**Bigdata Assignment 2.8**

* entered into hive shell by using command -

**start-all.sh**

**hive**

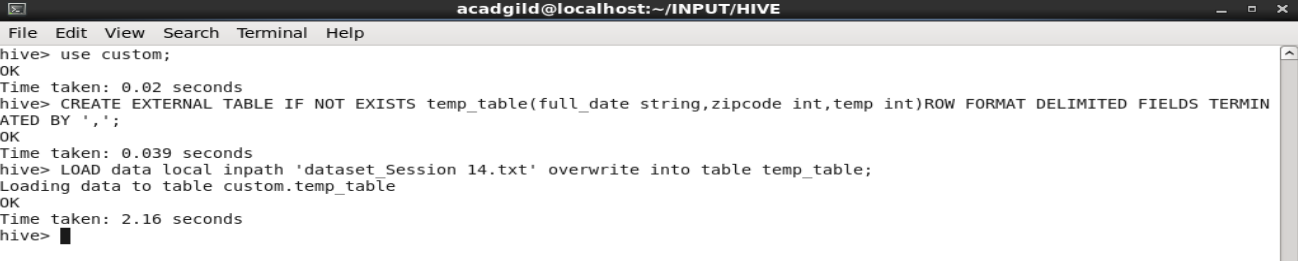
* Then in the 'custom' database , we created a external temporary table named temp\_table and loaded the data into the tablefrom the text file loacted in the local file system.

Command used -

**use custom;**

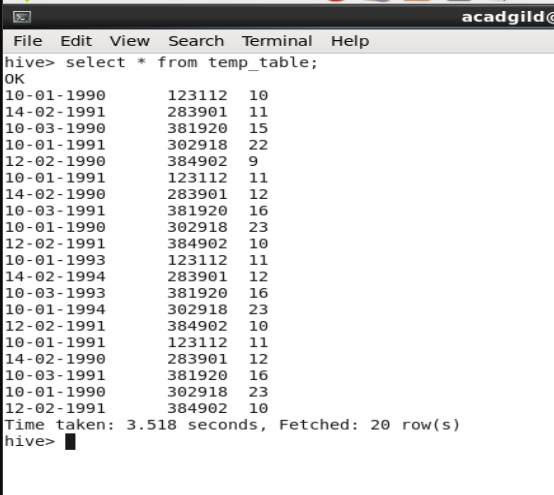
**CREATE EXTERNAL TABLE IF NOT EXISTS temp\_table(full\_date string,zipcode , temp int) ROW FORMAT DELIMITED FIELDS TERMIBATED BY ',' ;**

**LOAD data local inpath 'dataset\_Session 14.txt' overwrite into table temp\_table;**



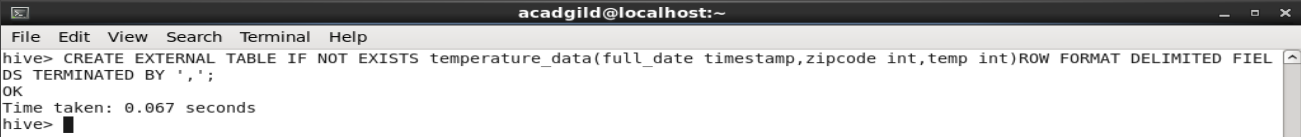
* Then we checked the contents of the table as it can be ssen int he below screenshot the date format is dd-MM-yyyy.

