Q1)

1)Load data into the 2 tables (naming EX: STUDENTNAME\_TABLE1) created in database (BIGDATA) with the First.csv and Second.csv files provided.

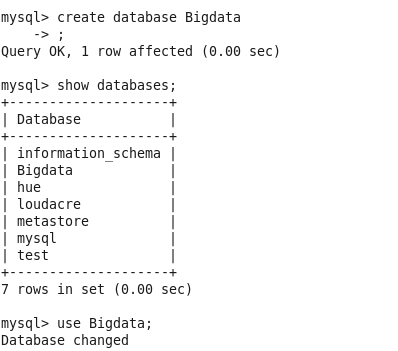
2)Using SQOOP perform DML operations (insert, update, delete of records) for any one of the tables created.

3) Now, Using SQOOP get the records from the tables into HDFS.

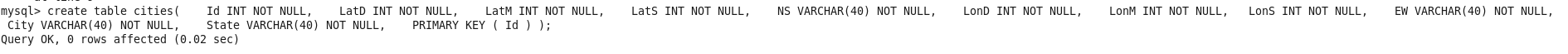
4)Using SQOOP merge both the files in HDFS and load the results into a new table.

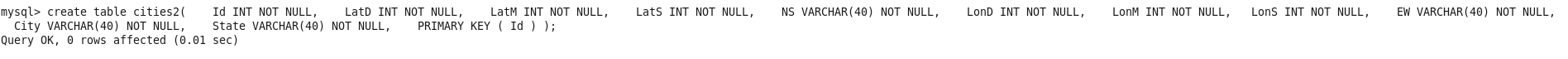
If there is any failure and you do not want partial data get loaded into Target table, find a solution for this (should be included in syntax)

Creating Database in MYSQL:

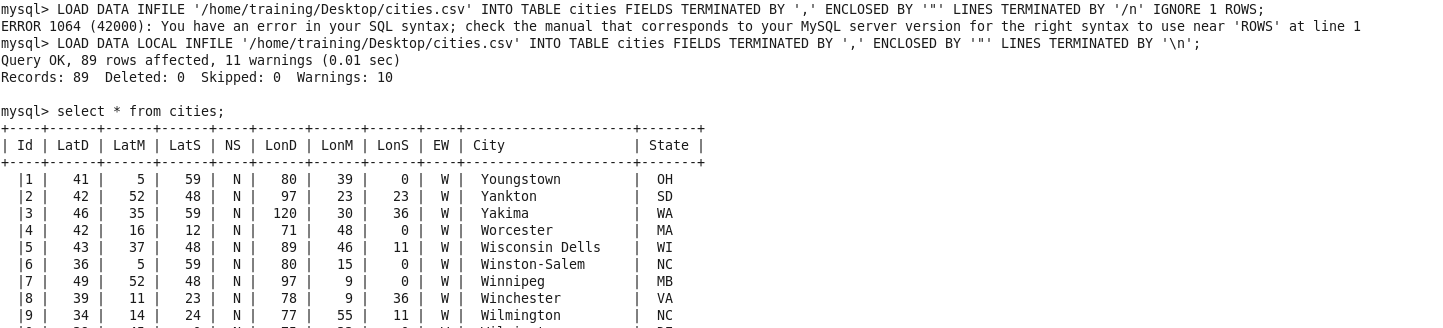


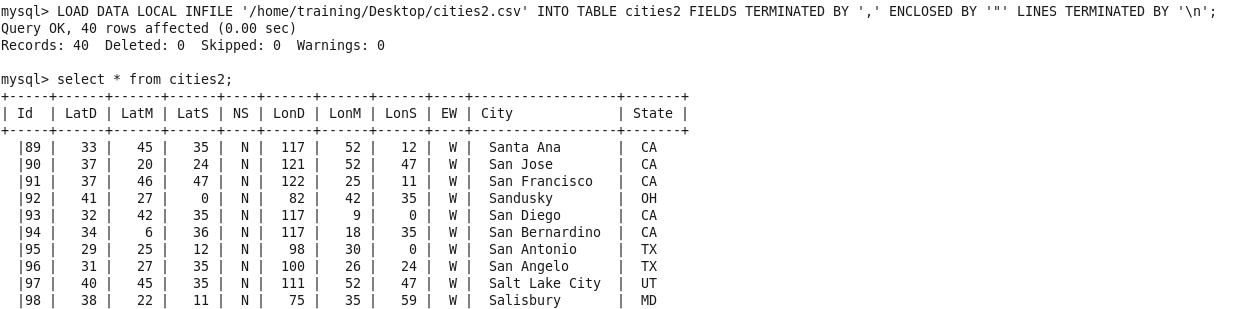
Create Tables:



