

## Task 1

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

```

hive> select country, sum(total) from olympic where sport="Swimming" group by country;
WARNING: Hive-on-HR 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 = acedgild_20180520220327_000001795-7704-4e30-bf4c-2270a2054400
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.reducers=<number>
Starting Job = job_1526792975751_0000, Tracking URL = http://localhost:8080/proxy/application_1526792975751_0000/
Kill Command = /home/acedgild/install/hadoop/hadoop-2.6.0/bin/hadoop job -kill job_1526792975751_0000
Hadoop job information for Stage-1: number of mappers: 1, number of reducers: 1
2018-05-20 22:03:40,330 Stage-1 map = 0%, reduce = 0%
2018-05-20 22:03:54,912 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.95 sec
2018-05-20 22:04:00,450 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.09 sec
MapReduce Total cumulative CPU time: 5 seconds 90 msec
Ended Job = job_1526792975751_0000
HadoopJob Launched:
Stage-Stage 1: Map: 1 Reducers: 1 Cumulative CPU: 5.09 sec HDFS Read: 528574 HDFS Write: 831 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 90 msec
OK
Argentina 1
Australia 153
Austria 3
Belarus 2
Brazil 8
Canada 5
China 30
Costa Rica 2
Croatia 1
Denmark 1
France 30
Germany 32
Great Britain 11
Hungary 9

```

```

Stage-Stage 1: Map: 1 Reducers: 1 Cumulative CPU: 5.09 sec HDFS Read: 528574 HDFS Write: 831 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 90 msec
OK
Argentina 1
Australia 153
Austria 3
Belarus 2
Brazil 8
Canada 5
China 30
Costa Rica 2
Croatia 1
Denmark 1
France 30
Germany 32
Great Britain 11
Hungary 9
Italy 10
Japan 43
Lithuania 1
Netherlands 45
Norway 2
Poland 3
Romania 6
Russia 20
Serbia 1
Slovakia 2
Slovenia 1
South Africa 11
South Korea 4
Spain 5
Sweden 9
Trinidad and Tobago 1
Tunisia 3
Ukraine 7
United States 257
Zimbabwe 7
Time taken: 30.88 seconds, Fetched: 34 row(s)
hive>

```

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_78188538728847_3675485-bc9a-4183-aa01-904126cc6709
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=numberbytes
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.reducers=number
Starting Job = job_1526792975751_0007, Tracking URL = http://localhost:8080/proxy/application_1526792975751_0007/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1526792975751_0007
Hadoop job information for Stage-1: number of mappers: 1, number of reducers: 1
2016-05-20 22:08:20,730 Stage-1 map = 0%, reduce = 0%
2016-05-20 22:09:32,000 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.95 sec
2016-05-20 22:09:32,530 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.95 sec
MapReduce Total cumulative CPU time: 4 seconds 950 msec
Ended Job = job_1526792975751_0007
MapReduce Jobs Launched:
  Stage-Stage-1: Map: 1 Reduc: 1 Cumulative CPU: 4.95 sec  HDFS Read: 528562 HDFS Write: 163 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 950 msec
OK
2008      1
2004      1
2008      3
2012      6
Time taken: 33.273 seconds, Fetched: 4 row(s)
hive>

```

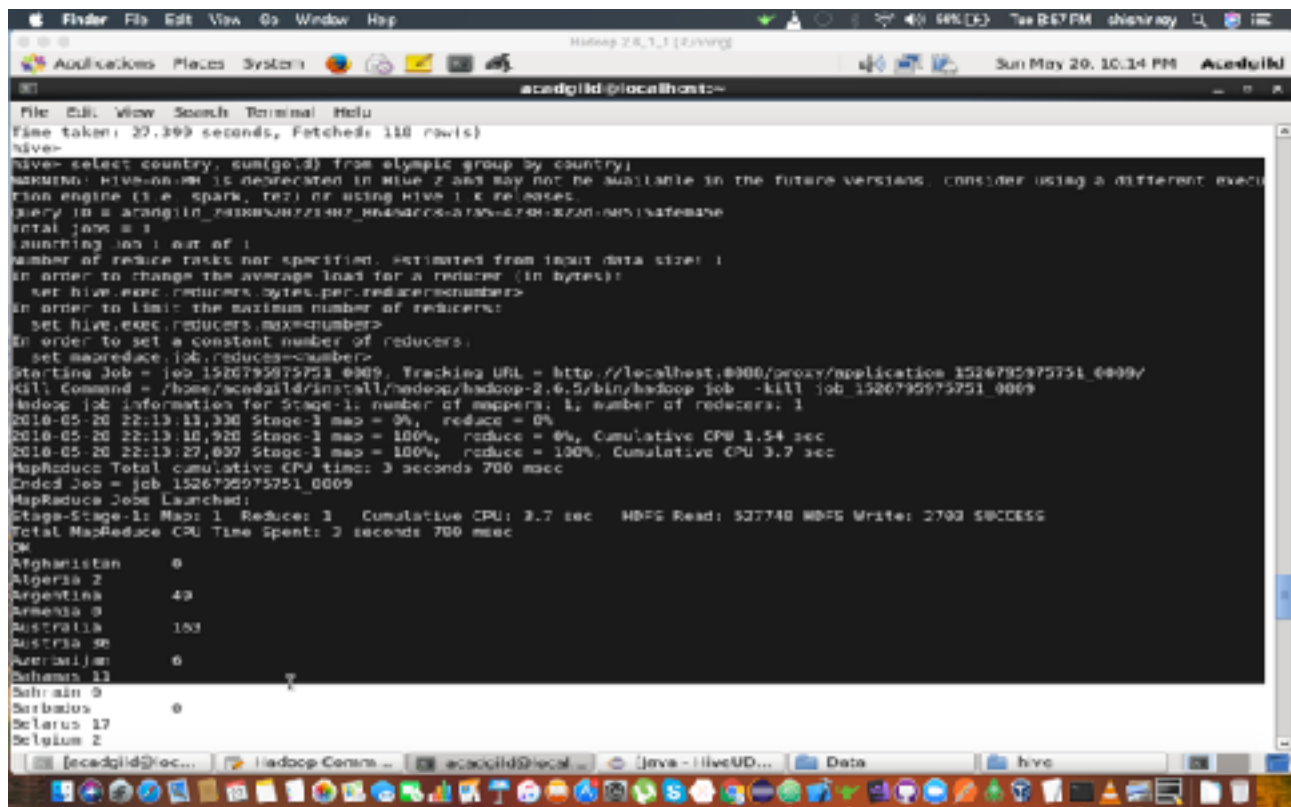
3. Write a Hive Program to find the total number of medals each country won.

