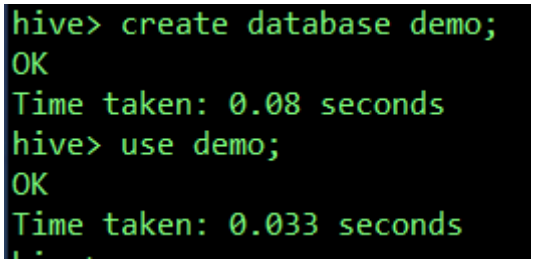
Module 3: Assignment – 2

1. Create a database called Demo and use it



2. Create a table called Suicides in it, matching with the schema of the data

```
hive> create table suicides
    > (state string, year int, type_code string, type string, gender string, age_group string, total int
)
    > row format delimited fields terminated by ',' stored as textfile
    > tblproperties("skip.header.line.count"="1");
OK
```

3. Load the given CSV file into the table

```
hive> load data local inpath '/home/bitnami/Suicides.csv' into table suicides;
Loading data to table demo.suicides
OK
Time taken: 1.159 seconds
```

```
hive> select * from suicides limit 10;
OK
A & N Islands   2001    Causes  Illness (Aids/STD)      Female  0-14    0
A & N Islands   2001    Causes  Bankruptcy or Sudden change in Economic Female  0-14    0
A & N Islands   2001    Causes  Cancellation/Non-Settlement of Marriage Female  0-14    0
A & N Islands   2001    Causes  Physical Abuse (Rape/Incest Etc.)       Female  0-14    0
A & N Islands   2001    Causes  Dowry Dispute   Female  0-14    0
A & N Islands   2001    Causes  Family Problems Female  0-14    0
A & N Islands   2001    Causes  Ideological Causes/Hero Worshipping     Female  0-14    0
A & N Islands   2001    Causes  Other Prolonged Illness Female  0-14    0
A & N Islands   2001    Causes  Property Dispute        Female  0-14    0
A & N Islands   2001    Causes  Fall in Social Reputation       Female  0-14    0
Time taken: 0.203 seconds, Fetched: 10 row(s)
```

4. Find out the most common suicide cause among females in India over the entire period 2001–2012

```
hive> select type, count(type) as cause_count from suicides
    > where gender = 'Female'
    > group by type
    > sort by cause_count desc
    > limit 1;
```

Running Map Reduce Job to view Output

```
Automatically selecting local only mode for query
Query ID = root_20211104175253_1e12d789-6d4a-497b-8476-017b3907d67e
Total jobs = 3
Launching Job 1 out of 3
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>
Job running in-process (local Hadoop)
2021-11-04 17:52:55,929 Stage-1 map = 100%,  reduce = 100%
Ended Job = job_local1249027422_0136
Launching Job 2 out of 3
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>
Selecting local mode for task: Stage-2
Job running in-process (local Hadoop)
2021-11-04 17:52:58,526 Stage-2 map = 100%,  reduce = 100%
Ended Job = job_local1836210998_0137
Launching Job 3 out of 3
Number of reduce tasks determined at compile time: 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>
Selecting local mode for task: Stage-3
Job running in-process (local Hadoop)
2021-11-04 17:53:01,163 Stage-3 map = 100%,  reduce = 100%
Ended Job = job_local2003615079_0138
MapReduce Jobs Launched:
Stage-Stage-1:  HDFS Read: 1604479480 HDFS Write: 58937522 SUCCESS
Stage-Stage-2:  HDFS Read: 1604495072 HDFS Write: 58953320 SUCCESS
Stage-Stage-3:  HDFS Read: 1604508885 HDFS Write: 58967700 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
```

Output:

```
Others (Please Specify) 3512
Time taken: 7.453 seconds, Fetched: 1 row(s)
```

5. Find out the state-wise most common cause among males over the entire period

```
hive> select state, type, count(type) as cause_count from suicides
    > where gender = 'Male'
    > group by state, type
    > sort by cause_count desc
    > limit 38;
```

Running Map Reduce Job to view Output:

```
Automatically selecting local only mode for query
Query ID = root_20211104175520_74f1cfc6-30ab-4847-9a49-3451db873679
Total jobs = 3
Launching Job 1 out of 3
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>
Job running in-process (local Hadoop)
2021-11-04 17:55:22,834 Stage-1 map = 100%,  reduce = 100%
Ended Job = job_local1274085815_0146
Launching Job 2 out of 3
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>
Selecting local mode for task: Stage-2
Job running in-process (local Hadoop)
2021-11-04 17:55:25,417 Stage-2 map = 100%,  reduce = 100%
Ended Job = job_local159793380_0147
Launching Job 3 out of 3
Number of reduce tasks determined at compile time: 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>
Selecting local mode for task: Stage-3
Job running in-process (local Hadoop)
2021-11-04 17:55:28,004 Stage-3 map = 100%,  reduce = 100%
Ended Job = job_local218665293_0148
MapReduce Jobs Launched:
Stage-Stage-1:  HDFS Read: 1728322288 HDFS Write: 59653107 SUCCESS
Stage-Stage-2:  HDFS Read: 1728578823 HDFS Write: 59791853 SUCCESS
Stage-Stage-3:  HDFS Read: 1728597271 HDFS Write: 59810755 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
```