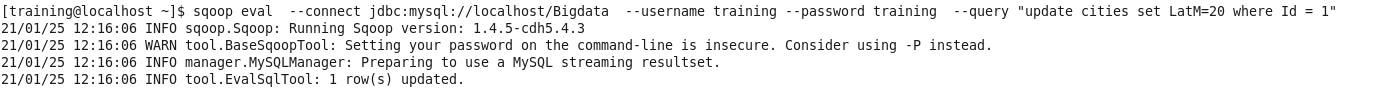


Loading the tables with data from files in local:



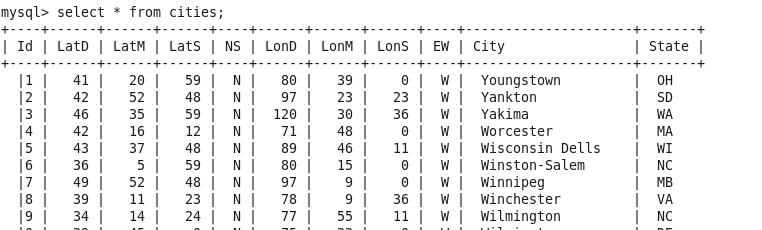


Updating the table using sqoop:

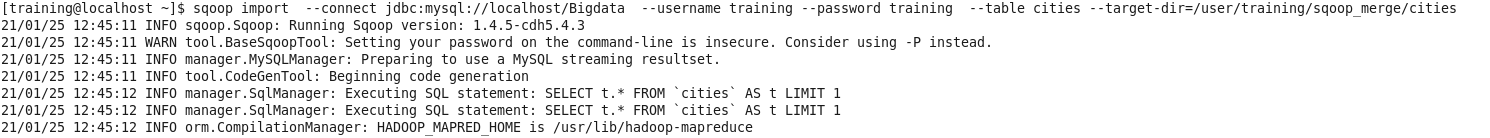


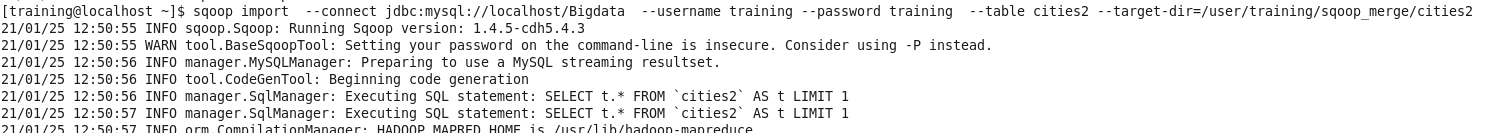
The LatM is changed to 20 from 5 for Id “1”

Validation:

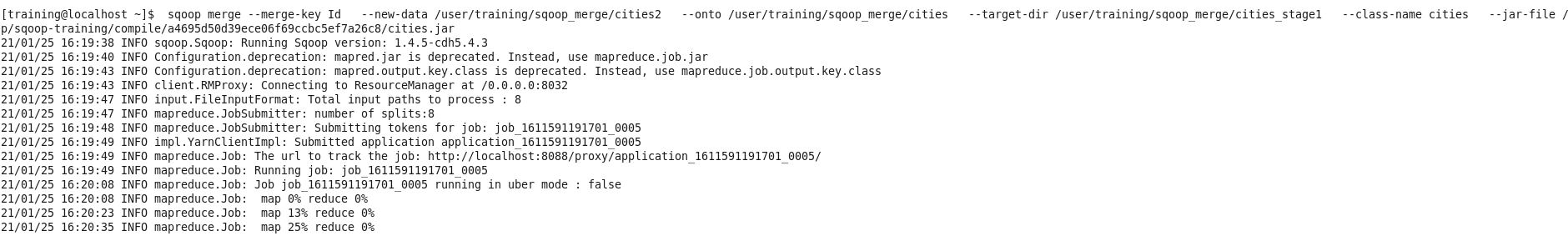


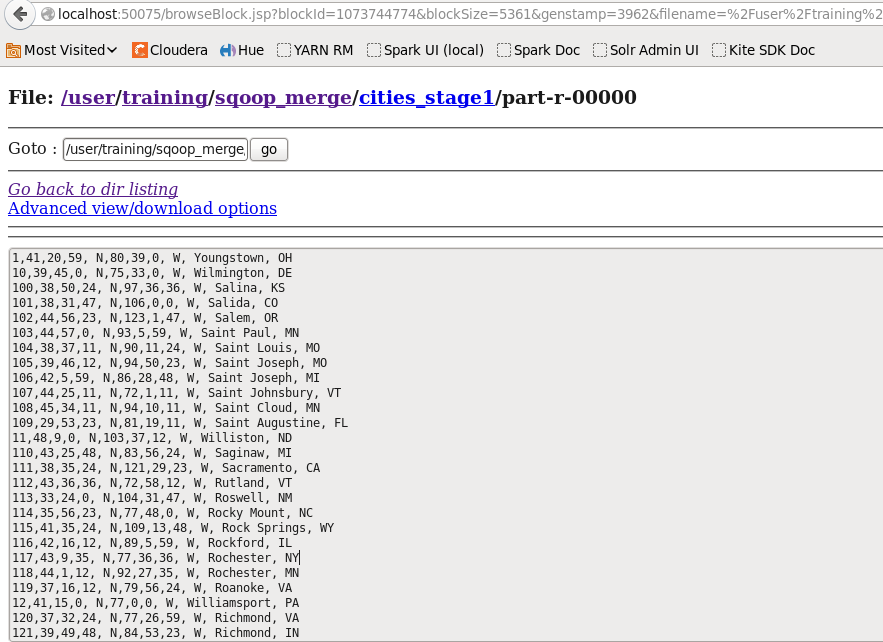
Importing the tables into HDFS:



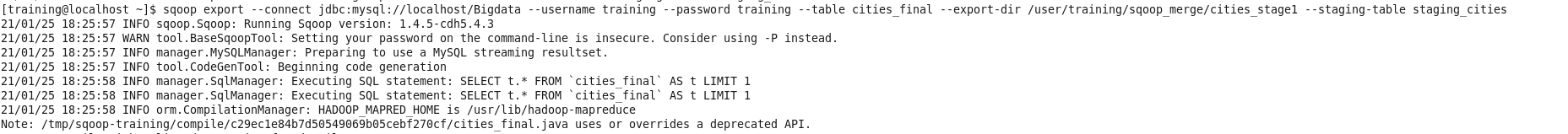


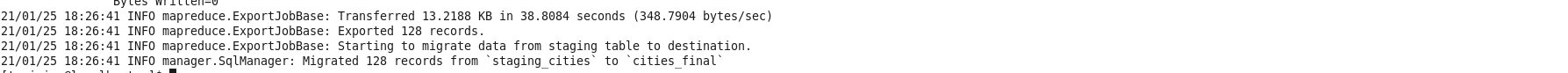
Merging the files:

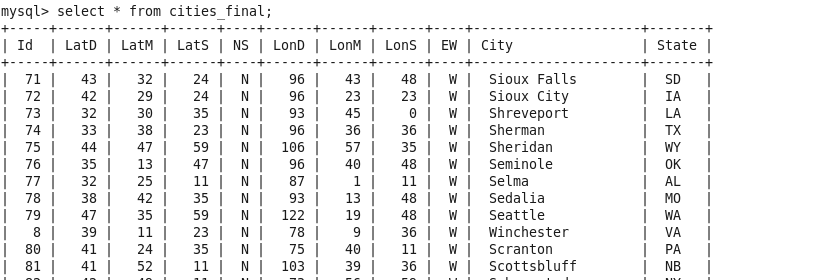


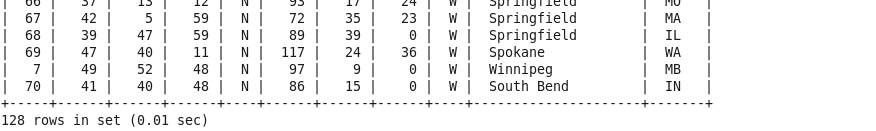


Exporting the data from Final file into staging and then into Final table:









The total 128 records are loaded.

