**Problem statement**

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

**Sample from Data :**

|  |  |
| --- | --- |
| 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 |  |

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.

**First to Load the data in HIVE table.**

[cloudera@quickstart ~]$ cd hive

[cloudera@quickstart hive]$ ls

000000\_0 olympix\_data.csv test

CREATE TABLE olympics\_data

(

athlete STRING,

age INT,

country STRING,

year INT,

closing\_date STRING,

sport STRING,

gold INT,

silver INT,

bronze INT,

total INT

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ' '

COLLECTION ITEMS TERMINATED BY ':';

LOAD DATA LOCAL INPATH '/home/cloudera/hive/olympix\_data.csv'

INTO TABLE olympics\_data;