

# 第十二节、sqoop的安装与使用

## 第十二节、sqoop的安装与使用

### 一、功能概述

### 二、安装配置

1. 解压缩文件
2. 配置文件
3. 拷贝jar包
4. 环境变量

### 三、常用功能

1. 数据库配置-Oracle
2. 数据库配置-Mysql
  - ( 1 ) 界面操作
  - ( 2 ) 命令操作
3. 数据导入(使用-verbose查看更多信息)
4. 以文件的方式执行

## 一、功能概述

主要用于在Hadoop(Hive)与传统的数据库(mysql、oracle...)间进行数据的传递，可以将一个关系型数据库中的数据导入到HDFS ( Hive ) 中，也可以将HDFS的数据导进到关系型数据库中。

## 二、安装配置

### 1. 解压缩文件

```
1. tar -zxvf sqoop-1.4.6.bin__hadoop-2.0.4-alpha.tar.gz
```

### 2. 配置文件

- sqoop-env.sh(重命名conf下的sqoop-env-template.sh)

```
1. #Set path to where bin/hadoop is available
2. export HADOOP_COMMON_HOME=/home/hadoopadmin/hadoop-2.7.1
3.
4. #Set path to where hadoop-*-core.jar is available
5. export HADOOP_MAPRED_HOME=/home/hadoopadmin/hadoop-2.7.1
6.
7. #Set the path to where bin/hive is available
8. export HIVE_HOME=/home/hadoopadmin/apache-hive-1.2.1-bin
```

- configure-sqoop(bin下)

注释135-143行

```
1. ## Moved to be a runtime check in sqoop.
2. #if [ ! -d "${HCAT_HOME}" ]; then
3. #   echo "Warning: $HCAT_HOME does not exist! HCatalog jobs will fail."
4. #   echo 'Please set $HCAT_HOME to the root of your HCatalog installation.'
5. #fi
6.
7. #if [ ! -d "${ACCUMULO_HOME}" ]; then
8. #   echo "Warning: $ACCUMULO_HOME does not exist! Accumulo imports will fail."
9. #   echo 'Please set $ACCUMULO_HOME to the root of your Accumulo installation.'
10. #fi
```

### 3. 拷贝jar包

将hadoop-common-2.7.1.jar拷贝至sqoop的lib目录

### 4. 环境变量

```
1. export SQOOP_HOME=/home/hadoopadmin/sqoop-1.4.6.bin__hadoop-2.0.4-alpha
2. export PATH=$PATH:$SQOOP_HOME/bin
```

