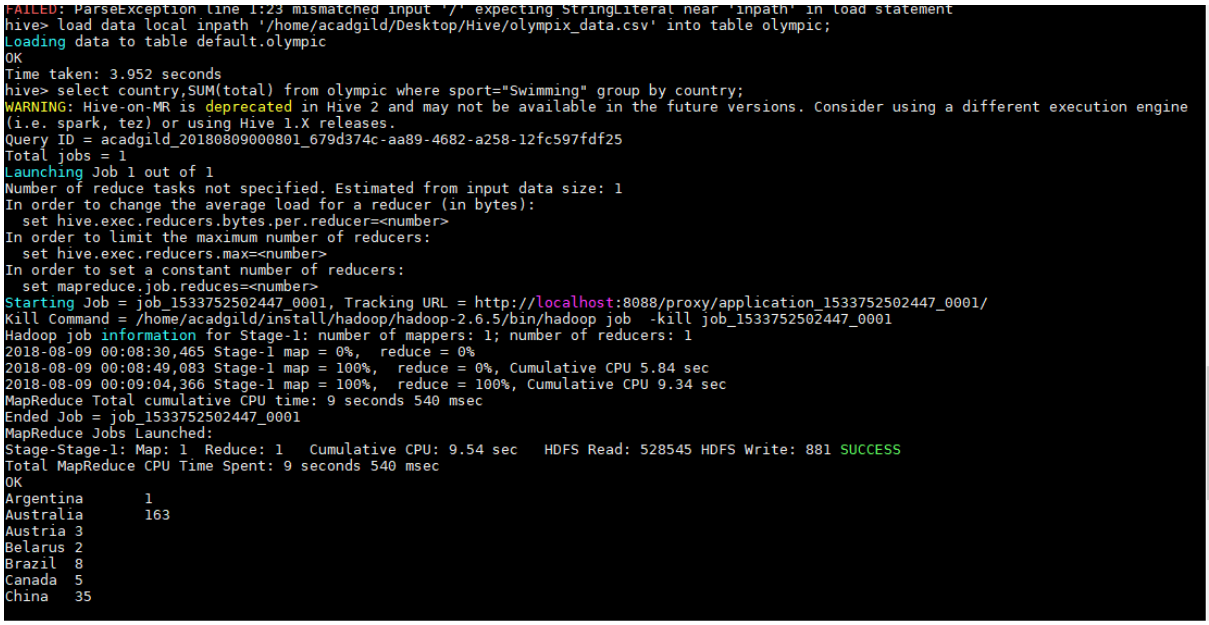
Session 9:
Advance Hive

Assignment 1

Task 1

1) Write a Hive program to find the number of medals won by each country in swimming.



```
Argentina       1
Australia       163
Austria 3
Belarus 2
Brazil  8
Canada  5
China   35
Costa Rica      2
Croatia 1
Denmark 1
France  39
Germany 32
Great Britain   11
Hungary 9
Italy   16
Japan   43
Lithuania       1
Netherlands     46
Norway  2
Poland  3
Romania 6
Russia  20
Serbia  1
Slovakia        2
Slovenia        1
South Africa    11
South Korea     4
Spain   3
Sweden  9
Trinidad and Tobago     1
Tunisia 3
Ukraine 7
United States   267
Zimbabwe        7
Time taken: 65.482 seconds, Fetched: 34 row(s)
```

2)Write a Hive program to find the number of medals that India won year wise.

```
hive> select year,SUM(total) from olympic where country = "India" GROUP BY year;
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_20180809001130_8a03f46b-c835-4e91-bae5-d089b9065300
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_1533752502447_0002, Tracking URL = http://localhost:8088/proxy/application_1533752502447_0002/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job  -kill job_1533752502447_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-08-09 00:11:45,930 Stage-1 map = 0%,  reduce = 0%
2018-08-09 00:11:58,799 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.44 sec
2018-08-09 00:12:13,674 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 6.32 sec
MapReduce Total cumulative CPU time: 6 seconds 320 msec
Ended Job = job_1533752502447_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 6.32 sec   HDFS Read: 528541 HDFS Write: 163 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 320 msec
OK
2000    1
2004    1
2008    3
2012    6
Time taken: 44.108 seconds, Fetched: 4 row(s)
hive>
```