**select \* from temp\_table;**



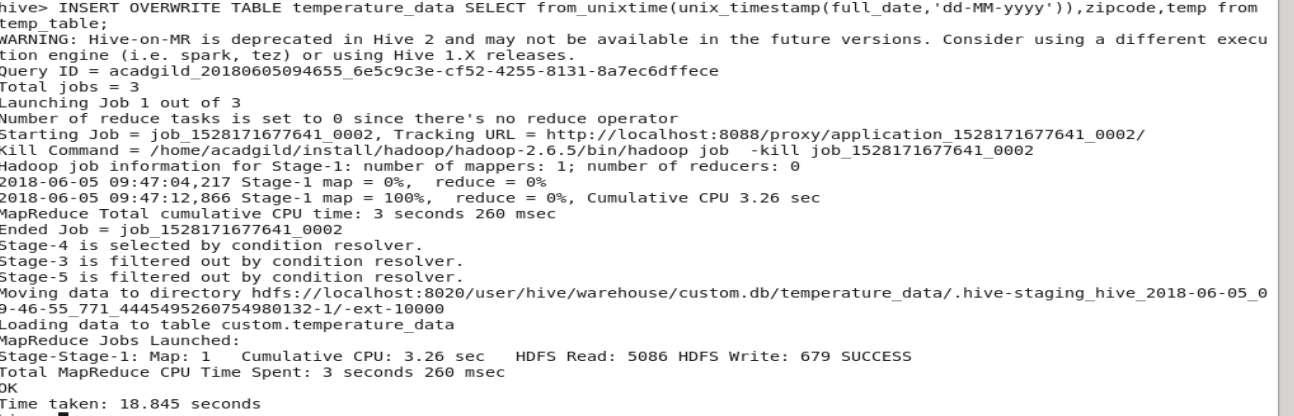
* We created actual table to store the correct date format data .

**CREATE EXTERNAL TABLE IF NOT EXISTS temperature\_data(full\_date timestamp, zipcode int, temp int) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' ;**



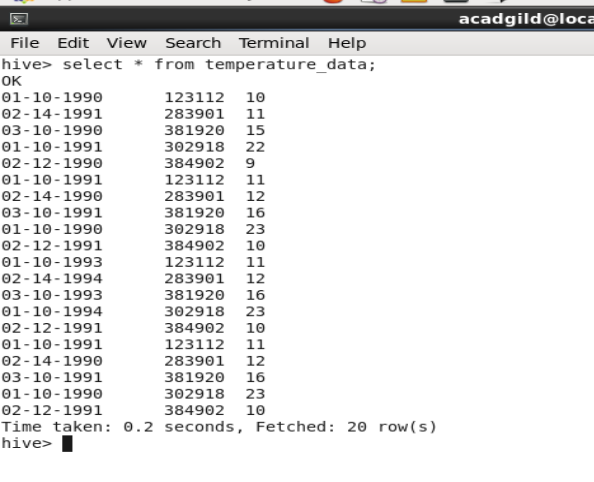
* Inserted the data into new table 'temperature\_data' from the previous table temp\_table using unix\_timestamp function.

**INSERT OVERWRITE TABLE temperature\_data SELECT from\_unixtime(unix\_timestamp(full\_date,'dd-MM-yyyy')), zipcode , temp from temp\_table;**



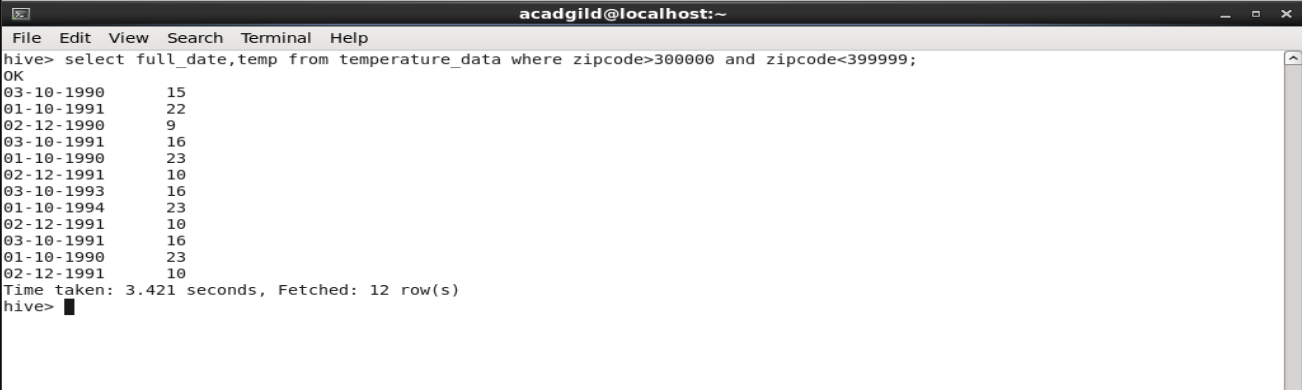
* After inserting the data , we checked the content of the table

