**安装及配置Sqoop**

# 一、实验目的

* 熟悉Sqoop
* 掌握Sqoop的基本配置

# 二、实验内容

本实验主要是为了在hadoop-master节点上安装并配置Sqoop。Sqoop是一个用来将Hadoop和关系型数据库中的数据相互转移的工具，可以将一个关系型数据库（例如 ：MySQL 、Oracle 、Postgres等）中的数据导入到Hadoop的HDFS中，也可以将HDFS的数据导入到关系型数据库中。通过本实验，学生可以增加对Sqoop的了解，为后续的大数据处理做准备。具体来说，学生可以将MySQL数据转换为hadoop文件，从而可以使用map/reduce进行大数据的分析，然后再将分析结果转换回MySQL，供业务查询调用。

# 三、实验知识点

* Sqoop的安装和配置
* Sqoop的基本操作

# 四、实验环境

* Centos 7.9.2009 64位
* Sqoop 1.4.7-cdh6.3.2

# 五、实验步骤

## 1. 安装Sqoop

本实验利用老师给的sqoop-1.4.7-cdh6.3.2.tar.gz进行安装配置。

（1）安装Sqoop

***cd /home***

***mkdir sqoop***

然后，将sqoop-1.4.7-cdh6.3.2.tar.gz复制到/home/sqoop目录下。执行如下命令：

***tar -zxvf sqoop-1.4.7-cdh6.3.2.tar.gz***

这样Sqoop就安装完成了。

（2）下载mysql驱动程序

使用Sqoop还需要添加MySql的驱动程序到Sqoop的lib目录下。因此，需要下载mysql-connector-java-5.1.48-bin.jar文件，并将文件复制到/home/sqoop/ sqoop-1.4.7-cdh6.3.2/lib目录下。具体执行命令如下：

***cd ~***

***wget https://cdn.mysql.com/archives/mysql-connector-java-5.1/mysql-connector-java-5.1.48.tar.gz***

***tar zxvf mysql-connector-java-5.1.48.tar.gz***

***cd mysql-connector-java-5.1.48***

***cp mysql-connector-java-5.1.48-bin.jar /home/sqoop/sqoop-1.4.7-cdh6.3.2/lib***

***cd ~***

（3）下载java-json.jar库

使用Sqoop还需要java-json.jar库。因此，需要下载java-json.jar文件，并将文件复制到/home/sqoop/ sqoop-1.4.7-cdh6.3.2/lib目录下。具体执行命令如下：

***cd ~***

***wget http://www.java2s.com/Code/JarDownload/java-json/java-json.jar.zip***

***unzip java-json.jar.zip***

***cp java-json.jar /home/sqoop/sqoop-1.4.7-cdh6.3.2/lib/***

***cd ~***

（4）配置环境变量

使用命令***vim /etc/profile***，编辑/etc/profile，在文件的最后添加如下内容：

***export SQOOP\_HOME=/home/sqoop/sqoop-1.4.7-cdh6.3.2/***

***export PATH=$SQOOP\_HOME/bin:$PATH***

然后，使用***source /etc/profile***命令使环境变量的设置生效。

（5）修改sqoop-env.sh文件

首先，进入/home/sqoop/sqoop-1.4.7-cdh6.3.2/conf目录，根据模板创建sqoop-env.sh文件。具体执行如下命令：

***cd /home/sqoop/sqoop-1.4.7-cdh6.3.2/conf***

***cp sqoop-env-template.sh sqoop-env.sh***

然后，修改sqoop-env.sh文件，文件位于/home/sqoop/sqoop-1.4.7-cdh6.3.2/conf目录下，添加如下内容：

***export HADOOP\_COMMON\_HOME=/home/hadoop/hadoop-3.1.3***

***export HADOOP\_MAPRED\_HOME=/home/hadoop/hadoop-3.1.3***

***export HIVE\_HOME=/home/hive/apache-hive-3.1.2-bin***

至此，Sqoop的安装和配置完成。

## 2. 测试和使用Sqoop

### 2.1 测试准备

（1）创建sqooptest数据库

首先，登录MariaDB创建数据库“sqooptest”。

***mysql -u root -p***

***create database sqooptest default charset utf8mb4 collate utf8mb4\_unicode\_ci;***

然后，创建sqooptest数据库对应的用户，并授予访问权限。

***grant all on sqooptest.\* to 'sqoop'@'%' IDENTIFIED BY 'sqooppwd';***

***grant all on sqooptest.\* to 'sqoop'@'hadoop-master' IDENTIFIED BY 'sqooppwd';***

***grant all on sqooptest.\* to 'sqoop'@'localhost' IDENTIFIED BY 'sqooppwd';***

***flush privileges;***

***exit;***

（2）在sqooptest数据创建表格，并添加数据

先利用sqoop用户登录MariaDB：

***mysql -u sqoop -p***

然后执行如下SQL命令：

***show databases;***

***use sqooptest;***

***create table student(***

***id int primary key auto\_increment,***

***name varchar(50),***

***gender char(1),***

***age int,***

***birth datetime);***

***insert into student(name, gender, age, birth) values('zhangsan', '0', 25, '2018-01-01 12:12:12');***

***insert into student(name, gender, age, birth) values('lisi', '0', 2, '2019-01-01 12:12:12');***

***insert into student(name, gender, age, birth) values('wangwu', '0', 1, '2020-01-01 12:12:12');***

***exit;***

至此，测试准备工作结束。

### 2.2 测试Mysql转HDFS

执行如下命令，将mysql的数据存储到hdfs中：

***sqoop import --connect jdbc:mysql://hadoop-master:3306/sqooptest --username sqoop --password sqooppwd --table student -m 1 --target-dir /user/sqooptest/input/student/***

