**安装**

下载地址：   
<http://archive.cloudera.com/cdh5/cdh/5/sqoop-1.4.6-cdh5.5.2.tar.gz>   
下载 sqoop-1.4.6-cdh5.5.2.tar.gz

解压后命名为sqoop-1.4.6

修改环境变量：   
export SQOOP\_HOME=/home/xuyao/下载/sqoop-1.4.6   
export PATH= $PATH:$SQOOP\_HOME/bin   
export HIVE\_CONF\_DIR=/home/xuyao/下载/hive-2.0.1/conf   
export HADOOP\_CLASSPATH=$HADOOP\_CLASSPATH:$HIVE\_HOME/lib/\*

修改sqoop-env.sh：   
export HADOOP\_COMMON\_HOME=/home/xuyao/下载/hadoop-2.6.5   
export HADOOP\_MAPRED\_HOME=/home/xuyao/下载/hadoop-2.6.5   
export HIVE\_HOME=/home/xuyao/下载/hive-2.0.1

修改bin/configure-sqoop：注释掉HCAT\_HOME、ACCUMULO\_HOME、ZOOKEEPER\_HOME的检查。

## Moved to be a runtime check in sqoop.

#if [ ! -d "${HCAT\_HOME}" ]; then

# echo "Warning: $HCAT\_HOME does not exist! HCatalog jobs will fail."

# echo 'Please set $HCAT\_HOME to the root of your HCatalog installation.'

#fi

#if [ ! -d "${ACCUMULO\_HOME}" ]; then

# echo "Warning: $ACCUMULO\_HOME does not exist! Accumulo imports will fail."

# echo 'Please set $ACCUMULO\_HOME to the root of your Accumulo installation.'

#fi

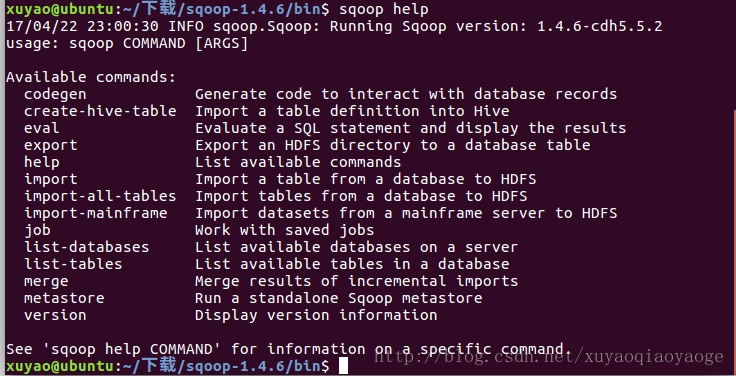
#if [ ! -d "${ZOOKEEPER\_HOME}" ]; then

# echo "Warning: $ZOOKEEPER\_HOME does not exist! Accumulo imports will fail."

# echo 'Please set $ZOOKEEPER\_HOME to the root of your Zookeeper installation.'

#fi

* 1
* 2
* 3
* 4
* 5
* 6
* 7
* 8
* 9
* 10
* 11
* 12
* 13
* 14

命令行输入sqoop help后：   


将mysql-connector-java-5.1.40-bin.jar复制到sqoop的lib下面。

安装mysql：   
sudo apt-get update   
sudo apt-get install mysql-server mysql-client   
在弹出的对话框中设置root的密码，就123好了。

sudo netstat -tap | grep mysql   
检查是否有端口在监听

mysql -u root -p   
输入密码123

create database mydatabase;

use mydatabase;

create table student

(

id int not null,

name varchar(50) not null,

age int not null,

primary key (id)

);

insert into student values(1,'zhangsan',20);

insert into student values(2,'lisi',21);

insert into student values(3,'wangwu',22);

select \* from student;

* 1
* 2
* 3
* 4
* 5
* 6
* 7
* 8
* 9
* 10
* 11
* 12
* 13
* 14

启动Hadoop，yarn

start-dfs.sh

start-yarn.sh

* 1
* 2

**导入**

将mysql导入HDFS中：

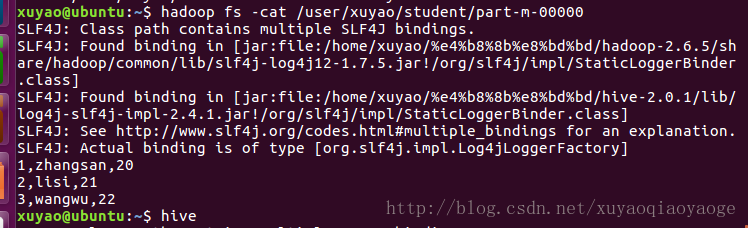
sqoop import --connect jdbc:mysql://localhost/mydatabase --username root --password 123 --table student -m 1

* 1

-m 表示启动N个map来并行导入数据，默认是4个，最好不要将数字设置为高于集群的节点数   
默认放在/user/用户名/   
查看：

hadoop fs -cat /user/xuyao/student/part-m-00000

* 1



在hive中建立表，注意用逗号：

Create Table student (id Int, name String, age Int) Row format delimited fields terminated By ',';

* 1

将HDFS中的这个表导入HIVE中：

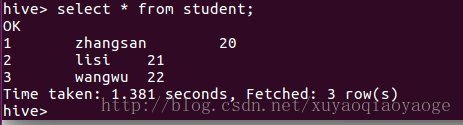
load data inpath '/user/xuyao/student' into table student;

* 1

查看：

select \* from student;

* 1



还有种非常方便的把上面的过程结合起来的，但是会出现BUG，目前还没解决：

sqoop import --connect jdbc:mysql://localhost/mydatabase --username root --password 123 --table student -m 1 --hive-import

* 1

**导出**

在mysql中建立：

Create Table dept (deptno Int,dname varchar(50));

* 1

导出：

sqoop export --connect jdbc:mysql://localhost/mydatabase --username root --password 123 --table dept --export-dir /xy/hive/warehouse/dept --input-fields-terminated-by '\t'

* 1

查看mysql：

select \* from dept;

* 1
* 2

