Assignment 15.2

Associated Data Files

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

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

**DATA 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

Load Data

CREATE EXTERNAL TABLE olympic\_data(

athlete\_name 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 '\t' STORED AS TEXTFILE;

hadoop fs -put /home/ec2-user/hive\_input/olympix\_data.csv /user/ec2-user/hive\_input\_files/olympix\_data.csv

load data inpath '/user/ec2-user/hive\_input\_files/olympix\_data.csv' into table olympic\_data;

**5. Problem Statement**

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

select country,year,sum(total) as total\_medals from olympic\_data where sport='Swimming' group by year,country;

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

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

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

select country, sum(total) as total\_medals from olympic\_data group by country;

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

select country, sum(gold) as number\_of\_gold\_medals from olympic\_data group by country;