执行如下命令，查看hdfs中存储的数据：

***hadoop fs -cat /user/sqooptest/input/student/\****

输出的hdfs数据为如下内容则将Mysql数据导入HDFS正确：

***1,zhangsan,0,25,2018-01-01 12:12:12.0***

***2,lisi,0,2,2019-01-01 12:12:12.0***

***3,wangwu,0,1,2020-01-01 12:12:12.0***

### 2.3 测试HDFS转Mysql

（1）在HDFS中创建文件并存入内容

具体执行如下命令：

***hadoop fs -mkdir -p /user/sqooptest/output/***

***hadoop fs -touchz /user/sqooptest/output/student.txt***

***echo "null,liuyi,0,4,2016-01-01 12:12:12.0" | hadoop fs -appendToFile - /user/sqooptest/output/student.txt***

***echo "null,chener,0,3,2017-01-01 12:12:12.0" | hadoop fs -appendToFile - /user/sqooptest/output/student.txt***

***hadoop fs -cat /user/sqooptest/output/student.txt***

可以看到存入到hdfs student.txt的内容为：

***null,liuyi,0,4,2016-01-01 12:12:12.0***

***null,chener,0,3,2017-01-01 12:12:12.0***

（2）从hdfs导入数据到mysql

执行如下命令，将hdfs数据导出到mysql中：

***sqoop export --connect jdbc:mysql://hadoop-master:3306/sqooptest --username sqoop --password sqooppwd --table student -m 1 --export-dir /user/sqooptest/output/***

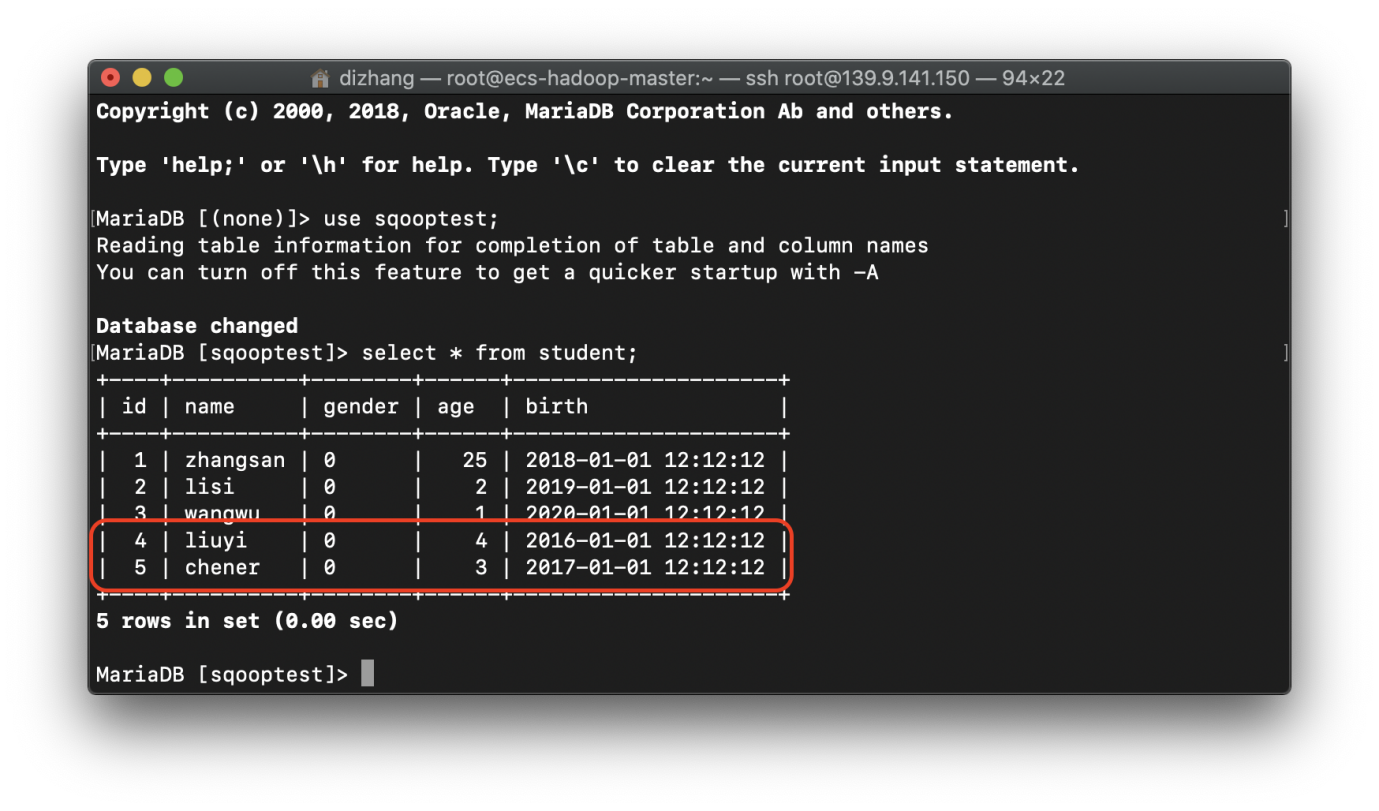
然后，查看导入是否成功，执行如下命令：

***mysql -u sqoop -p***

***use sqooptest;***

***select \* from student;***

如果成功，则可以看到student表中存储的数据如下图所示。



至此，Sqoop的安装配置和测试完成。

## 3. 使用技巧

从HDFS转换回mysql 的时候，自动增长的主键字段处，可以留空，也即null。具体上述每个命令中参数的含义可以参考官方Sqoop的文档。