**Assignment 2.8**

hive> LOAD DATA LOCAL INPATH '/home/cloudera/hive/dataset.txt'

> INTO TABLE temperature\_data;

Loading data to table custom.temperature\_data

Table custom.temperature\_data stats: [numFiles=2, totalSize=654]

OK

Time taken: 0.53 seconds

hive> select \* from temperature\_data ;

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

10-01-1990 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: 0.152 seconds, Fetched: 30 row(s)

hive>

**Problem Statement**

 Fetch date and temperature from temperature\_data where zip code is greater than

300000 and less than 399999.

 Calculate maximum temperature corresponding to every year from temperature\_data

table.

 Calculate maximum temperature from temperature\_data table corresponding to those

years which have at least 2 entries in the table.

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

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

file is '|' delimited.

** Fetch date and temperature from temperature\_data where zip code is greater than**

**300000 and less than 399999.**

hive> select \* from temperature\_data where zip>300000 & zip<399999 ;

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

10-01-1990 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: 0.496 seconds, Fetched: 30 row(s)

** Calculate maximum temperature corresponding to every year from temperature\_data**

**table.**

hive> select date, MAX(temperature) FROM temperature\_data GROUP BY date ;

Total MapReduce CPU Time Spent: 4 seconds 110 msec

OK

10-01-1990 23

10-01-1991 22

10-01-1993 11

10-01-1994 23

10-03-1990 15

10-03-1991 16

10-03-1993 16

12-02-1990 9

12-02-1991 10

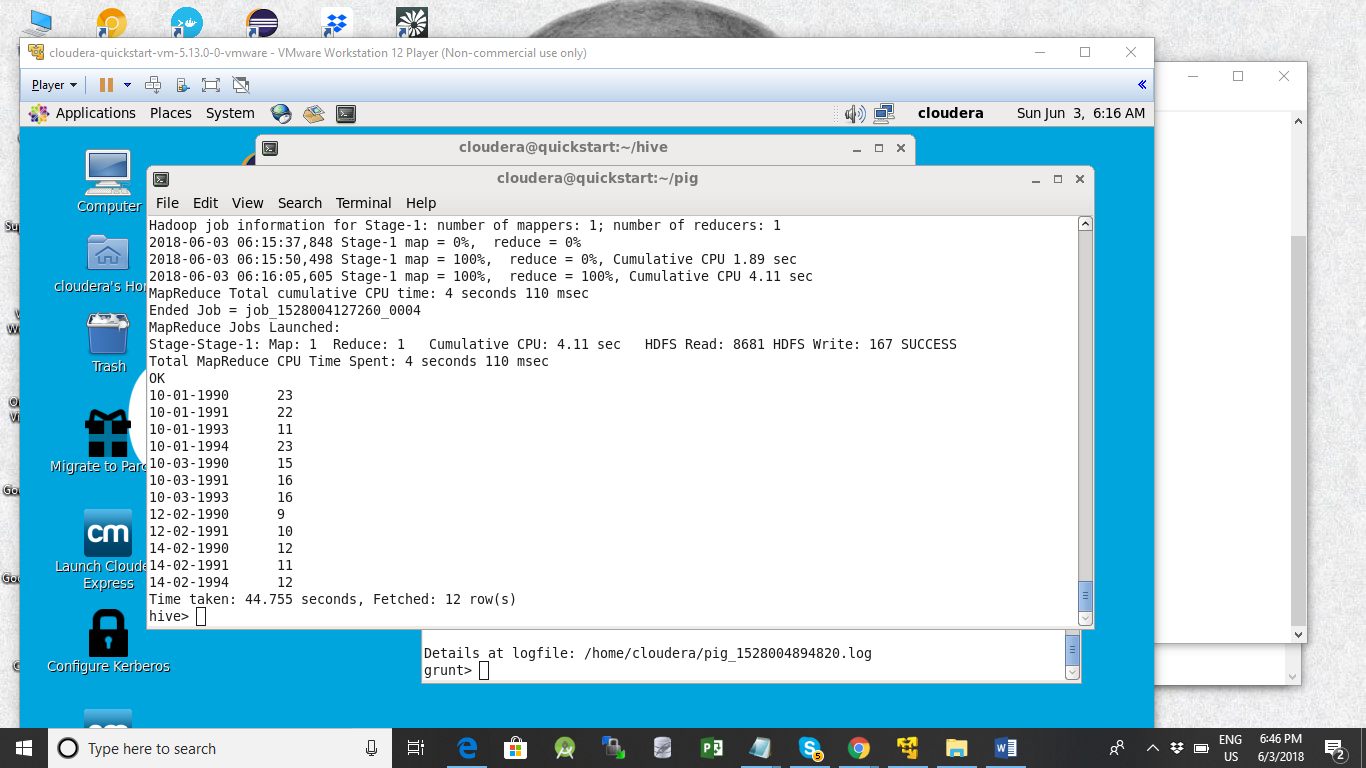
14-02-1990 12

14-02-1991 11

14-02-1994 12

Time taken: 44.755 seconds, Fetched: 12 row(s)

