

Assignment 15.2

Table created by copying data from olympix_data.csv file in HDFS to a table in hive

```
CREATE EXTERNAL TABLE IF NOT EXISTS olympic_details(athlete_name STRING,age INT,country  
STRING, year int,closing_date STRING,sport STRING,gold INT,silver INT,bronze INT, total INT) COMMENT  
'Olympic Details' ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t'STORED AS TEXTFILE LOCATION  
'/user/acadgild/olympic_data';
```

Select * from olympic_details;

The screenshot shows a terminal window titled 'Acadgild_64bit' within a VMware Workstation environment. The terminal displays the execution of a Hive SQL query to create an external table named 'olympic_details' and a subsequent query to select all data from it. The output shows 8618 rows of data. Below the terminal output, a table of archery results is displayed, listing athletes, their countries, years, and medal counts.

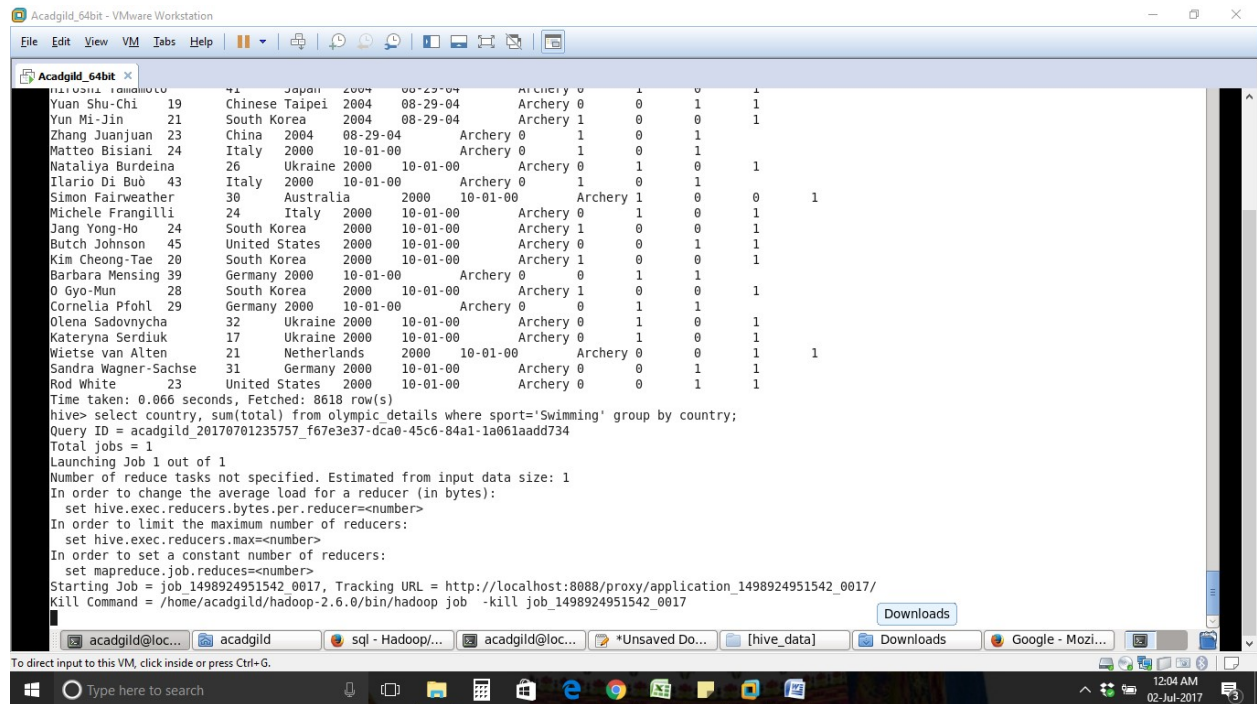
Rank	Age	Country	Year	Sport	Gold	Silver	Bronze	Total
1	20	Ukraine	2004	Archery	0	1	1	2
2	19	South Korea	2004	Archery	1	0	0	1
3	28	South Korea	2004	Archery	1	0	0	1
4	26	China	2004	Archery	0	1	0	1
5	19	Chinese Taipei	2004	Archery	0	1	0	1
6	28	South Korea	2004	Archery	1	0	0	1
7	23	Ukraine	2004	Archery	0	1	1	2
8	26	Ukraine	2004	Archery	0	1	1	2
9	17	Chinese Taipei	2004	Archery	0	1	0	1
10	32	Great Britain	2004	Archery	0	1	0	1
11	21	Chinese Taipei	2004	Archery	0	1	0	1
12	41	Japan	2004	Archery	0	1	0	1
13	19	Chinese Taipei	2004	Archery	0	1	0	1
14	21	South Korea	2004	Archery	1	0	0	1
15	23	China	2004	Archery	0	1	0	1
16	24	Italy	2000	Archery	0	1	0	1
17	26	Ukraine	2000	Archery	0	1	0	1
18	43	Italy	2000	Archery	0	1	0	1
19	30	Australia	2000	Archery	1	0	0	1
20	24	Italy	2000	Archery	0	1	0	1
21	24	South Korea	2000	Archery	1	0	0	1
22	45	United States	2000	Archery	0	1	0	1
23	20	South Korea	2000	Archery	1	0	0	1
24	39	Germany	2000	Archery	0	1	1	2
25	28	South Korea	2000	Archery	1	0	0	1
26	29	Germany	2000	Archery	0	1	1	2
27	32	Ukraine	2000	Archery	0	1	0	1
28	17	Ukraine	2000	Archery	0	1	0	1
29	21	Netherlands	2000	Archery	0	1	1	2
30	31	Germany	2000	Archery	0	1	1	2
31	23	United States	2000	Archery	0	1	1	2

The screenshot shows a terminal window titled 'Acadgild_64bit' within a VMware Workstation environment. The terminal displays the execution of a Hive SQL query to create an external table named 'olympic_details' and a subsequent query to select all data from it. The output shows 8618 rows of data. Below the terminal output, a table of archery results is displayed, listing athletes, their countries, years, and medal counts.

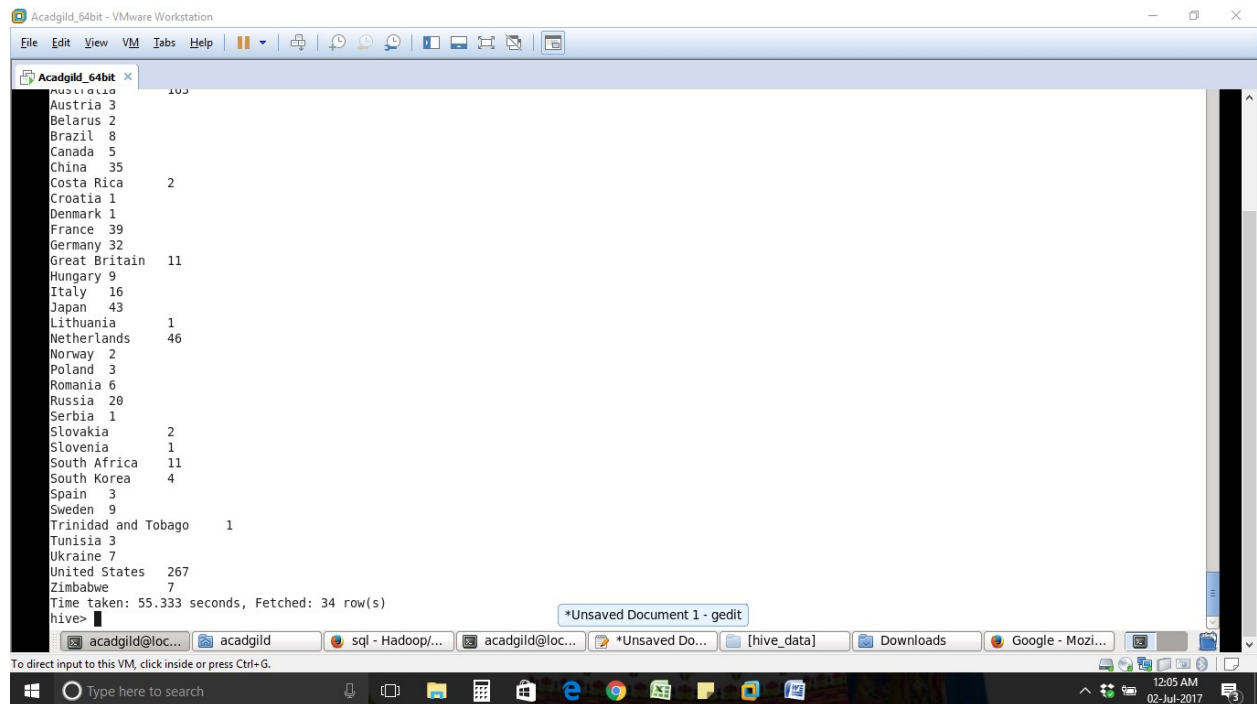
Rank	Age	Country	Year	Sport	Gold	Silver	Bronze	Total
1	20	Ukraine	2004	Archery	0	1	1	2
2	19	South Korea	2004	Archery	1	0	0	1
3	28	South Korea	2004	Archery	1	0	0	1
4	26	China	2004	Archery	0	1	0	1
5	19	Chinese Taipei	2004	Archery	0	1	0	1
6	28	South Korea	2004	Archery	1	0	0	1
7	23	Ukraine	2004	Archery	0	1	1	2
8	26	Ukraine	2004	Archery	0	1	1	2
9	17	Chinese Taipei	2004	Archery	0	1	0	1
10	32	Great Britain	2004	Archery	0	1	0	1
11	21	Chinese Taipei	2004	Archery	0	1	0	1
12	41	Japan	2004	Archery	0	1	0	1
13	19	Chinese Taipei	2004	Archery	0	1	0	1
14	21	South Korea	2004	Archery	1	0	0	1
15	23	China	2004	Archery	0	1	0	1
16	24	Italy	2000	Archery	0	1	0	1
17	26	Ukraine	2000	Archery	0	1	0	1
18	43	Italy	2000	Archery	0	1	0	1
19	30	Australia	2000	Archery	1	0	0	1
20	24	Italy	2000	Archery	0	1	0	1
21	24	South Korea	2000	Archery	1	0	0	1
22	45	United States	2000	Archery	0	1	0	1
23	20	South Korea	2000	Archery	1	0	0	1
24	39	Germany	2000	Archery	0	1	1	2
25	28	South Korea	2000	Archery	1	0	0	1
26	29	Germany	2000	Archery	0	1	1	2
27	32	Ukraine	2000	Archery	0	1	0	1
28	17	Ukraine	2000	Archery	0	1	0	1
29	21	Netherlands	2000	Archery	0	1	1	2
30	31	Germany	2000	Archery	0	1	1	2
31	23	United States	2000	Archery	0	1	1	2