```

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_78188538728847_3675485-bc9a-4aa2-bf79-73hav0scdab
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=numberbytes
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.reducers=number
Starting Job = job_1526792975751_0008, Tracking URL = http://localhost:8080/proxy/application_1526792975751_0008/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1526792975751_0008
Hadoop job information for Stage-1: number of mappers: 1, number of reducers: 1
2016-05-20 22:11:02,130 Stage-1 map = 0%, reduce = 0%
2016-05-20 22:11:09,730 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.54 sec
2016-05-20 22:11:10,522 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.45 sec
MapReduce Total cumulative CPU time: 3 seconds 450 msec
Ended Job = job_1526792975751_0008
MapReduce Jobs Launched:
  Stage-Stage-1: Map: 1 Reduc: 1 Cumulative CPU: 3.45 sec  HDFS Read: 527742 HDFS Write: 2743 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 450 msec
OK
Afghanistan      2
Algeria           8
Argentina         143
Armenia           10
Australia         609
Austria           93
Azerbaijan       23
Bahamas           24
Bahrain           1
Belarus           1
Belgium           97
Bolivia           16

```

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



```
accedgild@localhost:~$
File Edit View Search Terminal Help
Hadoop 2.6.1 (running)
Applications Places System
accedgild@localhost:~$
Time taken: 27.399 seconds, Fetched: 110 row(s)
hive>
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 = a7ad91d_781885d7214b7_864d4cc5-3735-474b-8720-0b513af6a59e
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.reducers=<number>
Starting Job = job_1526795975751_0009, Tracking URL = http://localhost:8080/proxy/application_1526795975751_0009/
Kill Command = /home/accedgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1526795975751_0009
Hadoop job information for Stage-1: number of mappers: 1, number of reducers: 1
2016-05-20 22:13:13,330 Stage-1 map = 0%, reduce = 0%
2016-05-20 22:13:16,920 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.54 sec
2016-05-20 22:13:27,007 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.7 sec
MapReduce Total cumulative CPU time: 3 seconds 700 msec
Ended Job = job_1526795975751_0009
MapReduce Job Launched:
Stage-Stage-1: Map: 1 Reducer: 1 Cumulative CPU: 3.7 sec HDFS Read: 527740 HDFS Write: 2700 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 700 msec
OK
Afghanistan      0
Algeria          2
Argentina        49
Armenia          0
Australia        153
Austria          38
Azerbaijan       6
Bahamas          11
Bahrain          0
Barbados         0
Belarus          17
Belgium          2
```