**select \* from temperature\_data;**



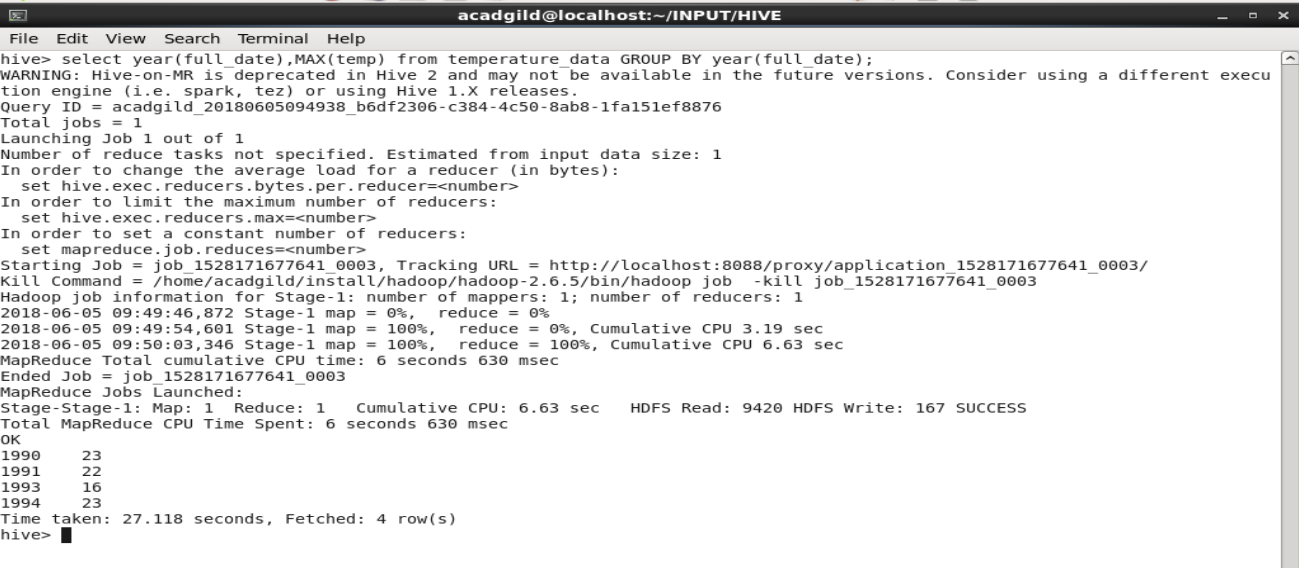
1. Fetch date and temperature from temperature\_data where zip code is greater than 300000 and less than 399999.

Ans - **select full\_date , temp from tempearture\_data where zipcode>300000 and zipcode<399999;**



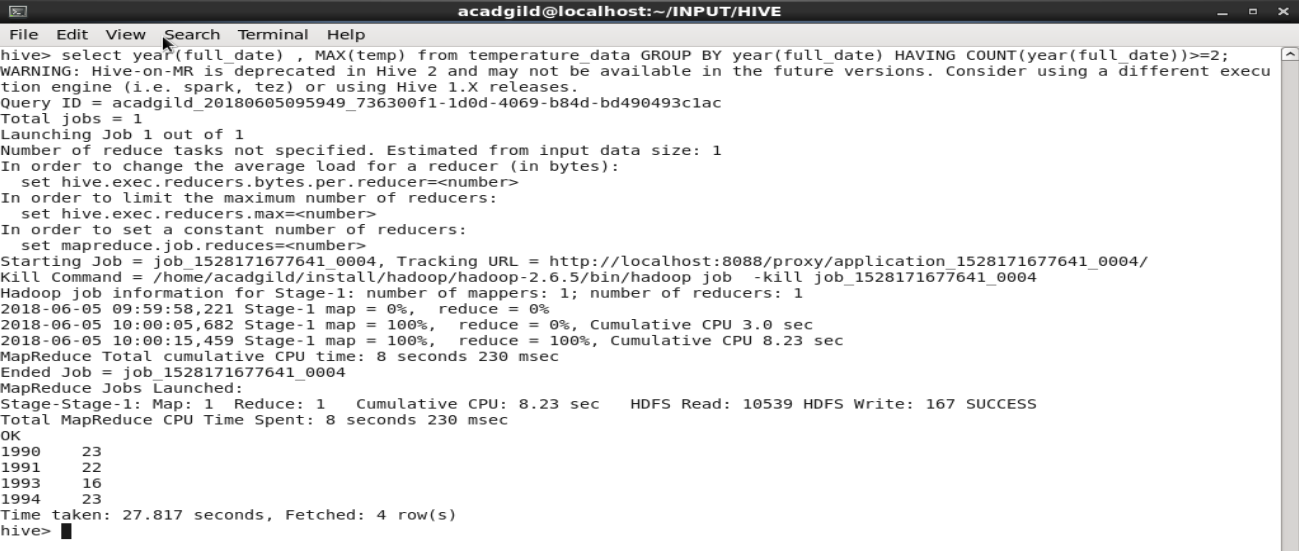
2. Calculate maximum temperature corresponding to every year from temperature\_data table.

