**Problem Statement:** In this project we are required to load the customers data (csv file provided) into the MySql and then loading the data to Hive using Sqoop.

**Solution:**

Step 1: Login into SQL in your VM using below commands

mysql -u root -p

password cloudera

Step 2: Create a Data base using below commands

create database antwak;

use antwak;

Step 3: Create the table in MySql

CREATE TABLE customers (customer\_id INT NOT NULL , customer\_fname VARCHAR(255) NOT NULL, customer\_lname VARCHAR(255) NOT NULL, customer\_street VARCHAR(255) NOT NULL,customer\_city VARCHAR(255) NOT NULL,customer\_state VARCHAR(255) NOT NULL,PRIMARY KEY (customer\_id));

Step 4: Insert records into MySql

LOAD DATA LOCAL INFILE '/home/cloudera/Desktop/customers.csv' INTO TABLE customers FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' (customer\_id, customer\_fname, customer\_lname, customer\_street, customer\_city,customer\_state);

Now we need to import the data from the MySql to Hive using Sqoop import command.

Command used:

sqoop import --connect "jdbc:mysql://quickstart.cloudera:3306/antwak" \

--username root \

--password cloudera \

--table customers \

--target-dir /user/cloudera/ingest/raw/antwak\_customers \

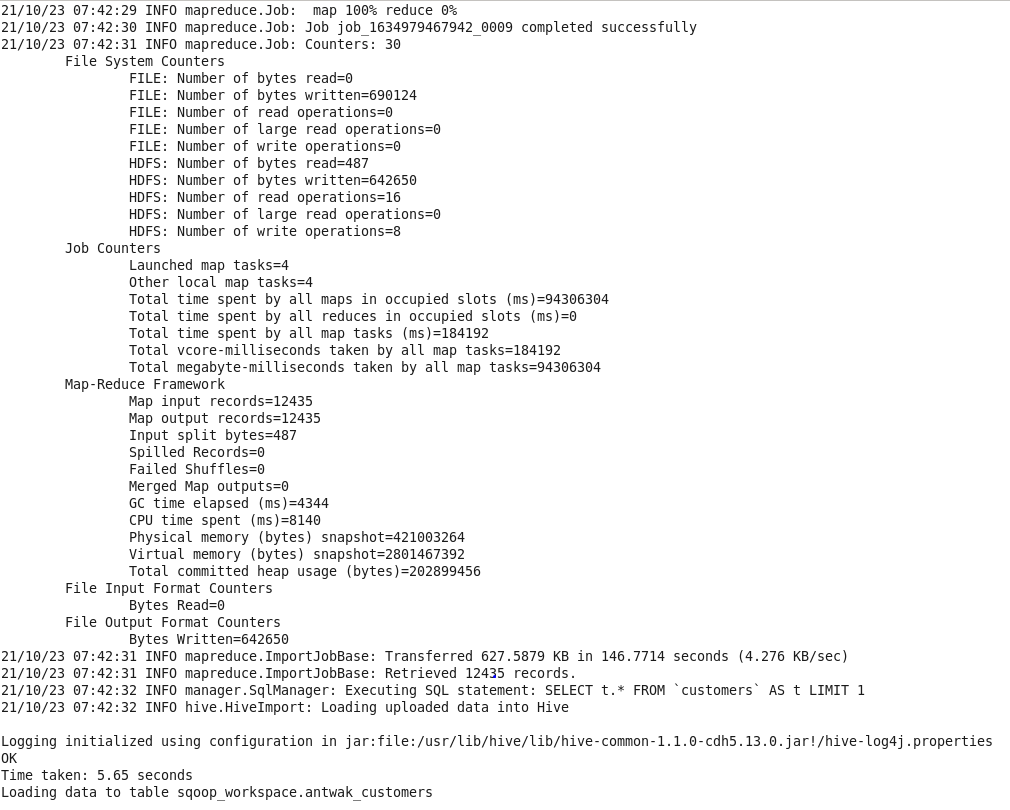
--fields-terminated-by "," \

--hive-import \

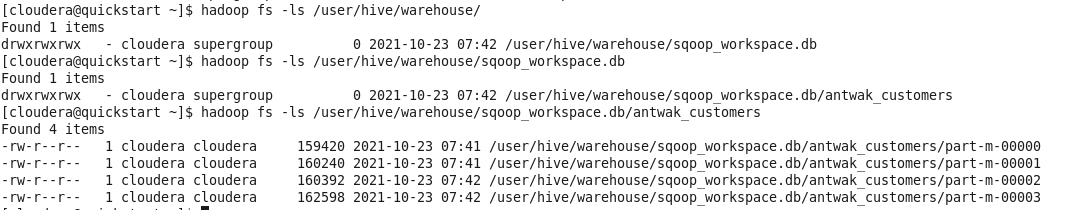
--create-hive-table \

--hive-table sqoop\_workspace.antwak\_customers

Sqoop job processing:



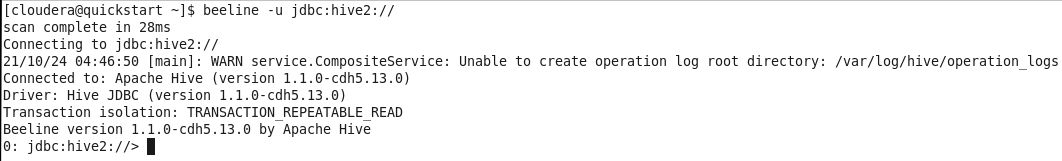
