

## Hive Assignment 6.2

Solution 1 :

**select dates, temperature from temperature\_data where zipcode > '300000' and zipcode < '399999';**

```
hive> select dates, 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: 0.27 seconds, Fetched: 12 row(s)
hive>
```

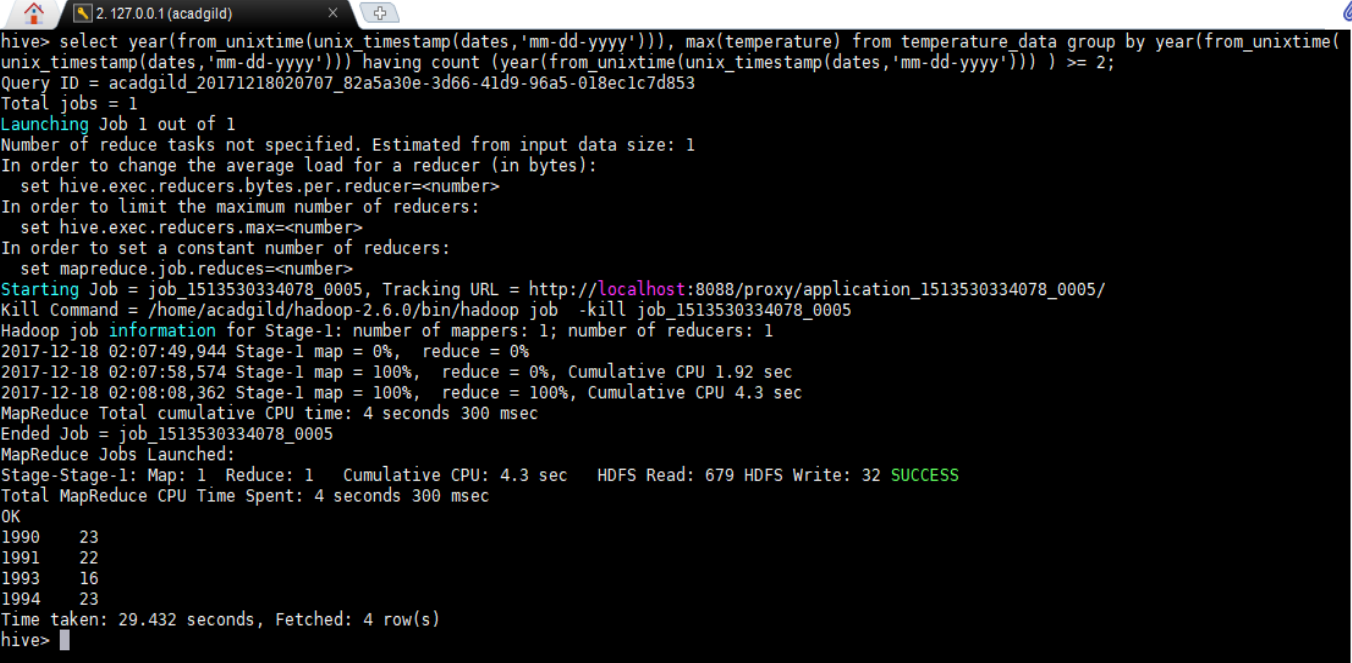
2) Solution 2:

**select year(from\_unixtime(unix\_timestamp(dates,'mm-dd-yyyy'))),max(temperature) from temperature\_data group by year(from\_unixtime(unix\_timestamp(dates,'mm-dd-yyyy')));**

```
2.127.0.0.1 (acadgild) x
> select year(from_unixtime(unix_timestamp(dates,'mm-dd-yyyy'))),max(temperature) from temperature_data group by year(from_unixtime(unix_timestamp(dates,'mm-dd-yyyy')));
Query ID = acadgild_20171218015959_d19c959c-5c4e-4c6d-9d01-fd08f4f4ebf30
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>
Starting Job = job_1513530334078_0003, Tracking URL = http://localhost:8088/proxy/application_1513530334078_0003/
Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job -kill job_1513530334078_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2017-12-18 01:59:34,971 Stage-1 map = 0%, reduce = 0%
2017-12-18 01:59:43,950 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.9 sec
2017-12-18 01:59:53,793 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.32 sec
MapReduce Total cumulative CPU time: 3 seconds 320 msec
Ended Job = job_1513530334078_0003
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.32 sec HDFS Read: 679 HDFS Write: 32 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 320 msec
OK
1990      23
1991      22
1993      16
1994      23
Time taken: 32.204 seconds, Fetched: 4 row(s)
```

3) Solution 3 :

```
select year(from_unixtime(unix_timestamp(dates,'mm-dd-yyyy'))), max(temperature) from
temperature_data group by year(from_unixtime(unix_timestamp(dates,'mm-dd-yyyy')))) having
count (year(from_unixtime(unix_timestamp(dates,'mm-dd-yyyy')))) ) >= 2;
```

A screenshot of a terminal window with a dark background. The window title is '2. 127.0.0.1(acadgild)'. The terminal shows the execution of a Hive query. The query is: `hive> select year(from_unixtime(unix_timestamp(dates,'mm-dd-yyyy'))), max(temperature) from temperature_data group by year(from_unixtime(unix_timestamp(dates,'mm-dd-yyyy')))) having count (year(from_unixtime(unix_timestamp(dates,'mm-dd-yyyy')))) ) >= 2;` The output shows the query ID, total jobs (1), and job launching information. It then displays the progress of the job, including the number of reduce tasks, estimated input data size, and various configuration settings. The job is identified as `job_1513530334078_0005`. The terminal shows the job's progress, including the number of mappers and reducers, and the cumulative CPU time. The job is completed successfully, and the results are displayed as a table with two columns: year and max(temperature). The results are: 1990 23, 1991 22, 1993 16, 1994 23. The time taken is 29.432 seconds, and 4 rows are fetched. The prompt `hive>` is visible at the bottom.

Solution : 4

```
create view temperature_data_vw as select year(from_unixtime(unix_timestamp(dates,'mm-dd-
yyyy'))), max(temperature)
```

```
from temperature_data group by year(from_unixtime(unix_timestamp(dates,'mm-dd-yyyy')))
```

```
having count (year(from_unixtime(unix_timestamp(dates,'mm-dd-yyyy')))) ) >= 2;
```

```
select * from temperature_data_vw;
```

```

hive> create view temperature_data_vw as select year(from_unixtime(unix_timestamp(dates,'mm-dd-yyyy'))), max(temperature) from temperatur
e_data group by year(from_unixtime(unix_timestamp(dates,'mm-dd-yyyy')))) having count (year(from_unixtime(unix_timestamp(dates,'mm-dd-yyyy
'))) ) >= 2;
OK
Time taken: 0.43 seconds
hive> select * from temperature_data_vw;
Query ID = acadgild_20171218021212_909409bb-3c71-485f-bc71-32101988aac7
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>
Starting Job = job_1513530334078_0006, Tracking URL = http://localhost:8088/proxy/application_1513530334078_0006/
Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job -kill job_1513530334078_0006
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2017-12-18 02:12:34,322 Stage-1 map = 0%, reduce = 0%
2017-12-18 02:12:42,004 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.7 sec
2017-12-18 02:12:52,868 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.08 sec
MapReduce Total cumulative CPU time: 4 seconds 80 msec
Ended Job = job_1513530334078_0006
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.08 sec HDFS Read: 679 HDFS Write: 32 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 80 msec
OK
1990 23
1991 22
1993 16
1994 23
Time taken: 28.734 seconds, Fetched: 4 row(s)
hive> █

```

Solution : 5

insert overwrite local directory '/home/acadgild/Downloads/' row format delimited fields terminated by '|' select \* from temperature\_data\_vw;

```

hive> insert overwrite local directory '/home/acadgild/Downloads/' row format delimited fields terminated by '|' select * from temperatur
e_data vw;
Query ID = acadgild_20171218021717_07ae1ea7-20f0-4e1a-bbfa-f8ad5bb16cf9
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>
Starting Job = job_1513530334078_0007, Tracking URL = http://localhost:8088/proxy/application_1513530334078_0007/
Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job -kill job_1513530334078_0007
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2017-12-18 02:17:53,382 Stage-1 map = 0%, reduce = 0%
2017-12-18 02:18:01,986 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.92 sec
2017-12-18 02:18:10,653 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.89 sec
MapReduce Total cumulative CPU time: 3 seconds 890 msec
Ended Job = job_1513530334078_0007
Copying data to local directory /home/acadgild/Downloads
Copying data to local directory /home/acadgild/Downloads
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.89 sec HDFS Read: 679 HDFS Write: 32 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 890 msec
OK
Time taken: 27.833 seconds
hive> █

```

```
Last login: Sun Dec 17 22:33:43 2017 from 10.0.2.2
[acadgild@localhost ~]$ cd /home/acadgild/Downloads/
[acadgild@localhost Downloads]$ ls -ltrh
total 4.0K
-rw-r--r--. 1 acadgild acadgild 32 Dec 18 02:18 000000_0
[acadgild@localhost Downloads]$ cat 000000_0
1990|23
1991|22
1993|16
1994|23
[acadgild@localhost Downloads]$ █
```