

## 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-22f9a2954299
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         39
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         39
Germany        32
Great Britain  11
Hungary        9
Italy          16
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: 38.88 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-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-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_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-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          97
Belgium          16

```

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

```

File Edit View Search Terminal Help
Hadoop 2.6.1 [2/3/2016]
Acadgild@localhost:~$
File Edit View Search Terminal Help
Time taken: 20.399 seconds, Fetched: 118 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 = acadgild_20160520101407_hiveacdc3-afac-47d6-8720-00013afefame
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/acadgild/install/hadoop/hadoop-2.6.0/bin/hadoop job -kill job_1526795975751_0009
Hadoop job information for Stage-1: number of mappers: 1, number of reducers: 1
2016-05-20 22:13:18,330 Stage-3 map = 0%, reduce = 0%
2016-05-20 22:13:18,920 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 1.54 sec
2016-05-20 22:13:27,007 Stage-3 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 Reduces: 1 Cumulative CPU: 3.7 sec HDFS Read: 527740 HDFS Write: 2700 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 700 msec
country gold
Afghanistan 0
Algeria 2
Argentina 43
Armenia 0
Australia 183
Austria 36
Azerbaijan 0
Bahamas 11
Belarus 0
Belgium 2

```

## Task 2

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.

```

File Edit View Search Terminal Help
Hadoop 2.8.1 [2x]
Applications Places System
acsdgild@localhost:~$
File Edit View Search Terminal Help

hive> set hive.support.concurrency = true;
hive> set hive.enforce.bucketing = true;
hive> set hive.exec.dynamic.partition.mode = nonstrict;
hive> set hive.tbl.manager = org.apache.hadoop.hive.q1.lockmgr.DataSourceManager;
hive> create table college(crig in int, c1g_name string, c1g_location string) clustered by (c1g_in) into 5 buckets stored as rc
rc (format=keys,transactional) = true;
hive>
time taken: 0.444 seconds
hive> show tables;
hive>
college
dynamic
temperature_data
temperature_data_mv
time taken: 0.159 seconds, fetched: 4 row(s)
hive> insert into table college values(1,'nit','bangalore'), (2,'nit','raipet'), (3,'nit','ranchi'), (4,'nit','pune'), (5,'vit','vellore'), (6,'arm','chennai'), (7,'rv','bangalore');
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 = acsdgild_20100520230747_dcf60ec-5307-4530-9731-c12710db2dc
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 5
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_0000, Tracking URL = http://localhost:8086/proxy/application_1526795975751_0010/
Kill Command = /home/acsdgild/install/hadoop/hadoop-2.8.5/bin/hadoop job -kill job_1526795975751_0010
MapJob job information for Stage1: number of mappers: 1; number of reducers: 5
2018-05-20 23:57:58.417 Stage-1 map = 0%, reduce = 0%
2018-05-20 23:58:08.326 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.84 sec
2018-05-20 23:58:38.027 Stage-1 map = 100%, reduce = 13%, Cumulative CPU 3.88 sec
2018-05-20 23:58:46.834 Stage-1 map = 100%, reduce = 33%, Cumulative CPU 8.34 sec
2018-05-20 23:58:42.330 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 9.3 sec
2018-05-20 23:58:39.823 Stage-1 map = 100%, reduce = 89%, Cumulative CPU 18.95 sec
[acsdgild@localhost ~]$ [Hadoop Commands (...)] [acsdgild@localhost ~]$ [java - HiveUOF Demo(...)]

```

