**Problem Objective**

This Data set is about Olympics. You can download the data set from the below link:

https://drive.google.com/open?id=0ByJLBTmJojjzV1czX3Nha0R3bTQ

**DATE 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

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.

Code/Script

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

-- create a database custom

create database custom;

-- entering to database custom

use custom;

-- create table OLYMPIC

create table olympic

(athlete string,age int,country string,year int,closingdate date,sport string,goldmedals int,silvermedals int,bronzemedals int, totalmedals int)

row format serde 'org.apache.hadoop.hive.serde2.OpenCSVSerde'

with serdeproperties (

"separatorChar"="\t","quoteChar" = "'","escapeChar" ="\\"

) stored as textfile;

-- load data into OLYMPIC

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

-- count medals from olympic where sport = swimming group by country

select country,sum(goldmedals+silvermedals+bronzemedals) from olympic where sport = 'Swimming' group by country;

Output

[acadgild@localhost ~]$ hive

Logging initialized using configuration in jar:file:/usr/local/hive/lib/hive-common-0.14.0.jar!/hive-log4j.properties

SLF4J: Class path contains multiple SLF4J bindings.

SLF4J: Found binding in [jar:file:/usr/local/hive/lib/hive-jdbc-0.14.0-standalone.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: Found binding in [jar:file:/usr/local/hadoop-2.6.0/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: See http://www.slf4j.org/codes.html#multiple\_bindings for an explanation.

SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]

hive> show databases;

OK

b1

custom

default

test

Time taken: 6.151 seconds, Fetched: 4 row(s)

hive> use custom;

OK

Time taken: 0.342 seconds

hive> show tables;

OK

temp

temperature\_data

temperature\_data\_vw

Time taken: 0.294 seconds, Fetched: 3 row(s)

hive> create table olympic

> (athlete string,age int,country string,year int,closingdate date,sport string,goldmedals int,silvermedals int,bronzemedals int, totalmedals int)

> row format serde 'org.apache.hadoop.hive.serde2.OpenCSVSerde'

> with serdeproperties (

> "separatorChar"="\t","quoteChar" = "'","escapeChar" ="\\"

> ) stored as textfile;

OK

Time taken: 3.141 seconds

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

Loading data to table custom.olympic

Table custom.olympic stats: [numFiles=1, totalSize=518669]

OK

Time taken: 10.792 seconds

hive> select country,sum(goldmedals+silvermedals+bronzemedals) from olympic where sport = 'Swimming' group by country;

Query ID = acadgild\_20170401221414\_bb677680-dd75-4f54-807a-004b15ab5e81

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\_1491017400247\_0010, Tracking URL = http://localhost:8088/proxy/application\_1491017400247\_0010/

Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job -kill job\_1491017400247\_0010

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1

2017-04-01 22:17:20,263 Stage-1 map = 0%, reduce = 0%

2017-04-01 22:18:20,965 Stage-1 map = 0%, reduce = 0%

2017-04-01 22:19:22,141 Stage-1 map = 0%, reduce = 0%

2017-04-01 22:20:23,472 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 16.6 sec

2017-04-01 22:21:24,744 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 16.6 sec

2017-04-01 22:22:25,487 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 16.6 sec

2017-04-01 22:23:26,715 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 19.47 sec

2017-04-01 22:23:31,361 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 20.57 sec

2017-04-01 22:24:11,200 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 29.66 sec

MapReduce Total cumulative CPU time: 29 seconds 660 msec

Ended Job = job\_1491017400247\_0010

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 32.24 sec HDFS Read: 518899 HDFS Write: 454 SUCCESS

Total MapReduce CPU Time Spent: 32 seconds 240 msec

OK

Argentina 1.0

Australia 158.0

Austria 3.0

Belarus 2.0

Brazil 8.0

Canada 5.0

China 35.0

Costa Rica 2.0

Croatia 1.0

Denmark 1.0

France 39.0

Germany 32.0

Great Britain 11.0

Hungary 9.0

Italy 16.0

Japan 43.0

Lithuania 1.0

Netherlands 46.0

Norway 2.0

Poland 3.0

Romania 6.0

Russia 20.0

Serbia 1.0

Slovakia 2.0

Slovenia 1.0

South Africa 11.0

South Korea 4.0

Spain 3.0

Sweden 9.0

Trinidad and Tobago 1.0

Tunisia 3.0

Ukraine 7.0

United States 267.0

Zimbabwe 7.0

Time taken: 565.501 seconds, Fetched: 34 row(s)

Code/Script

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

-- create a database custom

create database custom;

-- entering to database custom

use custom;

-- create table OLYMPIC

create table olympic

(athlete string,age int,country string,year int,closingdate date,sport string,goldmedals int,silvermedals int,bronzemedals int, totalmedals int)

row format serde 'org.apache.hadoop.hive.serde2.OpenCSVSerde'

with serdeproperties (

"separatorChar"="\t","quoteChar" = "'","escapeChar" ="\\"

) stored as textfile;

create table olympic

(athlete string,age int,country string,year int,closingdate date,sport string,goldmedals int,silvermedals int,bronzemedals int, totalmedals int)

row format serde 'org.apache.hadoop.hive.serde2.OpenCSVSerde'

stored as textfile;

-- load data into OLYMPIC

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

-- no of medals won by india by year

select country,year,sum(goldmedals+silvermedals+bronzemedals) from olympic where country = 'India' group by year,country;

Output

hive> select country,year,sum(goldmedals+silvermedals+bronzemedals) from olympic where country = 'India' group by year,country;

Query ID = acadgild\_20170401223131\_5bb4b2c7-228f-45d3-9885-832e4a9b1223

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\_1491017400247\_0011, Tracking URL = http://localhost:8088/proxy/application\_1491017400247\_0011/

Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job -kill job\_1491017400247\_0011

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1

2017-04-01 22:35:08,458 Stage-1 map = 0%, reduce = 0%

2017-04-01 22:36:09,214 Stage-1 map = 0%, reduce = 0%

2017-04-01 22:37:10,147 Stage-1 map = 0%, reduce = 0%

2017-04-01 22:38:20,477 Stage-1 map = 0%, reduce = 0%

2017-04-01 22:39:23,069 Stage-1 map = 0%, reduce = 0%, Cumulative CPU 7.38 sec

2017-04-01 22:40:20,125 Stage-1 map = 67%, reduce = 0%, Cumulative CPU 21.04 sec

2017-04-01 22:40:23,084 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 22.01 sec

2017-04-01 22:41:26,791 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 22.01 sec

2017-04-01 22:42:27,600 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 22.01 sec

2017-04-01 22:43:31,133 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 22.01 sec

2017-04-01 22:44:31,164 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 22.01 sec

2017-04-01 22:45:31,459 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 22.01 sec

2017-04-01 22:46:31,680 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 22.01 sec

2017-04-01 22:47:32,377 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 22.01 sec

2017-04-01 22:47:33,818 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 25.2 sec

2017-04-01 22:47:57,880 Stage-1 map = 100%, reduce = 83%, Cumulative CPU 34.3 sec

2017-04-01 22:48:04,610 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 37.48 sec

MapReduce Total cumulative CPU time: 37 seconds 480 msec

Ended Job = job\_1491017400247\_0011

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 38.23 sec HDFS Read: 518899 HDFS Write: 60 SUCCESS

Total MapReduce CPU Time Spent: 38 seconds 230 msec

OK

India 2000 1.0

India 2004 1.0

India 2008 3.0

India 2012 6.0

Time taken: 988.491 seconds, Fetched: 4 row(s)

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

-- create a database custom

create database custom;

-- entering to database custom

use custom;

-- create table OLYMPIC

create table olympic

(athlete string,age int,country string,year int,closingdate date,sport string,goldmedals int,silvermedals int,bronzemedals int, totalmedals int)

row format serde 'org.apache.hadoop.hive.serde2.OpenCSVSerde'

with serdeproperties (

"separatorChar"="\t","quoteChar" = "'","escapeChar" ="\\"

) stored as textfile;

create table olympic

(athlete string,age int,country string,year int,closingdate date,sport string,goldmedals int,silvermedals int,bronzemedals int, totalmedals int)

row format serde 'org.apache.hadoop.hive.serde2.OpenCSVSerde'

stored as textfile;

-- load data into OLYMPIC

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

-- no of medals won by each country

select country,sum(totalmedals) from olympic group by country;

Output

hive> select country,sum(totalmedals) from olympic group by country;

Query ID = acadgild\_20170401225858\_07fcc878-ac8e-4779-b6e0-9222390d8fc1

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\_1491017400247\_0012, Tracking URL = http://localhost:8088/proxy/application\_1491017400247\_0012/

Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job -kill job\_1491017400247\_0012

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1

2017-04-01 23:02:50,094 Stage-1 map = 0%, reduce = 0%

2017-04-01 23:03:50,380 Stage-1 map = 0%, reduce = 0%

2017-04-01 23:04:50,763 Stage-1 map = 0%, reduce = 0%

2017-04-01 23:05:51,210 Stage-1 map = 0%, reduce = 0%

2017-04-01 23:06:08,838 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 14.11 sec

2017-04-01 23:07:09,491 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 14.11 sec

2017-04-01 23:08:09,675 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 18.29 sec

2017-04-01 23:08:22,172 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 26.11 sec

MapReduce Total cumulative CPU time: 26 seconds 110 msec

Ended Job = job\_1491017400247\_0012

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 26.64 sec HDFS Read: 518899 HDFS Write: 1541 SUCCESS

Total MapReduce CPU Time Spent: 26 seconds 640 msec

OK

NULL NULL

Afghanistan 2.0

Algeria 8.0

Argentina 139.0

Armenia 10.0

Australia 601.0

Austria 91.0

Azerbaijan 25.0

Bahamas 24.0

Bahrain 1.0

Barbados 1.0

Belarus 97.0

Belgium 18.0

Botswana 1.0

Brazil 221.0

Bulgaria 41.0

Cameroon 17.0

Canada 369.0

Chile 22.0

China 530.0

Chinese Taipei 20.0

Colombia 13.0

Costa Rica 2.0

Croatia 81.0

Cuba 188.0

Cyprus 1.0

Czech Republic 81.0

Denmark 89.0

Dominican Republic 5.0

Ecuador 1.0

Egypt 8.0

Eritrea 1.0

Estonia 18.0

Ethiopia 29.0

Finland 118.0

France 317.0

Gabon 1.0

Georgia 21.0

Germany 629.0

Great Britain 322.0

Greece 59.0

Grenada 1.0

Guatemala 1.0

Hong Kong 3.0

Hungary 145.0

Iceland 15.0

India 11.0

Indonesia 22.0

Iran 24.0

Ireland 7.0

Israel 3.0

Italy 329.0

Jamaica 80.0

Japan 282.0

Kazakhstan 42.0

Kenya 39.0

Kuwait 2.0

Kyrgyzstan 3.0

Latvia 17.0

Lithuania 30.0

Macedonia 1.0

Malaysia 3.0

Mauritius 1.0

Mexico 38.0

Moldova 5.0

Mongolia 10.0

Montenegro 14.0

Morocco 11.0

Mozambique 1.0

Netherlands 318.0

New Zealand 52.0

Nigeria 39.0

North Korea 21.0

Norway 192.0

Panama 1.0

Paraguay 17.0

Poland 80.0

Portugal 9.0

Puerto Rico 2.0

Qatar 3.0

Romania 123.0

Russia 768.0

Saudi Arabia 5.0

Serbia 31.0

Serbia and Montenegro 38.0

Singapore 7.0

Slovakia 35.0

Slovenia 25.0

South Africa 25.0

South Korea 308.0

Spain 204.0

Sri Lanka 1.0

Sudan 1.0

Sweden 180.0

Switzerland 93.0

Syria 1.0

Tajikistan 3.0

Thailand 18.0

Togo 1.0

Trinidad and Tobago 19.0

Tunisia 4.0

Turkey 28.0

Uganda 1.0

Ukraine 141.0

United Arab Emirates 1.0

United States 1301.0

Uruguay 1.0

Uzbekistan 19.0

Venezuela 4.0

Vietnam 2.0

Zimbabwe 7.0

Time taken: 616.217 seconds, Fetched: 111 row(s)

Code/Script

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

-- create a database custom

create database custom;

-- entering to database custom

use custom;

-- create table OLYMPIC

create table olympic

(athlete string,age int,country string,year int,closingdate date,sport string,goldmedals int,silvermedals int,bronzemedals int, totalmedals int)

row format serde 'org.apache.hadoop.hive.serde2.OpenCSVSerde'

with serdeproperties (

"separatorChar"="\t","quoteChar" = "'","escapeChar" ="\\"

) stored as textfile;

create table olympic

(athlete string,age int,country string,year int,closingdate date,sport string,goldmedals int,silvermedals int,bronzemedals int, totalmedals int)

row format serde 'org.apache.hadoop.hive.serde2.OpenCSVSerde'

stored as textfile;

-- load data into OLYMPIC

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

-- no of gold medals won by each country

select country,sum(goldmedals) from olympic group by country;

Output

hive> select country,sum(goldmedals) from olympic group by country;

Query ID = acadgild\_20170401230909\_be74bb24-63c4-493b-aac8-5637d362d26d

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\_1491017400247\_0014, Tracking URL = http://localhost:8088/proxy/application\_1491017400247\_0014/

Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job -kill job\_1491017400247\_0014

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1

2017-04-01 23:18:56,202 Stage-1 map = 0%, reduce = 0%

2017-04-01 23:19:56,676 Stage-1 map = 0%, reduce = 0%

2017-04-01 23:20:53,772 Stage-1 map = 67%, reduce = 0%, Cumulative CPU 17.45 sec

2017-04-01 23:20:58,151 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 17.89 sec

2017-04-01 23:21:58,859 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 17.89 sec

2017-04-01 23:22:59,211 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 17.89 sec

2017-04-01 23:23:20,151 Stage-1 map = 100%, reduce = 33%, Cumulative CPU 20.67 sec

2017-04-01 23:23:22,960 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 22.64 sec

2017-04-01 23:23:37,003 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 29.85 sec

MapReduce Total cumulative CPU time: 29 seconds 850 msec

Ended Job = job\_1491017400247\_0014

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 30.97 sec HDFS Read: 518899 HDFS Write: 1502 SUCCESS

Total MapReduce CPU Time Spent: 30 seconds 970 msec

OK

NULL NULL

Afghanistan 0.0

Algeria 2.0

Argentina 48.0

Armenia 0.0

Australia 161.0

Austria 36.0

Azerbaijan 6.0

Bahamas 11.0

Bahrain 0.0

Barbados 0.0

Belarus 17.0

Belgium 2.0

Botswana 0.0

Brazil 46.0

Bulgaria 8.0

Cameroon 17.0

Canada 168.0

Chile 3.0

China 234.0

Chinese Taipei 2.0

Colombia 2.0

Costa Rica 0.0

Croatia 35.0

Cuba 57.0

Cyprus 0.0

Czech Republic 14.0

Denmark 46.0

Dominican Republic 3.0

Ecuador 0.0

Egypt 1.0

Eritrea 0.0

Estonia 6.0

Ethiopia 13.0

Finland 11.0

France 108.0

Gabon 0.0

Georgia 6.0

Germany 223.0

Great Britain 124.0

Greece 12.0

Grenada 1.0

Guatemala 0.0

Hong Kong 0.0

Hungary 77.0

Iceland 0.0

India 1.0

Indonesia 5.0

Iran 10.0

Ireland 1.0

Israel 1.0

Italy 86.0

Jamaica 24.0

Japan 57.0

Kazakhstan 13.0

Kenya 11.0

Kuwait 0.0

Kyrgyzstan 0.0

Latvia 3.0

Lithuania 5.0

Macedonia 0.0

Malaysia 0.0

Mauritius 0.0

Mexico 19.0

Moldova 0.0

Mongolia 2.0

Montenegro 0.0

Morocco 2.0

Mozambique 1.0

Netherlands 101.0

New Zealand 18.0

Nigeria 6.0

North Korea 6.0

Norway 97.0

Panama 1.0

Paraguay 0.0

Poland 20.0

Portugal 1.0

Puerto Rico 0.0

Qatar 0.0

Romania 57.0

Russia 234.0

Saudi Arabia 0.0

Serbia 1.0

Serbia and Montenegro 11.0

Singapore 0.0

Slovakia 10.0

Slovenia 5.0

South Africa 10.0

South Korea 110.0

Spain 19.0

Sri Lanka 0.0

Sudan 0.0

Sweden 57.0

Switzerland 21.0

Syria 0.0

Tajikistan 0.0

Thailand 6.0

Togo 0.0

Trinidad and Tobago 1.0

Tunisia 2.0

Turkey 9.0

Uganda 1.0

Ukraine 31.0

United Arab Emirates 1.0

United States 545.0

Uruguay 0.0

Uzbekistan 5.0

Venezuela 1.0

Vietnam 0.0

Zimbabwe 2.0

Time taken: 827.904 seconds, Fetched: 111 row(s)

hive>