

## **USECASE: Import single table to hdfs**

### **#Verify the target folder exists or not.**

```
hadoop fs -ls /user/cloudera/orders
```

### **#Deleting the target folder**

```
hadoop fs -rm -R /user/cloudera/orders
```

### **#Import data from MySQL table orders into HDFS location /user/cloudera/orders**

#### **#As Textfile format**

#### **#Default field delimiter is comma**

```
sqoop import \  
--connect jdbc:mysql://quickstart.cloudera:3306/retail_db \  
--username retail_dba \  
--password cloudera \  
--table orders \  
--as-textfile \  
--target-dir=/user/cloudera/orders
```

### **#Import data from MySQL table orders into HDFS location /user/cloudera/orders**

#### **#Let sqoop delete target dir if exists**

```
sqoop import \  
--connect jdbc:mysql://quickstart.cloudera:3306/retail_db \  
--username retail_dba \  
--password cloudera \  
--table orders \  
--delete-target-dir \  
--as-textfile \  
--target-dir=/user/cloudera/orders
```

### **#Import data from MySQL table orders into HDFS location /user/cloudera/orders**

#### **#Overriding the fields delimiter to \$**

```
sqoop import \  
--connect jdbc:mysql://quickstart.cloudera:3306/retail_db \  
--username retail_dba
```

```
--password cloudera \  
--table orders \  
--delete-target-dir \  
--as-textfile \  
--target-dir=/user/cloudera/orders \  
--fields-terminated-by '$'
```

### **#Verification Script (using hadoop fs command)**

```
hadoop fs -cat /user/cloudera/orders/part* | wc -l
```

### **#Verification Script (using sqoop eval)**

```
sqoop eval \  
--connect jdbc:mysql://quickstart.cloudera:3306/retail_db \  
--username retail_dba \  
--password cloudera \  
--query "select count(1) from orders"
```

### **#Conditional Import**

```
sqoop import \  
--connect jdbc:mysql://quickstart.cloudera:3306/retail_db \  
--username retail_dba \  
--password cloudera \  
--table orders \  
--delete-target-dir \  
--as-textfile \  
--target-dir=/user/cloudera/orders \  
--where "order_status='COMPLETE'"
```

### **#Conditional Import & Append to existing file**

```
sqoop import \  

```

```
--connect jdbc:mysql://quickstart.cloudera:3306/retail_db \  
--username retail_dba \  
--password cloudera \  
--table orders \  
--as-textfile \  
--target-dir=/user/cloudera/orders \  
--where "order_status='CANCELED'" \  
--append
```

### **#Sqoop Import using Serial way**

#### **#This will have impact in performance**

```
sqoop import \  
--connect jdbc:mysql://quickstart.cloudera:3306/retail_db \  
--username retail_dba \  
--password cloudera \  
--table orders \  
--delete-target-dir \  
--as-textfile \  
--target-dir=/user/cloudera/orders \  
--where "order_status='COMPLETE'" \  
--m 1
```

#### **#Free form import**

#### **#\$CONDITIONS must be passed in the where clause and also split-by field must be specified**

```
sqoop import \  
--connect jdbc:mysql://quickstart.cloudera:3306/retail_db \  
--username retail_dba \  
--password cloudera \  
--delete-target-dir \  
--as-textfile \  
--target-dir=/user/cloudera/orders \  
--query "select * from orders where \${CONDITIONS}" \  
--split-by order_id
```

### **#Free form import with user-defined conditions**

```
sqoop import \  
--connect jdbc:mysql://quickstart.cloudera:3306/retail_db \  
--username retail_dba \  
--password cloudera \  
--delete-target-dir \  
--as-textfile \  
--target-dir=/user/cloudera/orders \  
--query "select * from orders where \${CONDITIONS} and order_status='COMPLETE'" \  
--split-by order_id
```

### **#Built-in validator**

#### **#Validator works with single table only... and cannot use where criteria**

```
sqoop import \  
--connect jdbc:mysql://quickstart.cloudera:3306/retail_db \  
--username retail_dba \  
--password cloudera \  
--table orders \  
--target-dir /user/cloudera/orders \  
--delete-target-dir \  
--validate
```

### **#Import mysql table data into HDFS using Sequence File format**

```
sqoop import \  
--connect jdbc:mysql://quickstart.cloudera:3306/retail_db \  
--username retail_dba \  
--password cloudera \  
--table orders \  
--target-dir /user/cloudera/orders \  
--delete-target-dir \  
--as-sequencefile
```

### **#Import mysql table data into HDFS using Avro File format**

**#This will create sqoop import <<Table>>.avsc file under current directory where sqoop command is executed.It is the schema file**

```
sqoop import \  
--connect jdbc:mysql://quickstart.cloudera:3306/retail_db \  
--username retail_dba \  
--password cloudera \  
--table orders \  
--target-dir /user/cloudera/orders \  
--delete-target-dir \  
--as-avrodatafile
```

### **#We can use the .avsc file to import data into Hive**

**hadoop fs -put sqoop\_import\_orders.avsc /user/cloudera**

```
CREATE EXTERNAL TABLE orders ROW FORMAT SERDE  
'org.apache.hadoop.hive.serde2.avro.AvroSerDe' STORED AS INPUTFORMAT  
'org.apache.hadoop.hive.q1.io.avro.AvroContainerInputFormat' OUTPUTFORMAT  
'org.apache.hadoop.hive.q1.io.avro.AvroContainerOutputFormat' LOCATION  
'hdfs:///user/cloudera/orders' TBLPROPERTIES  
('avro.schema.url'='hdfs://quickstart.cloudera/user/cloudera/sqoop_import_orders.avsc');
```

### **#Using boundary query, selected list of columns**

**#When you specify list of columns, make sure there is no white-space between field name and comma**

```
sqoop import \  
--connect jdbc:mysql://quickstart.cloudera:3306/retail_db \  
--username retail_dba \  
--password cloudera \  
--table orders \  
--target-dir /user/cloudera/orders \  
--delete-target-dir \  
--boundary-query "select 100, 200 from orders limit 1" \  
--columns order_id,order_status
```

### **#Incremental Load**

**#You need to specify the check-column (mostly the primary key column) and its last value**

```
sqoop import \  
--connect jdbc:mysql://quickstart.cloudera:3306/retail_db \  
--username retail_dba \  
--password cloudera \  
--table orders \  
--target-dir /user/cloudera/orders \  
--append \  
--where "order_id < 600" \  
--check-column "order_id" \  
--last-value 99 \  
--incremental append
```