Assignment 2.7

HIVE

1. **Create a database named 'custom'.**

Create database custom;

```
hive> show databases;
OK
anupam
default
Time taken: 7.073 seconds. Fetched: 2 row(s)
hive> create database custom;
OK
Time taken: 0.172 seconds
hive> show databases;
OK
anupam
custom
default
Time taken: 0.03 seconds, Fetched: 3 row(s)
hive>
```

2. **Create a table named temperature_data inside custom having below fields:**
   a. **date (mm-dd-yyyy) format**
   b. **zip code**
   c. **temperature**

   The table will be loaded from comma-delimited file.

   Load the dataset.txt (which is ',' delimited) in the table.

Using custom database for this activity;

```
hive> use custom;
OK
Time taken: 0.027 seconds
```

Creating temporary table

```
hive> CREATE EXTERNAL TABLE IF NOT EXISTS temp_table (the_date String,zipcode int,temperature int) ROW FORMAT DELIMIT
ED FIELDS TERMINATED BY ',';
OK
Time taken: 0.69 seconds
```

```
hive> LOAD data local inpath 'dataset.txt' overwrite into table temp_table;
Loading data to table custom.temp_table
OK
Time taken: 1.624 seconds
hive> select * from temp_table;
OK
10-01-1990      123112   10
14-02-1991      283901   11
10-03-1990      381920   15
10-01-1991      302918   22
12-02-1990      384902   9
10-01-1991      123112   11
14-02-1990      283901   12
10-03-1991      381920   16
10-01-1990      302918   23
12-02-1991      384902   10
10-01-1993      123112   11
14-02-1994      283901   12
10-03-1993      381920   16
10-01-1994      302918   23
12-02-1991      384902   10
10-01-1991      123112   11
14-02-1990      283901   12
10-03-1991      381920   16
10-01-1990      302918   23
12-02-1991      384902   10
Time taken: 2.607 seconds, Fetched: 20 row(s)
hive>
```

Creating the actual table

```
hive> CREATE EXTERNAL TABLE IF NOT EXISTS temparature_data (the_date string ,zipcode int,temperature int);
OK
Time taken: 0.068 seconds
hive>
```

Inserting data after converting to correct date format

```
hive> INSERT OVERWRITE TABLE temparature_data SELECT date_format(from_unixtime(unix_timestamp(the_date,'dd-MM-yyyy'))
,"MM-dd-YYYY"),zipcode,temperature from temp_table;
```

```
hive> select * from temparature_data;
OK
01-10-1990      123112  10
02-14-1991      283901  11
03-10-1990      381920  15
01-10-1991      302918  22
02-12-1990      384902  9
01-10-1991      123112  11
02-14-1990      283901  12
03-10-1991      381920  16
01-10-1990      302918  23
02-12-1991      384902  10
01-10-1993      123112  11
02-14-1994      283901  12
03-10-1993      381920  16
01-10-1994      302918  23
02-12-1991      384902  10
01-10-1991      123112  11
02-14-1990      283901  12
03-10-1991      381920  16
01-10-1990      302918  23
02-12-1991      384902  10
Time taken: 0.147 seconds, Fetched: 20 row(s)
hive>
```