

Sqoop Tutorial

Pasos previos

- Desplegar los contenedores
Ir a la ruta donde se encuentra el archivo docker-compose.yml

>_ docker-compose up
- Instalar sqoop en el contenedor datanode

>_ docker exec -it datanode bash

>_ curl -s https://archive.apache.org/dist/sqoop/1.4.7/sqoop-1.4.7.bin__hadoop-2.6.0.tar.gz |
tar -xz -C /usr/local

>_ export HADOOP_HOME=/opt/hadoop-2.7.4

>_ export SQOOP_HOME=/usr/local/sqoop-1.4.7.bin__hadoop-2.6.0

>_ export PATH=\$PATH:\$HADOOP_HOME/bin:\$SQOOP_HOME/bin

>_ rm /opt/hadoop-2.7.4/share/hadoop/common/lib/avro-1.7.4.jar
- Abrir otro terminal y copiar los jars. Ubicarse en la carpeta donde se encuentran los archivos.

>_ docker cp org.json-20120521.jar datanode:/opt/hadoop-2.7.4/share/hadoop/common/lib

>_ docker cp avro-1.8.1.jar datanode:/opt/hadoop-2.7.4/share/hadoop/common/lib

>_ docker cp org.json-20120521.jar datanode: /usr/local/sqoop-1.4.7.bin__hadoop-2.6.0/lib/

Sqoop

- Listar bases de datos

sqoop list-databases \
--connect "jdbc:mysql://192.168.2.20:3310/" \
--username root \
--password root
- Listar tablas

sqoop list-tables \
--connect "jdbc:mysql://192.168.2.20:3310/retail_db" \
--username root \
--password root

- Importando tablas como archivos de texto

```
sqoop import \  
--connect "jdbc:mysql://192.168.2.20:3310/retail_db" \  
--username=root \  
--password=root \  
--table customers \  
--as-textfile \  
--target-dir=/user/vagrant/datasets/customers
```

```
sqoop import \  
--connect "jdbc:mysql://192.168.220.165:3310/retail_db" \  
--username=root \  
--password=root \  
--table departments \  
--as-textfile \  
--target-dir=/user/vagrant/datasets/departments
```

```
sqoop import \  
--connect "jdbc:mysql://192.168.2.20:3310/retail_db" \  
--username=root \  
--password=root \  
--table categories \  
--as-textfile \  
--target-dir=/user/vagrant/datasets/categories
```

```
sqoop import \  
--connect "jdbc:mysql://192.168.2.20:3310/retail_db" \  
--username=root \  
--password=root \  
--table orders \  
--as-textfile \  
--target-dir=/user/vagrant/datasets/orders
```

```
sqoop import \  
--connect "jdbc:mysql://192.168.2.20:3310/retail_db" \  
--username=root \  
--password=root \  
--table order_items \  
--as-textfile \  
--target-dir=/user/vagrant/datasets/order_items
```

```
sqoop import \  
--connect "jdbc:mysql://192.168.2.20:3310/retail_db" \  
--username=root \  
--password=root \  
--table products \  
--as-textfile \  
--target-dir=/user/vagrant/datasets/products\  
--delete-target-dir
```

```
sqoop import \  
--connect "jdbc:mysql://192.168.2.20:3310/retail_db" \  
--username=root \  
--password=root \  
--table departments \  
--as-textfile \  
--target-dir=/user/vagrant/datasets/departments_test
```

- Importando tablas como archivos avros

```
sqoop import \  
--connect "jdbc:mysql://192.168.2.20:3310/retail_db" \  
--username=root \  
--password=root \  
--table customers \  
--as-avrodatafile \  
--target-dir=/user/vagrant/datasets/avro/customers \  
--delete-target-dir
```

```
sqoop import \  
--connect "jdbc:mysql://192.168.2.20:3310/retail_db" \  
--username=root \  
--password=root \  
--table departments \  
--as-avrodatafile \  
--target-dir=/user/vagrant/datasets/avro/departments \  
--delete-target-dir
```

```
sqoop import \  
--connect "jdbc:mysql://192.168.2.20:3310/retail_db" \  
--username=root \  
--password=root \  
--table categories \  
--as-avrodatafile \  
--target-dir=/user/vagrant/datasets/avro/categories \  
--delete-target-dir
```

```
sqoop import \  
--connect "jdbc:mysql://192.168.2.20:3310/retail_db" \  
--username=root \  
--password=root \  
--table orders \  
--as-avrodatafile \  
--target-dir=/user/vagrant/datasets/avro/orders \  
--delete-target-dir
```

```
sqoop import \  
--connect "jdbc:mysql://192.168.2.20:3310/retail_db" \  
--username=root \  
--password=root \  
--table order_items \  
--as-avrodatafile \  
--target-dir=/user/vagrant/datasets/avro/order_items \  
--delete-target-dir
```

- ```

sqoop import \
--connect "jdbc:mysql://192.168.2.20:3310/retail_db" \
--username=root \
--password=root \
--table products \
--as-avrodatafile \
--target-dir=/user/vagrant/datasets/avro/products

```
- Uso de eval. Nos permite ejecutar queries en la bd
- ```

sqoop eval \
--connect "jdbc:mysql://192.168.2.20:3306/retail_db" \
--username root \
--password root \
--query "select count(1) from customers"

```
- Importamos todas las tablas
- ```

sqoop import-all-tables \
--connect "jdbc:mysql://192.168.2.20:3310/retail_db" \
--username root \
--password root \
--warehouse-dir /user/vagrant/retail_db/import-all-tables \
--driver com.mysql.jdbc.Driver

```
- Importando con condiciones
- ```

sqoop import \
--connect "jdbc:mysql://192.168.2.20:3310/retail_db" \
--username=root \
--password=root \
--table departments \
--target-dir /user/vagrant/datasets/retail_ods.db/departments \
--append \
--fields-terminated-by '|' \
--lines-terminated-by '\n' \
--split-by department_id \
--where "department_id > 3" \
--outdir java_files

```
- Exportando a mysql
- ```

sqoop export --connect "jdbc:mysql://192.168.2.20:3310/retail_db" \
--username root \
--password root \
--table departments2 \
--export-dir /user/vagrant/datasets/retail_ods.db/departments \
--input-fields-terminated-by '|' \
--input-lines-terminated-by '\n' \
--num-mappers 2 \
--batch \
--outdir java_files

```

- Copiamos los archivos que contienen la estructura de nuestros avros

```
>_hdfs dfs -ls /user/vagrant/datasets/avro/
```