Assignment 3

Prathik B Shetty | 211058019

Hive Assignment

Problems

1. Create database in Hive

2. Create Hive table inside database.

3. Load data into hive table

4. Load data from local file into Hive table (Hint: use keyword localin load data command)

5. Run queries for customer table for following:

   a) Find unique states

   b) Find number of customers from each state

   c) List and count number of unique "fnames"

   d) List and count number of unique "cities"\

1. Create database in hive

```
#Create database in hive
hive> CREATE DATABASE assignment_db;
```

```
OK
Time taken: 0.427 seconds
```

2. Create Hive table inside database.

Create skeleton of `customers` table in hive.

```
hive>  CREATE TABLE assignment_db.customers(
    >  customer_id int,
    >  customer_fname varchar(45),
    >  customer_lname varchar(45),
    >  customer_email varchar(45),
    >  customer_password varchar(45),
    >  customer_street varchar(255),
```

```
    >  customer_city varchar(45),
    >  customer_state varchar(45),
    >  customer_zipcode varchar(45)) row format delimited fields terminated by ',' stored as textfile;
```

```
OK
Time taken: 1.195 seconds
```

3. Load Data into hive

Load data into `customers` **table from hdfs**

```
hive> load data inpath '/retail_db/customers' into table assignment_db.customers;
```

```
Loading data to table assignment_db.customers
OK
Time taken: 2.936 seconds
```

4. Load data from local file into Hive table (Hint: use keyword `load data local inpath` command)

```
hive> load data local inpath '/home/sois/Documents/file.csv' overwrite into table assignment_db.customers;
```

```
Loading data to table assignment_db.customers
OK
Time taken: 1.129 seconds
```

`file.csv` **- Local csv file used to load data from local machine to hive.**

```
102,Prathik,Shetty,prathik@gmail.com,12345pass,Manipal Road,Udupi,Karnataka,576124
103,Pradyumna,Alevoor,pradyumna@gmail.com,12345pass,Alevoor Road,Udupi,Karnataka,574118
```

Run queries for customer table for following:

a) Find unique states

```
hive> SELECT DISTINCT customer_state
    > FROM assignment_db.customers;
```

b) Find number of customers from each state

```
hive>  SELECT customer_state, count(*) AS COUNT
    >  FROM assignment_db.customers
    >  GROUP BY customer_state;
```

c) List and count number of unique `fnames` .

```
--Group people with same first name
hive> SELECT customer_fname, COUNT(*)
    > FROM assignment_db.customers
    > GROUP BY customer_fname;
```

```
--Total Unique fnames
hive> SELECT COUNT(DISTINCT customer_fname)
    > FROM assignment_db.customers;
```

d) List and count number of unique `cities` .

```
--Group people who are from same city
hive> SELECT customer_city, COUNT(*)
    > FROM assignment_db.customers
    > GROUP BY customer_city;
```

```
--Total Unique Cities
hive> SELECT COUNT(DISTINCT customer_city)
    > FROM assignment_db.customers;
```