

Sqoop

安装部署

sqoop相关配置

测试Sqoop

从mariadb导入到HDFS

从HDFS导出到mariadb

从mariadb导入HBase

从HBase导出到mariadb

从mariadb导入Hive

从Hive到导出到mariadb

Sqoop

安装部署

[sqoop --> 1.4.7](#)

hadoop --> 2.7.5

默认系统已经安装了mysql或者mariadb

将下载好的sqoop的二进制包解压到/usr/local下然后进行配置

sqoop相关配置

- 配置环境变量

```
# vim /etc/profile
export SQOOP_HOME=/usr/local
export PATH=$SQOOP_HOME/bin:$PATH
:wq
# source /etc/profile
```

- 配置sqoop相关参数

```
# cp conf/sqoop-env-template.sh conf/sqoop-env.sh
# vim conf/sqoop-env.sh
# Hadoop
export HADOOP_PREFIX=/usr/local/hadoop
export HADOOP_HOME=${HADOOP_PREFIX}
export PATH=$PATH:$HADOOP_PREFIX/bin:$HADOOP_PREFIX/sbin
export HADOOP_COMMON_HOME=${HADOOP_PREFIX}
export HADOOP_HDFS_HOME=${HADOOP_PREFIX}
export HADOOP_MAPRED_HOME=${HADOOP_PREFIX}
```

```
export HADOOP_YARN_HOME=${HADOOP_PREFIX}
# Native Path
export HADOOP_COMMON_LIB_NATIVE_DIR=${HADOOP_PREFIX}/lib/native
export HADOOP_OPTS="-Djava.library.path=$HADOOP_PREFIX/lib/native"
# Hadoop end

#Hive
export HIVE_HOME=/usr/local/hive
export PATH=$HIVE_HOME/bin:$PATH

#HBase
export HBASE_HOME=/usr/local/hbase
export PATH=$HBASE/bin:$PATH
```

- 下载mysql驱动

在conf/lib目录下

```
# wget http://central.maven.org/maven2/mysql/mysql-connector-java/5.1.38/mysql-connector-java-5.1.38.jar
```

测试Sqoop

从mariadb导入到HDFS

```
# sqoop import \
--connect jdbc:mysql://localhost/test \
--username root \
--password 111111 \
--table table1 \
-m 1 \
--target-dir /test
# hadoop fs -ls /test
# hadoop fs -cat /test/part*
```

命令解析：

- | | |
|---------------------------------------|-------------------------------------|
| • sqoop | # sqoop命令 |
| • import | # 表示导入 |
| • --connect jdbc:mysql://ip:port/test | # 告诉jdbc,连接mysql的url |
| • --username root | # 连接mysql/mariadb的用户名 |
| • --password 111111 | # 连接mysql/mariadb的密码 |
| • --table table1 | # 从数据库中导出的表的名字 |
| • --query | # 查询语句 |
| • --columns | # 指定列 |
| • --where | # 条件 |
| • --hive-import | # 指定为hive导入 |
| • --hive-table | # 指定hive表, 可以使用--target-dir /test替换 |
| • --fields-terminated-by '\t' | # 指定输出文件中的行的字段分隔符 |
| • -m 1 | # 复制整个过程使用一个map作业 |

从HDFS导出到mariadb

```
# sqoop export \  
--connect jdbc:mysql://localhost:3306/test \  
--driver com.mysql.jdbc.Driver \  
--username root \  
--password 111111 \  
--table table1 \  
--export-dir /test
```

从mariadb导入HBase

```
# hbase shell  
# create 'test', 'column'  
# sqoop import --connect jdbc:mysql://localhost:3306/test --username root --password 111111 --  
table table1 --hbase-table test --hbase-row-key id --column-family data --hbase-create-table --m  
1
```

其中--hbase-row-key的id是RDBMS中的一个列，作为行键，RDBMS的表中有逐渐，那么主键就为HBase的行键。

从HBase导出到mariadb

从mariadb导入Hive

在从RDBMS导入HIVE中的时候，出现如下错误：

```
SLF4J: Found binding in [jar:file:/usr/local/hadoop-2.7.5/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]  
SLF4J: Found binding in [jar:file:/usr/local/hbase/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]  
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.  
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]  
18/06/29 02:09:17 INFO hive.HiveImport: Loading uploaded data into Hive  
18/06/29 02:09:17 ERROR hive.HiveConfig: Could not load org.apache.hadoop.hive.conf.HiveConf. Make sure HIVE_CONF_DIR is set correctly.  
18/06/29 02:09:17 ERROR tool.CreateHiveTableTool: Encountered IOException running create table job: java.io.IOException: java.lang.ClassNotFoundException: org.apache  
hadoop.hive.conf.HiveConf  
    at org.apache.sqoop.hive.HiveConfig.getHiveConf(HiveConfig.java:50)  
    at org.apache.sqoop.hive.HiveImport.getHiveArgs(HiveImport.java:392)  
    at org.apache.sqoop.hive.HiveImport.executeExternalHiveScript(HiveImport.java:379)  
    at org.apache.sqoop.hive.HiveImport.executeScript(HiveImport.java:337)  
    at org.apache.sqoop.hive.HiveImport.importTable(HiveImport.java:241)  
    at org.apache.sqoop.tool.CreateHiveTableTool.run(CreateHiveTableTool.java:57)  
    at org.apache.sqoop.Sqoop.run(Sqoop.java:147)  
    at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:70)  
    at org.apache.sqoop.Sqoop.runSqoop(Sqoop.java:183)  
    at org.apache.sqoop.Sqoop.runTool(Sqoop.java:234)  
    at org.apache.sqoop.Sqoop.runTool(Sqoop.java:243)  
    at org.apache.sqoop.Sqoop.main(Sqoop.java:252)  
Caused by: java.lang.ClassNotFoundException: org.apache.hadoop.hive.conf.HiveConf  
    at java.net.URLClassLoader.findClass(URLClassLoader.java:381)  
    at java.lang.ClassLoader.loadClass(ClassLoader.java:424)  
    at sun.misc.Launcher$AppClassLoader.loadClass(Launcher.java:338)  
    at java.lang.ClassLoader.loadClass(ClassLoader.java:357)  
    at java.lang.Class.forName0(Native Method)  
    at java.lang.Class.forName(Class.java:264)  
    at org.apache.sqoop.hive.HiveConfig.getHiveConf(HiveConfig.java:44)  
    ... 11 more
```

只要sqoop找到hive-exec.jar即可，所以我们在配置了HIVE_CONF_DIR之后还要执行以下命令：

```
# ln -s $HIVE_HOME/lib/hive-exec.jar $SQOOP_HOME/lib/hive-exec.jar
```

以上在Hive的lib中如果jar包带有版本号，可根据情况进行链接创建。

```
# sqoop import --hive-import --connect jdbc:mysql://localhost/test --username root --password 111111 --table table1 --m 1 --delete-target-dir --verbose
```

从Hive到导出到mariadb