```
File Edit View Search Terminal Help
Hadoop 2.6.1, 1 (running)
Applications Places System
accdgild@localhost:~$
File Edit View Search Terminal Help
olympic
temperature_data
temperature_data_vw
time taken: 0.159 seconds, fetched: 4 row(s)
hive> insert into table college values(1,'ist','bangalore'), (2,'nit','raipuri'), (3,'nit','ranchi'), (4,'nit','pune'), (5,'vit','vellore'), (6,'srn','chennai'), (7,'rvu','bangalore');
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 = accdgild_2a885a2e7e2e7_nesrmer-5a97-ee58-7241-m271a8b7ndc
total jobs = 1
launching job 1 out of 1
number of reduce tasks determined at compile time: 5
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_1526795075751_0010, Tracking URL = http://localhost:8080/proxy/application_1526795075751_0010/
Kill Command = /home/accdgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1526795075751_0010
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 5
2016-05-20 23:57:50,417 Stage-1 map = 0%, reduce = 0%
2016-05-20 23:58:00,326 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.64 sec
2016-05-20 23:58:38,927 Stage-1 map = 100%, reduce = 13%, Cumulative CPU 2.90 sec
2016-05-20 23:58:40,824 Stage-1 map = 100%, reduce = 53%, Cumulative CPU 8.34 sec
2016-05-20 23:58:42,525 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 9.5 sec
2016-05-20 23:58:50,823 Stage-1 map = 100%, reduce = 80%, Cumulative CPU 18.05 sec
2016-05-20 23:59:01,507 Stage-1 map = 100%, reduce = 93%, Cumulative CPU 21.4 sec
2016-05-20 23:59:02,755 Stage-1 map = 100%, reduce = 99%, Cumulative CPU 25.70 sec
2016-05-20 23:59:03,825 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 26.19 sec
MapReduce Total Cumulative CPU Time: 26 seconds 190 msec
ended job = job_1526795075751_0010
Loading data to table custom_college
MapReduce Job Launched:
Stage-Stage-1: Map: 1 Reducer: 5 Cumulative CPU: 26.19 sec HOPS Read: 27155 HOPS Write: 4963 SUCCESS
Total MapReduce CPU Time Spent: 26 seconds 190 msec
OK
Time taken: 76.63 seconds
hive>
```

```
File Edit View Search Terminal Help
Hadoop 2.6.1, 1 (running)
Applications Places System
accdgild@localhost:~$
File Edit View Search Terminal Help
hive> select * from college;
OK
5      vit      vellore
6      srn      chennai
1      ist      bangalore
7      rvu      bangalore
2      nit      raipuri
3      nit      ranchi
4      nit      pune
time taken: 0.243 seconds, fetched: 7 row(s)
hive>
```



```
File Edit View Search Terminal Help
Hadoop 2.8.1_1 [2/3/10]
Applications Places System
acacgild@localhost:~$
acacgild@localhost:~$
File Edit View Search Terminal Help
hive> select * from college;
OK
+----+-----+
| cld | value |
+----+-----+
| 1   | 100000 |
| 2   | 200000 |
| 3   | 300000 |
| 4   | 400000 |
| 5   | 500000 |
| 6   | 600000 |
| 7   | 700000 |
| 8   | 800000 |
+----+-----+
time taken: 0.271 seconds, Fetched: 7 row(s)
hive> update college set cld_id = 8 where cld_id = 7;
FAILED: SemanticException [Error 10002]: Updating values of bucketing columns is not supported. Column cld_id.
hive> update college set cld_name = '111 mamad' where cld_id = 0;
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 = acacgild_20180521090344_95e3f2c1-ccc1-47c0-a377-de000190ccc0
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 5
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_0011, Tracking URL = http://localhost:8080/proxy/application_1526795975751_0011/
Kill Command = /home/acacgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1526795975751_0011
Hadoop job information for Stage-1: number of mappers: 5; number of reducers: 5
2018-05-21 09:03:55,424 Stage-1 map = 0%, reduce = 0%, Cumulative CPU 2.94 sec
2018-05-21 09:04:26,578 Stage-1 map = 20%, reduce = 0%, Cumulative CPU 12.83 sec
2018-05-21 09:04:37,483 Stage-1 map = 40%, reduce = 0%, Cumulative CPU 14.27 sec
2018-05-21 09:04:58,793 Stage-1 map = 60%, reduce = 0%, Cumulative CPU 14.27 sec
```

