

Assignment 7.2

Problem Statement 1 :-

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

Solution:-

use custom;

```
CREATE TABLE olympix_data
```

```
(
```

```
  Aname STRING,
```

```
  Age INT,
```

```
  Country STRING,
```

```
  Year INT,
```

```
  Closing_Datre STRING,
```

```
  Sport STRING,
```

```
  Gold INT,
```

```
  Silver INT,
```

```
  Bronze INT,
```

```
  Total_medals INT
```

```
)
```

```
ROW FORMAT DELIMITED
```

```
FIELDS TERMINATED BY '\t';
```

LOAD DATA LOCAL INPATH '/home/acadgild/hadoop/olympix_data.csv' INTO
TABLE custom.olympix_data;

```
hive> use custom;
OK
Time taken: 0.04 seconds
hive> CREATE TABLE olympix_data
> (
>   Aname STRING,
>   Age INT,
>   Country STRING,
>   Year INT,
>   Closing_Datre STRING,
>   Sport STRING,
>   Gold INT,
>   Silver INT,
>   Bronze INT,
>   Total_medals INT
> )
> ROW FORMAT DELIMITED
> FIELDS TERMINATED BY '\t';
OK
Time taken: 0.377 seconds
hive> LOAD DATA LOCAL INPATH '/home/acadgild/hadoop/olympix_data.csv' INTO TABLE custom.olympix_data;
Loading data to table custom.olympix_data
OK
Time taken: 0.794 seconds
hive> █
```

SELECT Country,COUNT(Total_medals) FROM olympix_data WHERE Sport =
"Swimming" GROUP BY Country;

```
hive>
>
>
> SELECT Country,COUNT(Total_medals) FROM olympix_data WHERE Sport = "Swimming" 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_20171109234043_dc9ba205-5a31-4a68-b06a-edea0782291f
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_1510243495509_0002, Tracking URL = http://localhost:8088/proxy/application_1510243495509_0002/
Kill Command = /home/acadgild/hadoop-2.7.2/bin/hadoop job -kill job_1510243495509_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2017-11-09 23:41:13,167 Stage-1 map = 0%, reduce = 0%
2017-11-09 23:41:36,833 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 6.27 sec
2017-11-09 23:41:55,706 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 9.74 sec
MapReduce Total cumulative CPU time: 9 seconds 740 msec
Ended Job = job_1510243495509_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 9.74 sec HDFS Read: 528925 HDFS Write: 878 SUCCESS
Total MapReduce CPU Time Spent: 9 seconds 740 msec
OK
```

```
Argentina      1
Australia      92
Austria 2
Belarus 1
Brazil 7
Canada 5
China 29
Costa Rica     1
Croatia 1
Denmark 1
France 26
Germany 27
Great Britain  9
Hungary 7
Italy 13
Japan 30
Lithuania      1
Netherlands    32
Norway 2
Poland 1
Romania 4
Russia 19
Serbia 1
Slovakia       1
Slovenia       1
South Africa   8
South Korea    2
Spain 2
Sweden 7
Trinidad and Tobago 1
Tunisia 2
Ukraine 4
United States  145
Zimbabwe       2
Time taken: 74.618 seconds, Fetched: 34 row(s)
hive>
```

Problem Statement 2 :-

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

Solution:-

```
SELECT Year,COUNT(Total_medals) FROM olympix_data WHERE  
Country="India" GROUP BY Year;
```

```
hive> SELECT Year,COUNT(Total_medals) FROM olympix_data 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_20171109234920_45e8e6db-7f70-4d2c-888a-0b4a85771adb  
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_1510243495509_0003, Tracking URL = http://localhost:8088/proxy/application_1510243495509_0003/  
Kill Command = /home/acadgild/hadoop-2.7.2/bin/hadoop job -kill job_1510243495509_0003  
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1  
2017-11-09 23:49:49,483 Stage-1 map = 0%, reduce = 0%  
2017-11-09 23:50:12,808 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 6.17 sec  
2017-11-09 23:50:33,577 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 9.67 sec  
MapReduce Total cumulative CPU time: 9 seconds 670 msec  
Ended Job = job_1510243495509_0003  
MapReduce Jobs Launched:  
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 9.67 sec HDFS Read: 528915 HDFS Write: 163 SUCCESS  
Total MapReduce CPU Time Spent: 9 seconds 670 msec  
OK  
2000 1  
2004 1  
2008 3  
2012 6  
Time taken: 74.262 seconds, Fetched: 4 row(s)  
hive> █
```

Problem Statement 3 :-

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

Solution:-

```
SELECT Country,COUNT(Total_medals) FROM olympix_data GROUP BY
Country;
```

```
hive> SELECT Country,COUNT(Total_medals) FROM olympix_data 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_20171109235533_9970cb36-6f6f-4d44-992b-7a18dac654e1
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_1510243495509_0004, Tracking URL = http://localhost:8088/proxy/application_1510243495509_0004/
Kill Command = /home/acadgild/hadoop-2.7.2/bin/hadoop job -kill job_1510243495509_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2017-11-09 23:56:01,715 Stage-1 map = 0%, reduce = 0%
2017-11-09 23:56:22,313 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.74 sec
2017-11-09 23:56:42,096 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 8.03 sec
MapReduce Total cumulative CPU time: 8 seconds 30 msec
Ended Job = job_1510243495509_0004
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 8.03 sec HDFS Read: 528105 HDFS Write: 2741 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 30 msec
OK
```

Afghanistan	2	
Algeria	8	
Argentina	141	
Armenia	10	
Australia	524	
Austria	70	
Azerbaijan	25	
Bahamas	23	
Bahrain	1	
Barbados	1	
Belarus	92	
Belgium	18	
Botswana	1	
Brazil	220	
Bulgaria	36	
Cameroon	20	
Canada	351	
Chile	20	
China	450	
Chinese Taipei	20	
Colombia	13	
Costa Rica	1	
Croatia	76	
Cuba	188	
Cyprus	1	
Czech Republic	75	
Denmark	89	
Dominican Republic		5
Ecuador	1	
Egypt	8	
Eritrea	1	
Estonia	16	
Ethiopia	24	
Finland	112	
France	287	
Gabon	1	
Georgia	23	
Germany	552	
Great Britain	296	
Greece	59	

Grenada	1
Guatemala	1
Hong Kong	3
Hungary	132
Iceland	15
India	11
Indonesia	22
Iran	24
Ireland	9
Israel	4
Italy	307
Jamaica	61
Japan	259
Kazakhstan	42
Kenya	38
Kuwait	2
Kyrgyzstan	3
Latvia	17
Lithuania	30
Macedonia	1
Malaysia	3
Mauritius	1
Mexico	38
Moldova	5
Mongolia	10
Montenegro	14
Morocco	10
Mozambique	1
Netherlands	286
New Zealand	51
Nigeria	39
North Korea	21
Norway	158
Panama	1
Paraguay	17
Poland	73
Portugal	9
Puerto Rico	2
Qatar	3
Romania	97
Russia	706

```
Saudi Arabia    6
Serbia    31
Serbia and Montenegro    38
Singapore      6
Slovakia      33
Slovenia      24
South Africa   22
South Korea   274
Spain    195
Sri Lanka     1
Sudan         1
Sweden    167
Switzerland   87
Syria        1
Tajikistan    3
Thailand      18
Togo         1
Trinidad and Tobago    16
Tunisia      3
Turkey       27
Uganda       1
Ukraine     137
United Arab Emirates    1
United States 1109
Uruguay      1
Uzbekistan   19
Venezuela    4
Vietnam      2
Zimbabwe     2
Time taken: 71.162 seconds, Fetched: 110 row(s)
hive> █
```


Problem Statement 4 :-

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

Solution:-

```
SELECT Country,COUNT(Gold) FROM olympix_data GROUP BY Country;
```

```
hive> SELECT Country,COUNT(Gold) FROM olympix_data 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_20171110000209_f6a55ba7-9d7e-4dc7-b5c1-5be994e8dc6
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_1510243495509_0005, Tracking URL = http://localhost:8088/proxy/application_1510243495509_0005/
Kill Command = /home/acadgild/hadoop-2.7.2/bin/hadoop job -kill job_1510243495509_0005
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2017-11-10 00:02:38,277 Stage-1 map = 0%, reduce = 0%
2017-11-10 00:02:58,770 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.69 sec
2017-11-10 00:03:18,140 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 8.01 sec
MapReduce Total cumulative CPU time: 8 seconds 10 msec
Ended Job = job_1510243495509_0005
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 8.01 sec HDFS Read: 528055 HDFS Write: 2741 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 10 msec
OK
```

Afghanistan	2	
Algeria	8	
Argentina	141	
Armenia	10	
Australia	524	
Austria	70	
Azerbaijan	25	
Bahamas	23	
Bahrain	1	
Barbados	1	
Belarus	92	
Belgium	18	
Botswana	1	
Brazil	220	
Bulgaria	36	
Cameroon	20	
Canada	351	
Chile	20	
China	450	
Chinese Taipei	20	
Colombia	13	
Costa Rica	1	
Croatia	76	
Cuba	188	
Cyprus	1	
Czech Republic	75	
Denmark	89	
Dominican Republic	5	
Ecuador	1	
Egypt	8	
Eritrea	1	
Estonia	16	
Ethiopia	24	
Finland	112	
France	287	
Gabon	1	
Georgia	23	
Germany	552	
Great Britain	296	
Greece	59	

Greece	59
Grenada	1
Guatemala	1
Hong Kong	3
Hungary	132
Iceland	15
India	11
Indonesia	22
Iran	24
Ireland	9
Israel	4
Italy	307
Jamaica	61
Japan	259
Kazakhstan	42
Kenya	38
Kuwait	2
Kyrgyzstan	3
Latvia	17
Lithuania	30
Macedonia	1
Malaysia	3
Mauritius	1
Mexico	38
Moldova	5
Mongolia	10
Montenegro	14
Morocco	10
Mozambique	1
Netherlands	286
New Zealand	51
Nigeria	39
North Korea	21
Norway	158
Panama	1
Paraguay	17
Poland	73
Portugal	9
Puerto Rico	2
Qatar	3
Romania	97

```
Romania 97
Russia 706
Saudi Arabia 6
Serbia 31
Serbia and Montenegro 38
Singapore 6
Slovakia 33
Slovenia 24
South Africa 22
South Korea 274
Spain 195
Sri Lanka 1
Sudan 1
Sweden 167
Switzerland 87
Syria 1
Tajikistan 3
Thailand 18
Togo 1
Trinidad and Tobago 16
Tunisia 3
Turkey 27
Uganda 1
Ukraine 137
United Arab Emirates 1
United States 1109
Uruguay 1
Uzbekistan 19
Venezuela 4
Vietnam 2
Zimbabwe 2
Time taken: 71.192 seconds, Fetched: 110 row(s)
hive> █
```

Submitted By:-

Hardik Kaushik