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

Select country ,sum(total) from olympic_details where sport='Swimming' group by country



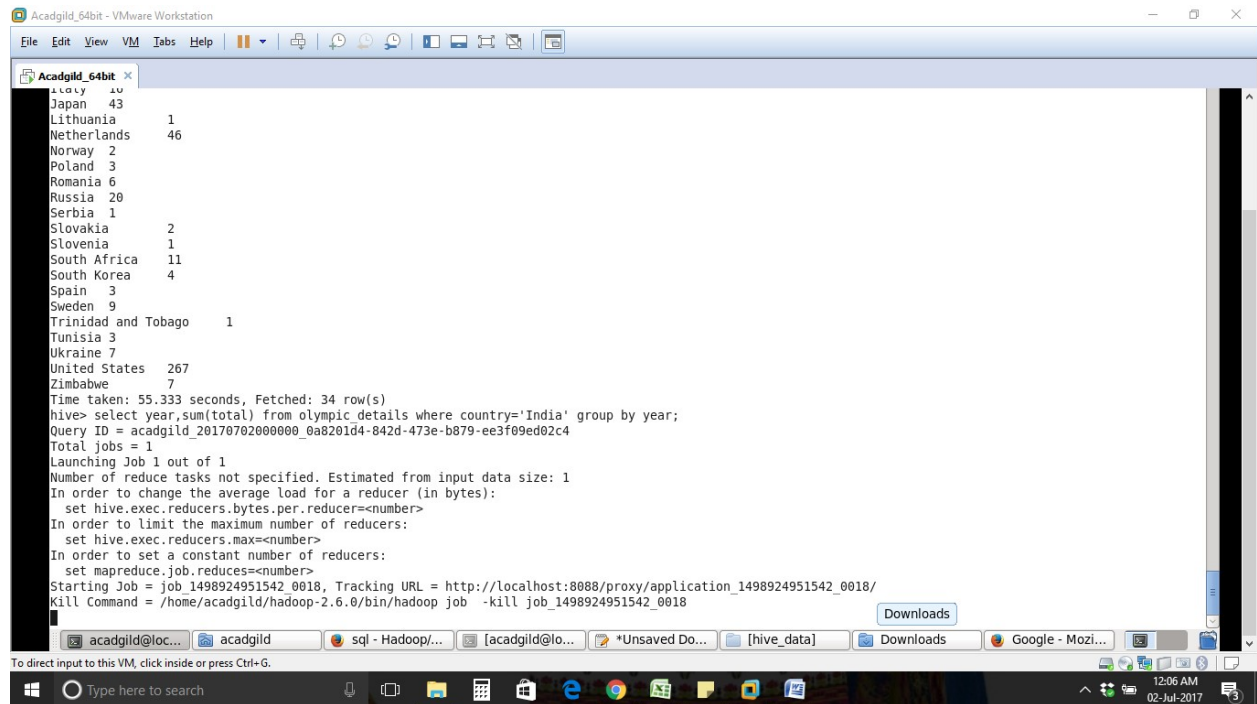
```
Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help
Acadgild_64bit x
Yuan Shu-Chi 19 Chinese Taipei 2004 08-29-04 Archery 0 1 0 1
Yun Mi-Jin 21 South Korea 2004 08-29-04 Archery 1 0 0 1
Zhang Juanjuan 23 China 2004 08-29-04 Archery 0 1 0 1
Matteo Bisiani 24 Italy 2000 10-01-00 Archery 0 1 0 1
Nataliya Burdeina 26 Ukraine 2000 10-01-00 Archery 0 1 0 1
Ilario Di Buò 43 Italy 2000 10-01-00 Archery 0 1 0 1
Simon Fairweather 30 Australia 2000 10-01-00 Archery 1 0 0 1
Michele Frangilli 24 Italy 2000 10-01-00 Archery 0 1 0 1
Jang Yong-Ho 24 South Korea 2000 10-01-00 Archery 1 0 0 1
Butch Johnson 45 United States 2000 10-01-00 Archery 0 0 1 1
Kim Cheong-Tae 20 South Korea 2000 10-01-00 Archery 1 0 0 1
Barbara Mensing 39 Germany 2000 10-01-00 Archery 0 0 1 1
O Gyo-Mun 28 South Korea 2000 10-01-00 Archery 1 0 0 1
Cornelia Pfuhl 29 Germany 2000 10-01-00 Archery 0 0 1 1
Olena Sadovnycha 32 Ukraine 2000 10-01-00 Archery 0 1 0 1
Kateryna Serdiuk 17 Ukraine 2000 10-01-00 Archery 0 1 0 1
Wietse van Alten 21 Netherlands 2000 10-01-00 Archery 0 0 1 1
Sandra Wagner-Sachse 31 Germany 2000 10-01-00 Archery 0 0 1 1
Rod White 23 United States 2000 10-01-00 Archery 0 0 1 1
Time taken: 0.066 seconds, Fetched: 8618 row(s)
hive> select country, sum(total) from olympic_details where sport='Swimming' group by country;
Query ID = acadgild_20170701235757_f67e3e37-dca0-45c6-84a1-1a061aadd734
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_1498924951542_0017, Tracking URL = http://localhost:8088/proxy/application_1498924951542_0017/
Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job -kill job_1498924951542_0017
Downloads
acadgild@loc... acadgild sql - Hadoop... acadgild@loc... *Unsaved Do... [hive_data] Downloads Google - Mozi...
To direct input to this VM, click inside or press Ctrl+G.
Type here to search 12:04 AM 02-Jul-2017
```



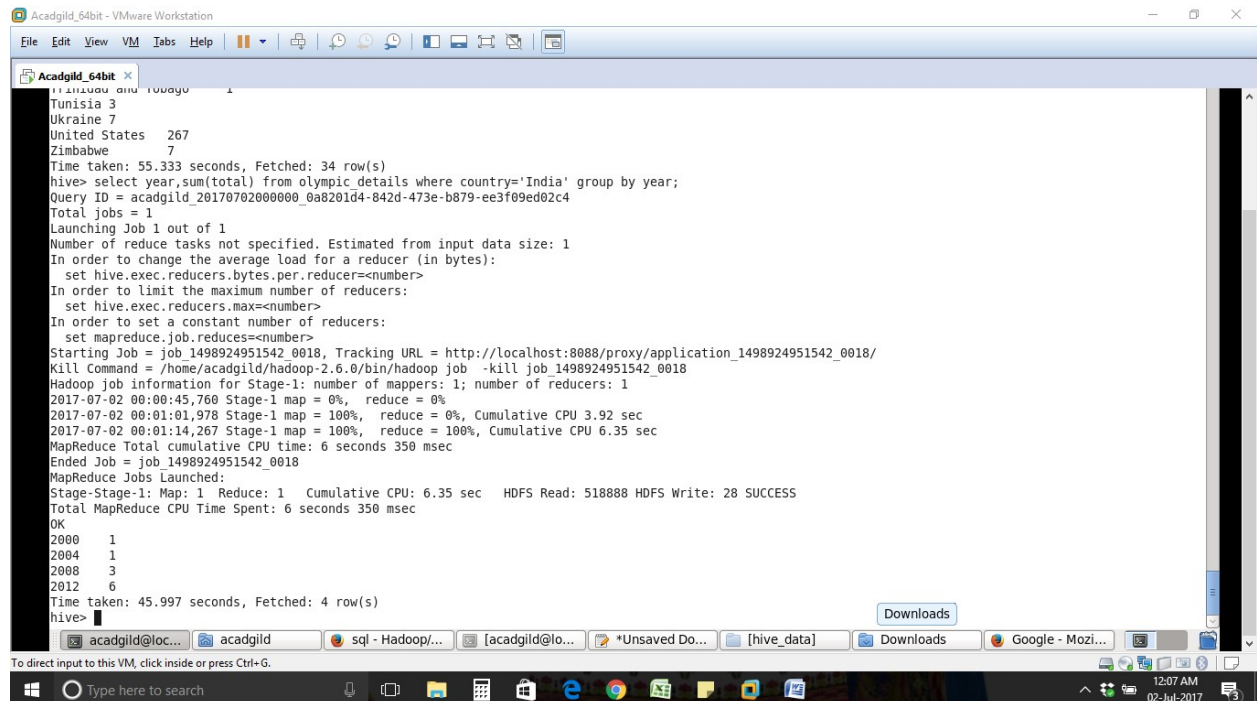
```
Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help
Acadgild_64bit x
Austria 3
Belarus 2
Brazil 8
Canada 5
China 35
Costa Rica 2
Croatia 1
Denmark 1
France 39
Germany 32
Great Britain 11
Hungary 9
Italy 16
Japan 43
Lithuania 1
Netherlands 46
Norway 2
Poland 3
Romania 6
Russia 20
Serbia 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: 55.333 seconds, Fetched: 34 row(s)
hive>
*Unsaved Document 1 - gedit
acadgild@loc... acadgild sql - Hadoop... acadgild@loc... *Unsaved Do... [hive_data] Downloads Google - Mozi...
To direct input to this VM, click inside or press Ctrl+G.
Type here to search 12:05 AM 02-Jul-2017
```

