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

300000 and less than 399999.

hive> SELECT full\_date, temperature FROM custom1.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: 0.338 seconds, Fetched: 12 row(s)

hive>

1. Calculate maximum temperature corresponding to every year from temperature\_data

table.

hive> set hive.cli.print.current.db=true;

hive (custom1)> SELECT SUBSTR(full\_date, 7, 4), MAX(temperature) FROM temperature\_data GROUP BY SUBSTR(full\_date, 7, 4);

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 = acadgild\_20170725144223\_d0d102cf-cf70-4bf6-898d-a9c82d1a524a

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\_1500962087858\_0003, Tracking URL = http://localhost:8088/proxy/application\_1500962087858\_0003/

Kill Command = /home/acadgild/hadoop-2.7.2/bin/hadoop job -kill job\_1500962087858\_0003

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

2017-07-25 14:42:41,887 Stage-1 map = 0%, reduce = 0%

2017-07-25 14:42:52,865 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.76 sec

2017-07-25 14:43:04,437 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.27 sec

MapReduce Total cumulative CPU time: 5 seconds 270 msec

Ended Job = job\_1500962087858\_0003

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.27 sec HDFS Read: 9264 HDFS Write: 167 SUCCESS

Total MapReduce CPU Time Spent: 5 seconds 270 msec

OK

1990 23

1991 22

1993 16

1994 23

Time taken: 42.925 seconds, Fetched: 4 row(s)

hive (custom1)>

1. Calculate maximum temperature from temperature\_data table corresponding to those

years which have at least 2 entries in the table.

hive (custom1)> SELECT SUBSTR(full\_date, 7, 4), MAX(temperature) FROM temperature\_data GROUP BY SUBSTR(full\_date, 7, 4) HAVING COUNT(\*) > 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 = acadgild\_20170725163452\_878145fd-6350-401b-9aaf-1ebfc68757bb

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\_1500962087858\_0006, Tracking URL = http://localhost:8088/proxy/application\_1500962087858\_0006/

Kill Command = /home/acadgild/hadoop-2.7.2/bin/hadoop job -kill job\_1500962087858\_0006

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

2017-07-25 16:35:06,883 Stage-1 map = 0%, reduce = 0%

2017-07-25 16:35:17,551 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.66 sec

2017-07-25 16:35:29,560 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.88 sec

MapReduce Total cumulative CPU time: 5 seconds 880 msec

Ended Job = job\_1500962087858\_0006

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.88 sec HDFS Read: 10256 HDFS Write: 127 SUCCESS

Total MapReduce CPU Time Spent: 5 seconds 880 msec

OK

1990 23

1991 22

Time taken: 38.436 seconds, Fetched: 2 row(s)

hive (custom1)>

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

hive (custom1)> CREATE VIEW temperature\_data\_view

> AS

> SELECT SUBSTR(full\_date, 7, 4), MAX(temperature)

> FROM temperature\_data

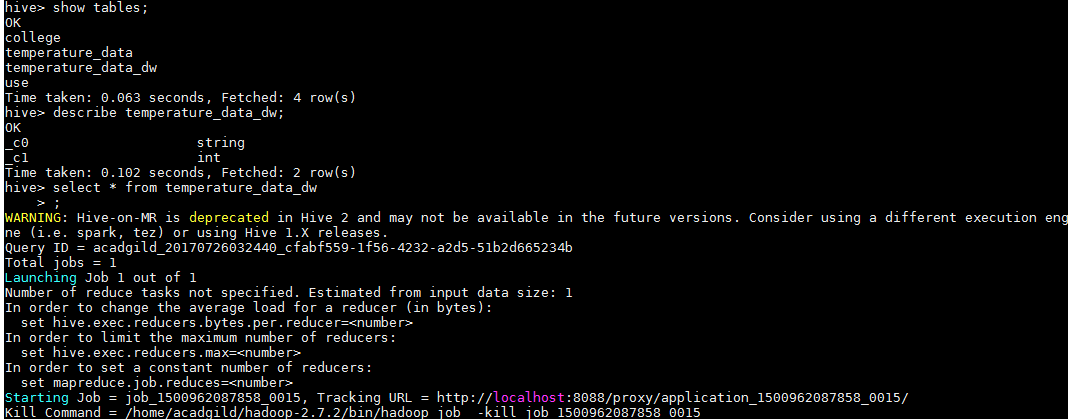
> GROUP BY SUBSTR(full\_date, 7, 4)

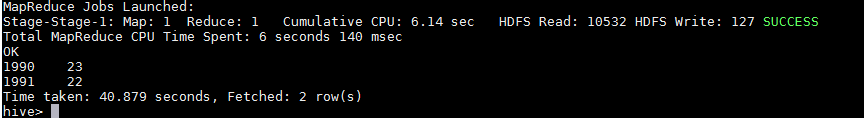
> HAVING COUNT(\*) > 2;

OK

Time taken: 0.216 seconds

hive (custom1)>





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

hive> INSERT OVERWRITE LOCAL DIRECTORY '/home/acadgild/leena/hive-test/'

> ROW FORMAT DELIMITED

> FIELDS TERMINATED BY '|'

> SELECT \* FROM temperature\_data\_dw;

