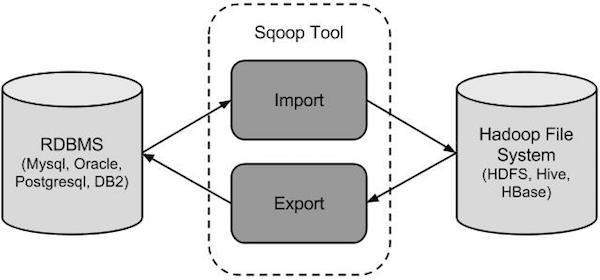
How Sqoop Works?

The following image describes the workflow of Sqoop.



Sqoop Import

The import tool imports individual tables from RDBMS to HDFS. Each row in a table is treated as a record in HDFS. All records are stored as text data in text files or as binary data in Avro and Sequence files.

Sqoop Export

The export tool exports a set of files from HDFS back to an RDBMS. The files given as input to Sqoop contain records, which are called as rows in table. Those are read and parsed into a set of records and delimited with user-specified delimiter.

sqoop list-databases --connect jdbc:oracle:thin:@INLT356:1521:xe --username SASI --password sasi

sqoop list-tables --connect jdbc:oracle:thin:@INLT356:1521:xe --username SASI --password sasi

* --target-dir /user/XXX/result
* --where “city=’bglr’”

Syntax:

--incremental <mode>

--check-column <column name>

--last value <last check column value>

Example

$ sqoop import \

--connect jdbc:mysql://localhost/userdb \

--username root \

--table emp \

--m 1 \

--incremental append \

--check-column id \

--last value 1101

sqoop import-all-tables \

--connect jdbc:mysql://localhost/userdb \

--username root

Export

In HDFS We have a directory emp1234.

It contains a file emp1

Content of emp1 is

1,sasi,5000

2,siva,3000

3,ram,4000

Emp1234 contains one more file emp2.

Content of Emp2 is

4,sasi,5000

5,siva,3000

6,ram,4000

CREATE TABLE employee (

id INT NOT NULL PRIMARY KEY,

name VARCHAR2(20),

salary INT);

sqoop export --connect jdbc:oracle:thin:@INLT356:1521:xe --username SASI --password sasi --table EMPLOYEE --export-dir /user/sbkt/emp1234

select \* from employee;

ID NAME SALARY

1 sasi 5000

2 siva 3000

3 ram 4000

4 sasi 5000

5 siva 3000

6 ram 4000

Job creation

sqoop job --create tbl1job \

--import \

--connect jdbc:mysql://localhost/db \

--username root \

--table employee --m 1

sqoop job --list

Available jobs:

Tbl1job

sqoop job --show tbl1job

sqoop job --exec tbl1job

**Eval**

sqoop eval \

--connect jdbc:mysql://localhost/db \

--username root \

--query “SELECT \* FROM employee LIMIT 3”

sqoop eval \

--connect jdbc:mysql://localhost/db \

--username root \

-e “INSERT INTO employee VALUES(1207,‘Raju’,‘UI dev’,15000,‘TP’)”

sqoop import \

--query 'SELECT a.\*, b.\* FROM a JOIN b on (a.id == b.id) WHERE $CONDITIONS' \

-m 1 --target-dir /user/foo/joinresults

an escape character (--fields-terminated-by \t). Supported escape characters are:

* \b (backspace)
* \n (newline)
* \r (carriage return)
* \t (tab)
* \" (double-quote)
* \\' (single-quote)
* \\ (backslash)