hive>



**Calculate maximum temperature from temperature\_data table corresponding to thoseyears which have at least 2 entries in the table.**

hive> select date, MAX(temperature), count(\*) FROM temperature\_data GROUP BY date ;

10-01-1990 23 5

10-01-1991 22 5

10-01-1993 11 1

10-01-1994 23 1

10-03-1990 15 2

10-03-1991 16 3

10-03-1993 16 1

12-02-1990 9 2

12-02-1991 10 4

14-02-1990 12 3

14-02-1991 11 2

14-02-1994 12 1

Time taken: 44.755 seconds, Fetched: 12 row(s)

hive> Select a, b

> from

> (select date AS a, MAX(temperature) AS b, count(\*) AS c FROM temperature\_data GROUP BY date)

> Temp

> Where c >1 ;

Query ID = cloudera\_20180603151414\_44efabc0-bef6-44e1-b8b5-9e2acfba835a

Total jobs = 1

Output:

Total MapReduce CPU Time Spent: 5 seconds 200 msec

OK

10-01-1990 23

10-01-1991 22

10-03-1990 15

10-03-1991 16

12-02-1990 9

12-02-1991 10

14-02-1990 12

14-02-1991 11

Time taken: 49.506 seconds, Fetched: 8 row(s)

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

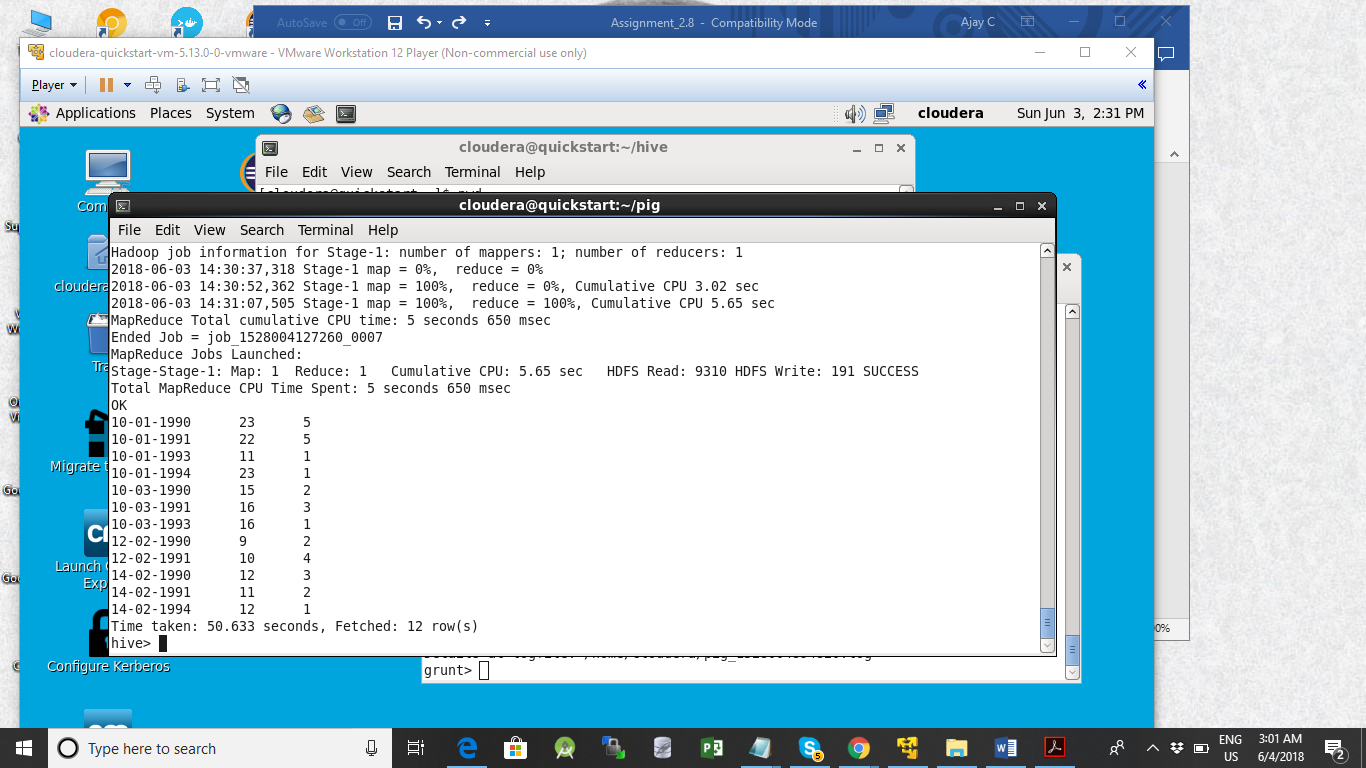
CREATE VIEW v1 select date, MAX(temperature), count(\*) FROM temperature\_data GROUP BY date ;

