*Simple import (hdfs)*

$ ls

$ hadoop fs -ls

$ hadoop fs -ls /user/cloudera

$ hadoop fs -mkdir /user/cloudera/MS

$ sqoop import --connect "jdbc:mysql://localhost:3306/MB" --username root --password cloudera --table test1 -m 1 --target-dir "/user/cloudera/MS/test1111"

$ hadoop fs -ls

$ hadoop fs -ls

$ hadoop fs -ls "/user/cloudera/MS"

$ hadoop fs -ls "/user/cloudera/MS/test1111"

$ hadoop fs -cat "/user/cloudera/MS/test1111/part-m-00000"

*Import split-by*

$ sqoop import --connect "jdbd:mysql://localhost:3306/MB" --username root --password cloudera --table test1 --split-by id --target-dir "/user/cloudera/MBB/test2"

$ hadoop fs -ls

$ hadoop fs -ls "/user/cloudera/MBB"

$ hadoop fs -ls "/user/cloudera/MBB/test26"

$ hadoop fs -cat "/user/cloudera/MBB/test26/part-m-00000"

*Import boundary*

\*compulsary to add --spilt-by with --boundary-query if the primary key not declaired & mapper is not compulsary.

\* if using mapper then compulsary to add --split-by

$sqoop import --connect "jdbc:mysql://localhost:3306/m" --username root --password cloudera --table pra -m 1 --split-by id --boundary-query "select 2,4 from pra" --target-dir "/user/cloudera/s/pawan14"

$ hadoop fs -ls

$ hadoop fs -ls "/user/cloudera/MBB"

$ hadoop fs -ls "/user/cloudera/MBB/test27"

$ hadoop fs -cat "/user/cloudera/MBB/test27/part-m-00000"

\* if primary key present then, no need to set --split-by and also mapper is not compulsary.

$ sqoop import --connect 'jdbc:mysql://localhost:3306/sql\_data' --username root --password cloudera --table emp --boundary-query "select 2,4 from emp" --target-dir '/user/cloudera/sql\_data1/boundary3'

###### where not working.

$ sqoop import --connect "jdbc:mysql://localhost:3306/MB" --username root --password cloudera --table test1 -m 1 --split-by id --boundary-query "select 1,5 from test1 where addr='pune'" --target-dir "/user/cloudera/MBB/test27"

*Import append – incremental append*

\* if primay key is not present then use --split-by

$ sqoop import --connect "jdbc:mysql://localhost:3306/MB" --username root --password cloudera --table test1 --split-by id -m 2 --target-dir "/user/cloudera/MSS/test" --check-column id --incremental append --last-value 10

$ hadoop fs -ls /user/cloudera/MSS/

$ hadoop fs -ls /user/cloudera/MSS/test/

$ hadoop fs -cat /user/cloudera/MSS/test/part-m-00004

example:

[cloudera@quickstart ~]$ hadoop fs -ls /user/cloudera/p

Found 1 items

drwxr-xr-x - cloudera cloudera 0 2019-06-06 15:11 /user/cloudera/p/pranjal

[cloudera@quickstart ~]$ hadoop fs -ls /user/cloudera/p/pranjal

Found 3 items

-rw-r--r-- 1 cloudera cloudera 0 2019-06-06 15:03 /user/cloudera/p/pranjal/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 357 2019-06-06 15:03 /user/cloudera/p/pranjal/part-m-00000

-rw-r--r-- 1 cloudera cloudera 71 2019-06-06 15:11 /user/cloudera/p/pranjal/part-m-00001

[cloudera@quickstart ~]$ hadoop fs -cat /user/cloudera/p/pranjal/part-m-00000

1,manish,jodhpur,2019-06-03 18:39:25.0

2,datta,pune,2019-06-03 18:39:52.0

3,ak,mumbai,2019-06-03 18:40:21.0

4,nik,delhi,2019-06-03 18:48:54.0

5,prachi,delhi,2019-06-04 12:50:41.0

6,pranajl,jaipur,2019-06-04 12:51:48.0

7,am,ajmer,2019-06-04 12:52:21.0

8,pawan,mumbai,2019-06-04 12:52:53.0

9,vijay,vapi,2019-06-04 12:53:15.0

10,ajay,goa,2019-06-04 12:53:57.0

[cloudera@quickstart ~]$ hadoop fs -cat /user/cloudera/p/pranjal/part-m-00001

11,max,delhi,2019-06-06 15:04:44.0

12,pawan,pune,2019-06-06 15:05:15.0

[cloudera@quickstart ~]$

\*\*\*

\* if primary key is present then no need to add --split-by & also mapper's is depend on programmer.

$ sqoop import --connect 'jdbc:mysql://localhost:3306/sql\_data' --username root --password cloudera --table student -m 2 --target-dir '/user/cloudera/sqoop\_data1/append1' --check-column id --incremental append --last-value 110

*Import last modified*

\* lastmodified is only done on the append file which has already appended..

append = test = lastmodified

$ sqoop import --connect "jdbc:mysql://localhost:3306/MB" --username root --password cloudera --table test1 -m 1 --target-dir "/user/cloudera/MSS/test" --check-column dt --incremental lastmodified --merge-key id

$ hadoop fs -ls /user/cloudera/MSS

$ hadoop fs -ls /user/cloudera/MSS/test/

$ hadoopfs -cat /user/cloudera/MSS/test/part-r-00000

*Export sqoop*

for export data from linux to mysql..

we need to empty table in mysql as well as we that data sends..

1. first create table in mysql..

mysql> create table test2(id int, name varchar(20), addr varchar(20), primary key(id), dt DATETIME);

2. then sends the data

$ sqoop export --connect "jdbc:mysql://localhost:3306/MB" --username root --password cloudera --table test2 --export-dir "/user/cloudera/MSS/test"

\*if same data again exported into same table then job will be fail there are primary key.

mysql> select \* from test2;