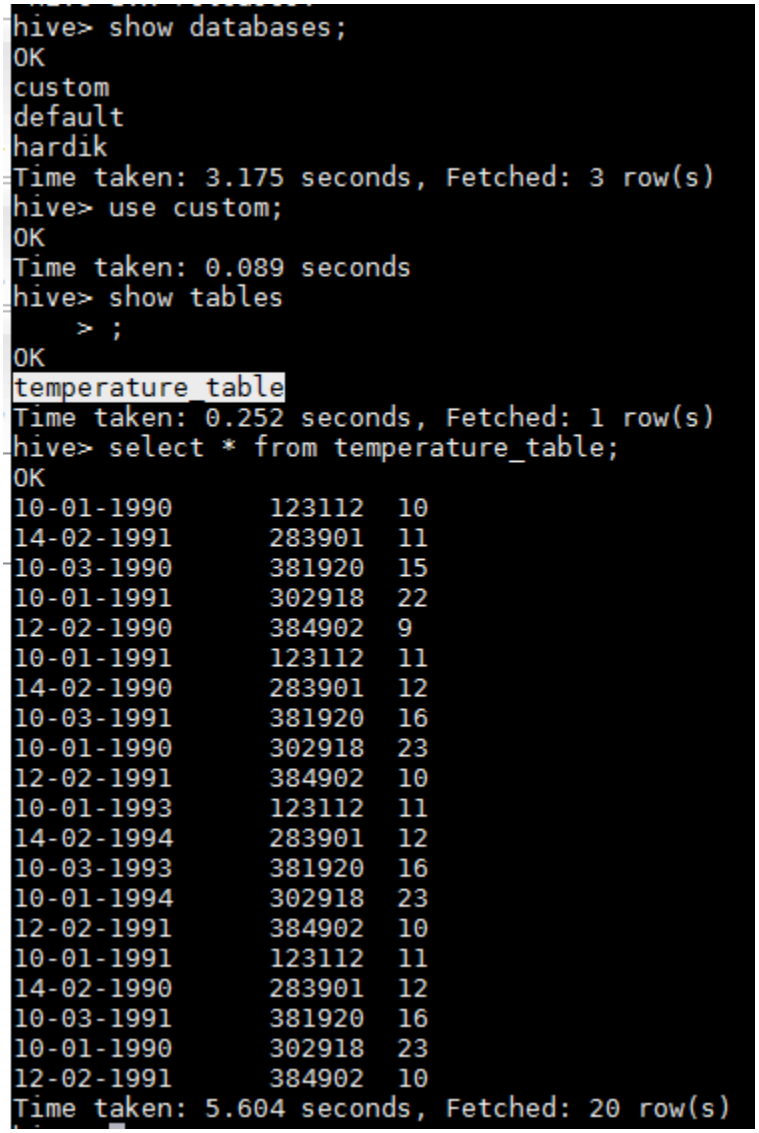
Assignment 6.2

Problem Statement 1:-

Fetch date and temperature from temperature_data where zip code is greater than 300000 and less than 399999.

Solution:-



```
hive> show databases;
OK
custom
default
hardik
Time taken: 3.175 seconds, Fetched: 3 row(s)
hive> use custom;
OK
Time taken: 0.089 seconds
hive> show tables
    > ;
OK
temperature_table
Time taken: 0.252 seconds, Fetched: 1 row(s)
hive> select * from temperature_table;
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: 5.604 seconds, Fetched: 20 row(s)
```

SELECT full_date, temperature FROM temperature_table WHERE zip>300000 AND zip<399999;

```
hive> SELECT full_date, temperature FROM temperature_table WHERE zip>300000 AND zip<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: 1.612 seconds, Fetched: 12 row(s)
hive>
```

Problem Statement 2:-

Calculate maximum temperature corresponding to every year from temperature_data table.

