

Bigdata Assignemnt 2.7

Create a database named 'custom'.

Create a table named temperature_data inside custom having below fields:

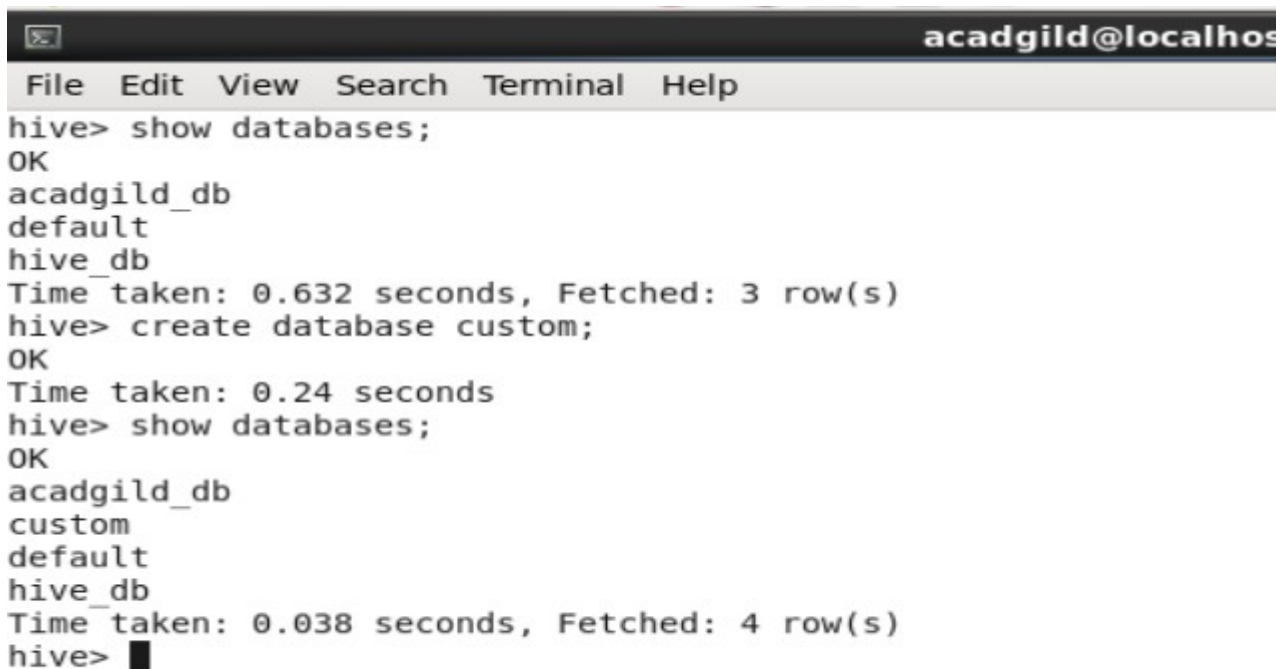
1. date (mm-dd-yyyy) format
2. zip code
3. temperature

The table will be loaded from comma-delimited file.

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

Solution -

- entered into hive shell by using command -
start-all.sh
hive
- Checked the databases and then created a database named as 'custom' and then gain checked if it is created or not.
Command used :-
show databases;
create database custom;
show databases;



```
acadgild@localhos
File Edit View Search Terminal Help
hive> show databases;
OK
acadgild_db
default
hive_db
Time taken: 0.632 seconds, Fetched: 3 row(s)
hive> create database custom;
OK
Time taken: 0.24 seconds
hive> show databases;
OK
acadgild_db
custom
default
hive_db
Time taken: 0.038 seconds, Fetched: 4 row(s)
hive> █
```

- Then in the 'custom' database , we created a external table named

temp_table and loaded the data into the table from the text file located in the local file system.

Command used -

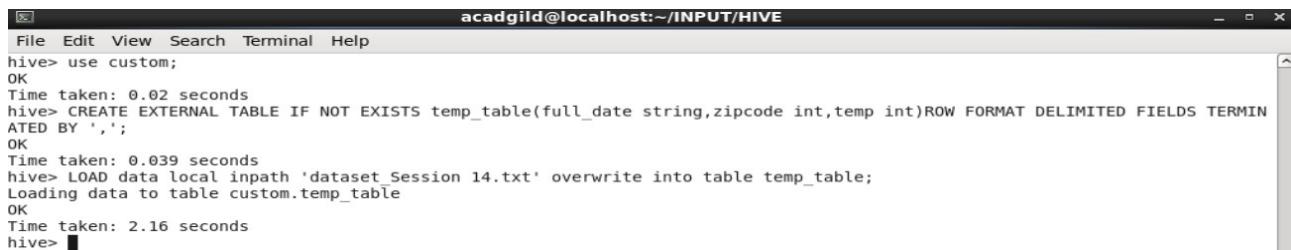
use custom;

CREATE EXTERNAL TABLE IF NOT EXISTS

temp_table(full_date string, zipcode int, temp int) ROW FORMAT

DELIMITED FIELDS TERMINATED BY ',' ;

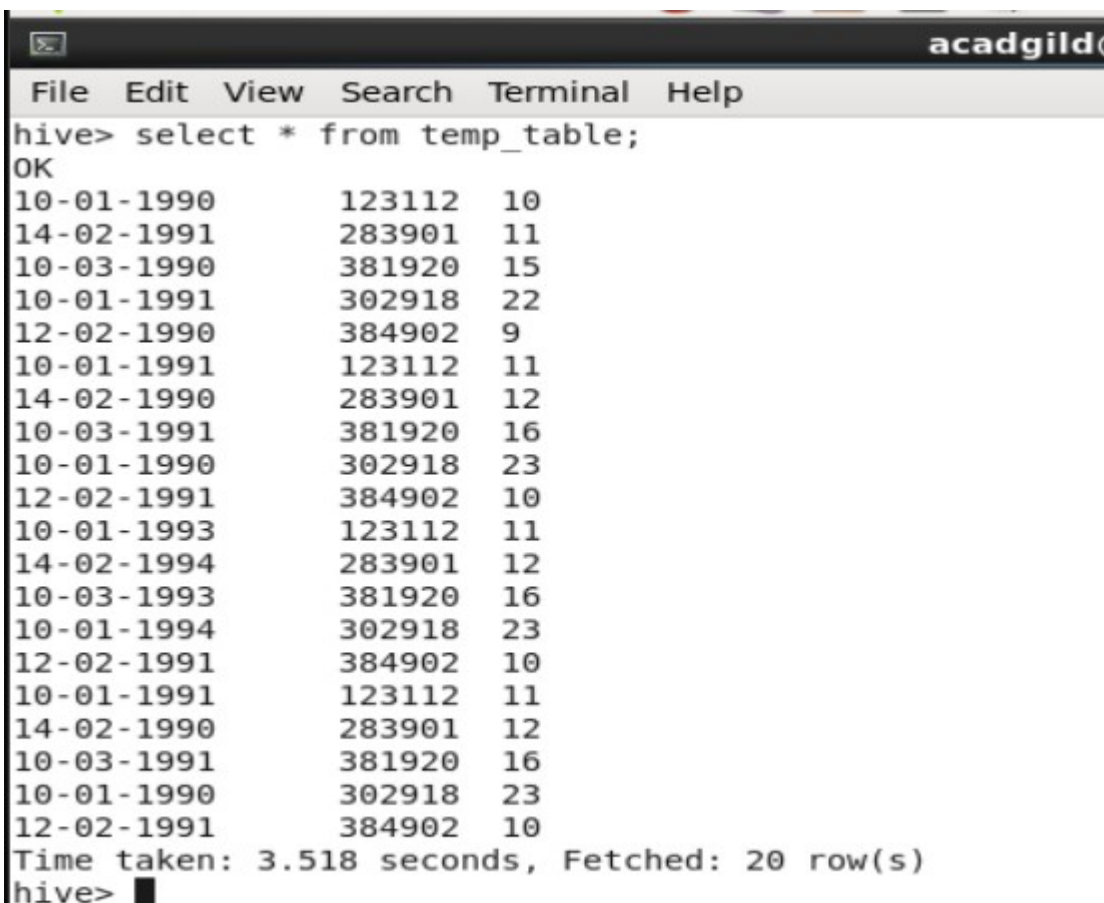
LOAD data local inpath 'dataset_Session 14.txt' overwrite into table temp_table;



```
acadgild@localhost:~/INPUT/HIVE
File Edit View Search Terminal Help
hive> use custom;
OK
Time taken: 0.02 seconds
hive> CREATE EXTERNAL TABLE IF NOT EXISTS temp_table(full_date string, zipcode int, temp int) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' ;
OK
Time taken: 0.039 seconds
hive> LOAD data local inpath 'dataset_Session 14.txt' overwrite into table temp_table;
Loading data to table custom.temp_table
OK
Time taken: 2.16 seconds
hive>
```

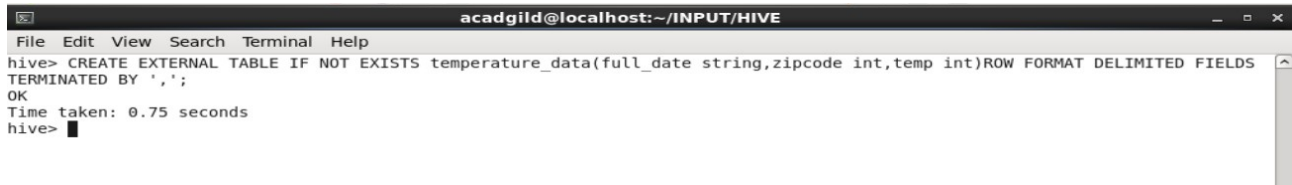
- Then we checked the contents of the table as it can be seen in the below screenshot the date format is dd-MM-yyyy.

select * from temp_table;



```
acadgild@localhost:~/INPUT/HIVE
File Edit View Search Terminal Help
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: 3.518 seconds, Fetched: 20 row(s)
hive>
```

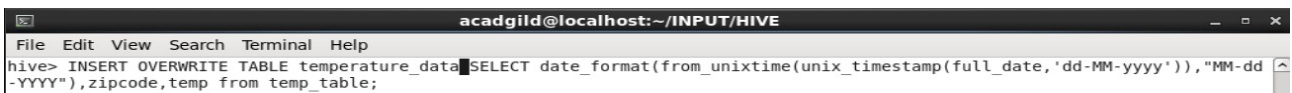
- We created another table to store the correct date format data .
**CREATE EXTERNAL TABLE IF NOT EXISTS
temperature_data(full_date string, zipcode int, temp int) ROW
FORMAT DELIMITED FIELDS TERMINATED BY ',' ;**



```
acadgild@localhost: ~/INPUT/HIVE
File Edit View Search Terminal Help
hive> CREATE EXTERNAL TABLE IF NOT EXISTS temperature_data(full_date string,zipcode int,temp int)ROW FORMAT DELIMITED FIELDS
TERMINATED BY ',';
OK
Time taken: 0.75 seconds
hive>
```

- Inserted the data into new table 'temperature_data' from the previous table temp_table with changed data format(MM-dd-YYYY) using unix_timestamp function.

**INSERT OVERWRITE TABLE temperature_data SELECT
date_format(from_unixtime(unix_timestamp(full_date,'dd-MM-
yyyy')),'MM-dd-YYYY'), zipcode , temp from temp_table;**



```
acadgild@localhost: ~/INPUT/HIVE
File Edit View Search Terminal Help
hive> INSERT OVERWRITE TABLE temperature_data SELECT date_format(from_unixtime(unix_timestamp(full_date,'dd-MM-yyyy')),'MM-dd
-YYYY'),zipcode,temp from temp_table;
```

- After inserting the data , we checked the content of the table
select * from temperature_data;

```
acadgild@local:~$ File Edit View Search Terminal Help
hive> select * from temperature_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: 0.2 seconds, Fetched: 20 row(s)
hive> █
```

Conclusion -

As in the table's content we can see the date format changed to (MM-dd-YYYY) and the rest column as per the question