Assignment 9

Task1: Hive program to find the number of medals won by each country in swimming.

Creating the main table and loading the olympx.csv in to the table

```
Create Database and Tables
--------------------------
create database simplidb;
show databases;
use simplidb;
show tables;


CREATE TABLE olympics(
Athlete STRING,
Age INT,
Country STRING,
Year INT,
Closingdate INT,
Sport STRING,
Gold INT,
Silver INT,
Bronze INT,
Total INT
)
row format delimited fields terminated by '\t';

LOAD DATA LOCAL INPATH 'olympix_data.csv' into TABLE olympics;
```

```
create table swim (country STRING, total INT)
partitioned by (sport STRING)
row format delimited fields terminated by '\t';


alter table swim add partition(sport = 'Swimming');


FROM olympics oly INSERT OVERWRITE TABLE swim PARTITION(sport) select oly.country,oly.total,
oly.sport DISTRIBUTE BY sport;
```

Task 2 Hive program to find the number of medals that India won year wise.

```
CREATE TABLE olympics
(
athlete STRING,
age INT,
country STRING,
year INT,
closingdate INT,
Sport STRING,
Gold INT,
Silver INT,
Bronze INT,
Total INT
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY '\t';

select * from olympics where country = 'India'
```

```
hive> select * from olympics where country = 'India';
OK
Yogeshwar Dutt  29      India   2012    NULL    Wrestling       0       0       1       1
Sushil Kumar    29      India   2012    NULL    Wrestling       0       1       0       1
Sushil Kumar    25      India   2008    NULL    Wrestling       0       0       1       1
Karnam Malleswari       25      India   2000    NULL    Weightlifting   0       0       1       1
Vijay Kumar     26      India   2012    NULL    Shooting        0       1       0       1
Gagan Narang    29      India   2012    NULL    Shooting        0       0       1       1
Abhinav Bindra  25      India   2008    NULL    Shooting        1       0       0       1
Rajyavardhan Rathore    34      India   2004    NULL    Shooting        0       1       0       1
M. C. Mary Kom  29      India   2012    NULL    Boxing  0       0       1       1
Vijender Singh  22      India   2008    NULL    Boxing  0       0       1       1
Saina Nehwal    22      India   2012    NULL    Badminton       0       0       1       1
Time taken: 1.273 seconds, Fetched: 11 row(s)
```

Medals won by India.

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

```
hive> describe olympics;
OK
athlete                 string
age                     int
country                 string
year                    int
closingdate             int
sport                   string
gold                    int
silver                  int
bronze                  int
total                   int
Time taken: 0.134 seconds, Fetched: 10 row(s)
hive> SELECT country, SUM(total) FROM olympics group by country;
```

```
Afghanistan     2
Algeria 8
Argentina       141
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
```

```
Gabon   1
Georgia 23
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
```

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

```
hive> SELECT country, SUM(Gold) FROM olympics group by country;
```

```
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
Mozambique      1
```

```
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
```

TASK3:

Creating a table called college

```
hive> set hive.support.concurrency = true;
hive> set hive.enforce.bucketing = true;
hive> set hive.exec.dynamic.partition.mode = nonstrict;
hive> set hive.txn.manager = org.apache.hadoop.hive.ql.lockmgr.DbTxnManager;
hive> set hive.compactor.inititor.on = true;
Query returned non-zero code: 1, cause: hive configuration hive.compactor.inititor.on does not exists.
hive> set hive.compactor.worker.threads = a positive numberon at least one instance of the Thrift metastore se
Query returned non-zero code: 1, cause: 'SET hive.compactor.worker.threads=a positive numberon at least one in
hrift metastore service' FAILED because hive.compactor.worker.threads expects INT type value.
hive> CREATE TABLE college(clg_id int,clg_name string,clg_loc string) clustered by (clg_id) into 5 buckets sto
ROPERTIES('transactional'='true');
OK
Time taken: 1.464 seconds
hive> show tables;
OK
buck_users
college
csv_table
```

Loading DATA

```
,'JNTUA','atp'),(7,'cambridge','us');
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different exe
tion engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = acadgild_20180919004320_19141fc6-de51-4ffe-aa9f-0c0e4e28093b
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.reduces=<number>
Starting Job = job_1537247019915_0022, Tracking URL = http://localhost:8088/proxy/application_1537247019915_0022/
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job  -kill job_1537247019915_0022
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 5
2018-09-19 00:43:32,590 Stage-1 map = 0%,  reduce = 0%
2018-09-19 00:43:42,769 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.79 sec
2018-09-19 00:44:11,645 Stage-1 map = 100%,  reduce = 13%, Cumulative CPU 3.76 sec
2018-09-19 00:44:17,496 Stage-1 map = 100%,  reduce = 27%, Cumulative CPU 5.28 sec
2018-09-19 00:44:20,336 Stage-1 map = 100%,  reduce = 40%, Cumulative CPU 6.35 sec
2018-09-19 00:44:21,960 Stage-1 map = 100%,  reduce = 53%, Cumulative CPU 6.86 sec
2018-09-19 00:44:24,692 Stage-1 map = 100%,  reduce = 67%, Cumulative CPU 9.48 sec
2018-09-19 00:44:33,751 Stage-1 map = 100%,  reduce = 73%, Cumulative CPU 13.73 sec
2018-09-19 00:44:37,817 Stage-1 map = 100%,  reduce = 80%, Cumulative CPU 16.67 sec
2018-09-19 00:44:39,077 Stage-1 map = 100%,  reduce = 87%, Cumulative CPU 19.54 sec
2018-09-19 00:44:40,194 Stage-1 map = 100%,  reduce = 93%, Cumulative CPU 22.45 sec
2018-09-19 00:44:41,234 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 25.23 sec
MapReduce Total cumulative CPU time: 25 seconds 230 msec
Ended Job = job_1537247019915_0022
Loading data to table simplidb.college
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 5   Cumulative CPU: 25.23 sec   HDFS Read: 27118 HDFS Write: 4006 SUCCESS
Total MapReduce CPU Time Spent: 25 seconds 230 msec
OK
Time taken: 83.44 seconds
hive>
```

Table contents

```
Time taken: 83.44 seconds
hive> select * from college;
OK
5       stanford        uk
6       JNTUA   atp
1       nec     nlr
7       cambridge       us
2       vit     vlr
3       srm     chen
4       lpu     del
Time taken: 0.406 seconds, Fetched: 7 row(s)
hive>
```

Inserting again

```
Time taken: 86.035 seconds
hive> select * from college;
OK
5       stanford        uk
5       stanford        uk
6       JNTUA   atp
1       nec     nlr
6       JNTUA   atp
1       nec     nlr
7       cambridge       us
2       vit     vlr
7       cambridge       us
2       vit     vlr
3       srm     chen
3       srm     chen
4       lpu     del
4       lpu     del
Time taken: 0.317 seconds, Fetched: 14 row(s)
hive>
```

Data update

```
5       stanford        uk
5       stanford        uk
6       IIT     atp
1       nec     nlr
6       IIT     atp
1       nec     nlr
7       cambridge       us
2       vit     vlr
7       cambridge       us
2       vit     vlr
3       srm     chen
3       srm     chen
4       lpu     del
4       lpu     del
Time taken: 0.337 seconds, Fetched: 14 row(s)
hive>
```

Deleting a row

```
6       IIT     atp
1       nec     nlr
6       IIT     atp
1       nec     nlr
7       cambridge       us
2       vit     vlr
7       cambridge       us
2       vit     vlr
3       srm     chen
3       srm     chen
4       lpu     del
4       lpu     del
Time taken: 0.299 seconds, Fetched: 12 row(s)
hive>
```