Ans – **select year(full\_date) , MAX(temp) from temperature\_data GROUP BY year(full\_date);**



3. Calculate maximum temperature from temperature\_data table corresponding to those years which have at least 2 entries in the table.

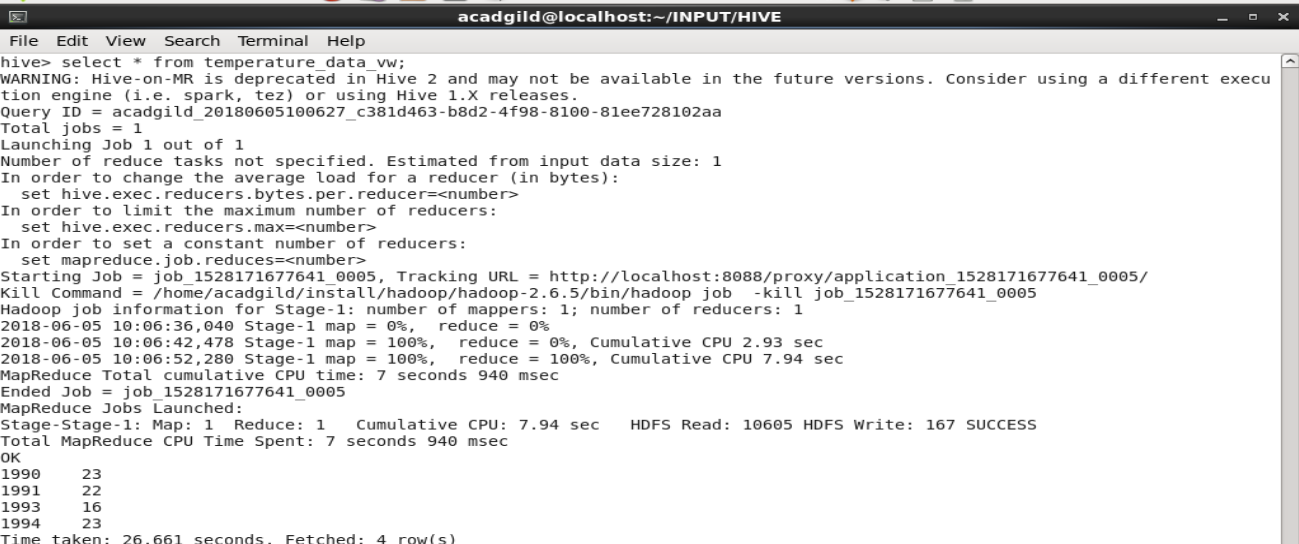
Ans - **select year(full\_date) , MAX(temperature)\_data GROUP BY year(full\_date) HAVING COUNT(year(full\_date))>=2;**



4. Create a view on the top of last query, name it temperature\_data\_vw.

Ans – **CREATE VIEW tempearture\_data\_vw as select year(full\_date) , MAX(temp) from tempearture\_data GROUP BY year(full\_date) HAVING COUNT(year(full\_date))>=2;**





5.Export contents from temperature\_data\_vw to a file in local file system, such that each file is '|' delimited.

Ans – **INSERT OVERWRITE LOCAL DIRECTORY 'home/acadgild/INPUT/HIVE/tempearture\_data\_vw.txt' ROW FORMAT DELIMITED FIELDS TERMINATED BT '|' select \* from temperature\_data\_vw;**



**The content of the file is shown in the below screenshot.**

