

# Sqoop安装及指令

2016年1月20日 18:41

## Sqoop介绍

sqoop是Apache 提供的工具

用于hdfs和关系型数据库之间数据的导入和导出

可以从hdfs导出数据到关系型数据库，也可以从关系型数据库导入数据到hdfs。

## 实现步骤：

- 1.准备sqoop安装包，官网地址：<http://sqoop.apache.org>
- 2.配置jdk环境变量和Hadoop的环境变量。因为sqoop在使用是会去找环境变量对应的路径，从而完成工作。
- 3.sqoop解压即可使用（前提是环境变量都配好了）
- 4.需要将要连接的数据库的驱动包加入sqoop的lib目录下（本例中用的是mysql数据库）
- 5.利用指令操作sqoop

## Sqoop基础指令（在Sqoop的bin目录下执行下列指令）

说明	指令示例
查看mysql所有数据库	<pre>sh sqoop list-databases --connect jdbc:mysql://192.168.234.21:3306/ -username root -password root</pre> 
查看指定数据库下的所有表	<pre>sh sqoop list-tables --connect jdbc:mysql://192.168.234.21:3306/hive -username root -password root</pre>
关系型数据库==>hdfs  sh sqoop import -help（查看import的帮助指令）	<p><b>实现步骤：</b></p> <p>1.现在mysql数据库的test数据下建立一张tabx表，并插入测试数据 建表：create table tabx (id int,name varchar(20)); 插入：insert into tabx (id,name) values (1,'aaa'),(2,'bbb'),(3,'ccc'), (1,'ddd'),(2,'eee'),(3,'fff');</p> <p>2.进入到sqoop的bin目录下，执行导入语句 导入：</p> <pre>sh sqoop import --connect jdbc:mysql://192.168.234.21:3306/test --username root --password root --table tabx --target-dir '/sqoop/tabx' --fields-terminated-by ' ' --split-by id;</pre> <p><b>说明：</b>注意标红处对应的含义。</p> <p>此外，--split-by id表示的是tabx表在导入hdfs时分区的原则。本例是按id字段切分。 如果不想分区，则把--split-by id 替换成 -m 1即可。</p> <pre>sh sqoop import --connect jdbc:mysql://192.168.234.21:3306/test --username root --password root --table tabx --target-dir '/sqoop/tabx' --fields-terminated-by ' ' -m 1;</pre>
hdfs==>关系型数据库	<p>执行：sh sqoop export --connect jdbc:mysql://192.168.234.21:3306/test --username root --password root --export-dir '/sqoop/tabx' --table tabx -m 1 --fields-terminated-by ' '</p> <p><b>注：</b>sqoop只能导出数据，不能自动建表。所以在导出之前，要现在mysql数据库里建好对应的表</p>
关系型数据库==>hive	<p>执行：sh sqoop import --connect jdbc:mysql://192.168.234.21:3306/test --username root --password root --table tabx --hive-table tabx --fields-terminated-by ' ' -m 1</p> <p><b>注：</b>关系型数据库==&gt;hive，注意表的存放路径</p>

```

  ▾ 📁 user (2)
    ▸ 📁 hive (1)
    ▾ 📁 root (2)
      📁 .Trash (0)
      ▸ 📁 taby (2)

```

## Sqoop高级指令

说明	指令示例
在导入数据时，如果不想导入所有字段数据，可以增加where条件，注意，条件必须用引号引起来	<code>sqoop import --connect jdbc:mysql://192.168.1.10:3306/tehu --username root --password 123 --table trade_detail --where 'id&gt;3' --target-dir '/sqoop/td2'</code>
如果觉得where条件不够，可以直接写query语句。 注意：如果使用--query这个命令的时候，需要注意的是where后面的参数，AND \$CONDITIONS这个参数必须加上 而且存在单引号与双引号的区别，如果--query后面使用的是双引号，那么需要在\$CONDITIONS前加上\即\$CONDITIONS 如果设置map数量为1个时即-m 1，不用加上--split-by \${tablename.column}，否则需要加上	<code>sqoop import --connect jdbc:mysql://192.168.1.10:3306/tehu --username root --password 123 --query 'SELECT * FROM trade_detail where id &gt; 2 AND \$CONDITIONS' --split-by trade_detail.id --target-dir '/sqoop/td3'</code>
从关系型数据库导入数据到hive	<code>sqoop import --connect jdbc:mysql://192.168.1.10:3306/tehu --username root --password 123 --table trade_detail --hive-import-hive-table hivetab1 -m 1</code>