## 三、常用功能

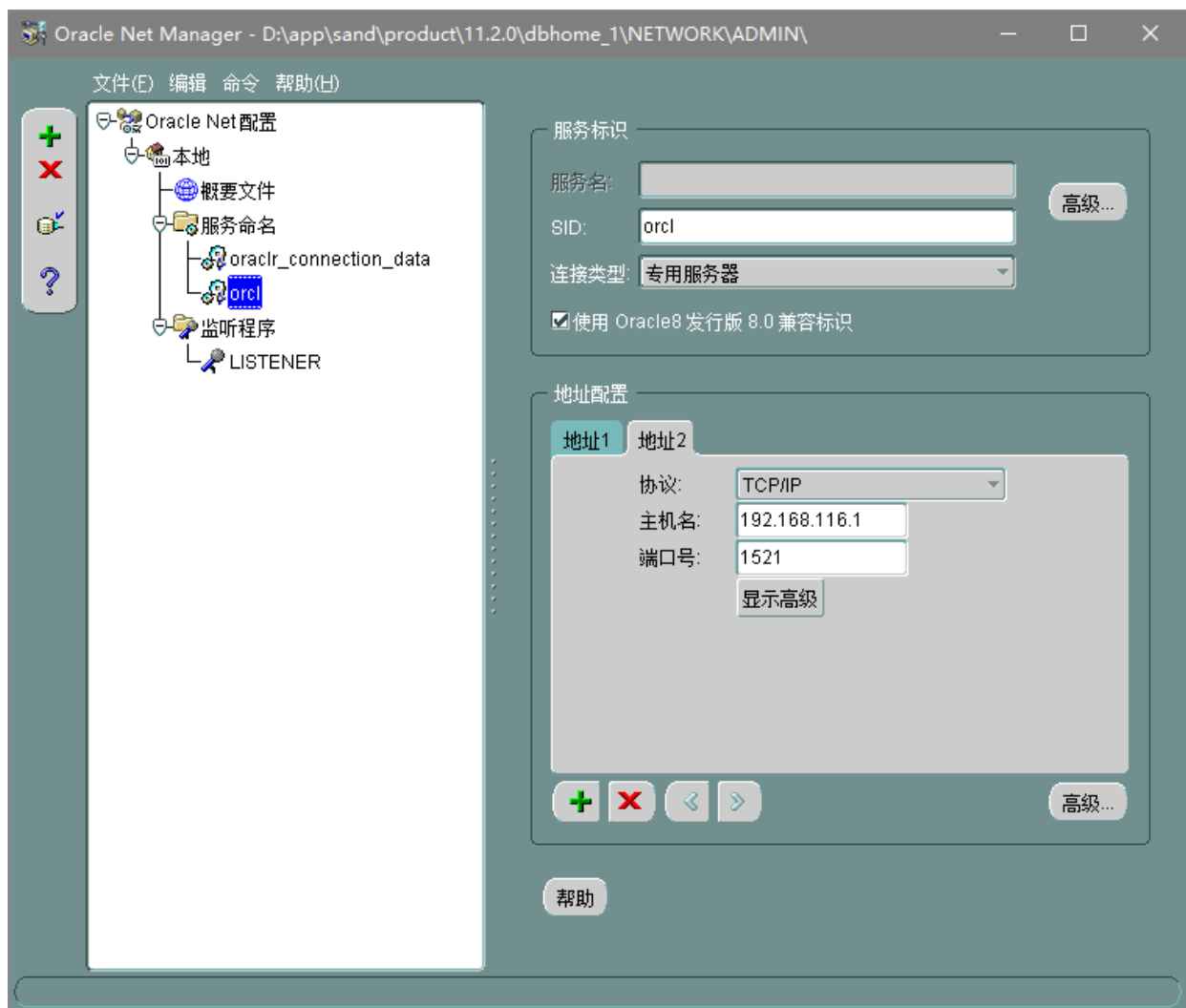
# 1. 数据库配置-Oracle

将Oracle数据库的驱动jar包拷贝至sqoop的lib目录下

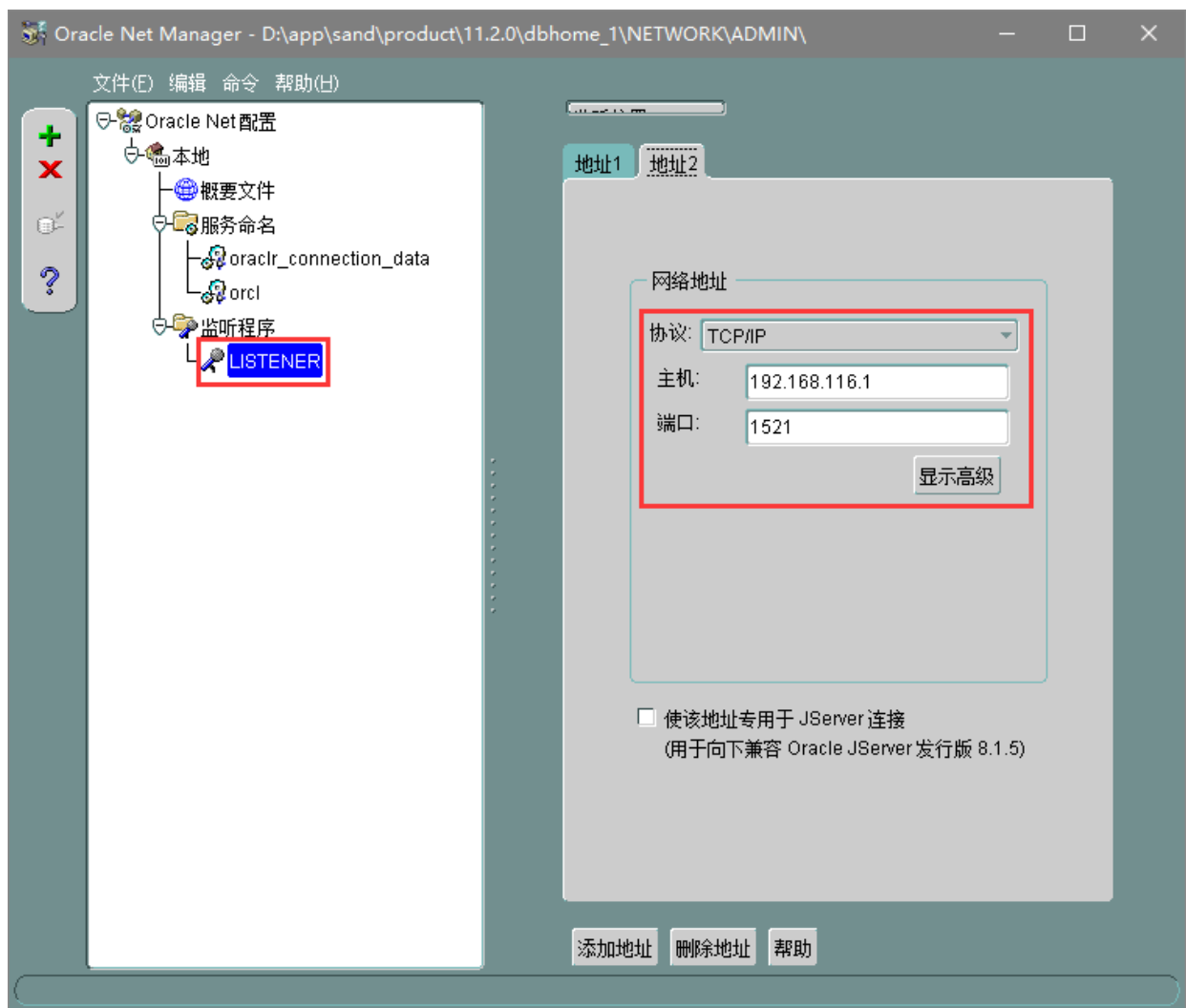
- Oracle Net Manager ( 配置完成后重启监听服务 )



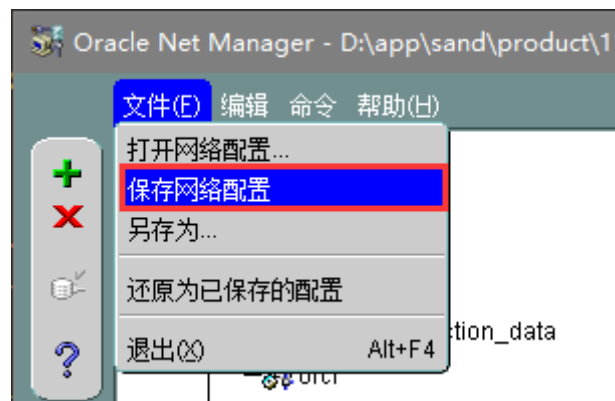
添加数据库服务地址



IP地址填写VMnet1网卡地址



监听器中添加相应的配置



保存配置

- 测试连接

connect后接连接数据库字符串，username后接连接数据库用户名，password后接连接数据库密码，list-databases展示数据库列表，list-table展示连接数据库下的所有表

