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

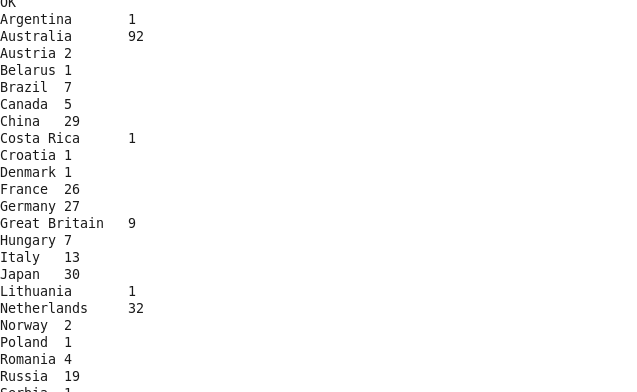
Bronze Medals: No. of Bronze medals

Total Medals: Consists of total no. of medals

5. Problem Statement

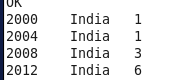
1. Write a Hive program to find the number of medals won by each country in swimming.

Select country , count(total) from Olympic\_data where sports=”Swimming” group by country;



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

Select year,country,count(total) from Olympic where country = “India” group by year,country;



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

Create view country\_medals as Select country,count(total) from Olympic group by country;

Insert overwrite local directory ‘/home/acadgild/Azhar/ASS7.2/output’ row format delimited fields terminated by ‘,’ select \* from country\_medals;

Output: Refer 7.3.3 document in github

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

Create view country\_gold as Select country,count(gold) from Olympic group by country;

Insert overwrite local directory ‘/home/acadgild/Azhar/ASS7.2/ASS7.2.4’ row format delimited fields terminated by ‘,’ select \* from country\_gold;

Output: Refer 7.3.4 document in github