**Sqoop**

Soop is a data ingestion tool used to ingest or import data from traditional RDBMS to HDFS.

Details of sqoop command parameters:

1. --connect – used to connect to the DB. E.g. mysql
2. --username – username to connect to database
3. --password – password to connect to database
4. --query – the query to execute on the DB
5. --validate – to validate the data imported from DB. E.g. it gives the number of rows imported.
6. --sqoop export – to export data from hdfs to RDBMS
7. --hive-import – to import data to hive table, --hive-table, --hive-overwrite
8. --incremental append – to import data after specified ID and append in existing data
9. --incremental lastmodified – to get the last value from DB and import records after the specified record
10. --merge-key – used with –incremental lastmodified parameter. Used to update the data that already exists.
11. --compress –compression-codec – to compress the data imported. E.g. snappy codec used to compress
12. --target-dir – to specify the directory where data should be imported
13. --warehouse-dir – creates a new direct
14. --m – specify the number of mappers for the job
15. --sqoop job – to create a job to run

Below are few widely used commands in sqoop.

1. **Listing data from table**

sqoop eval --connect jdbc:mysql://192.168.0.101/db1 --username test --query "select \* from emp" -P;

1. **Insert data into Mysql emp table**

sqoop eval --connect jdbc:mysql://192.168.0.104/db1 --username test --query "insert into emp values(4,'ZMO')" -P;

1. **Import Column Specific data**

sqoop import --connect jdbc:mysql://192.168.0.104/test --table EMPLOYEES --columns "employee\_id,first\_name,last\_name,job\_title"

1. **Validator**

sqoop import --connect jdbc:mysql://192.168.0.104/test1 --table mytable2 --validate --username test -P;

1. **Importing the data to Hive:**

sqoop import --connect jdbc:mysql://192.168.0.104/test1 --table mytable2 --hive-import --target-dir /home/mytab --username test -P;

1. **Specifying only the subset of data**

sqoop import --connect jdbc:mysql://192.168.30.1/db100 --table student \

--where "marks > '70'"

--username root \

--target-dir /subsetdata -P -m 1

1. **Importing all your tables**

sqoop import-all-tables --connect jdbc:mysql://192.168.30.1/db1 \

--username root --target-dir /alltables -P -m 1

1. **Incremental import**

sqoop import --connect jdbc:mysql://192.168.30.1/pri --username root -P - -table persons --incremental append --check-column P\_Id --last-value 1 -- target-dir /increment2

1. **Importing the table data with JOIN function**

sqoop import --connect jdbc:mysql://192.168.30.1/sstechnosystems \

--username root -P \

--query 'SELECT persons.P\_Id, persons1.city, FROM persons JOIN persons USING (P\_Id) WHERE $CONDITIONS' \

--split-by P\_Id --target-dir /join

1. **Sqoop –Merge**

sqoop import --connect jdbc:mysql://localhost/test --table emp \

--username hive -password hive --incremental lastmodified --merge-key empid --check-column cr\_date \

--target-dir /sqoop/empdata/

1. **overwrite the data in hive using sqoop**

sqoop import --connect jdbc:mysql://localhost/test --username root --password 'hr' --table sample --hive-import --hive-overwrite --hive-table sqoophive -m 1 --fields-terminated-by '\t' --lines-terminated-by '\n'

1. **Compression**

sqoop import --connect  
jdbc:mysql://192.168.0.104/test1 --username test –password test  
--table TableA -m 5 -split-by ID --hive-table output.TableACompress  
--hive-import --compress --compression-codec  
org.apache.hadoop.io.compress.SnappyCodec -- --schema COMMERCIAL  
--as-sequencefile

1. **Creating saved jobs**

sqoop job --create myjob -- import --connect jdbc:mysql://example.com/db --table mytable

sqoop job --create myjob1 -- import --connect jdbc:mysql://127.0.0.1/mysql --username root --password root --table test1 --incremental append --check-column id --last-value 1 --target-dir /home/sample -m 1

1. **Executing Job**

sqoop job --exec myjob -- --username someuser –P