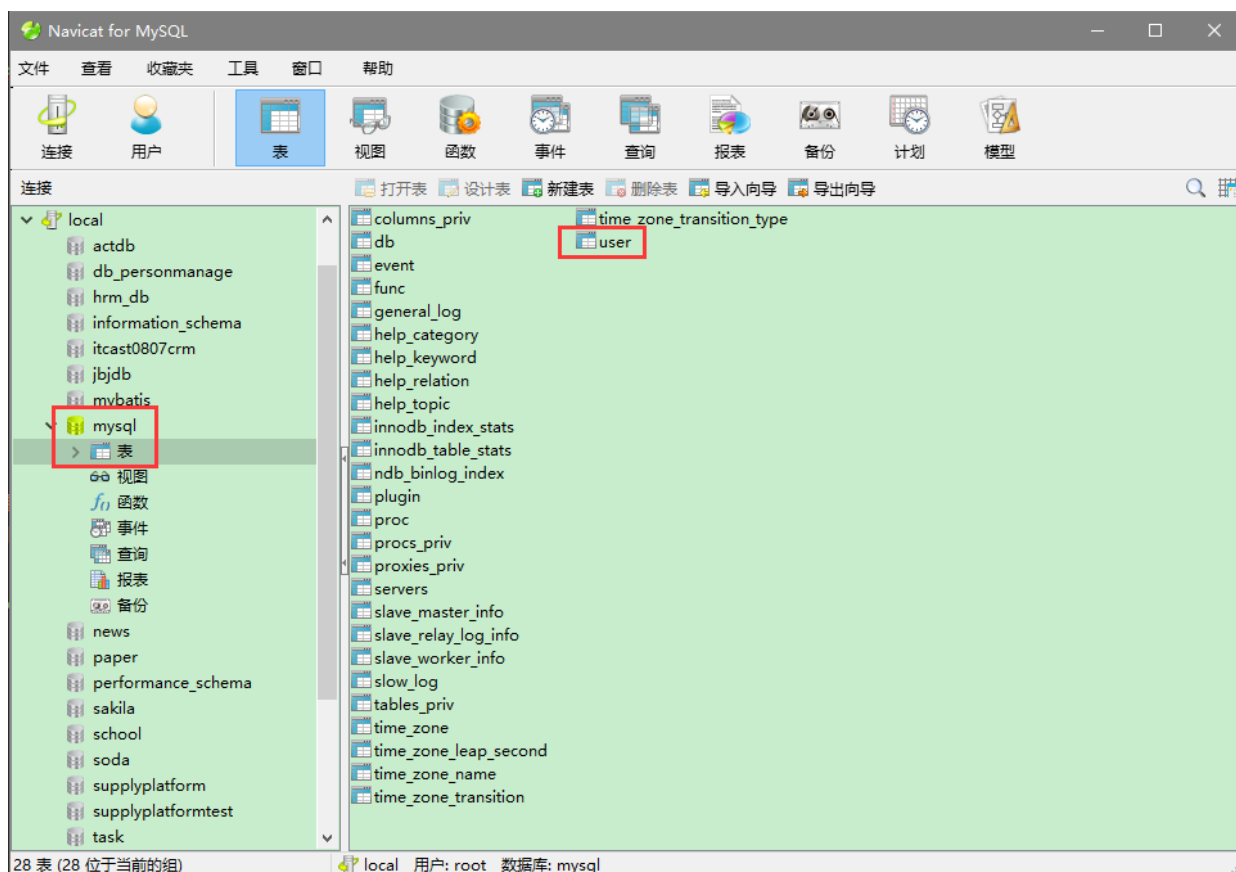
1. `sqoop list-databases --connect jdbc:oracle:thin:@host:1521:orcl --username sand --password sandTower`
2. `sqoop list-tables --connect jdbc:oracle:thin:@host:1521:orcl --username sand --password sandTower`