Q2)

1)Create a table with the data from the old File.

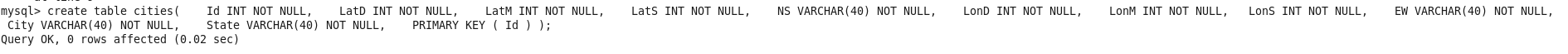
2)Transfer the data from table to HDFS (file) using Sqoop.

3)Add the new data from new file to the existing table.

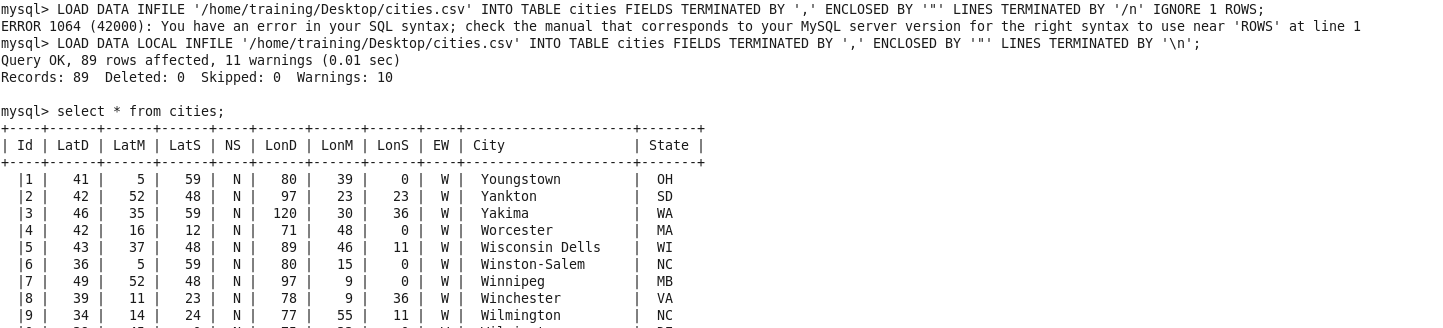
4)Create a Job using Sqoop to append the added data into the previous HDFS (file).

After loading the data, there may be changes to the data. Instead of overwriting the data find a solution to overcome this (should be included in syntax)

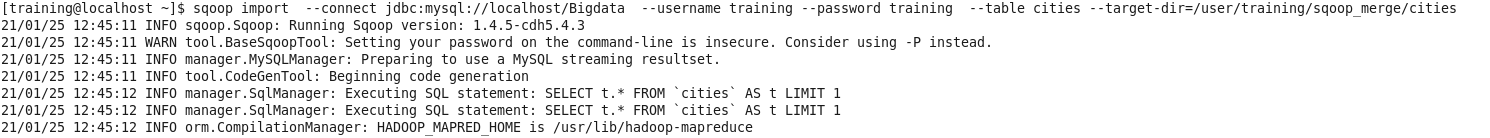
Creating table:



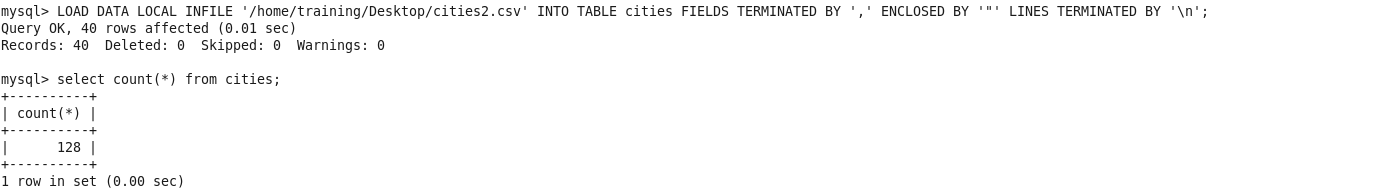
Loading the table:



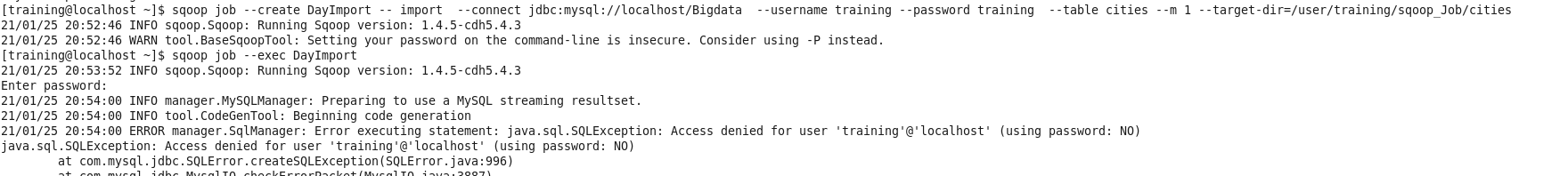
Importing into HDFS:

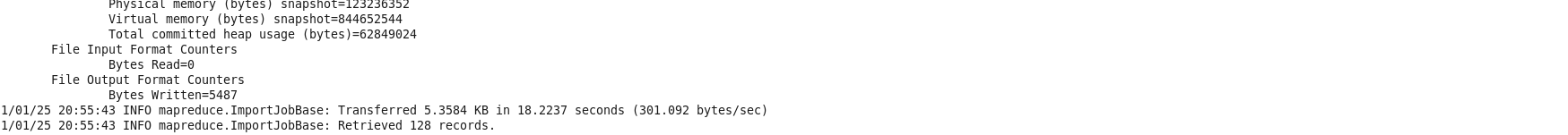


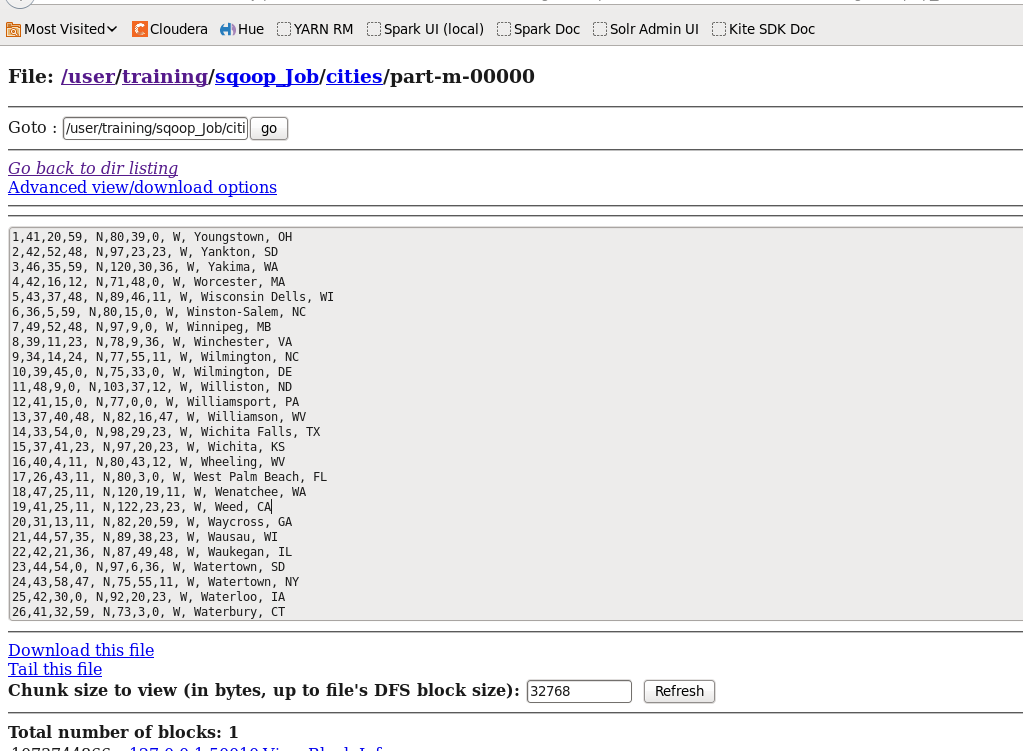
Adding data from new file to table:



Creating Job using Sqoop:







Exporting using upsert:

