Hive

Problem Statement

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.

Below are steps followed to complete the assignment.

Steps:

1) Created a database named 'custom', using the below command..

   CREATE DATABASE custom;

2) To start using the database for creating a table and etc, we have to use the below command.

   USE custom;

Screenshot of Mobaxterm for creating and using the DB:

```
hive> CREATE DATABASE custom;
OK
Time taken: 0.338 seconds
hive> USE custom;
OK
Time taken: 0.043 seconds
```

3) Next we have to create a table named temperature_data, with the fields mentioned in the problem statement. The command used is as below.

```
CREATE TABLE temperature_data
(
full_date STRING,
zip INT,
temperature INT
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ',';
```

Screenshot of Mobaxterm for creating the table temperature_data:

```
hive (custom)> CREATE TABLE temperature_data
             > (
             > full_date STRING,
             > zip INT,
             > temperature INT
             > )
             > ROW FORMAT DELIMITED
             > FIELDS TERMINATED BY ',';
OK
```

4) Next we have to load the contents of dataset.txt. I have actutally renamed the dataset.txt to temp_dataset.txt and stored the file in the path '/home/acadgild/hive/temp_dataset.txt'

So we are loading the contents of the temp_dataset.txt into the table temperature_data using the below commands.

```
LOAD DATA
LOCAL INPATH '/home/acadgild/hive/temp_dataset.txt'
INTO TABLE temperature_data;
```

Screenshot of Mobaxterm for loading temp_dataset.txt into temperature_data:

```
hive (custom)> LOAD DATA
             > LOCAL INPATH '/home/acadgild/hive/temp_dataset.txt'
             > INTO TABLE temperature_data;
Loading data to table custom.temperature_data
Table custom.temperature_data stats: [numFiles=1, totalSize=437]
OK
```

5) Now we can check if the contents are loaded into the temperature_data table using the below command.

select * from temperature_data;

Output :

```
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
```

Screenshot of Mobaxterm for select * from temperature_data

```
hive (custom)> select * from temperature_data;
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: 0.054 seconds, Fetched: 20 row(s)
```