Task-1

Input File temperature_data.hive

```
create database IF NOT EXISTS custom;
use custom;
create table if not exists temperature_data
(date1 string,
zipcode int,
temperature int)
row format delimited
fields terminated by ','
lines terminated by '\n'
stored as textfile;
describe temperature_data;
load data local inpath '/home/acadgild/hive/dataset.txt' overwrite into table temperature_data;
select * from temperature_data;
~
~
```

[acadgild@localhost hive]$ hive -f temperature_data.hive

```
4j2.properties Async: true
OK
Time taken: 11.538 seconds
OK
Time taken: 0.045 seconds
OK
Time taken: 1.03 seconds
OK
date1                   string
zipcode                 int
temperature             int
Time taken: 0.31 seconds, Fetched: 3 row(s)
Loading data to table custom.temperature_data
OK
Time taken: 2.186 seconds
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: 3.37 seconds, Fetched: 20 row(s)
```

Task-2

Query1:

Fetch date and temperature from temperature_data where zip code is greater than 300000 and less than 399999..?

SOL:

select zipcode from temperature_data where zipcode between 300000 and 399999;

output:

```
hive> select zipcode from temperature_data where zipcode between 300000 and 399999;
OK
zipcode
381920
302918
384902
381920
302918
384902
381920
302918
384902
381920
302918
384902
```

Query2:

Calculate maximum temperature corresponding to every year from temperature_data table..?

SOL:

select FROM_UNIXTIME(UNIX_TIMESTAMP(date1,'DD-MM-YYYY'),'YYYY') as year,

max(temperature) as maxtemp **from** temperature_data

group by FROM_UNIXTIME(UNIX_TIMESTAMP(date1,'DD-MM-YYYY'),'YYYY');

output:

```
OK
year    maxtemp
1990    23
1991    22
1993    16
1994    23
```

Query3:

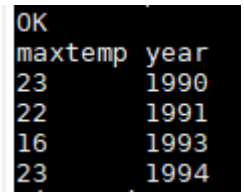
Calculate maximum temperature from temperature_data table corresponding to those years which have at least 2 entries in the table..?

Sol:

select max(temperature) as maxtemp,FROM_UNIXTIME(UNIX_TIMESTAMP(date1,'DD-MM-YYYY'),'YYYY') as year from temperature_data **group by** FROM_UNIXTIME(UNIX_TIMESTAMP(date1,'DD-MM-YYYY'),'YYYY')

having count(FROM_UNIXTIME(UNIX_TIMESTAMP(date1,'DD-MM-YYYY'),'YYYY'))>=2 ;

output:



Query-4:

Create a view on the top of last query, name it temperature_data_vw..?

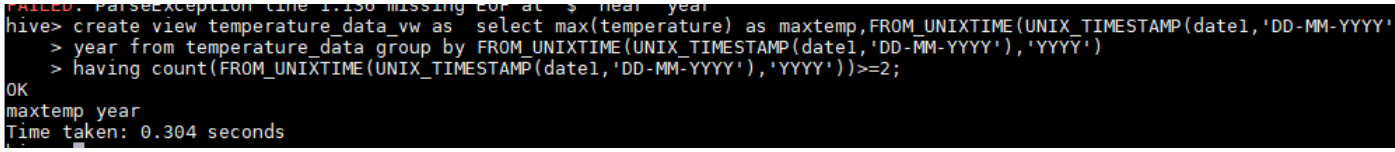
SOL:

create view temperature_data_vw as select max(temperature) as maxtemp,FROM_UNIXTIME(UNIX_TIMESTAMP(date1,'DD-MM-YYYY'),'YYYY') as year $

group by FROM_UNIXTIME(UNIX_TIMESTAMP(date1,'DD-MM-YYYY'),'YYYY')

having count(FROM_UNIXTIME(UNIX_TIMESTAMP(date1,'DD-MM-YYYY'),'YYYY'))>=2 ;

output:



Query-5:

Export contents from temperature_data_vw to a file in local file system, such that each file is '|' delimited

SOL:

insert overwrite local directory '/home/acadgild/hive/**viewdata1'** row format delimited fields terminated by '|' select * from temperature_data_vw;

output: