Session 7

Assignment 2

|  |  |
| --- | --- |
| **Prepared For:** | AcadGild |
|  |  |
| **Document Approval:** | **AcadGild** |
|  |  |
|  |  |
|  |  |
|  |  |
| **Project Title:** | Session 7 – Assignment 2 |
|  |  |
| **Prepared By:** | Duncan Burgess |
|  |  |
|  | dburgess@duncb.com |
|  |  |
| **Primary Engineer:** | Duncan Burgess |
|  |  |
| **Document Reference:** | **Session 7 – Assignment 2** |
|  |  |
| **Start Date:** | 21/09/2017 |
|  |  |
|  |  |

# 

# Contents

[Contents 2](#_Toc493768005)

[Change History 3](#_Toc493768006)

[1. Problem Statement 4](#_Toc493768007)

[2. Create Table and Load Data 4](#_Toc493768008)

[3. Solutions 5](#_Toc493768009)

[3.1. medals won by each country in swimming 5](#_Toc493768010)

[3.2. Number of medals that India won year wise 6](#_Toc493768011)

[3.3. Total number of medals each country won 6](#_Toc493768012)

[3.4. Number of gold medals each country won 8](#_Toc493768013)

[3.4.1. Enhanced to show only countries that have won a medal 10](#_Toc493768014)

[3.4.2. Another enhancement show in order of most gold medals won. 12](#_Toc493768015)

# Change History

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Document Revision** | **Date** | **Authored By** | **Authorised By** | **Sections Affected** | **Reason for Change** |
| Rev 01 | 21/09/2017 | Duncan Burgess |  | All | Initial release. |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

# Problem Statement

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

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

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

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

**Dataset set description**

The data set consists of the following fields.

* Athlete: This field consists of the athlete name
* Age: This field consists of athlete ages
* Country: This fields consists of the country names which participated in Olympics
* Year: This field consists of the year
* Closing Date: This field consists of the closing date of ceremony
* Sport: Consists of the sports name
* Gold Medals: No. of Gold medals
* Silver Medals: No. of Silver medals
* Bronze Medals: No. of Bronze medals
* Total Medals: Consists of total no. of medals

**Dataset (sample)**

Michael Phelps 23 United States 2008 08-24-08 Swimming 8 0 0 8

Michael Phelps 19 United States 2004 08-29-04 Swimming 6 0 2 8

Michael Phelps 27 United States 2012 08-12-12 Swimming 4 2 0 6

Natalie Coughlin 25 United States 2008 08-24-08 Swimming 1 2 3 6

Aleksey Nemov 24 Russia 2000 10-01-00 Gymnastics 2 1 3 6

Alicia Coutts 24 Australia 2012 08-12-12 Swimming 1 3 1 5

Missy Franklin 17 United States 2012 08-12-12 Swimming 4 0 1 5

Ryan Lochte 27 United States 2012 08-12-12 Swimming 2 2 1 5

# Create Table and Load Data

CREATE TABLE olympic\_data(athelete string,age int,country string,year int,closing\_date date,

sport string,gold int,silver int,bronze int,tot\_medal int)

row format delimited fields

terminated by '\t';

hive> load data local inpath '/home/acadgild/Datasets/olympix\_data.csv' into table olympic\_data;

Test data is loaded.

hive (Demo)> select \* from olympic\_data limit 5;

OK

Michael Phelps 23 United States 2008 NULL Swimming 8 0 0 8

Michael Phelps 19 United States 2004 NULL Swimming 6 0 2 8

Michael Phelps 27 United States 2012 NULL Swimming 4 2 0 6

Natalie Coughlin 25 United States 2008 NULL Swimming 1 2 3 6

Aleksey Nemov 24 Russia 2000 NULL Gymnastics 2 1 3 6

Time taken: 0.037 seconds, Fetched: 5 row(s)

To display the database working in I set

***Set hive.cli.print.current.db = true***

# Solutions

## medals won by each country in swimming

**Query run**

*select country ,sum(tot\_medal) from olympic\_data where sport = 'Swimming' group by country;*

**Results**

*OK*

*Argentina 1*

*Australia 163*

*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: 20.693 seconds, Fetched: 34 row(s)*

## Number of medals that India won year wise

**Query run**

*select year ,sum(tot\_medal) from olympic\_data where country = 'India' group by year;*

**Result**

*OK*

*2000 1*

*2004 1*

*2008 3*

*2012 6*

*Time taken: 23.08 seconds, Fetched: 4 row(s)*

## Total number of medals each country won

**Query run**

*select country ,sum(tot\_medal) from olympic\_data group by country;*

*OK*

*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*

*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*

*Norway 192*

*Panama 1*

*Paraguay 17*

*Poland 80*

*Portugal 9*

*Puerto Rico 2*

*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: 21.678 seconds, Fetched: 110 row(s)*

## Number of gold medals each country won

**Query run**

select country ,sum(gold) from olympic\_data where group by country;

**Results**

*OK*

*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*

*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*

*Netherlands 101*

*New Zealand 18*

*Nigeria 6*

*North Korea 6*

*Norway 97*

*Panama 1*

*Paraguay 0*

*Poland 20*

*Portugal 1*

*Puerto Rico 0*

*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: 21.512 seconds, Fetched: 110 row(s)*

### Enhanced to show only countries that have won a medal

**Query run**

select country ,sum(gold) from olympic\_data where gold >0 group by country;

*OK*

*Algeria 2*

*Argentina 49*

*Australia 163*

*Austria 36*

*Azerbaijan 6*

*Bahamas 11*

*Belarus 17*

*Belgium 2*

*Brazil 46*

*Bulgaria 8*

*Cameroon 20*

*Canada 168*

*Chile 3*

*China 234*

*Chinese Taipei 2*

*Colombia 2*

*Croatia 35*

*Cuba 57*

*Czech Republic 14*

*Denmark 46*

*Dominican Republic 3*

*Egypt 1*

*Estonia 6*

*Ethiopia 13*

*Finland 11*

*France 108*

*Georgia 6*

*Germany 223*

*Great Britain 124*

*Greece 12*

*Grenada 1*

*Hungary 77*

*India 1*

*Indonesia 5*

*Iran 10*

*Ireland 1*

*Israel 1*

*Italy 86*

*Jamaica 24*

*Japan 57*

*Kazakhstan 13*

*Kenya 11*

*Latvia 3*

*Lithuania 5*

*Mexico 19*

*Mongolia 2*

*Morocco 2*

*Mozambique 1*

*Netherlands 101*

*New Zealand 18*

*Nigeria 6*

*North Korea 6*

*Norway 97*

*Panama 1*

*Poland 20*

*Portugal 1*

*Romania 57*

*Russia 234*

*Serbia 1*

*Serbia and Montenegro 11*

*Slovakia 10*

*Slovenia 5*

*South Africa 10*

*South Korea 110*

*Spain 19*

*Sweden 57*

*Switzerland 21*

*Thailand 6*

*Trinidad and Tobago 1*

*Tunisia 2*

*Turkey 9*

*Uganda 1*

*Ukraine 31*

*United Arab Emirates 1*

*United States 552*

*Uzbekistan 5*

*Venezuela 1*

*Zimbabwe 2*

*Time taken: 20.293 seconds, Fetched: 78 row(s)*

### Another enhancement show in order of most gold medals won.

**Query run**

*select country ,sum(gold) as cnt from olympic\_data where gold >0 group by country order by cnt desc;*

*OK*

*United States 552*

*Russia 234*

*China 234*

*Germany 223*

*Canada 168*

*Australia 163*

*Great Britain 124*

*South Korea 110*

*France 108*

*Netherlands 101*

*Norway 97*

*Italy 86*

*Hungary 77*

*Romania 57*

*Sweden 57*

*Japan 57*

*Cuba 57*

*Argentina 49*

*Denmark 46*

*Brazil 46*

*Austria 36*

*Croatia 35*

*Ukraine 31*

*Jamaica 24*

*Switzerland 21*

*Cameroon 20*

*Poland 20*

*Spain 19*

*Mexico 19*

*New Zealand 18*

*Belarus 17*

*Czech Republic 14*

*Ethiopia 13*

*Kazakhstan 13*

*Greece 12*

*Bahamas 11*

*Finland 11*

*Serbia and Montenegro 11*

*Kenya 11*

*Iran 10*

*Slovakia 10*

*South Africa 10*

*Turkey 9*

*Bulgaria 8*

*Estonia 6*

*Georgia 6*

*North Korea 6*

*Thailand 6*

*Nigeria 6*

*Azerbaijan 6*

*Indonesia 5*

*Slovenia 5*

*Lithuania 5*

*Uzbekistan 5*

*Dominican Republic 3*

*Chile 3*

*Latvia 3*

*Zimbabwe 2*

*Tunisia 2*

*Morocco 2*

*Mongolia 2*

*Colombia 2*

*Chinese Taipei 2*

*Belgium 2*

*Algeria 2*

*India 1*

*Ireland 1*

*Israel 1*

*Mozambique 1*

*Panama 1*

*Venezuela 1*

*Portugal 1*

*Serbia 1*

*Trinidad and Tobago 1*

*Egypt 1*

*Uganda 1*

*United Arab Emirates 1*

*Grenada 1*

*Time taken: 45.413 seconds, Fetched: 78 row(s)*