**Sqoop部署实验手册**

目录

1 实验介绍 4

1.1 实验内容 4

1.2 实验知识点 4

1.3 实验环境 4

1.4 实验资源 4

1.5 适合人群 4

1.6 实验步骤清单 5

2 实训架构 5

3 实验环境准备 5

4 实验步骤 5

4.1 Sqoop安装部署 5

4.1.1 安装Sqoop软件 5

4.1.2 修改Sqoop配置文件 6

4.2 Sqoop数据迁移 7

4.2.1 检查环境 7

4.2.2 数据迁移 8

5 实验总结 14

6 参考文档 14

# 实验介绍

## 实验内容

本次实验包括sqoop安装部署及利用Sqoop在mysql数据库与hive之间进行数据迁移。

## 实验知识点

* Sqoop安装
* mysql到Hive数据迁移

## 实验环境

* Sqoop-1.4.7
* 网易云平台

## 实验资源

|  |  |
| --- | --- |
| 资源名称 | 存储目录 |
| Sqoop安装包 | /opt/package/software |

## 适合人群

本课程难度为一般，属于初级课程，适合具有linux、mysql、hadoop及hive基础的用户。

## 实验步骤清单

* sqoop安装部署（安装sqoop、修改sqoop配置文件）
* sqoop数据迁移（检查环境、数据迁移）

# 实训架构

|  |  |  |  |
| --- | --- | --- | --- |
| **序号** | **机器名** | **类型** | **目录** |
| 1 | master | Sqoop | /usr/local |
| 2 | slave1 |  |
| 3 | slave2 |  |

# 实验环境准备

进入实验列表，找到sqoop，点击开启实验，启动master、slave1及salve2。

# 实验步骤

## Sqoop安装部署

### 安装Sqoop软件

#解压sqoop安装包

[root@master ~]# tar xzvf /opt/package/software/sqoop-1.4.7.bin\_\_hadoop-2.6.0.tar.gz -C /usr/local/

#重命名sqoop目录

[root@master ~]# mv /usr/local/sqoop-1.4.7.bin\_\_hadoop-2.6.0/ /usr/local/sqoop/

#添加环境变量

[root@master ~]# vi /etc/profile

export SQOOP\_HOME=/usr/local/sqoop

export PATH=${SQOOP\_HOME}/bin:$PATH

export HADOOP\_CLASSPATH=$HADOOP\_CLASSPATH:$HIVE\_HOME/lib/\*

#环境变量生效

[root@master ~]# source /etc/profile

### 修改Sqoop配置文件

#编辑sqoop-env.sh文件

[root@master ~]# cp /usr/local/sqoop/conf/sqoop-env-template.sh /usr/local/sqoop/conf/sqoop-env.sh

[root@master ~]# vi /usr/local/sqoop/conf/sqoop-env.sh

export HADOOP\_COMMON\_HOME=/usr/local/hadoop

export HADOOP\_MAPRED\_HOME=/usr/local/hadoop

export HIVE\_HOME=/usr/local/hive

export HIVE\_CONF\_DIR=/usr/local/hive/conf

#上传mysql-connector jar包到sqoop lib目录下

[root@master ~]# cp /opt/package/software/mysql-connector-java-5.1.46-bin.jar /usr/local/sqoop/lib/

## Sqoop数据迁移

### 检查环境

#查看Sqoop版本

[root@master ~]# sqoop version

Warning: /usr/local/sqoop/../hcatalog does not exist! HCatalog jobs will fail.

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

Warning: /usr/local/sqoop/../accumulo does not exist! Accumulo imports will fail.

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

18/04/18 12:07:32 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7

Sqoop 1.4.7

git commit id 2328971411f57f0cb683dfb79d19d4d19d185dd8

Compiled by maugli on Thu Dec 21 15:59:58 STD 2017

See 'sqoop help COMMAND' for information ona specific command.

#启动mysql

[root@master ~]# service mysql start

#查看mysql的表

[root@master ~]# sqoop list-tables --username root --password 'root' --connect jdbc:mysql://localhost:3306/metastore?characterEncoding=UTF-8

AUX\_TABLE

BUCKETING\_COLS

CDS

COLUMNS\_V2

COMPACTION\_QUEUE

COMPLETED\_COMPACTIONS

COMPLETED\_TXN\_COMPONENTS

DATABASE\_PARAMS

DBS

DB\_PRIVS

DELEGATION\_TOKENS

### 数据迁移

1、查看用于测试的MySQL数据库表及数据

#登录mysql

[root@master ~]# mysql -uroot -proot

#创建testdb数据库

mysql> create database testdb;

Query OK, 1 row affected (0.00 sec)

mysql>use testdb;

#创建表及插入数据脚本

mysql> create table t1 (id int(11),name varchar(10));

Query OK, 0 rows affected (0.30 sec)

mysql> insert into t1 values(10,'tom');

Query OK, 1 row affected (0.01 sec)

mysql> insert into t1 values(20,'jerry');

