**Session 7: HIVE Operations**

Assignment 7.2

Student Name: Abarajithan SA

Course: Big Data Hadoop & Spark Training

Start Date:  2017-09-09

End Date:  2017-11-26

**Assignment 7.2**– This Data set is about Olympics. Write HIVE programs to finish the provided tasks

Contents

[Introduction 2](#_Toc498098237)

[Associated Data files 2](#_Toc498098238)

[Problem Statement 2](#_Toc498098239)

[HIVE Queries create a Database and a table providing the Input Dataset. 2](#_Toc498098240)

[Task1 3](#_Toc498098241)

[HIVE Query 3](#_Toc498098242)

[TASK2 4](#_Toc498098244)

[HIVE Query 4](#_Toc498098245)

[TASK3 4](#_Toc498098247)

[HIVE Query 4](#_Toc498098248)

[Task4 6](#_Toc498098249)

[HIVE Query 6](#_Toc498098250)

# Introduction

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

# Associated Data files

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

# 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.

# HIVE Queries create a Database and a table providing the Input Dataset.

***CREATE DATABASE olympic;***

***USE olympic;***

***CREATE TABLE olympic\_data***

***(Athlete STRING, Age INT, Country STRING, Year INT, Closing\_Date STRING, Sport STRING, Gold\_Medals INT, Silver\_Medals INT, Bronze\_Medals INT,Total\_Medals INT)***

***ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t';***

***LOAD DATA LOCAL INPATH '/home/acadgild/hadoop/olympic\_data.json'***

***INTO TABLE olympic\_data;***

# Task1

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

## HIVE Query

***SELECT country, SUM(total\_medals) as Total\_Medals FROM olympic\_data WHERE sport<=>'Swimming' GROUP BY country;***



Expected Output



# TASK2

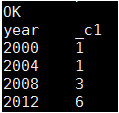
Write a Hive program to find the number of medals that India won year wise.

## HIVE Query

***SELECT year, SUM(total\_medals) FROM olympic\_data Where country<=>'India' GROUP BY year;***



Expected Output



# TASK3

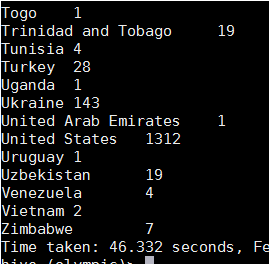
Write a Hive Program to find the total number of medals each country won.

## HIVE Query

***SELECT country, SUM(total\_medals) as Total\_Medals FROM olympic\_data GROUP BY country;***

Expected Output





# Task4

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

## HIVE Query

***SELECT country, SUM(gold\_medals) as GOLD\_Medals FROM olympic\_data GROUP BY country;***

Expect Output