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

Select year,sum(total) from olympic_details where country='India' group by year;



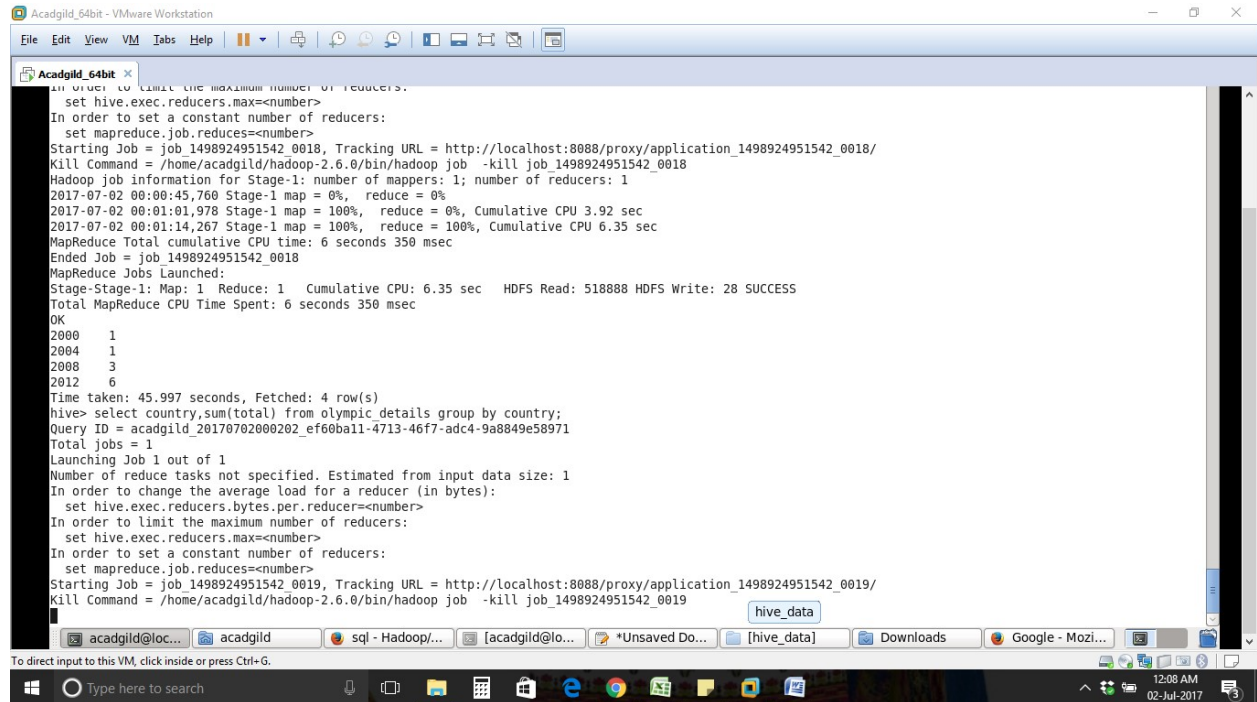
```
Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help
Acadgild_64bit
Japan 43
Lithuania 1
Netherlands 46
Norway 2
Poland 3
Romania 6
Russia 20
Serbia 1
Slovakia 2
Slovenia 1
South Africa 11
South Korea 4
Spain 3
Sweden 9
Trinidad and Tobago 1
Tunisia 3
Ukraine 7
United States 267
Zimbabwe 7
Time taken: 55.333 seconds, Fetched: 34 row(s)
hive> select year,sum(total) from olympic_details where country='India' group by year;
Query ID = acadgild_20170702000000_0a8201d4-842d-473e-b879-ee3f09ed02c4
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_1498924951542_0018, Tracking URL = http://localhost:8088/proxy/application_1498924951542_0018/
Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job -kill job_1498924951542_0018
Downloads
acadgild@loc... acadgild sql - Hadoop... [acadgild@lo... *Unsaved Do... [hive_data] Downloads Google - Mozi...
To direct input to this VM, click inside or press Ctrl+G.
Type here to search 12:06 AM 02-Jul-2017
```