3)Write a Hive Program to find the total number of medals each country won.

```
Time taken: 44.168 seconds, Fetched: 4 row(s)
hive> select country,SUM(total) from olympic GROUP BY country;
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_20180809001335_9334c6a7-00b3-41a1-871a-61229a976d04
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_1533752502447_0003, Tracking URL = http://localhost:8088/proxy/application_1533752502447_0003/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job  -kill job_1533752502447_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-08-09 00:13:50,358 Stage-1 map = 0%,  reduce = 0%
2018-08-09 00:14:03,046 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.51 sec
2018-08-09 00:14:20,472 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.6 sec
MapReduce Total cumulative CPU time: 5 seconds 600 msec
Ended Job = job_1533752502447_0003
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 5.6 sec   HDFS Read: 527721 HDFS Write: 2742 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 600 msec
OK
Afghanistan     2
Algeria 8
Argentina       141
Armenia 10
Australia       609
```

```
Armenia 10
Australia       609
Austria 91
Azerbaijan      25
Bahamas 24
Bahrain 1
Barbados        1
Belarus 97
Belgium 18
Botswana        1
Brazil  221
Bulgaria        41
Cameroon        20
Canada  370
Chile   22
China   530
Chinese Taipei  20
Colombia        13
Costa Rica      2
Croatia 81
Cuba    188
Cyprus  1
Czech Republic  81
Denmark 89
Dominican Republic      5
Ecuador 1
Egypt   8
Eritrea 1
Estonia 18
Ethiopia        29
Finland 118
France  318
Gabon   1
Georgia 23
Germany 629
Great Britain   322
Greece  59
Grenada 1
```

```
Germany 629
Great Britain   322
Greece  59
Grenada 1
Guatemala       1
Hong Kong       3
Hungary 145
Iceland 15
India   11
Indonesia       22
Iran    24
Ireland 9
Israel  4
Italy   331
Jamaica 80
Japan   282
Kazakhstan      42
Kenya   39
Kuwait  2
Kyrgyzstan      3
Latvia  17
Lithuania       30
Macedonia       1
Malaysia        3
Mauritius       1
Mexico  38
Moldova 5
Mongolia        10
Montenegro      14
Morocco 11
Mozambique      1
Netherlands     318
New Zealand     52
Nigeria 39
North Korea     21
```

```
Nigeria 39
North Korea     21
Norway  192
Panama  1
Paraguay        17
Poland  80
Portugal        9
Puerto Rico     2
Qatar   3
Romania 123
Russia  768
Saudi Arabia    6
Serbia  31
Serbia and Montenegro   38
Singapore       7
Slovakia        35
Slovenia        25
South Africa    25
South Korea     308
Spain   205
Sri Lanka       1
Sudan   1
Sweden  181
Switzerland     93
Syria   1
Tajikistan      3
Thailand        18
Togo    1
Trinidad and Tobago     19
Tunisia 4
Turkey  28
Uganda  1
```

```
Spain   205
Sri Lanka       1
Sudan   1
Sweden  181
Switzerland     93
Syria   1
Tajikistan      3
Thailand        18
Togo    1
Trinidad and Tobago     19
Tunisia 4
Turkey  28
Uganda  1
Ukraine 143
United Arab Emirates    1
United States   1312
Uruguay 1
Uzbekistan      19
Venezuela       4
Vietnam 2
Zimbabwe        7
Time taken: 46.26 seconds, Fetched: 110 row(s)
hive>
```

