Session 06: Hive Introduction

Assignment 2

**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.

**Loading data –**

**hive> create table if not exists temperature\_data ( date string, zip int, temperature int) row format delimited fields terminated by ',';**

OK

Time taken: 0.293 seconds

**hive> load data local inpath '/home/acadgild/Acadgild/Assignemnt6\_2/dataset\_Session 14.txt' into table temperature\_data;**

Loading data to table custom.temperature\_data

Table custom.temperature\_data stats: [numFiles=1, totalSize=437]

OK

Time taken: 0.987 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

10-01-1993 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: 0.463 seconds, Fetched: 20 row(s)

**Solution:**

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

300000 and less than 399999.

**hive>** select date, temperature from temperature\_data where zip > 300000 and zip<399999;

**output-**



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

table.

**hive>** select from\_unixtime(unix\_timestamp(date , 'dd-MM-yyyy'),'yyyy') , max(temperature) from temperature\_data group by from\_unixtime(unix\_timestamp(date , 'dd-MM-yyyy'),'yyyy');

Query ID = acadgild\_20171013004747\_86d5ccbe-787d-4a46-a46e-1749bf77741f

Total jobs = 1

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

2017-10-13 00:48:07,248 Stage-1 map = 0%, reduce = 0%

2017-10-13 00:48:22,575 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.21 sec

2017-10-13 00:48:43,044 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 5.92 sec

2017-10-13 00:48:48,456 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 8.6 sec

MapReduce Total cumulative CPU time: 8 seconds 600 msec

Ended Job = job\_1507824426583\_0001

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 8.6 sec HDFS Read: 682 HDFS Write: 32 SUCCESS

Total MapReduce CPU Time Spent: 8 seconds 600 msec

OK

**1990 23**

**1991 22**

**1993 16**

**1994 23**

Time taken: 68.675 seconds, Fetched: 4 row(s)

**Output-**



1. 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(date , 'dd-MM-yyyy'),'yyyy') , max(temperature) from temperature\_data group by from\_unixtime(unix\_timestamp(date , 'dd-MM-yyyy'),'yyyy')

having count(temperature)>=2;

Query ID = acadgild\_20171013005555\_ebaf3a69-46a7-432a-a787-d77df8a1d001

Total jobs = 1

Launching Job 1 out of 1

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

2017-10-13 00:55:36,681 Stage-1 map = 0%, reduce = 0%

2017-10-13 00:55:48,874 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.57 sec

2017-10-13 00:56:04,768 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.78 sec

MapReduce Total cumulative CPU time: 6 seconds 780 msec

Ended Job = job\_1507824426583\_0002

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 6.78 sec HDFS Read: 682 HDFS Write: 32 SUCCESS

Total MapReduce CPU Time Spent: 6 seconds 780 msec

OK

**1990 23**

**1991 22**

**1993 16**

**1994 23**

Time taken: 42.503 seconds, Fetched: 4 row(s)

**Output –**



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

**hive>** CREATE VIEW temperature\_data\_vw AS

select from\_unixtime(unix\_timestamp(date , 'dd-MM-yyyy'),'yyyy') , max(temperature) from temperature\_data group by from\_unixtime(unix\_timestamp(date , 'dd-MM-yyyy'),'yyyy')

having count(temperature)>=2;

OK

Time taken: 0.47 seconds

**hive>** select \* from temperature\_data\_vw;

Query ID = acadgild\_20171013010202\_2bc4bc42-5fef-43cc-b970-b2a0f0795a4d

Total jobs = 1

Launching Job 1 out of 1

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

2017-10-13 01:03:23,748 Stage-1 map = 0%, reduce = 0%

2017-10-13 01:03:37,294 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.4 sec

2017-10-13 01:03:50,858 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.09 sec

MapReduce Total cumulative CPU time: 6 seconds 90 msec

Ended Job = job\_1507824426583\_0004

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 6.09 sec HDFS Read: 682 HDFS Write: 32 SUCCESS

Total MapReduce CPU Time Spent: 6 seconds 90 msec

OK

**1990 23**

**1991 22**

**1993 16**

**1994 23**

Time taken: 62.504 seconds, Fetched: 4 row(s)

**Output-**



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/Acadgild/Assignemnt6\_2'

row format delimited fields terminated by '|' select \* from temperature\_data\_vw;

Query ID = acadgild\_20171013011010\_76fc71a2-3873-4581-a863-adc92737e663

Total jobs = 1

Launching Job 1 out of 1

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

2017-10-13 01:10:39,197 Stage-1 map = 0%, reduce = 0%

2017-10-13 01:10:51,332 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.64 sec

2017-10-13 01:11:08,400 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.15 sec

MapReduce Total cumulative CPU time: 6 seconds 150 msec

Ended Job = job\_1507824426583\_0005

Copying data to local directory /home/acadgild/Acadgild/Assignemnt6\_2

Copying data to local directory /home/acadgild/Acadgild/Assignemnt6\_2

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 6.15 sec HDFS Read: 682 HDFS Write: 32 SUCCESS

Total MapReduce CPU Time Spent: 6 seconds 150 msec

OK

Time taken: 41.798 seconds

**Output-**





Output file –