hive> CREATE VIEW v1 AS select date, MAX(temperature), count(\*) FROM temperature\_data GROUP BY date ;

OK

Time taken: 0.543 seconds

hive> select \* FROM v1 ;

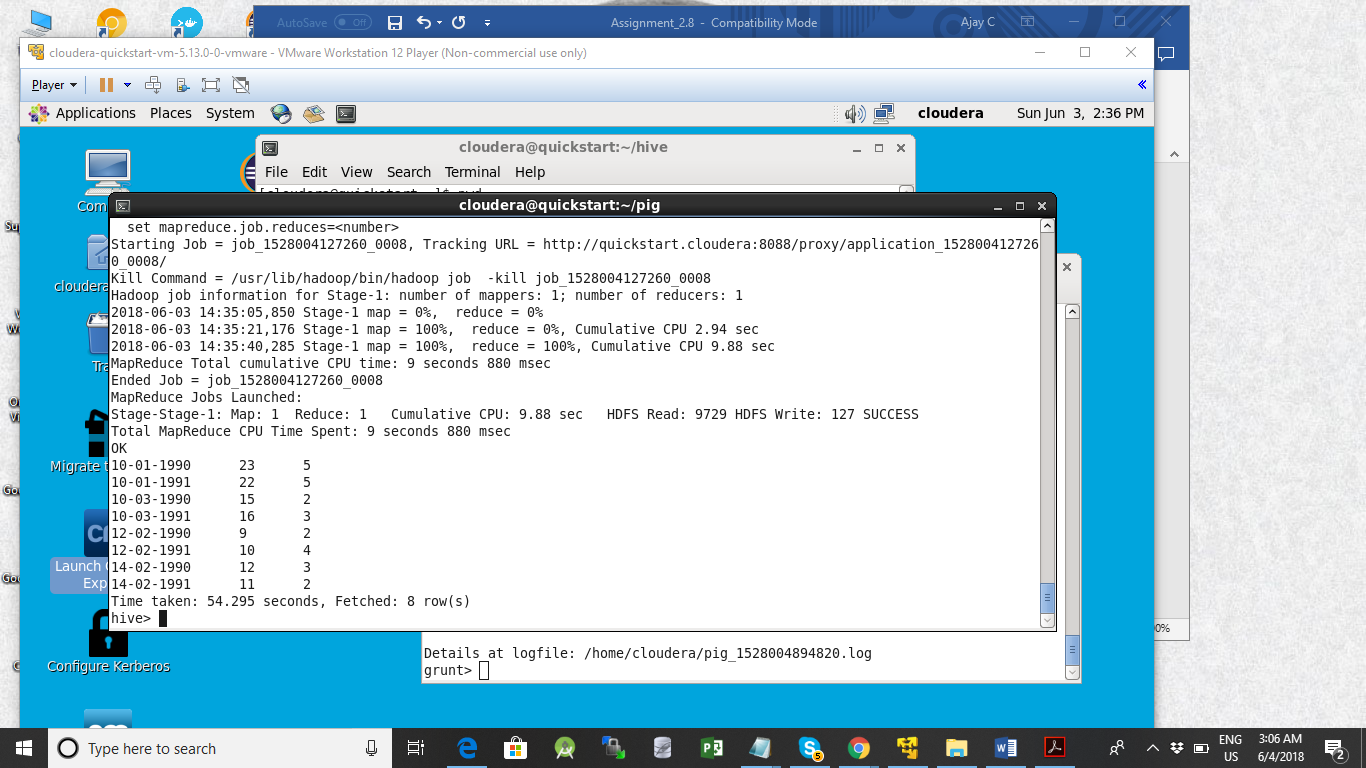


hive> CREATE VIEW v2 AS select date as a, MAX(temperature) as b, count(\*) as c FROM temperature\_data GROUP BY date ;