4)Write a Hive program to find the number of gold medals each country won

```
rk, tez) or using Hive 1.X releases.
hive> select country,SUM(gold) from olympic GROUP BY country;
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_20180809001412_41390f73-2b1d-4dd9-ba55-1a2b6b6ab44c
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_1533752502447_0004, Tracking URL = http://localhost:8088/proxy/application_1533752502447_0004/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job  -kill job_1533752502447_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2018-08-09 00:14:53,089 Stage-1 map = 0%,  reduce = 0%
2018-08-09 00:15:06,740 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.44 sec
2018-08-09 00:15:20,129 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.44 sec
MapReduce Total cumulative CPU time: 5 seconds 440 msec
Ended Job = job_1533752502447_0004
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 5.44 sec   HDFS Read: 527711 HDFS Write: 2703 SUCCESS
```

```
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 5.44 sec   HDFS Read: 527711 HDFS Write: 2703 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 440 msec
OK
Afghanistan     0
Algeria 2
Argentina       49
Armenia 0
Australia       163
Austria 36
Azerbaijan      6
Bahamas 11
Bahrain 0
Barbados        0
Belarus 17
Belgium 2
Botswana        0
Brazil  46
Bulgaria        8
Cameroon        20
Canada  168
Chile   3
China   234
Chinese Taipei  2
Colombia        2
Costa Rica      0
Croatia 35
Cuba    57
Cyprus  0
Czech Republic  14
Denmark 46
Dominican Republic      3
Ecuador 0
Egypt   1
Eritrea 0
Estonia 6
Ethiopia        13
```

```
Estonia 6
Ethiopia        13
Finland 11
France  108
Gabon   0
Georgia 6
Germany 223
Great Britain   124
Greece  12
Grenada 1
Guatemala       0
Hong Kong       0
Hungary 77
Iceland 0
India   1
Indonesia       5
Iran    10
Ireland 1
Israel  1
Italy   86
Jamaica 24
Japan   57
Kazakhstan      13
Kenya   11
Kuwait  0
Kyrgyzstan      0
Latvia  3
Lithuania       5
Macedonia       0
Malaysia        0
Mauritius       0
Mexico  19
Moldova 0
Mongolia        2
Montenegro      0
Morocco 2
```

```
Morocco 2
Mozambique      1
Netherlands     101
New Zealand     18
Nigeria 6
North Korea     6
Norway  97
Panama  1
Paraguay        0
Poland  20
Portugal        1
Puerto Rico     0
Qatar   0
Romania 57
Russia  234
Saudi Arabia    0
Serbia  1
Serbia and Montenegro   11
Singapore       0
Slovakia        10
Slovenia        5
South Africa    10
South Korea     110
Spain   19
Sri Lanka       0
Sudan   0
Sweden  57
Switzerland     21
Syria   0
Tajikistan      0
Thailand        6
Togo    0
Trinidad and Tobago     1
Tunisia 2
Turkey  9
Uganda  1
```

```
Tunisia 2
Turkey  9
Uganda  1
Ukraine 31
United Arab Emirates    1
United States   552
Uruguay 0
Uzbekistan      5
Venezuela       1
Vietnam 0
Zimbabwe        2
Time taken: 69.956 seconds, Fetched: 110 row(s)
hive>
```

Task 3

Link: https://acadgild.com/blog/transactions-in-hive/

Refer the above given link for transactions in Hive and implement the operations given in the

blog using your own sample data set and send us the screenshot.

Creating a Table That Supports Hive Transactions

```
rk, tez) or using Hive 1.X releases.
hive> CREATE TABLE student(student_name string,age int,gpa decimal(3,2)) clustered by (age) into 2 buckets stored as orc TBLPROPERTIES('t
ransactional'='true');
OK
Time taken: 10.617 seconds
hive> select * from student;
FAILED: SemanticException [Error 10265]: This command is not allowed on an ACID table default.student with a non-ACID transaction manager
. Failed command: select * from student
hive> show tables;
OK
college
olympic
student
Time taken: 0.217 seconds, Fetched: 3 row(s)
```

Inserting Data into a Hive Table

```
hive> INSERT INTO TABLE student VALUES ('jyoti', 32, 1.28), ('sadhana', 32, 2.32);
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_20180817003342_6704ed2e-a490-475e-b29d-dafa21d99e01
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 2
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_1534444133836_0004, Tracking URL = http://localhost:8088/proxy/application_1534444133836_0004/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job  -kill job_1534444133836_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 2
2018-08-17 00:34:00,268 Stage-1 map = 0%,  reduce = 0%
2018-08-17 00:34:13,315 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.24 sec
2018-08-17 00:34:34,521 Stage-1 map = 100%,  reduce = 33%, Cumulative CPU 5.77 sec
2018-08-17 00:34:35,799 Stage-1 map = 100%,  reduce = 67%, Cumulative CPU 8.45 sec
2018-08-17 00:34:38,160 Stage-1 map = 100%,  reduce = 83%, Cumulative CPU 10.29 sec
2018-08-17 00:34:39,241 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 11.94 sec
MapReduce Total cumulative CPU time: 11 seconds 940 msec
Ended Job = job_1534444133836_0004
Loading data to table default.student
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 2   Cumulative CPU: 11.94 sec   HDFS Read: 13680 HDFS Write: 1138 SUCCESS
Total MapReduce CPU Time Spent: 11 seconds 940 msec
OK
Time taken: 59.232 seconds
hive> select * from student;
OK
sadhana 32      2.32
jyoti   32      1.28
Time taken: 0.41 seconds, Fetched: 2 row(s)
hive>
```