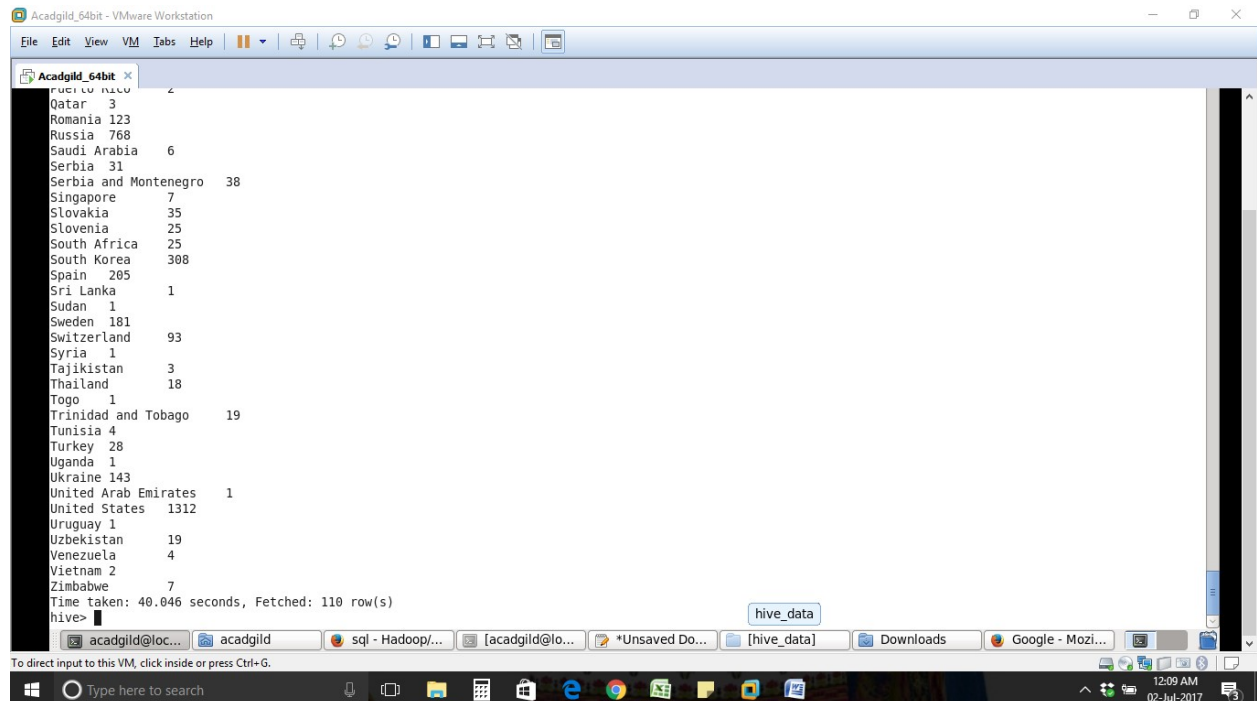
```
Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help
Acadgild_64bit
Tunisia 3
Ukraine 7
United States 267
Zimbabwe 7
Time taken: 55.333 seconds, Fetched: 34 row(s)
hive> select year,sum(total) from olympic_details where country='India' group by year;
Query ID = acadgild_20170702000000_0a8201d4-842d-473e-b879-ee3f09ed02c4
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_1498924951542_0018, Tracking URL = http://localhost:8088/proxy/application_1498924951542_0018/
Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job -kill job_1498924951542_0018
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2017-07-02 00:00:45,760 Stage-1 map = 0%, reduce = 0%
2017-07-02 00:01:01,978 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.92 sec
2017-07-02 00:01:14,267 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.35 sec
MapReduce Total cumulative CPU time: 6 seconds 350 msec
Ended Job = job_1498924951542_0018
MapReduce Jobs Launched:
Stage-Stage1: Map: 1 Reduce: 1 Cumulative CPU: 6.35 sec HDFS Read: 518888 HDFS Write: 28 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 350 msec
OK
2000 1
2004 1
2008 3
2012 6
Time taken: 45.997 seconds, Fetched: 4 row(s)
hive>
Downloads
acadgild@loc... acadgild sql - Hadoop... [acadgild@lo... *Unsaved Do... [hive_data] Downloads Google - Mozi...
To direct input to this VM, click inside or press Ctrl+G.
Type here to search 12:07 AM 02-Jul-2017
```

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

Select country,sum(total) from olympic_data group by country;



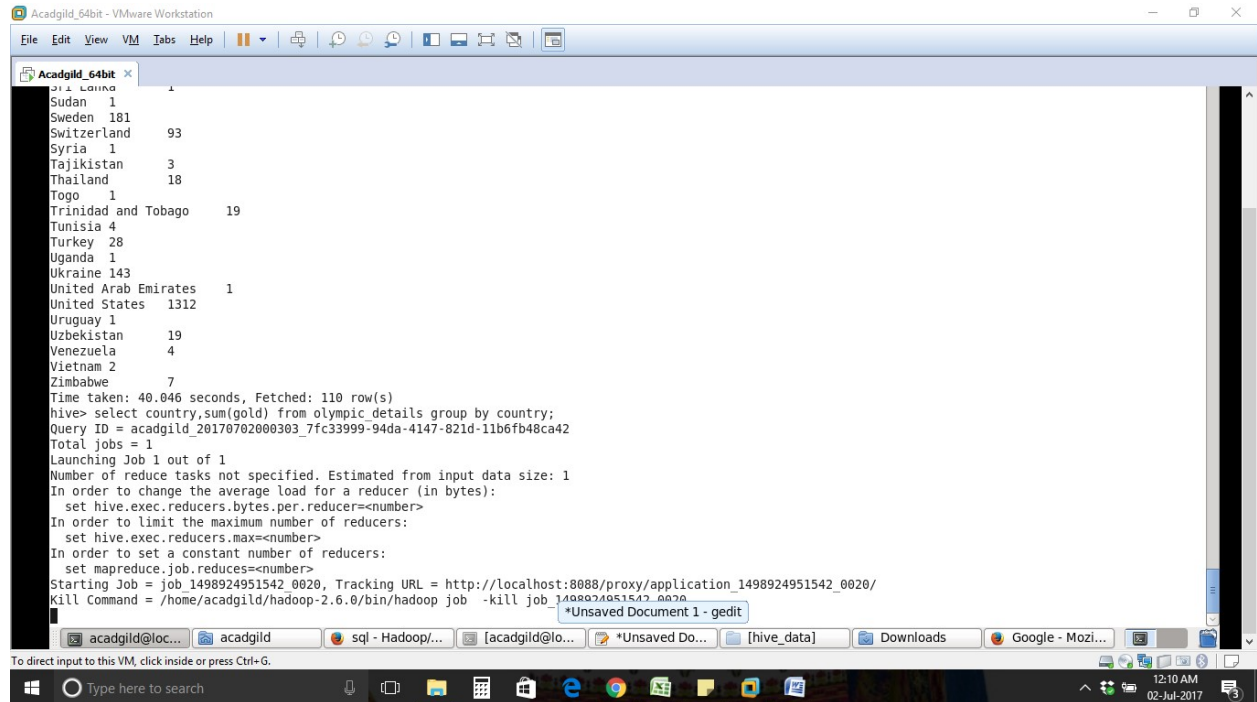
```
Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help
Acadgild_64bit
All other commands to limit the maximum number of reducers.
set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
set mapreduce.job.reducers=<number>
Starting Job = job_1498924951542_0018, Tracking URL = http://localhost:8088/proxy/application_1498924951542_0018/
Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job -kill job_1498924951542_0018
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2017-07-02 00:00:45,760 Stage-1 map = 0%, reduce = 0%
2017-07-02 00:01:01,978 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.92 sec
2017-07-02 00:01:14,267 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.35 sec
MapReduce Total cumulative CPU time: 6 seconds 350 msec
Ended Job = job_1498924951542_0018
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 6.35 sec HDFS Read: 518888 HDFS Write: 28 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 350 msec
OK
2000 1
2004 1
2008 3
2012 6
Time taken: 45.997 seconds, Fetched: 4 row(s)
hive> select country,sum(total) from olympic_details group by country;
Query ID = acadgild_20170702000202_ef60ba11-4713-46f7-adc4-9a8849e58971
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_1498924951542_0019, Tracking URL = http://localhost:8088/proxy/application_1498924951542_0019/
Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job -kill job_1498924951542_0019
hive_data
acadgild@loc... acadgild sql - Hadoop... [acadgild@lo... *Unsaved Do... [hive_data] Downloads Google - Mozi...
To direct input to this VM, click inside or press Ctrl+G.
Type here to search 12:08 AM 02-Jul-2017
```



```
Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help
Acadgild_64bit
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
Ukraine 143
United Arab Emirates 1
United States 1312
Uruguay 1
Uzbekistan 19
Venezuela 4
Vietnam 2
Zimbabwe 7
Time taken: 40.046 seconds, Fetched: 110 row(s)
hive>
hive_data
acadgild@loc... acadgild sql - Hadoop... [acadgild@lo... *Unsaved Do... [hive_data] Downloads Google - Mozi...
To direct input to this VM, click inside or press Ctrl+G.
Type here to search 12:09 AM 02-Jul-2017
```

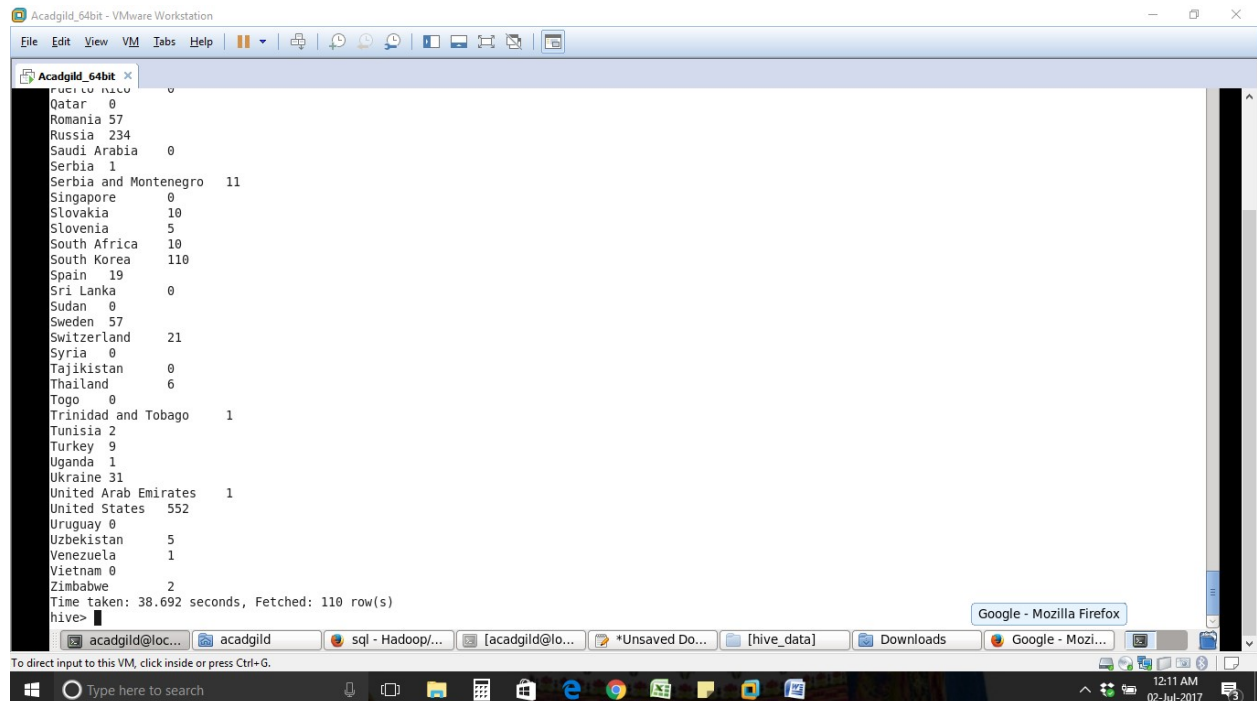
Write a Hive program to find the number of gold medals each country won.

Select country,sum(gold) from olympic_data group by country;



The screenshot shows a terminal window titled 'Acadgild_64bit - VMware Workstation'. The terminal displays the output of a Hive query. The query is: `hive> select country,sum(gold) from olympic_details group by country;`. The output lists 28 countries with their respective gold medal counts. The total number of jobs is 1, and the time taken is 40.046 seconds. The terminal also shows the Hive configuration and the job's tracking URL.

```
Acadgild_64bit
Sudan 1
Sweden 101
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: 40.046 seconds, Fetched: 110 row(s)
hive> select country,sum(gold) from olympic_details group by country;
Query ID = acadgild_20170702000303_7fc33999-94da-4147-821d-11b6fb48ca42
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_1498924951542_0020, Tracking URL = http://localhost:8088/proxy/application_1498924951542_0020/
Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job -kill job_1498924951542_0020
*Unsaved Document 1 - gedit
acadgild@loc... acadgild sql - Hadoop/... [acadgild@lo... *Unsaved Do... [hive_data] Downloads Google - Mozi...
```



The screenshot shows a terminal window titled 'Acadgild_64bit - VMware Workstation'. The terminal displays the output of a Hive query. The query is: `hive> select country,sum(gold) from olympic_data group by country;`. The output lists 28 countries with their respective gold medal counts. The total number of jobs is 1, and the time taken is 38.692 seconds. The terminal also shows the Hive configuration and the job's tracking URL.

```
Acadgild_64bit
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
Ukraine 31
United Arab Emirates 1
United States 552
Uruguay 0
Uzbekistan 5
Venezuela 1
Vietnam 0
Zimbabwe 2
Time taken: 38.692 seconds, Fetched: 110 row(s)
hive>
Google - Mozilla Firefox
acadgild@loc... acadgild sql - Hadoop/... [acadgild@lo... *Unsaved Do... [hive_data] Downloads Google - Mozi...
```