Solution:-

SELECT SUBSTR(full_date, 7, 10), MAX(temperature) FROM temperature_table GROUP BY SUBSTR(full_date, 7, 10);

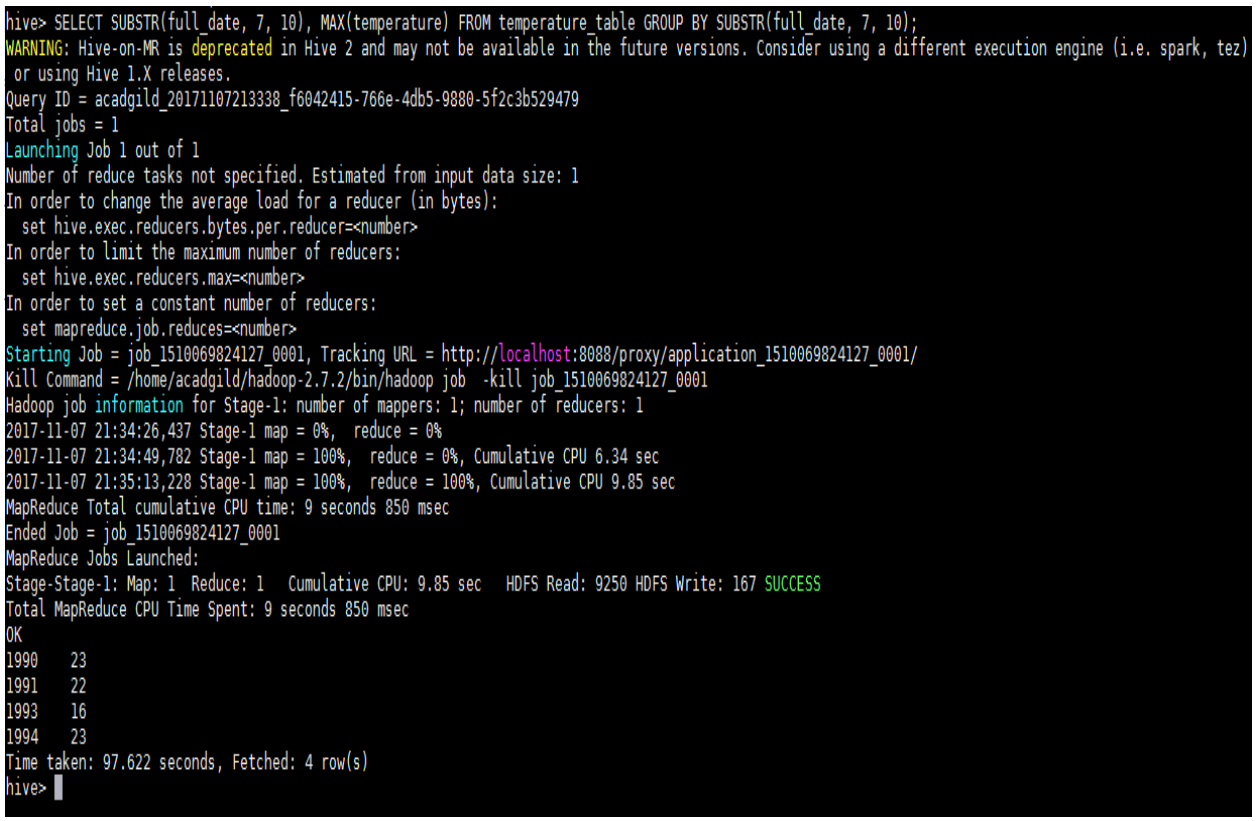
```
hive> SELECT SUBSTR(full_date, 7, 10), MAX(temperature) FROM temperature_table GROUP BY SUBSTR(full_date, 7, 10);
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez)
 or using Hive 1.X releases.
Query ID = acadgild_20171107213338_f6042415-766e-4db5-9880-5f2c3b529479
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_1510069824127_0001, Tracking URL = http://localhost:8088/proxy/application_1510069824127_0001/
Kill Command = /home/acadgild/hadoop-2.7.2/bin/hadoop job  -kill job_1510069824127_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2017-11-07 21:34:26,437 Stage-1 map = 0%,  reduce = 0%
2017-11-07 21:34:49,782 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 6.34 sec
2017-11-07 21:35:13,228 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 9.85 sec
MapReduce Total cumulative CPU time: 9 seconds 850 msec
Ended Job = job_1510069824127_0001
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 9.85 sec   HDFS Read: 9250 HDFS Write: 167 SUCCESS
Total MapReduce CPU Time Spent: 9 seconds 850 msec
OK
1990    23
1991    22
1993    16
1994    23
Time taken: 97.622 seconds, Fetched: 4 row(s)
hive>
```

Problem Statement 3:-

Calculate maximum temperature from temperature_data table corresponding to those years which have at least 2 entries in the table.