Re-insert the same data again, it will be appended to the previous data as shown below:

```
hive> INSERT INTO TABLE student VALUES ('jyoti', 32, 1.28), ('sadhana', 32, 2.32);
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_20180817003508_06e1b4ba-2a74-43e0-83bc-7396278b937b
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 2
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_1534444133836_0005, Tracking URL = http://localhost:8088/proxy/application_1534444133836_0005/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job  -kill job_1534444133836_0005
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 2
2018-08-17 00:35:20,813 Stage-1 map = 0%,  reduce = 0%
2018-08-17 00:35:31,189 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.73 sec
2018-08-17 00:35:50,291 Stage-1 map = 100%,  reduce = 33%, Cumulative CPU 5.4 sec
2018-08-17 00:35:52,694 Stage-1 map = 100%,  reduce = 67%, Cumulative CPU 8.06 sec
2018-08-17 00:35:53,832 Stage-1 map = 100%,  reduce = 83%, Cumulative CPU 9.7 sec
2018-08-17 00:35:55,053 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 11.26 sec
MapReduce Total cumulative CPU time: 11 seconds 260 msec
Ended Job = job_1534444133836_0005
Loading data to table default.student
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 2   Cumulative CPU: 11.26 sec   HDFS Read: 13594 HDFS Write: 1139 SUCCESS
Total MapReduce CPU Time Spent: 11 seconds 260 msec
OK
Time taken: 49.679 seconds
hive> select * from student;
OK
sadhana 32      2.32
jyoti   32      1.28
sadhana 32      2.32
jyoti   32      1.28
Time taken: 0.308 seconds, Fetched: 4 row(s)
hive>
```

Updating the Data in Hive Table

```
Time taken: 0.308 seconds, Fetched: 4 row(s)
hive> UPDATE student set age = 30 where age = 32;
FAILED: SemanticException [Error 10302]: Updating values of bucketing columns is not supported.  Column age.
hive> UPDATE student set gpa = 3.3 where student_name = 'sadhana';
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_20180817003911_1b315938-8597-4a42-a63d-a01bdb5558e0
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 2
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_1534444133836_0006, Tracking URL = http://localhost:8088/proxy/application_1534444133836_0006/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job  -kill job_1534444133836_0006
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 2
2018-08-17 00:39:25,484 Stage-1 map = 0%,  reduce = 0%
2018-08-17 00:39:47,680 Stage-1 map = 33%,  reduce = 0%, Cumulative CPU 3.92 sec
2018-08-17 00:39:48,879 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 7.93 sec
2018-08-17 00:40:10,573 Stage-1 map = 100%,  reduce = 91%, Cumulative CPU 13.27 sec
2018-08-17 00:40:11,894 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 14.42 sec
MapReduce Total cumulative CPU time: 14 seconds 420 msec
Ended Job = job_1534444133836_0006
Loading data to table default.student
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2  Reduce: 2   Cumulative CPU: 14.42 sec   HDFS Read: 22079 HDFS Write: 905 SUCCESS
Total MapReduce CPU Time Spent: 14 seconds 420 msec
OK
Time taken: 63.308 seconds
hive>
```

Output

```
OK
Time taken: 63.308 seconds
hive> select * from student;
OK
sadhana 32      3.30
jyoti   32      1.28
sadhana 32      3.30
jyoti   32      1.28
Time taken: 0.32 seconds, Fetched: 4 row(s)
hive>
```

Deleting a Row from Hive Table