Sqoop notes, samples, cook notebook.

Arturo Alatriste Trujillo.

Contents

[Sqoop commands 1](#_Toc506040311)

[Use a txt file with options 1](#_Toc506040312)

[IMPORT 1](#_Toc506040313)

[List all tables in the loudacre database in MySQL 1](#_Toc506040314)

[import-all-tables tool imports an entire database 1](#_Toc506040315)

[--warehouse-dir option to specify a different base directory 1](#_Toc506040316)

[This example Imports the accounts table 1](#_Toc506040317)

[This variation writes tab--‐delimited fields instead 1](#_Toc506040318)

[Sqoop’s Incremental lastmodified mode imports new and modified records 1](#_Toc506040319)

[Use Sqoop’s Incremental append mode to import only new records 1](#_Toc506040320)

[Import data and save it as avro 1](#_Toc506040321)

[Import data as parquet 1](#_Toc506040322)

[PARTIAL IMPORT 1](#_Toc506040323)

[Import Only specified columns from accounts table 1](#_Toc506040324)

[Import Only matching rows from accounts table 1](#_Toc506040325)

[Import using a complete query 1](#_Toc506040326)

[The --where Option is ignored in a free--‐form query 1](#_Toc506040327)

[IMPORT to HIVE 1](#_Toc506040328)

[Import one table from mySQL to Hive 1](#_Toc506040329)

[Import a complete mySQL database to a Hive database. 1](#_Toc506040330)

[EXPORT 1](#_Toc506040331)

[Exort from Hadoop to RDBM 1](#_Toc506040332)

[Execute a QUERY 1](#_Toc506040333)

[Execute a SQL in database 1](#_Toc506040334)

[Compression 1](#_Toc506040335)

[Use the compression codec flag 1](#_Toc506040336)

# Sqoop commands

## Use a txt file with options

We can set options in the txt file and also we can add more options in command line.

sqoop --options-file /users/homer/work/import.txt --table TEST

## IMPORT

### List all tables in the loudacre database in MySQL

**$ sqoop list-tables \**

**--connect jdbc:mysql://*dbhost*/loudacre \**

**--username *dbuser* \**

**--password *pw***

### import-all-tables tool imports an entire database

**$ sqoop import-all-tables \**

**--connect jdbc:mysql://*dbhost*/loudacre \**

**--username *dbuser*  \**

**--password *pw***

### --warehouse-dir option to specify a different base directory

**$ sqoop import-all-tables \**

**--connect jdbc:mysql://*dbhost*/loudacre \**

**--username *dbuser*  \**

**--password *pw* \**

**--warehouse-dir /loudacre**

### This example Imports the accounts table

– It Stores the data in HDFS As comma--‐delimited fields

$ **sqoop import \**

**--table accounts \**

**--connect jdbc:mysql://*dbhost*/loudacre \**

**--username *dbuser*  \**

**--password *pw***

### This variation writes tab--‐delimited fields instead

$ **sqoop import \**

**--table accounts \**

**--connect jdbc:mysql://*dbhost*/loudacre \**

**--username *dbuser*  \**

**--password *pw***

**--fields-terminated-by "\t"**

### Sqoop’s Incremental lastmodified mode imports new and modified records

**$ sqoop import \**

**--table invoices \**

**--connect jdbc:mysql://*dbhost*/loudacre \**

**--username *dbuser* \**

**--password *pw* \**

**--incremental lastmodified \**

**--check-column mod\_dt \**

**--last-value '2015-09-30 16:00:00'**

### Use Sqoop’s Incremental append mode to import only *new* records

**$ sqoop import \**

**--table invoices \**

**--connect jdbc:mysql://*dbhost*/loudacre \**

**--username *dbuser* \**

**--password *pw* \**

**--incremental append \**

**--check-column id \**

**--last-value 9478306**

### Import data and save it as avro

**$ sqoop import \**

**--connect jdbc:mysql://localhost/loudacre \**

**--username training \**

**--password training \**

**--table accounts \**

**--target-dir /loudacre/accounts\_avro \**

**--as-avrodatafile**

### Import data as parquet

**$ sqoop import \**

**--connect jdbc:mysql://localhost/loudacre \**

**--username training \**

**--password training \**

**--table accounts \**

**--target-dir /loudacre/accounts\_parquet \**

**--as-parquetfile**

## PARTIAL IMPORT

### Import Only specified columns from accounts table

**$ sqoop import \**

**--table accounts \**

**--connect jdbc:mysql://*dbhost*/loudacre \**

**--username *dbuser*  \**

**--password *pw* \**

**--columns "id,first\_name,last\_name,state"**

### Import Only matching rows from accounts table

**$ sqoop import \**

**--table accounts \**

**--connect jdbc:mysql://*dbhost*/loudacre \**

**--username *dbuser*  \**

**--password *pw* \**

**--where "state='CA'"**

### Import using a complete query

**$ sqoop import \**

**--connect jdbc:mysql://*dbhost*/loudacre \**

**--username *dbuser* \**

**--password *pw* \**

**--target-dir /data/loudacre/payable \**

**--split-by accounts.id \**

**--query 'SELECT accounts.id, first\_name,**

**last\_name, bill\_amount FROM accounts JOIN invoices ON**

**(accounts.id = invoices.cust\_id) WHERE $CONDITIONS'**

### The --where Option is ignored in a free--‐form query

– You Must specify your criteria using AND Following the WHERE clause

**$ sqoop import \**

**--connect jdbc:mysql://*dbhost*/loudacre \**

**--username *dbuser* \**

**--password *pw* \**

**--target-dir /data/loudacre/payable \**

**--split-by accounts.id \**

**--query 'SELECT accounts.id, first\_name,**

**last\_name, bill\_amount FROM accounts JOIN invoices ON**

**(accounts.id = invoices.cust\_id) WHERE $CONDITIONS AND**

**bill\_amount >= 40'**

## IMPORT to HIVE

### Import one table from mySQL to Hive

**$ sqoop import \**

**--connect jdbc:mysql://localhost/loudacre \**

**--username training \**

**--password training \**

**--fields-terminated-by '\t' \**

**--table employees \**

**--hive-import**

**--hive-table employees2**

The parameter hive-table set the name of the table in Hive. If this parameter is not specified the default name will be the name of the name of the table in database.

### Import a complete mySQL database to a Hive database.

sqoop import-all-tables

--connect jdbc:mysql://quickstart:3306/retail\_db

--username=retail\_dba

--password=cloudera

-m3

--hive-import

--hive-overwrite

--hive-database retail\_db

--compress

--compression-codec org.apache.hadoop.io.compress.SnappyCodec

--outdir java\_output

Parameter –-outdir is a local directory that will be created by sqoop.

## EXPORT

### Exort from Hadoop to RDBM

**sqoop export \**

**--connect jdbc:mysql://*dbhost*/loudacre \**

**--username *dbuser* \**

**--password *pw* \**

**--export-dir /loudacre/recommender\_output \**

**--update-mode allowinsert \**

**--table product\_recommendations**

## Execute a QUERY

### Execute a SQL in database

**sqoop eval \**

**--query "SELECT \* FROM webpage LIMIT 10" \**

**--connect jdbc:mysql://localhost/loudacre \**

**--username training \**

**--password training**

# Compression

## Use the compression codec flag

--compression-codec

org.apache.hadoop.io.compress.SnappyCodec