**Assignment 8.1(Hive Basics)**

**---------------------------------------**

**Task 1 :**

**-----------**

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.

**Query for table creation:**

**----------------------------------**

create table temperature\_data(dat String, zip Int,temp Int) row format delimited fields terminated by ',';

**Query for loading data into the table:**

**----------------------------------------------------**

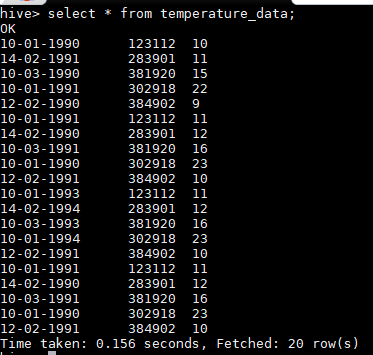
LOAD DATA LOCAL INPATH '/home/acadgild/dataset\_Session 14.txt into table temperature\_data;

**Output Screenshot:**

**---------------------------**

Table created containing the loaded data:

--------------------------------------------------------------



**Task 2**

**-----------**

1. Fetch date and temp from temperature\_data where zip code is greater than

300000 and less than 399999.

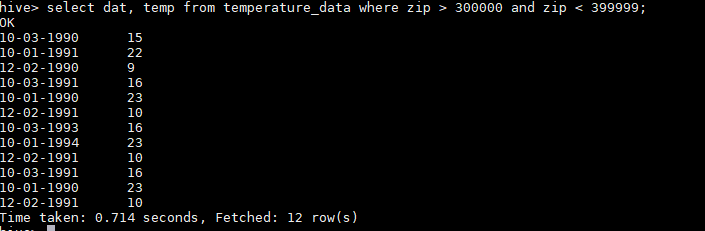
**Query for the above task:**

**---------------------------------------**

select dat, temperature from temperature\_data where zip > 300000 and zip < 399999.

**Output:**

**-----------**



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

table.

**Query for the above task:**

**--------------------------------------**

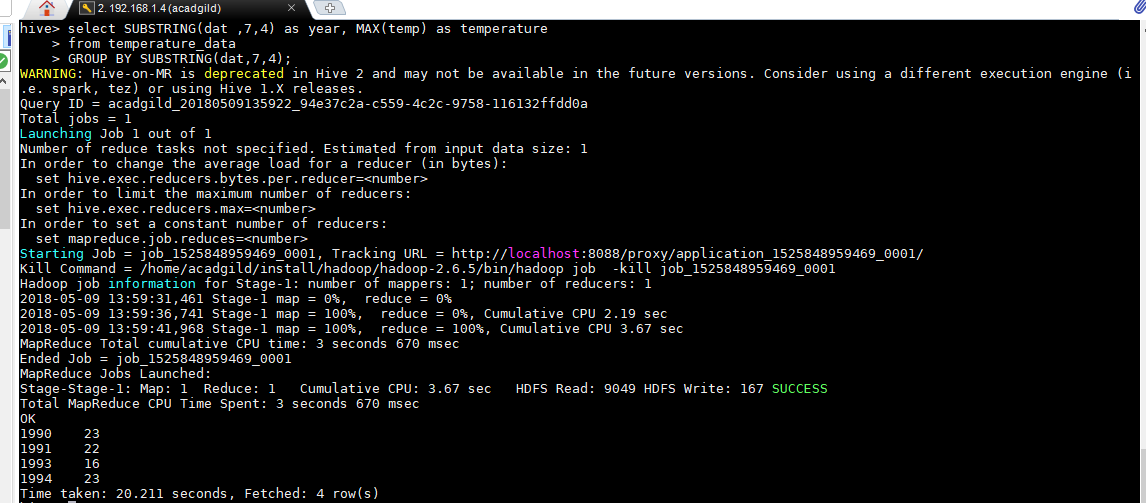
select SUBSTRING(dat ,7,4) as year, MAX(temp) as temperature

from temperature\_data

GROUP BY SUBSTRING(dat,7,4);

**Output:**

**-----------**



3. Calculate maximum temperature from temperature\_data table corresponding to those

years which have at least 2 entries in the table.

**Query for the above task:**

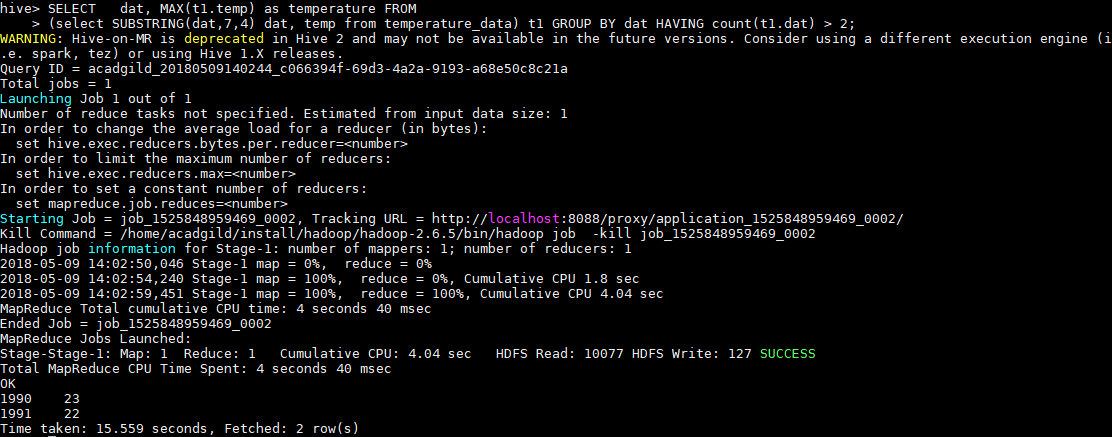
**--------------------------------------**

SELECT dat, MAX(t1.temp) as temperature FROM

(select SUBSTRING(dat,7,4) dat, temp from temperature\_data) t1 GROUP BY dat HAVING count(t1.dat) > 2;

**Output:**

**-----------**



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

**Query for the above task:**

**--------------------------------------**

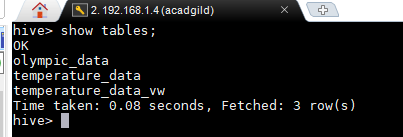
create view temperature\_data\_vw as SELECT dat,

MAX(t1.temp) as temperature FROM

(select SUBSTRING(dat,7,4) dat, temp from temperature\_data) t1 GROUP BY dat HAVING count(t1.dat) > 2;

**Output:**

**-----------**



4. Export contents from temperature\_data\_vw to a file in local file system, such that each

file is '|' delimited

**Query for the above task:**

**--------------------------------------**

INSERT OVERWRITE LOCAL DIRECTORY '/home/acadgild/temperature\_output' ROW FORMAT DELIMITED FIELDS TERMINATED BY '|'

**Output:**

**-----------**

