Task 1:

a. Creating database named custom in HIVE.

```
Time taken: 1.547 seconds
hive> show databases;
OK
acadgilddb
default
Time taken: 0.106 seconds, Fetched: 2 row(s)
hive> create database custom;
OK
Time taken: 0.383 seconds
hive> use custom;
OK
```

b. Creating table named Temparature_data inside Custom database having mentioned fields.

```
hive> create Table temperature_data(
    > tdate STRING,
    > zipcode int,
    > temperature int)
    > row format delimited fields terminated by ',';
OK
```

c. Loading dataset table into customer table in hive from local fs.

```
hive> LOAD DATA LOCAL INPATH '/home/acadgild/Desktop/a/dataset_Hive.txt' into table custom.temperature_data;
Loading data to table custom.temperature_data
OK
Time taken: 3.999 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
```

Task 2:

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

```
hive> select tdate, temperature from 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: 1.871 seconds, Fetched: 12 row(s)
```

b. Calculate maximum temperature corresponding to every year from **temperature_data** table.

```
hive> select SUBSTR(tdate,7,10) as year,MAX(temperature) from temperature_data group by substr(tdate,7,10);
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a differe
tion engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = acadgild_20180507091459_403ce2bf-1dbb-4687-865e-8aa53170f54a
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_1525608425288_0012, Tracking URL = http://localhost:8088/proxy/application_1525608425288_0012/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job  -kill job_1525608425288_0012
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-05-07 09:15:22,164 Stage-1 map = 0%,  reduce = 0%
2018-05-07 09:15:41,663 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 5.29 sec
2018-05-07 09:15:58,733 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 9.57 sec
MapReduce Total cumulative CPU time: 9 seconds 570 msec
Ended Job = job_1525608425288_0012
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 9.57 sec   HDFS Read: 9119 HDFS Write: 167 SUCCESS
Total MapReduce CPU Time Spent: 9 seconds 570 msec
OK
1990    23
1991    22
1993    16
1994    23
Time taken: 61.323 seconds, Fetched: 4 row(s)
hive>
```

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

```
hive> select SUBSTR(tdate,7,10) as year,MAX(temperature) from temperature_data group by substr(tdate,7,10) having count(subst
r(tdate,7,10)) >= 2;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execu
tion engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = acadgild_20180507095841_78665f40-fc68-49e2-8dc2-8a5f6e5f40f6
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_1525608425288_0014, Tracking URL = http://localhost:8088/proxy/application_1525608425288_0014/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job  -kill job_1525608425288_0014
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-05-07 09:59:05,086 Stage-1 map = 0%,  reduce = 0%
2018-05-07 09:59:26,028 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 6.2 sec
2018-05-07 09:59:44,022 Stage-1 map = 100%,  reduce = 67%, Cumulative CPU 11.44 sec
2018-05-07 09:59:45,460 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 12.47 sec
MapReduce Total cumulative CPU time: 12 seconds 470 msec
Ended Job = job_1525608425288_0014
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 12.47 sec   HDFS Read: 10148 HDFS Write: 167 SUCCESS
Total MapReduce CPU Time Spent: 12 seconds 470 msec
OK
1990    23
1991    22
1993    16
1994    23
Time taken: 66.848 seconds, Fetched: 4 row(s)
hive>
```

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

```
hive> create view temperature_data_vw as select SUBSTR(tdate,7,10) as year,MAX(temperature) from temperature_data group by su
bstr(tdate,7,10) having count(substr(tdate,7,10)) >= 2;
OK
Time taken: 1.16 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 execu
tion engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = acadgild_20180507102923_4a9e9519-5966-49bb-aaef-edd60c5fb145
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_1525608425288_0015, Tracking URL = http://localhost:8088/proxy/application_1525608425288_0015/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job  -kill job_1525608425288_0015
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-05-07 10:29:37,142 Stage-1 map = 0%,  reduce = 0%
2018-05-07 10:29:48,482 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.97 sec
2018-05-07 10:30:01,910 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 6.78 sec
MapReduce Total cumulative CPU time: 6 seconds 780 msec
Ended Job = job_1525608425288_0015
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 6.78 sec   HDFS Read: 10219 HDFS Write: 167 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 780 msec
OK
1990    23
1991    22
1993    16
1994    23
Time taken: 40.432 seconds, Fetched: 4 row(s)
hive>
```

e. Export content from **temeperature_data_vw** to a file in local fs, such that each file is '|' delimited.