## 2. 数据库配置-Mysql

将Mysql数据库的驱动jar包拷贝至sqoop的lib目录下

- 登录Mysql数据库，开启远程访问

### (1) 界面操作

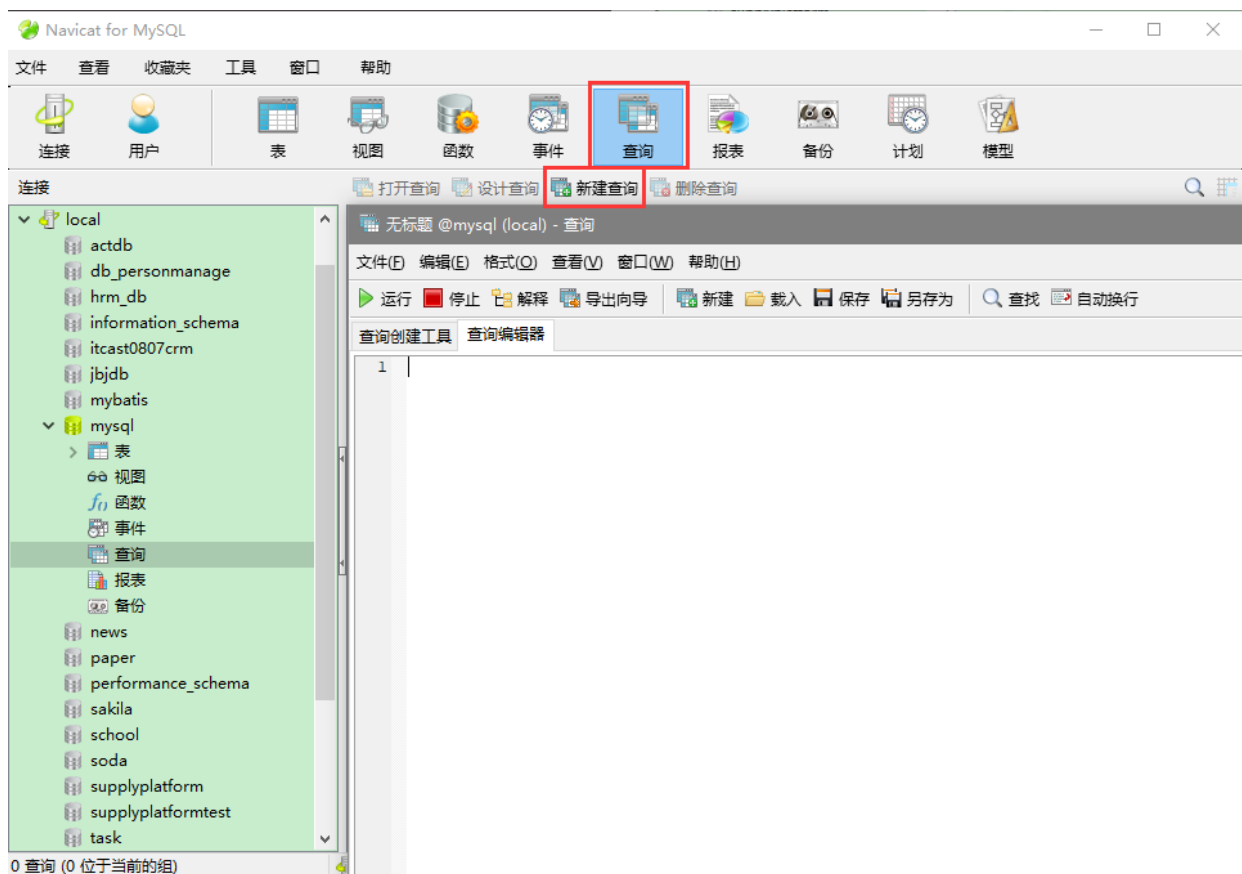


打开Mysql数据库的user表

The screenshot shows the 'user' table in the 'mysql' database. The table has columns: Host, User, Password, Select\_priv, Insert\_priv, Update\_priv, Delete\_priv, Create\_priv, Drop\_priv, and Reload\_priv. The data shows three users: 'localhost', '127.0.0.1', and '::1', all with 'root' as the user and a password of '\*E74858DB86EBA20BC33'. All users have 'Y' for all privilege columns.

Host	User	Password	Select_priv	Insert_priv	Update_priv	Delete_priv	Create_priv	Drop_priv	Reload_priv
localhost	root	*E74858DB86EBA20BC33	Y	Y	Y	Y	Y	Y	Y
127.0.0.1	root	*E74858DB86EBA20BC33	Y	Y	Y	Y	Y	Y	Y
::1	root	*E74858DB86EBA20BC33	Y	Y	Y	Y	Y	Y	Y