Output:

```
Andhra Pradesh  Others (Please Specify) 120
Rajasthan       Others (Please Specify) 120
Gujarat Others (Please Specify) 120
Chhattisgarh    Others (Please Specify) 120
Madhya Pradesh  Others (Please Specify) 120
Maharashtra     Others (Please Specify) 120
Haryana Others (Please Specify) 120
Bihar   Others (Please Specify) 120
Karnataka       Others (Please Specify) 120
Odisha  Others (Please Specify) 119
Jharkhand       Others (Please Specify) 119
Uttar Pradesh   Others (Please Specify) 118
Delhi (Ut)      Others (Please Specify) 118
Assam   Others (Please Specify) 118
Kerala  Others (Please Specify) 118
Tamil Nadu      Others (Please Specify) 116
Tripura Others (Please Specify) 116
Punjab  Others (Please Specify) 116
West Bengal     Others (Please Specify) 115
Himachal Pradesh        Others (Please Specify) 113
Goa     Others (Please Specify) 109
Uttarakhand     Others (Please Specify) 107
Jammu & Kashmir Others (Please Specify) 106
Puducherry      Others (Please Specify) 105
Mizoram Others (Please Specify) 100
Sikkim  Others (Please Specify) 99
Meghalaya       Others (Please Specify) 99
Chandigarh      Others (Please Specify) 91
A & N Islands   Others (Please Specify) 91
Daman & Diu     Others (Please Specify) 86
D & N Haveli    Others (Please Specify) 85
Nagaland        Others (Please Specify) 84
Manipur Others (Please Specify) 81
Arunachal Pradesh       Others (Please Specify) 80
Lakshadweep     Others (Please Specify) 62
A & N Islands   By Over Alcoholism      60
A & N Islands   Fall in Social Reputation       60
A & N Islands   Family Problems 60
Time taken: 7.071 seconds, Fetched: 38 row(s)
```

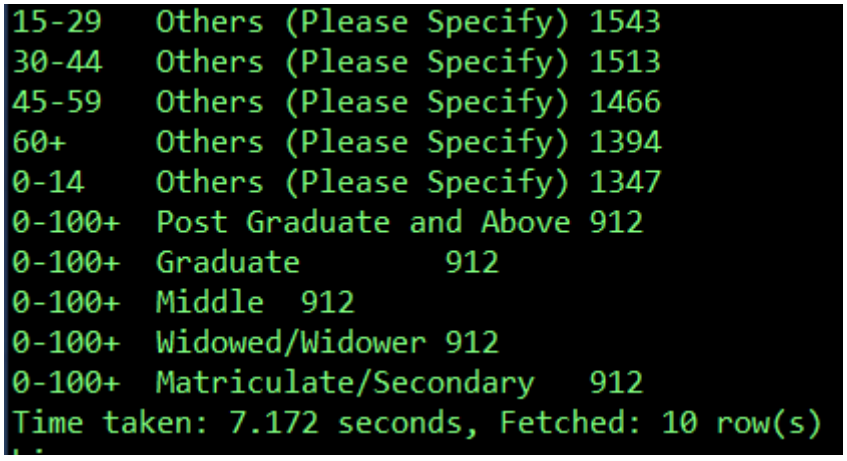
6. Find out the age group-wise most common cause among males and females

```
hive> select age_group, type, count(type) as cause_count from suicides
    > group by age_group, type
    > sort by cause_count desc
    > limit 10;
```

Running Map Reduce Job to view Output:

```
Automatically selecting local only mode for query
Query ID = root_20211104175945_7b1867ea-6b96-4915-b925-b2a812318b98
Total jobs = 3
Launching Job 1 out of 3
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>
Job running in-process (local Hadoop)
2021-11-04 17:59:47,146 Stage-1 map = 100%,  reduce = 100%
Ended Job = job_local1587071514_0149
Launching Job 2 out of 3
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>
Selecting local mode for task: Stage-2
Job running in-process (local Hadoop)
2021-11-04 17:59:49,750 Stage-2 map = 100%,  reduce = 100%
Ended Job = job_local1115743452_0150
Launching Job 3 out of 3
Number of reduce tasks determined at compile time: 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>
Selecting local mode for task: Stage-3
Job running in-process (local Hadoop)
2021-11-04 17:59:52,324 Stage-3 map = 100%,  reduce = 100%
Ended Job = job_local1787927152_0151
MapReduce Jobs Launched:
Stage-Stage-1:  HDFS Read: 1759417036 HDFS Write: 59826960 SUCCESS
Stage-Stage-2:  HDFS Read: 1759455697 HDFS Write: 59855246 SUCCESS
Stage-Stage-3:  HDFS Read: 1759471086 HDFS Write: 59871138 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
```

Output:


15-29 Others (Please Specify) 1543
30-44 Others (Please Specify) 1513
45-59 Others (Please Specify) 1466
60+ Others (Please Specify) 1394
0-14 Others (Please Specify) 1347
0-100+ Post Graduate and Above 912
0-100+ Graduate 912
0-100+ Middle 912
0-100+ Widowed/Widower 912
0-100+ Matriculate/Secondary 912
Time taken: 7.172 seconds, Fetched: 10 row(s)

7. Find out the total number of suicides per year per state

```
hive> select year, state, sum(total) as total_cases from suicides
    > group by year, state
    > sort by year,total_cases desc
    > limit 200;
```

Running Map Reduce Job to view Output:

```
Automatically selecting local only mode for query
Query ID = root_20211104180348_22747aaf-e3bd-4c5e-9dd8-6d1536a2b865
Total jobs = 3
Launching Job 1 out of 3
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>
Job running in-process (local Hadoop)
2021-11-04 18:03:50,339 Stage-1 map = 100%,  reduce = 100%
Ended Job = job_local1803110083_0155
Launching Job 2 out of 3
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>
Selecting local mode for task: Stage-2
Job running in-process (local Hadoop)
2021-11-04 18:03:52,913 Stage-2 map = 100%,  reduce = 100%
Ended Job = job_local50582443_0156
Launching Job 3 out of 3
Number of reduce tasks determined at compile time: 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>
Selecting local mode for task: Stage-3
Job running in-process (local Hadoop)
2021-11-04 18:03:54,414 Stage-3 map = 100%,  reduce = 100%
Ended Job = job_local122303796_0157
MapReduce Jobs Launched:
Stage-Stage-1:  HDFS Read: 1821187902 HDFS Write: 59975289 SUCCESS
Stage-Stage-2:  HDFS Read: 1821229609 HDFS Write: 60011712 SUCCESS
Stage-Stage-3:  HDFS Read: 1821258119 HDFS Write: 60040604 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
```

Output:

```
2001    Total (All India)       217012
2001    Total (States)  212950
2001    Maharashtra     73090
2001    West Bengal     68450
2001    Karnataka       59405
2001    Tamil Nadu      56450
2001    Andhra Pradesh  52610
2001    Kerala  47860
2001    Madhya Pradesh  34300
2001    Gujarat 23955
2001    Odisha  20254
2001    Chhattisgarh    20051
2001    Uttar Pradesh   17580
2001    Rajasthan       15975
2001    Assam   13230
2001    Haryana 10031
2001    Delhi (Ut)      6195
2001    Tripura 4270
2001    Total (Uts)     4062
2001    Punjab  3240
2001    Bihar   3015
2001    Puducherry      2645
2001    Uttarakhand     1555
2001    Himachal Pradesh        1535
2001    Goa     1280
2001    Jharkhand       1250
2001    Jammu & Kashmir 765
2001    A & N Islands   645
2001    Arunachal Pradesh       555
2001    Sikkim  470
2001    Meghalaya       435
2001    Chandigarh      350
2001    Mizoram 270
2001    D & N Haveli    250
2001    Manipur 205
2001    Nagaland        200
2001    Daman & Diu     69
2001    Lakshadweep     0
2002    Total (All India)       220834
2002    Total (States)  216998
2002    Maharashtra     72645
2002    West Bengal     65035
2002    Karnataka       61350
2002    Andhra Pradesh  58465
2002    Tamil Nadu      56134
2002    Kerala  49050
2002    Madhya Pradesh  34495
2002    Gujarat 23216
2002    Lakshadweep     0
2003    Total (All India)       221702
2003    Total (States)  217644
2003    Maharashtra     73800
2003    West Bengal     66400
2003    Karnataka       61805
2003    Tamil Nadu      59360
2003    Andhra Pradesh  57045
2003    Kerala  47190
2003    Madhya Pradesh  33810
2003    Gujarat 22830
2003    Odisha  22100
2003    Chhattisgarh    19595
2003    Uttar Pradesh   18315
2003    Rajasthan       18305
2003    Assam   12963
2003    Haryana 11135
2003    Delhi (Ut)      5765
2003    Tripura 4220
2003    Total (Uts)     4058
2003    Punjab  3155
2003    Bihar   2984
2003    Puducherry      2910
2003    Uttarakhand     1953
2003    Himachal Pradesh        1926
2003    Goa     1500
2003    Jharkhand       1360
2003    Jammu & Kashmir 690
2003    A & N Islands   565
2003    Sikkim  525
2003    Chandigarh      515
2003    Arunachal Pradesh       405
2003    D & N Haveli    260
2003    Mizoram 260
2003    Meghalaya       205
2003    Manipur 130
2003    Daman & Diu     118
2003    Nagaland        109
2003    Lakshadweep     10
2004    Total (All India)       227394
2004    Total (States)  223306
2004    Maharashtra     73645
2004    Andhra Pradesh  67630
2004    West Bengal     67035
2004    Tamil Nadu      64195
2004    Karnataka       59685
2004    Kerala  45265
2004    Madhya Pradesh  33975
```