OK

Time taken: 0.327 seconds

hive> select \* from v2 where c >1 ;



Final Output

OK

10-01-1990 23 5

10-01-1991 22 5

10-03-1990 15 2

10-03-1991 16 3

12-02-1990 9 2

12-02-1991 10 4

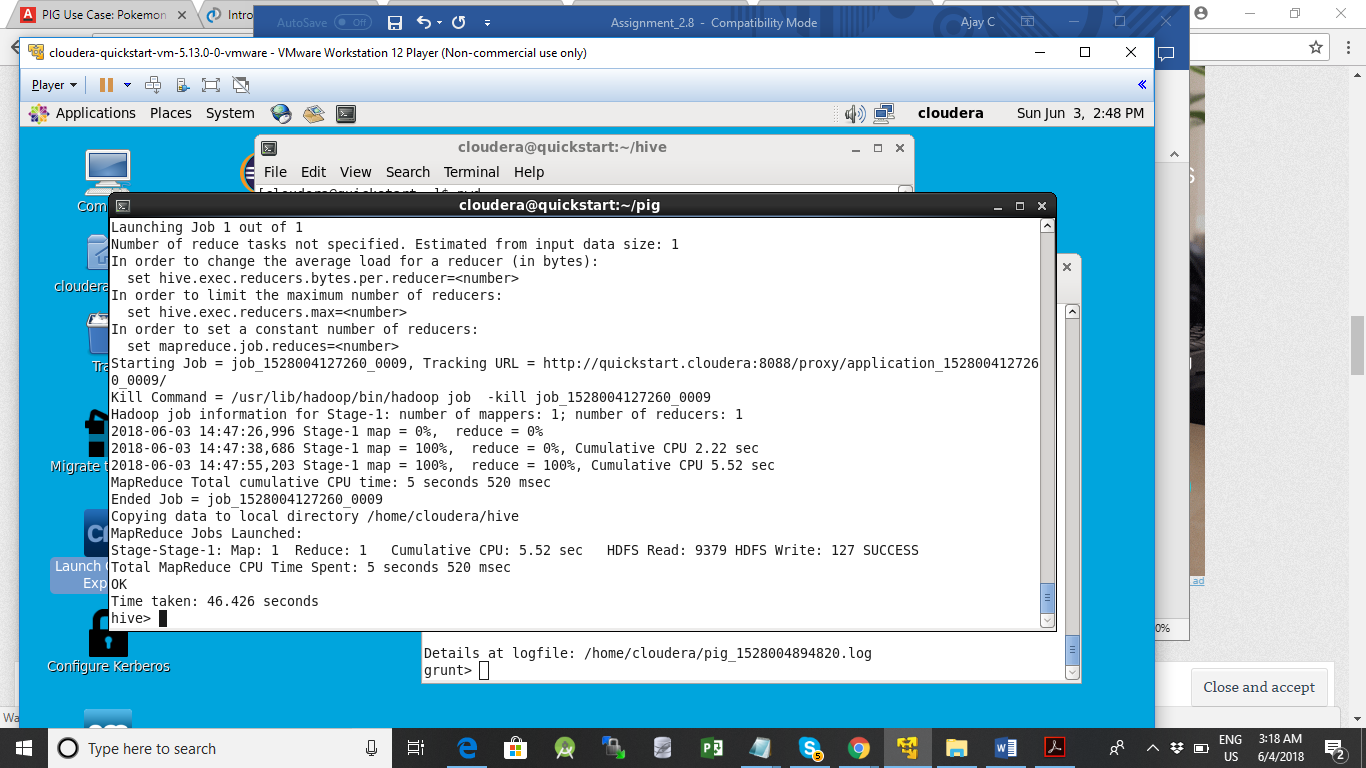
14-02-1990 12 3

14-02-1991 11 2

Time taken: 54.295 seconds, Fetched: 8 row(s)