```
File Edit View Search Terminal Help
Hadoop 2.8.1_1 [2/3/10]
Applications Places System
acacgild@localhost:~$
acacgild@localhost:~$
File Edit View Search Terminal Help
time taken: 0.271 seconds, Fetched: 7 row(s)
hive> update college set cld_id = 8 where cld_id = 7;
FAILED: SemanticException [Error 10002]: Updating values of bucketing columns is not supported. Column cld_id.
hive> update college set cld_name = '111 mamad' where cld_id = 0;
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 = acacgild_20180521090344_95e3f2c1-ccc1-47c0-a377-de000190ccc0
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 5
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_0011, Tracking URL = http://localhost:8080/proxy/application_1526795975751_0011/
Kill Command = /home/acacgild/install/hadoop/hadoop-2.6.5/bin/hadoop job -kill job_1526795975751_0011
Hadoop job information for Stage-1: number of mappers: 5; number of reducers: 5
2018-05-21 09:03:55,424 Stage-1 map = 0%, reduce = 0%, Cumulative CPU 2.94 sec
2018-05-21 09:04:26,578 Stage-1 map = 20%, reduce = 0%, Cumulative CPU 12.83 sec
2018-05-21 09:04:37,483 Stage-1 map = 40%, reduce = 0%, Cumulative CPU 14.27 sec
2018-05-21 09:04:58,793 Stage-1 map = 60%, reduce = 0%, Cumulative CPU 15.65 sec
2018-05-21 09:05:10,132 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 16.94 sec
2018-05-21 09:05:36,853 Stage-1 map = 100%, reduce = 27%, Cumulative CPU 18.04 sec
2018-05-21 09:05:10,132 Stage-1 map = 100%, reduce = 53%, Cumulative CPU 21.12 sec
2018-05-21 09:05:13,480 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 22.82 sec
2018-05-21 09:05:15,384 Stage-1 map = 100%, reduce = 87%, Cumulative CPU 25.82 sec
2018-05-21 09:05:16,484 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 27.51 sec
MapReduce Total Cumulative CPU time: 27 seconds 510 msec
Ending Job = job_1526795975751_0011
Loading data to table custom.college
MapReduce Job Launches:
Stage-Stage 1: Map: 5 Reducer: 5 Cumulative CPU: 27.51 sec HDFS Read: 52320 HDFS Write: 937 SUCCESS
Final MapReduce CPU Time Spent: 27 seconds 510 msec
OK
time taken: 30.499 seconds
hive>
```



```
Finder File Edit View Go Window Help
Hadoop 2.8.1_1 [Running]
Applications Places System
acacgild@localhost:~$
File Edit View Search Terminal Help
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 5
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/maxnumbers
In order to set a constant number of reducers:
  set mapreduce.job.reduces=number
Starting job = job_1526795975751_0012, tracking url = http://localhost:8080/proxy/application_1526795975751_0012/
kill: Command = /home/acacgild/installs/hadoop/hadoop-2.8.1/bin/hadoop job -kill job_1526795975751_0012
Hadoop job information for stage-1: number of mappers: 5; number of reducers: 5
2018-05-23 09:08:04,577 Stage-1 map = 0%, reduce = 0%
2018-05-23 09:08:42,630 Stage-1 map = 20%, reduce = 0%, Cumulative CPU 3.57 sec
2018-05-23 09:08:45,497 Stage-1 map = 40%, reduce = 0%, Cumulative CPU 6.68 sec
2018-05-23 09:08:46,090 Stage-1 map = 50%, reduce = 0%, Cumulative CPU 9.15 sec
2018-05-23 09:08:50,710 Stage-1 map = 90%, reduce = 0%, Cumulative CPU 12.35 sec
2018-05-23 09:08:53,937 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 12.99 sec
2018-05-23 09:09:17,256 Stage-1 map = 100%, reduce = 53%, Cumulative CPU 16.96 sec
2018-05-23 09:09:19,690 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 18.06 sec
2018-05-23 09:09:22,472 Stage-1 map = 100%, reduce = 87%, Cumulative CPU 21.46 sec
2018-05-23 09:09:23,594 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 22.92 sec
MapReduce Total cumulative CPU time: 22 seconds 920 msec
Ended Job = job_1526795975751_0012
Loading data to table custom.College
MapReduce Jobs Launched:
Stage-Stage-1: Map: 5 Reduce: 5 Cumulative CPU: 22.92 sec HDFS Read: 50677 HDFS Write: 747 SUCCESS
Total MapReduce CPU Time Spent: 22 seconds 920 msec
OK
Time taken: 22.779 seconds
hive> select * from college;
OK
5      EIT Mumbai      Chennai
1      EIT Bangalore
7      RV Bangalore
2      MIT Manipal
3      MIT Ranchi
4      MIT Pune
Time taken: 0.223 seconds, Fetched: 6 row(s)
hive>
```