : Table itcast.t_mysql stats: [numFiles=1, numRows=1, totalSize=11, rawDataSize=10]
INFO  : MapReduce Jobs Launched:
INFO  : Stage-Stage-1: Map: 1   Cumulative CPU: 3.62 sec   HDFS Read: 3782 HDFS Write: 81 SUCCESS
INFO  : Total MapReduce CPU Time Spent: 3 seconds 620 msec
INFO  : Completed executing command(queryId=root_20200809150202_c24eb5f6-baa7-4a60-82cc-95256f7e4
econds
INFO  : OK
No rows affected (21.368 seconds)
```

  - 此信息类似于计数器 用于告知用户插入数据的相关信息 但是不影响程序的正常执行。
- Hive新版本中 这是一个issues 临时解决方式如下
- https://community.cloudera.com/t5/Support-Questions/Hive-Metastore-Connection-Failure-then-Retry/td-p/151661
-



  - ==在mysql metastore中删除 PART_COL_STATS这张表即可==。
- Hive 连接 schemaTool -initSchema 报 mysql 连接 1045 错误

```
# 更新一下所有域访问的 root对应的密码
use mysql;
UPDATE user SET Password = PASSWORD('123456') where USER='root';
FLUSH PRIVILEGES;
exit;
```

hive连接不上查看日志可能是安全模式，关闭即可

```
hdfs dfsadmin -safemode forceExit
```

hive连接不上查看日志提示server.HiveServer2: Error starting HiveServer2 on attempt 23, will retry in 60000ms

查看日志其中显示

2022-02-09T09:27:03,331 INFO [main] server.HiveServer2: HS2 interactive HA not enabled. Starting tez sessions..

2022-02-09T09:27:03,331 INFO [main] server.HiveServer2: Starting/Reconnecting tez sessions..

显示HS2 interactive HA没有开启，就导致了开启tez session。那么我们进行如下配置开启HS2 interactive HA：

在hive中的修改hive-site.xml,将下面value改为true

```
<property>
    <name>hive.server2.active.passive.ha.enable</name>
    <value>true</value>
    <description>Whether HiveServer2 Active/Passive High Availability be enabled when Hive Interactive sessions are enabled.This will also require hive.server2.support.dynamic.service.discovery to be enabled.</description>
  </property>
```