**Assignment 6.2**

**--Problem 1**

hive> SELECT date,temperature FROM temperature\_data where zip\_code >300000 AND zip\_code < 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

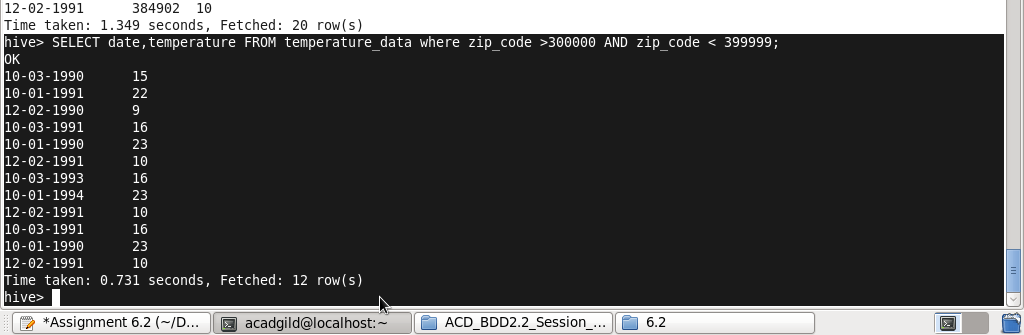
10-03-1991 16

10-01-1990 23

12-02-1991 10

Time taken: 0.731 seconds, Fetched: 12 row(s)

hive>



**--problem 2**

hive> SELECT year(cast(to\_date(from\_unixtime(unix\_timestamp(date, 'dd-mm-yyyy')))as date)), MAX(temperature) from temperature\_data group by year(cast(to\_date(from\_unixtime(unix\_timestamp(date, 'dd-mm-yyyy')))as date));

Query ID = acadgild\_20170916220101\_2de1f408-95ff-4a7d-9c4d-17d227fea2c1

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\_1505571178864\_0001, Tracking URL = http://localhost:8088/proxy/application\_1505571178864\_0001/

Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job -kill job\_1505571178864\_0001

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1

2017-09-16 22:01:49,986 Stage-1 map = 0%, reduce = 0%

2017-09-16 22:02:12,737 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.02 sec

2017-09-16 22:02:30,682 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.74 sec

MapReduce Total cumulative CPU time: 6 seconds 740 msec

Ended Job = job\_1505571178864\_0001

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 6.74 sec HDFS Read: 671 HDFS Write: 32 SUCCESS

Total MapReduce CPU Time Spent: 6 seconds 740 msec

OK

1990 23

1991 22

1993 16

1994 23

Time taken: 90.797 seconds, Fetched: 4 row(s)

hive>

**--problem 3**

hive> SELECT year(cast(to\_date(from\_unixtime(unix\_timestamp(date, 'dd-mm-yyyy')))as date)), MAX(temperature) from temperature\_data group by year(cast(to\_date(from\_unixtime(unix\_timestamp(date, 'dd-mm-yyyy')))as date)) having COUNT(year(cast(to\_date(from\_unixtime(unix\_timestamp(date, 'dd-mm-yyyy')))as date))) > 1;

Query ID = acadgild\_20170916222323\_f7d5591b-f433-409a-84ca-c772bf77c042

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\_1505571178864\_0003, Tracking URL = http://localhost:8088/proxy/application\_1505571178864\_0003/

Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job -kill job\_1505571178864\_0003

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1

2017-09-16 22:24:12,893 Stage-1 map = 0%, reduce = 0%

2017-09-16 22:24:29,855 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.29 sec

2017-09-16 22:24:48,216 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 7.9 sec

MapReduce Total cumulative CPU time: 7 seconds 900 msec

Ended Job = job\_1505571178864\_0003

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 7.9 sec HDFS Read: 671 HDFS Write: 32 SUCCESS

Total MapReduce CPU Time Spent: 7 seconds 900 msec

OK

1990 23

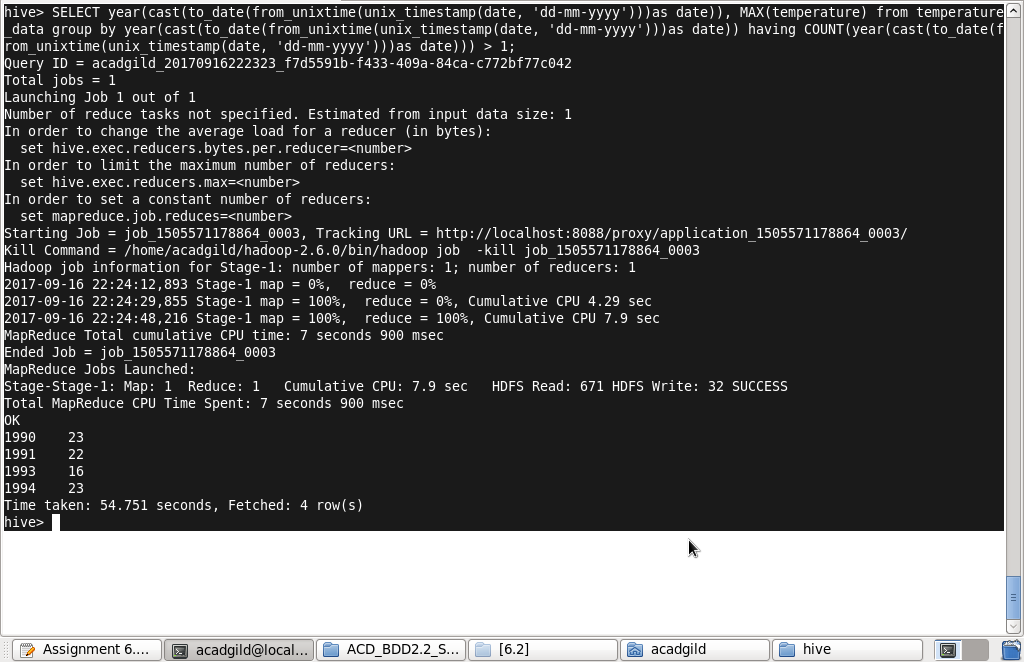
1991 22

1993 16

1994 23

Time taken: 54.751 seconds, Fetched: 4 row(s)

hive>



**--problem 4**

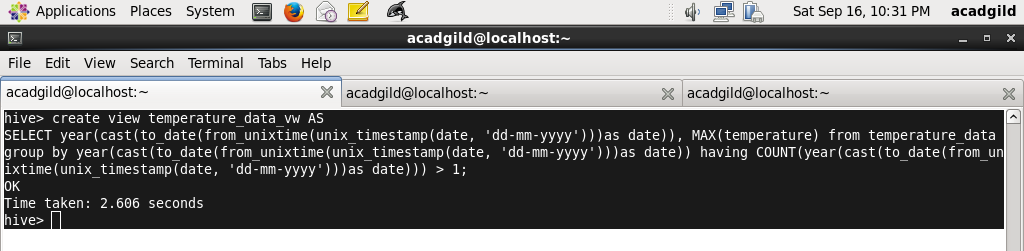
hive> create view temperature\_data\_vw AS

SELECT year(cast(to\_date(from\_unixtime(unix\_timestamp(date, 'dd-mm-yyyy')))as date)), MAX(temperature) from temperature\_data group by year(cast(to\_date(from\_unixtime(unix\_timestamp(date, 'dd-mm-yyyy')))as date)) having COUNT(year(cast(to\_date(from\_unixtime(unix\_timestamp(date, 'dd-mm-yyyy')))as date))) > 1;

OK

Time taken: 2.606 seconds

hive>



**--problem 5**

hive> INSERT OVERWRITE LOCAL DIRECTORY '/home/acadgild/hive/output/problem\_4' ROW FORMAT DELIMITED FIELDS TERMINATED BY '|'

SELECT \* FROM temperature\_data\_vw ;

Query ID = acadgild\_20170916231616\_ae5a2e76-cbc6-49c9-a3e3-4d6398bf8c52

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\_1505571178864\_0005, Tracking URL = http://localhost:8088/proxy/application\_1505571178864\_0005/

Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job -kill job\_1505571178864\_0005

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1

2017-09-16 23:16:41,673 Stage-1 map = 0%, reduce = 0%

2017-09-16 23:16:58,195 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.73 sec

2017-09-16 23:17:16,757 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 8.39 sec

MapReduce Total cumulative CPU time: 8 seconds 390 msec

Ended Job = job\_1505571178864\_0005

Copying data to local directory /home/acadgild/hive/output/problem\_4

Copying data to local directory /home/acadgild/hive/output/problem\_4

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 8.39 sec HDFS Read: 671 HDFS Write: 32 SUCCESS

Total MapReduce CPU Time Spent: 8 seconds 390 msec

OK

Time taken: 60.363 seconds

hive>