Query OK, 1 row affected (0.01 sec)

mysql> insert into t1 values(30,'rose');

Query OK, 1 row affected (0.01 sec)

mysql> commit;

Query OK, 0 rows affected (0.00 sec)

#查看表

mysql> show tables;

+------------------+

| Tables\_in\_testdb |

+------------------+

| t1 |

+------------------+

1 row in set (0.00 sec)

#查看表数据

mysql> select \* from t1;

+------+-------+

| id | name |

+------+-------+

| 10 | tom |

| 20 | jerry |

| 30 | rose |

+------+-------+

3 rows in set (0.01 sec)

#退出mysql客户端

mysql> exit;

Bye

2、基于MySQL的表t1创建Hive表t1

#查看mysql数据库

[root@master ~]# sqoop list-databases --connect jdbc:mysql://localhost:3306/ --username root --password root

information\_schema

metastore

mysql

performance\_schema

sys

testdb

#查看mysql数据库testdb中的表

[root@master ~]# sqoop list-tables --connect jdbc:mysql://localhost:3306/testdb --username root --password root

t1

#配置java安全策略文件

[root@master ~]# vi /usr/local/java/jre/lib/security/java.policy

添加以下权限，此权限表示在签名者或代码库中信任。

permission javax.management.MBeanTrustPermission "register";

#替换sqoop中的Jackson包

由于hive和sqoop jackson版本不一致，使用hive中的jackson包替换sqoop中的包。

[root@master ~]# rm -f /usr/local/sqoop/lib/jackson-\*

[root@master ~]# cp /usr/local/hive/lib/jackson-\* /usr/local/sqoop/lib/

#设置hosts

[root@master ~]# vi /etc/hosts

#hosts文件末尾加入如下内容(slave1ip及slave2ip是这两个容器的ip地址)

slave1ip slave1

slave2ip slave2

#让slave1及slave2的profile文件生效

[root@slave1 ~]# source /etc/profile

[root@slave2 ~]# source /etc/profile

#启动hadoop

[root@master ~]# start-all.sh

#创建hive表t1

在已经存在的testdb这个库中创建名为t1的表。

[root@master ~]# sqoop create-hive-table --connect jdbc:mysql://localhost:3306/testdb --table t1 --username root -password 'root' --hive-database testdb --hive-table t1

显示以下信息表示导入表完成。

22/05/30 07:45:27 INFO hive.HiveImport: Hive import complete.

#查看hive中的数据库和表

[root@master ~]# hive

#切换到testdb数据库

hive> use testdb;

OK

Time taken: 4.712 seconds

#查看表

hive> show tables;

OK

t1

Time taken: 0.17 seconds, Fetched: 1 row(s)

#查看表数据（此时数据为空）

hive> select \* from t1;

OK

Time taken: 1.955 seconds

#退出hive

hive> exit;

3、将MySQL中的数据导入到hive中

注意：hive是基于hadoop的HDFS的，在运行下面的导入命令请，请确保hadoop和hive都在正常运行，不需要事先在hive上创建对应的表。

#登录mysql

[root@master ~]# mysql -uroot -proot

#mysql用户root权限设置

mysql> GRANT ALL PRIVILEGES ON \*.\* TO 'root'@'%' IDENTIFIED BY 'Root123!';

Query OK, 0 rows affected, 1 warning (0.00 sec)

#刷新权限设置

mysql> flush privileges;

Query OK, 0 rows affected (0.01 sec)

#退出mysql

mysql> exit;

Bye

#清空hdfs数据

[root@master ~]# hdfs dfs -rm -R /user/hadoop/t1

#从mysql导入数据到hive

* [root@master ~]# sqoop import --connect jdbc:mysql://master:3306/testdb?characterEncoding=UTF-8 --table t1 --username root --password 'root' --hive-import --hive-database testdb -m 1

18/11/21 15:50:22 INFO hive.HiveImport: Hive import complete.

4、执行hive命令测试上面的操作是否成功

#启动hive

[root@master ~]# hive

#进入testdb数据库

hive> use testdb;

OK

Time taken: 0.036 seconds

#显示表

hive> show tables;

OK

t1

Time taken: 0.173 seconds, Fetched: 1 row(s)

#查看t1表里的数据

hive> select \* from t1;

OK

10 tom

20 jerry

30 rose

Time taken: 0.262 seconds, Fetched: 3 row(s)

# 实验总结

Sqoop是一款开源的工具，主要用于在Hadoop Hive与传统的数据库mysql间进行数据的传递，可以将一个关系型数据库 MySQL中的数据导进到Hadoop的HDFS中，也可以将HDFS的数据导进到关系型数据库中。

本次实验主要完成了Sqoop的安装及验证通过Sqoop将MySQL数据导入到Hive数据仓库中。

# 参考文档

本实验参考下列文档内容制作：

* <http://sqoop.apache.org/docs/1.4.6/SqoopUserGuide.html>