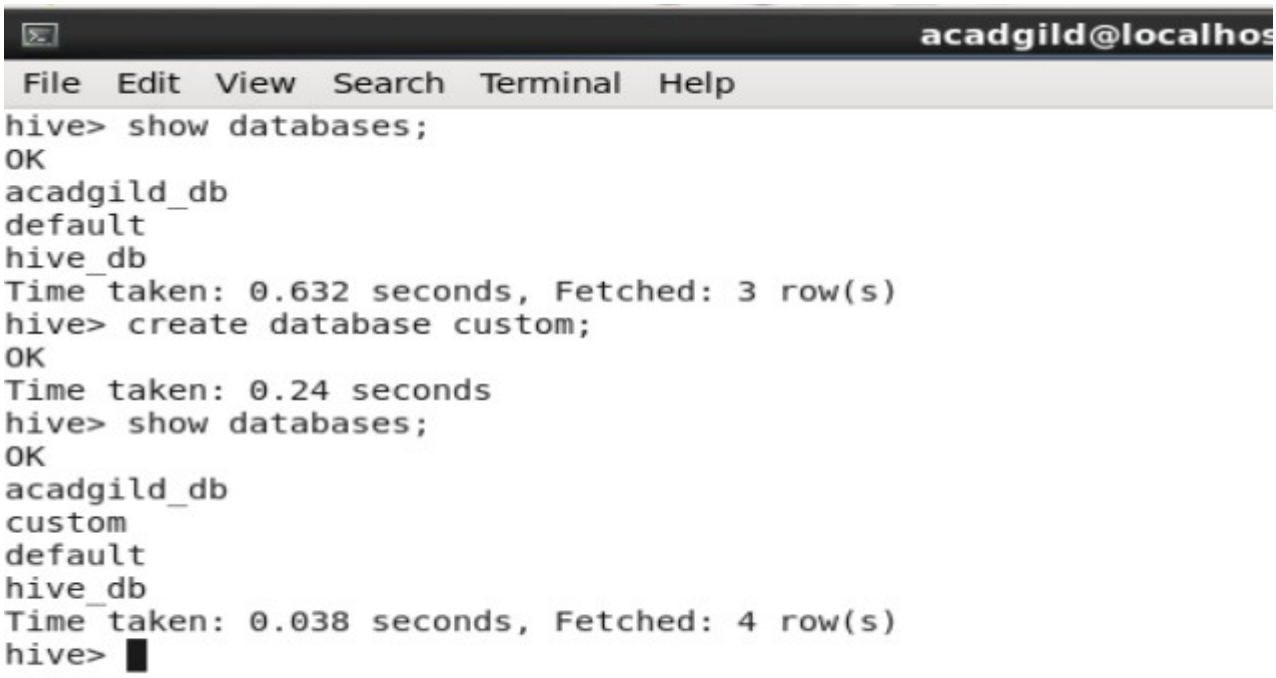
Bigdata Assignemnt 2.7

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.

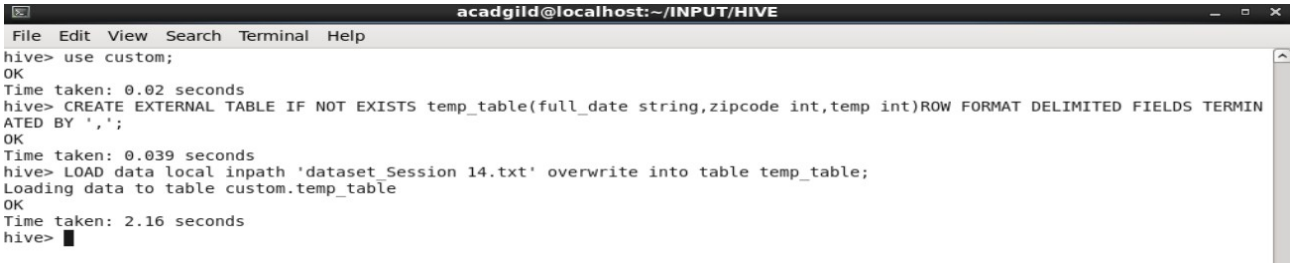
Solution -

- 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;

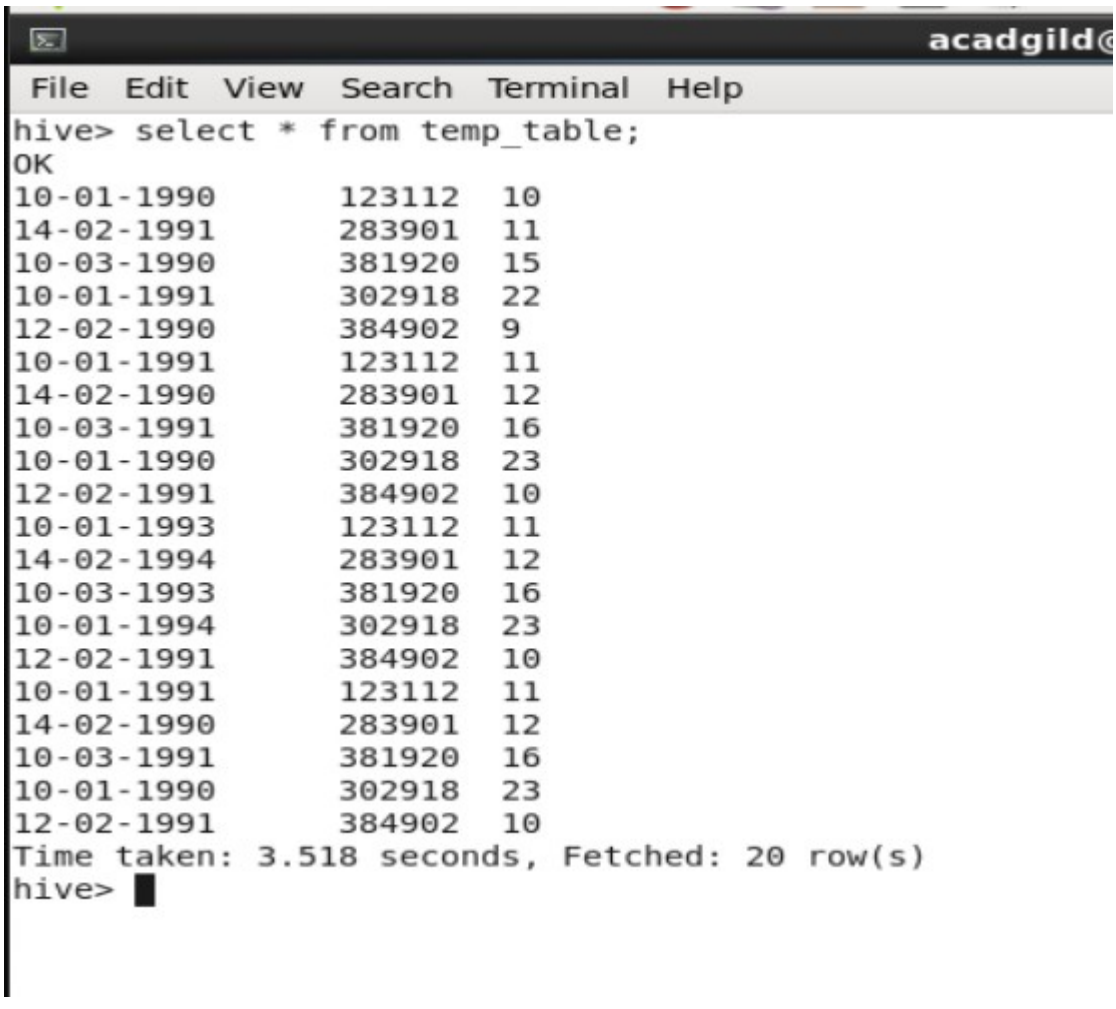


- Then in the 'custom' database , we created a external table named

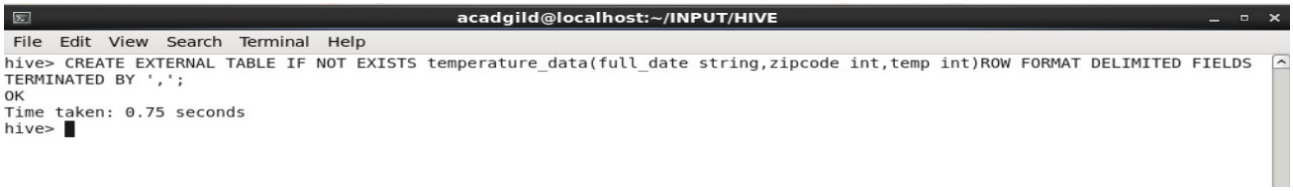
temp_table and loaded the data into the tablefrom the text file loacted in the local file system.
Command used -
use custom;
CREATE EXTERNAL TABLE IF NOT EXISTS temp_table(full_date string,zipcode , temp int) ROW FORMAT DELIMITED FIELDS TERMIBATED BY ',' ;
LOAD data local inpath 'dataset_Session 14.txt' overwrite into table temp_table;



- Then we checked the contents of the table as it can be ssen inthe below screenshot the date format is dd-MM-yyyy.
select * from temp_table;



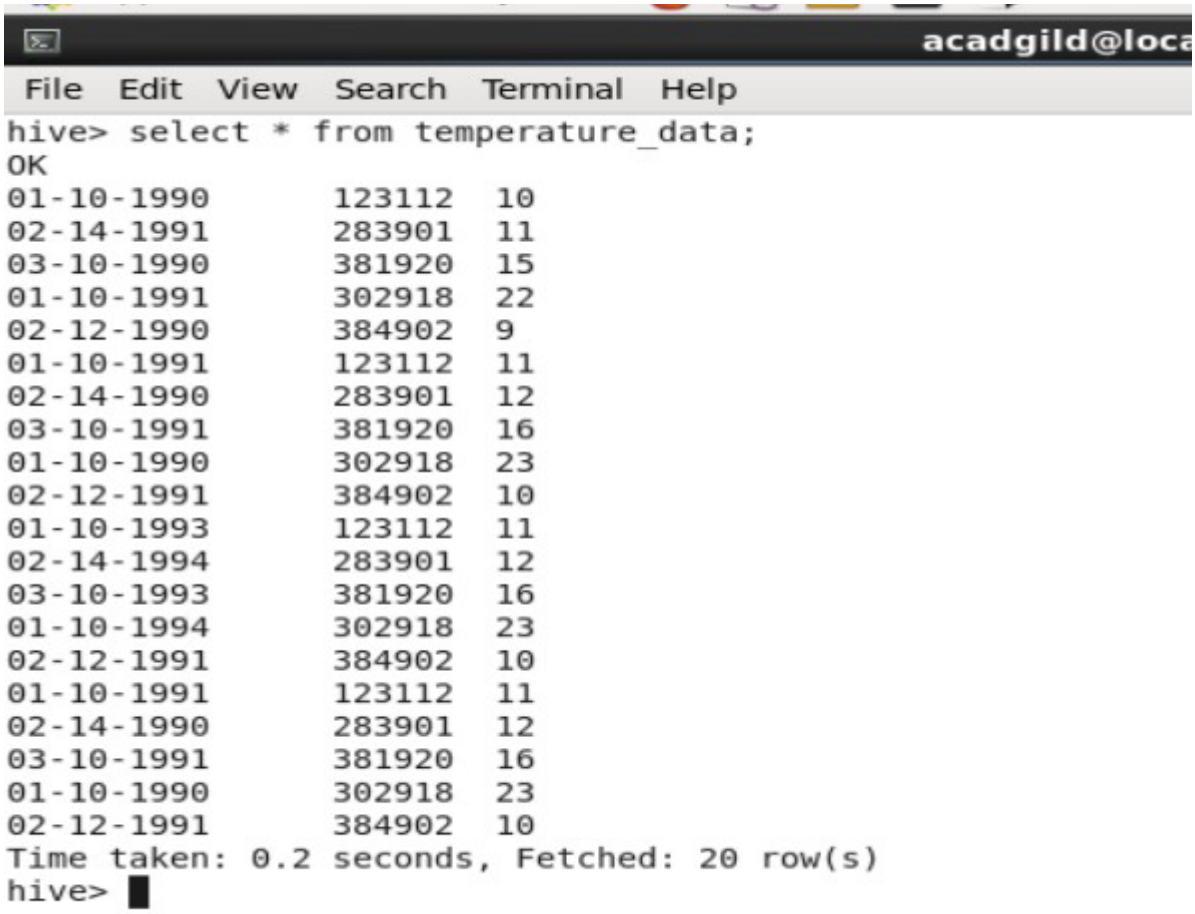
- We created another table to store the correct date format data .
  CREATE EXTERNAL TABLE IF NOT EXISTS temperature_data(full_date string, zipcode int, temp int) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' ;



- 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;

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



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