可查看当前的连接权限

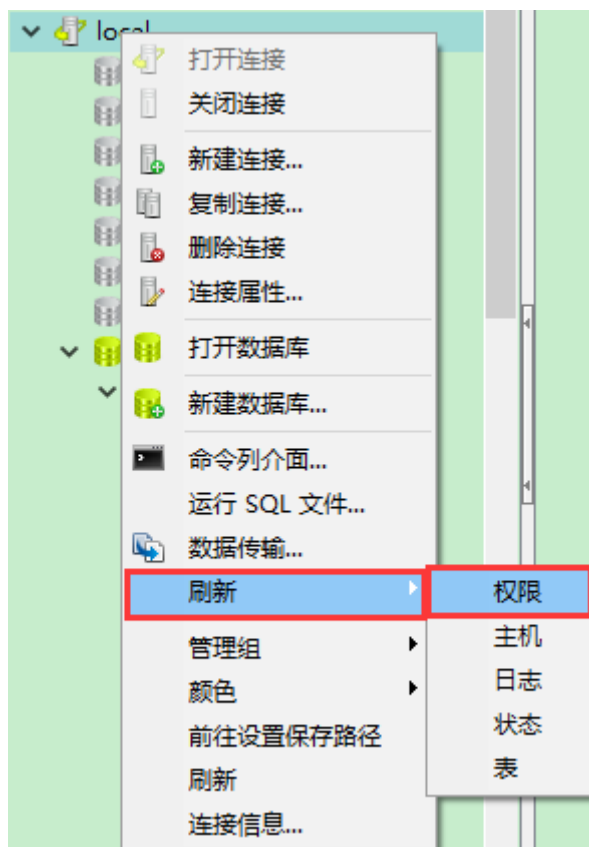


新建查询窗口

1. /\* 填写主机名，并在hosts文件中添加ip映射或直接使用ip地址 \*/
2. `GRANT ALL PRIVILEGES ON *.* TO 'root'@'etc01' IDENTIFIED BY 'mysql' WITH GRANT OPTION;`

etc01	root	*E7485DB86EBA20BC33	Y	Y	Y	Y	Y	Y	Y
-------	------	---------------------	---	---	---	---	---	---	---

user表中出现数据证明成功，password为加密后的密码



刷新权限

## (2) 命令操作

```
[hadoopadmin@etc01 ~]$ mysql -u root -p
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 467
Server version: 5.5.30 MySQL Community Server (GPL)

Copyright (c) 2000, 2013, Oracle and/or its affiliates. All rights reserved.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> use mysql;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Database changed
mysql> GRANT ALL PRIVILEGES ON *.* TO 'root'@'etc01' IDENTIFIED BY 'root' WITH GRANT OPTION;
```

登录数据库，开启远程访问

1. /\* 刷新权限 \*/
2. **FLUSH PRIVILEGES;**

- 测试连接

connect后接连接数据库字符串，username后接连接数据库用户名，password后接连接数据库密码，list-databases展示数据库列表，list-table展示连接数据库下的所有表



1. `sqoop list-databases --connect jdbc:mysql://host:3306 --username root --password mysql`
2. `sqoop list-tables --connect jdbc:mysql://host:3306/test --username root --password mysql`

### 3. 数据导入(使用-verbose查看更多信息)

- 导入到HDFS中

单表导入，可以指定某几列，或用where筛选某些数据(需要使用引号)，columns指定需要导入的列，table指定需要导入的表，delete-target-dir用于判断目标路径是否存在，存在则先删除，m用于指定最大mapTask个数

1. `sqoop import --connect jdbc:oracle:thin:@host:orcl --username sand --password sandTower --columns user_id,user_name --table user1 --where 'user_id > 5' --as-textfile --delete-target-dir --target-dir /input/sqoop/ -m 1`
2. `sqoop import --connect jdbc:mysql://host:3306/test --username root --password mysql --columns user_id,user_name --table user1 --where 'user_id > 5' --as-textfile --delete-target-dir --target-dir /input/sqoop/ -m 1`

- 导入到Hive表中

多表导入，使用query参数导入查询生成的结果集（多表，需要使用引号），如果有where条件需要添加and \$CONDITIONS

1. `sqoop import --connect jdbc:oracle:thin:@host:1521:orcl --username sand --password sandTower --query 'xxx' --hive-import --hive-database default --hive-table test --target-dir /input/sqoop/ -m 1`
2. `sqoop import --connect jdbc:mysql://host:3306 --username root --password root --query 'xxx' --hive-import --hive-database test --hive-table test --target-dir /input/sqoop/ -m 1`

### 4. 以文件的方式执行

可将参数部分写在文件中，通过--options-file命令进行调用