Solution:-

SELECT SUBSTR(full_date, 7, 10), MAX(temperature) FROM temperature_table GROUP BY SUBSTR(full_date, 7, 10);



Problem Statement 4:-

Create a view on the top of last query, name it temperature_data_vw.

Solution:-

CREATE VIEW temperature_data_vw (YEAR,TEMPERATURE) AS SELECT SUBSTR(full_date, 7, 10), MAX(temperature) FROM temperature_table GROUP BY SUBSTR(full_date, 7, 10);

```
hive> CREATE VIEW temperature_data_vw (YEAR,TEMPERATURE) AS SELECT SUBSTR(full_date, 7, 10), MAX(temperature) FROM temperature_table GROUP BY SUBSTR(full_
date, 7, 10);
OK
Time taken: 0.972 seconds
hive> SELECT * FROM temperature_data_vw;
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_20171107214926_5f519755-81f4-4446-9f59-f0eac60df60a
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_1510069824127_0002, Tracking URL = http://localhost:8088/proxy/application_1510069824127_0002/
Kill Command = /home/acadgild/hadoop-2.7.2/bin/hadoop job  -kill job_1510069824127_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2017-11-07 21:49:58,151 Stage-1 map = 0%,  reduce = 0%
2017-11-07 21:50:23,049 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 6.1 sec
2017-11-07 21:50:43,542 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 9.61 sec
MapReduce Total cumulative CPU time: 9 seconds 610 msec
Ended Job = job_1510069824127_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 9.61 sec   HDFS Read: 9250 HDFS Write: 167 SUCCESS
Total MapReduce CPU Time Spent: 9 seconds 610 msec
OK
1990    23
1991    22
1993    16
1994    23
Time taken: 78.331 seconds, Fetched: 4 row(s)
hive>
```

Problem Statement 5:-

Export contents from temperature_data_vw to a file in local file system, such that each file is '|' delimited.

Solution:-

INSERT OVERWRITE LOCAL DIRECTORY '/home/acadgild/hadoop/temperature_data_vw.txt' ROW FORMAT DELIMITED FIELDS TERMINATED BY '|' SELECT * FROM temperature_data_vw;

```
hive> INSERT OVERWRITE LOCAL DIRECTORY '/home/acadgild/hadoop/temperature_data_vw.txt' ROW FORMAT DELIMITED FIELDS TERMINATED BY '|' SELECT * FROM tempera
ture_data_vw;
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_20171107233751_94af3b68-abc0-4d5f-a1be-3feac8d81442
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_1510069824127_0003, Tracking URL = http://localhost:8088/proxy/application_1510069824127_0003/
Kill Command = /home/acadgild/hadoop-2.7.2/bin/hadoop job  -kill job_1510069824127_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2017-11-07 23:38:23,495 Stage-1 map = 0%,  reduce = 0%
2017-11-07 23:38:48,688 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 6.36 sec
2017-11-07 23:39:13,963 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 10.27 sec
MapReduce Total cumulative CPU time: 10 seconds 270 msec
Ended Job = job_1510069824127_0003
Moving data to local directory /home/acadgild/hadoop/temperature_data_vw.txt
Failed with exception Unable to move source hdfs://localhost:9000/tmp/hive/acadgild/c1531357-8119-454a-a5f1-c4f99ba801c7/hive_2017-11-07_23-37-51_590_1832
2138134544192 51-1/-mr-10000 to destination /home/acadgild/hadoop/temperature_data_vw.txt
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.MoveTask. Unable to move source hdfs://localhost:9000/tmp/hive/acadgild/c153135
7-8119-454a-a5f1-c4f99ba801c7/hive_2017-11-07_23-37-51_590_1832213813454419251-1/-mr-10000 to destination /home/acadgild/hadoop/temperature_data_vw.txt
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 10.27 sec   HDFS Read: 8979 HDFS Write: 32 SUCCESS
Total MapReduce CPU Time Spent: 10 seconds 270 msec
hive>
```

Submitted By

Hardik Kaushik