Check the files created in Hive Storage: Since there were 4 mappers so number of part files created is 4.



To query this Hive table, we can access having Hive using Beeline interface.

Step 1:

Command used: beeline -u jdbc:hive2://



Now we can start writing the HQL queries to get the data.

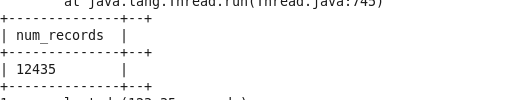
1. How many records are present?

Query:

select count(1) as NUM\_RECORDS from sqoop\_workspace.antwak\_customers;



Output:



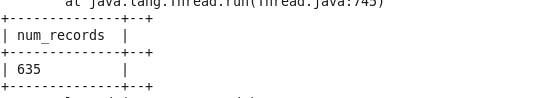
1. How many customers stay in the state TX

Query:

select count(1) as NUM\_RECORDS from sqoop\_workspace.antwak\_customers where customer\_state = 'TX';



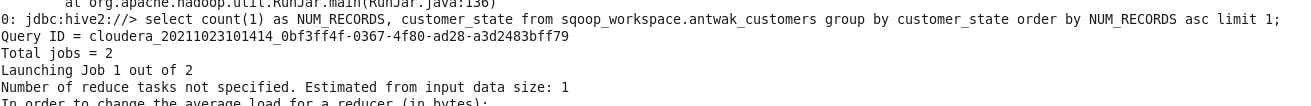
Output:



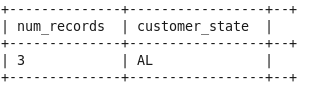
1. Which state has the lowest customer count?

Query:

select count(1) as NUM\_RECORDS, customer\_state from sqoop\_workspace.antwak\_customers group by customer\_state order by NUM\_RECORDS asc limit 1;



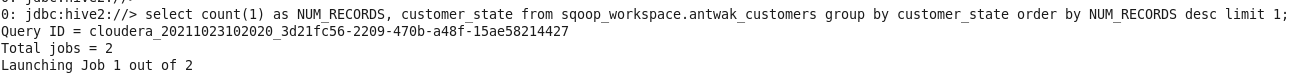
Output:



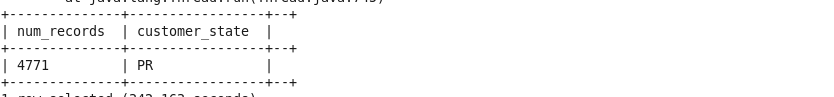
1. Which state has the highest customer count?

Query:

select count(1) as NUM\_RECORDS, customer\_state from sqoop\_workspace.antwak\_customers group by customer\_state order by NUM\_RECORDS desc limit 1;



Output:



1. How many customers are present with the First Name 'Mary'?

Query:

select count(1) as NUM\_RECORDS, customer\_fname from sqoop\_workspace.antwak\_customers where customer\_fname ='Mary' group by customer\_fname;



Output:

