

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

```
[hive> select dt, temperature from temperature_data where zipcode > 300000 and zipcode < 399999;
OK
10-03-1990      15
10-01-1991      22
12-02-1990       9
10-03-1991      16
10-01-1990      23
12-02-1991      10
10-03-1993      16
10-01-1994      23
12-02-1991      10
10-03-1991      16
10-01-1990      23
12-02-1991      10
Time taken: 3.788 seconds, Fetched: 12 row(s)
```

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

```
[hive> select from_unixtime(unix_timestamp(dt, 'MM-DD-YYYY'), 'YYYY') , Max(temperature) from temperature_data group by from_unixtime(unix_timestamp(dt, 'MM-DD-YYYY'), 'YYYY');
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = Disha_20170915171715_1ec7e70b-b57e-45f2-9abd-91ecf6fcf234
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Job running in-process (local Hadoop)
2017-09-15 17:17:19,711 Stage-1 map = 0%,   reduce = 0%
2017-09-15 17:17:20,726 Stage-1 map = 100%, reduce = 100%
Ended Job = job_local1565302829_0001
MapReduce Jobs Launched:
Stage-Stage-1:  HDFS Read: 1676 HDFS Write: 0 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
1990      23
1991      22
1993      16
1994      23
Time taken: 5.503 seconds, Fetched: 4 row(s)
```

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

```
[hive> select from_unixtime(unix_timestamp(dt, 'MM-DD-YYYY'), 'YYYY') , Max(temperature) from temperature_data group by from_unixtime(unix_timestamp(dt, 'MM-DD-YYYY'), 'YYYY') having count(from_unixtime(unix_timestamp(dt, 'MM-DD-YYYY'), 'YYYY')) >= 2;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = Disha_20170916224104_dd4e9a5d-3ca1-44e5-b9ee-c134c962bac0
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Job running in-process (local Hadoop)
2017-09-16 22:41:07,764 Stage-1 map = 100%,   reduce = 0%
2017-09-16 22:41:08,771 Stage-1 map = 100%,   reduce = 100%
Ended Job = job_local1575588027_0002
MapReduce Jobs Launched:
Stage-Stage-1:  HDFS Read: 1676 HDFS Write: 0 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
1990      23
1991      22
1993      16
1994      23
Time taken: 4.358 seconds, Fetched: 4 row(s)
hive>
```

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

```

Time taken: 4.212 seconds
hive> CREATE VIEW temperature_data_vw as select from_unixtime(unix_timestamp(dt,'MM-DD-YYYY'),'YYYY'), Max(temperature) from temperature_data group by from_unixtime(unix_timestamp(dt,'MM-DD-YYYY'),'YYYY') having count(from_unixtime(unix_timestamp(dt,'MM-DD-YYYY'),'YYYY')) >= 2;
OK
Time taken: 1.548 seconds
hive> EXPLAIN VIEW temperature_data_vw TO '/Users/Disha/temperature_data_vw.txt'

[hive> select * from temperature_data_vw;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = Disha_20170916230335_25458d57-ef21-41a2-a51f-ef8ddd148373
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
set mapreduce.job.reduces=<number>
Job running in-process (local Hadoop)
2017-09-16 23:03:39,434 Stage-1 map = 100%,  reduce = 100%
Ended Job = job_local1492575314_0003
MapReduce Jobs Launched:
Stage-Stage-1:  HDFS Read: 2514 HDFS Write: 0 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
1990      23
1991      22
1993      16
1994      23
Time taken: 3.84 seconds, Fetched: 4 row(s)

```

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

```

hive> INSERT OVERWRITE LOCAL DIRECTORY '/Users/Disha/temperature_data' ROW FORMAT DELIMITED FIELDS TERMINATED BY '|' SELECT * FROM temperature_data_vw;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = Disha_20170916230802_c6327a80-b9e2-4abf-bbf5-128939f41a20
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
set mapreduce.job.reduces=<number>
Job running in-process (local Hadoop)
2017-09-16 23:08:05,599 Stage-1 map = 100%,  reduce = 100%
Ended Job = job_local1119667805_0005
Moving data to local directory /Users/Disha/temperature_data
MapReduce Jobs Launched:
Stage-Stage-1:  HDFS Read: 4190 HDFS Write: 0 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
Time taken: 3.557 seconds
hive>

```

```

Chaitanyas-MacBook-Pro:~ Disha$ mkdir temperature_data
Chaitanyas-MacBook-Pro:~ Disha$ cd temperature_data
Chaitanyas-MacBook-Pro:temperature_data Disha$ pwd
/Users/Disha/temperature_data
Chaitanyas-MacBook-Pro:temperature_data Disha$ ls -lrt
total 8
-rw-r--r--  1 Disha  staff   32 Sep 16 23:08 000000_0
Chaitanyas-MacBook-Pro:temperature_data Disha$ vi 000000_0
Chaitanyas-MacBook-Pro:temperature_data Disha$ cat 000000_0
1990|23
1991|22
1993|16
1994|23
Chaitanyas-MacBook-Pro:temperature_data Disha$

```