**Export contents from temperature\_data\_vw to a file in local file system, such that eachfile is '|' delimited.**

hive> Insert overwrite local directory '/home/cloudera/hive' select \* from v2 where c >1 ;



View

[cloudera@quickstart ~]$ cat hive/000000\_0

10-01-1990235

10-01-1991225

10-03-1990152

10-03-1991163

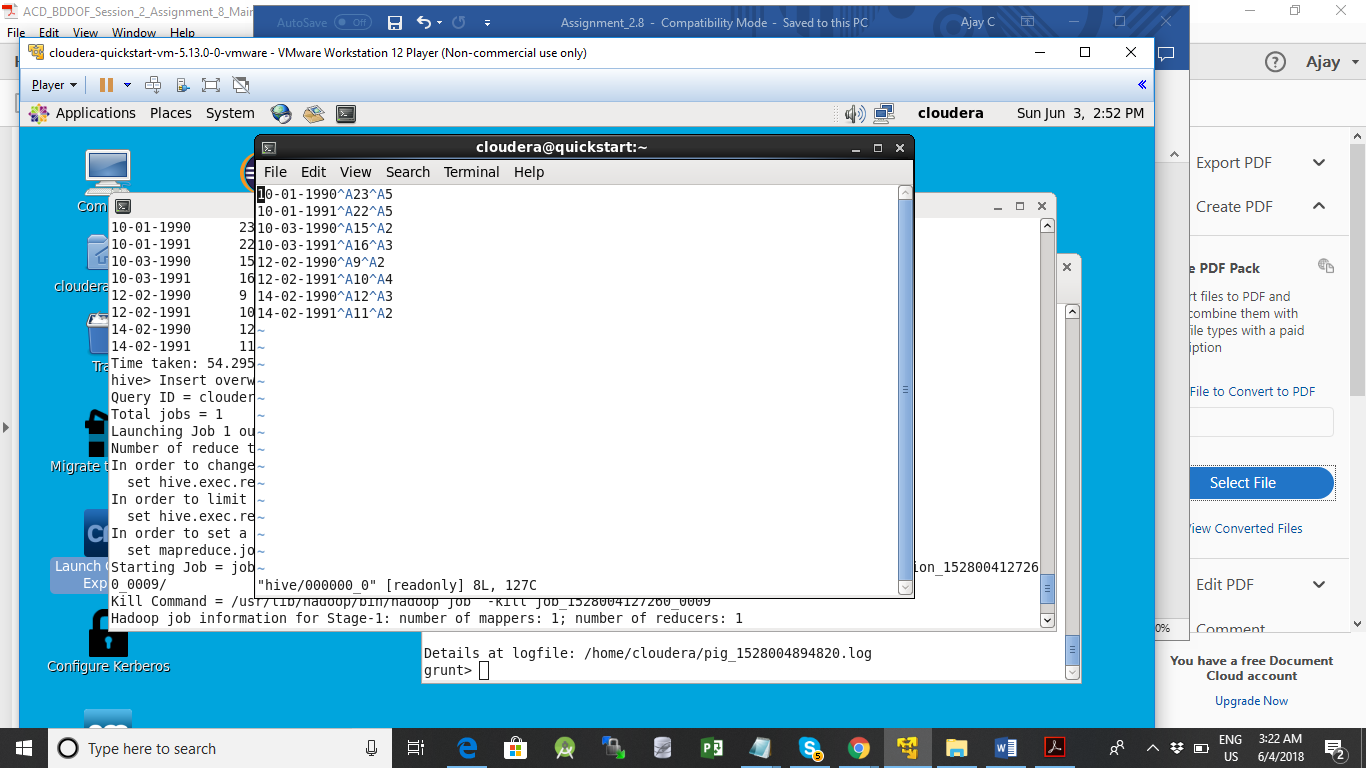
12-02-199092

12-02-1991104

14-02-1990123

14-02-1991112

[cloudera@quickstart ~]$ view hive/000000\_0



INSERT OVERWRITE LOCAL DIRECTORY '/home/cloudera/hive/test'

ROW FORMAT DELIMITED

FIELDS TERMINATED BY '|'

STORED AS TEXTFILE

select \* from v2 where c >1 ;

[cloudera@quickstart hive]$ cat test/000000\_0

10-01-1990|23|5

10-01-1991|22|5

10-03-1990|15|2

10-03-1991|16|3

12-02-1990|9|2

12-02-1991|10|4

14-02-1990|12|3

14-02-1991|11|2

[cloudera@quickstart hive]$

