## 1 Sqoop安装

### 1.1 下载并解压

1）下载地址：<http://mirrors.hust.edu.cn/apache/sqoop/1.4.6/>

2）上传安装包sqoop-1.4.6.bin\_\_hadoop-2.0.4-alpha.tar.gz到hadoop102的/opt/software路径中

3）解压sqoop安装包到指定目录，如：

[user1@hadoop102 software]$ tar -zxf sqoop-1.4.6.bin\_\_hadoop-2.0.4-alpha.tar.gz -C /opt/module/

4）解压sqoop安装包到指定目录，如：

[user1@hadoop102 module]$ mv sqoop-1.4.6.bin\_\_hadoop-2.0.4-alpha/ sqoop

### 1.2 修改配置文件

1) 进入到/opt/module/sqoop/conf目录，重命名配置文件

[user1@hadoop102 conf]$ mv sqoop-env-template.sh sqoop-env.sh

2) 修改配置文件

[user1@hadoop102 conf]$ vim sqoop-env.sh

增加如下内容

export HADOOP\_COMMON\_HOME=/opt/module/hadoop-3.1.3

export HADOOP\_MAPRED\_HOME=/opt/module/hadoop-3.1.3

export HIVE\_HOME=/opt/module/hive

export ZOOKEEPER\_HOME=/opt/module/zookeeper-3.5.7

export ZOOCFGDIR=/opt/module/zookeeper-3.5.7/conf

### 1.3 拷贝JDBC驱动

1）将mysql-connector-java-5.1.48.jar 上传到/opt/software路径

2）进入到/opt/software/路径，拷贝jdbc驱动到sqoop的lib目录下。

[user1@hadoop102 software]$ cp mysql-connector-java-5.1.48.jar /opt/module/sqoop/lib/

### 1.4 验证Sqoop

我们可以通过某一个command来验证sqoop配置是否正确：

[user1@hadoop102 sqoop]$ bin/sqoop help

出现一些Warning警告（警告信息已省略），并伴随着帮助命令的输出：

Available commands:

codegen Generate code to interact with database records

create-hive-table Import a table definition into Hive

eval Evaluate a SQL statement and display the results

export Export an HDFS directory to a database table

help List available commands

import Import a table from a database to HDFS

import-all-tables Import tables from a database to HDFS

import-mainframe Import datasets from a mainframe server to HDFS

job Work with saved jobs

list-databases List available databases on a server

list-tables List available tables in a database

merge Merge results of incremental imports

metastore Run a standalone Sqoop metastore

version Display version information

### 1.5 测试Sqoop是否能够成功连接数据库

[user1@hadoop102 sqoop]$ bin/sqoop list-databases --connect jdbc:mysql://hadoop102:3306/ --username root --password 000000

出现如下输出：

information\_schema

metastore

mysql

oozie

performance\_schema

### 1.6 mysql数据导入HDFS基本使用

**0）常用参数解释**

|  |  |
| --- | --- |
| --connect | 指定mysql连接url |
| --username | 指定mysql连接用户名 |
| --password | 指定mysql连接密码 |
| --table | 指定要导的表 |
| --columns | 指定要导的表字段，用逗号分隔 |
| --where | 指定过滤条件 |
| --query | 指定sql语句，注意$CONDITIONS |
| --target-dir | 指定数据导入到hdfs的路径 |
| --delete-target-dir | 如果指定的hdfs路径存在，进行删除 |
| --num-mappers | 指定启动几个mapTask，默认4个 |
| --split-by | 指定切片的字段 |
| --fields-terminated-by | 指定字段之间的分隔符 |

1. **案例一:**

需求: 将user\_info表中的id大于等于10并且小于等于100的数据导入到hdfs的/test路径下。注:只需导入id、login\_name、nick\_name 这三个字段。

bin/sqoop import \

--connect jdbc:mysql://hadoop202:3306/gmall \

--username root \

--password 123456 \

--table user\_info \

--columns id,login\_name \

--where "id >= 10 and id <=100" \

--target-dir /test \

--delete-target-dir \

--num-mappers 1 \

--split-by id \

--fields-terminated-by "\t"

1. **案例二**

需求: 将user\_info表中的id大于等于10并且小于等于100的数据导入到hdfs的/test路径下。注:只需导入id、login\_name、nick\_name 这三个字段。

要求: 通过sql的方式导入。

bin/sqoop import \

--connect jdbc:mysql://hadoop202:3306/gmall \

--username root \

--password 123456 \

--query "select id,login\_name from user\_info where id >=10 and id <=100 and $CONDITIONS" \

--target-dir /test \

--delete-target-dir \

--num-mappers 2 \

--split-by id \

--fields-terminated-by "\t"