Assignment - 8.1

Task1: Create a database named 'custom'. Create a table named temperature_data inside custom having below fields:

1. date (mm-dd-yyyy) format

2. zip code

3. temperature

The table will be loaded from comma-delimited file.

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

Answer1: Here first we create the database named – 'custom'

```
hive> create database custom;
OK
Time taken: 0.051 seconds
hive>
```

Use custome database to get in and perform all the following actions in the DB.

```
hive> use custom;
OK
Time taken: 0.029 seconds
hive>
```

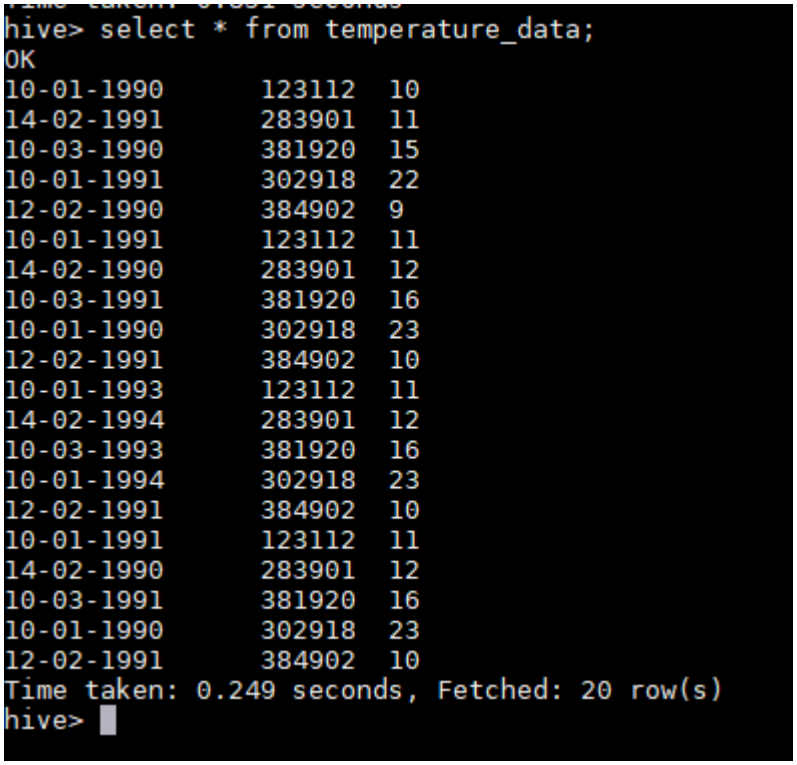
Now use below query to create the 'temperature_data' table to accept comma-delimited records.

```
hive> CREATE TABLE temperature_data(
    > tdate string,
    > zip_code INT, temp INT)
    > row format delimited fields terminated by ',';
OK
Time taken: 0.103 seconds
hive>
```

Now Load the table using LOAD command as mentioned in the screenshot.

```
hive> LOAD DATA LOCAL INPATH '/home/acadgild/dataset_Session14.txt' into table temperature_data;
Loading data to table custom.temperature_data
OK
Time taken: 0.851 seconds
hive>
```

Check the loaded records in the table using below mentioned command.



Task2:

● Fetch date and temperature from temperature_data where zip code is greater than 300000 and less than 399999.

● Calculate maximum temperature corresponding to every year from temperature_data table.

● Calculate maximum temperature from temperature_data table corresponding to those years which have at least 2 entries in the table.

● Create a view on the top of last query, name it temperature_data_vw.

● Export contents from temperature_data_vw to a file in local file system, such that each file is '|' delimited.

Answer2:

- Fetch date and temperature from temperature_data where zip code is greater than 300000 and less than 399999.
  Using Between clause we can get the desired output.

```
hive> select tdate, temp from temperature_data
    > where zip_code BETWEEN 300000 and 399999;
OK
10-03-1990      15
10-01-1991      22
12-02-1990      9
10-03-1991      16
10-01-1990      23
12-02-1991      10
10-03-1993      16
10-01-1994      23
12-02-1991      10
10-03-1991      16
10-01-1990      23
12-02-1991      10
Time taken: 0.729 seconds, Fetched: 12 row(s)
hive>
```

- Calculate maximum temperature corresponding to every year from temperature_data table.

```
hive> select substring(tdate,7,4), max(temp)
    > from temperature_data
    > group by substring(tdate,7,4);
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be avai
ne (i.e. spark, tez) or using Hive 1.X releases.
```

Output-

```
Total MapReduce CPU Time Spent: 4 seconds 220 msec
OK
1990    23
1991    22
1993    16
1994    23
Time taken: 34.568 seconds, Fetched: 4 row(s)
hive>
```

- Calculate maximum temperature from temperature_data table corresponding to those years which have at least 2 entries in the table.

```
hive> select substring(tdate,7,4), max(temp)
    > from temperature_data
    > group by substring(tdate,7,4)
    > having count(substring(tdate,7,4)) >= 2;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be avai
```

Output –

```
Total MapReduce CPU Time Spent: 5 seconds 460 msec
OK
1990    23
1991    22
1993    16
1994    23
Time taken: 38.141 seconds, Fetched: 4 row(s)
hive>
```

- Create a view on the top of last query, name it temperature_data_vw.

```
hive> create view temperature_data_vw as
    > select substring(tdate,7,4), max(temp)
    > from temperature_data
    > group by substring(tdate,7,4)
    > having count(substring(tdate,7,4)) >= 2;
OK
Time taken: 0.63 seconds
hive> show views;
OK
temperature_data_vw
Time taken: 0.068 seconds, Fetched: 1 row(s)
```

- Export contents from temperature_data_vw to a file in local file system, such that each file is '|' delimited.

```
hive> INSERT OVERWRITE LOCAL DIRECTORY '/home/acadgild/hivetohdfs'
    > row format delimited fields terminated by '|'
    > select * from temperature_data_vw;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available
```

Output:

```
[acadgild@localhost ~]$ ls -l /home/acadgild/hivetohdfs
total 4
-rw-r--r--. 1 acadgild acadgild 32 Apr  1 18:14 000000_0
[acadgild@localhost ~]$ cat /home/acadgild/hivetohdfs/000000_0
1990|23
1991|22
1993|16
1994|23
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```