*Big Data and Hadoop Development*

Session 07: Hive Operations

Assignment 2

**A C A D G I L D Page 1**

*Big Data and Hadoop Development*

**A C A D G I L D**

*Session 15: Hive Operations*

*Assignment –Hive*

**Table of Contents**

Table of Contents .......................................................................................................................................... 2

1. Introduction .......................................................................................................................................... 3

2. Objective ............................................................................................................................................... 3

3. Prerequisites ......................................................................................................................................... 3

4. Associated Data Files ............................................................................................................................ 3

5. Problem Statement ............................................................................................................................... 4

6. Expected output .................................................................................................................................... 4

7. Approximate Time to Complete Task ................................................................................................... 4

**Page 2**

*Big Data and Hadoop Development*

**A C A D G I L D**

**1. Introduction**

In this assignment you need to write hive programs for given problem statements.

**2. Objective**

This assignment will help you to understand hiveql and implement it with the given use case.

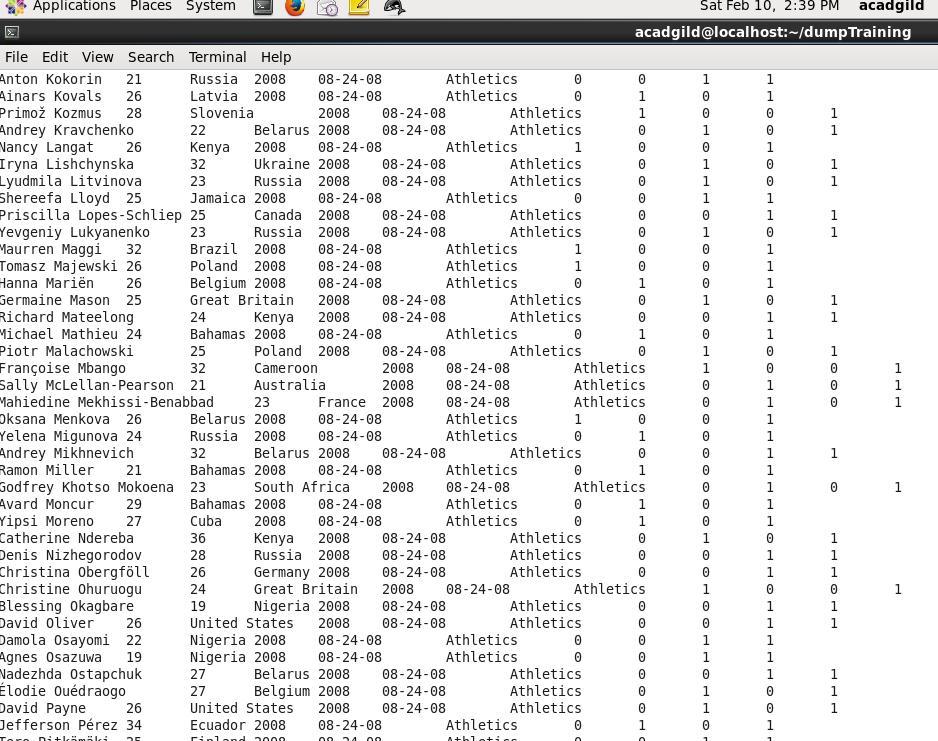
**3. Prerequisites**

Acadgild’s VM, or Linux operating system with Hadoop installed in it.

**4. 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



**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

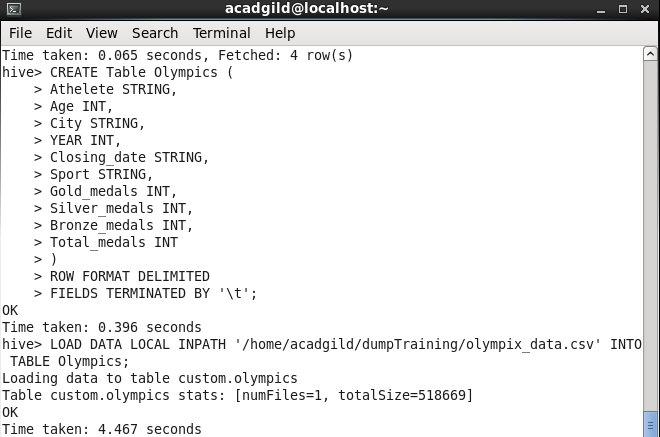
Sport: Consists of the sports name

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

****

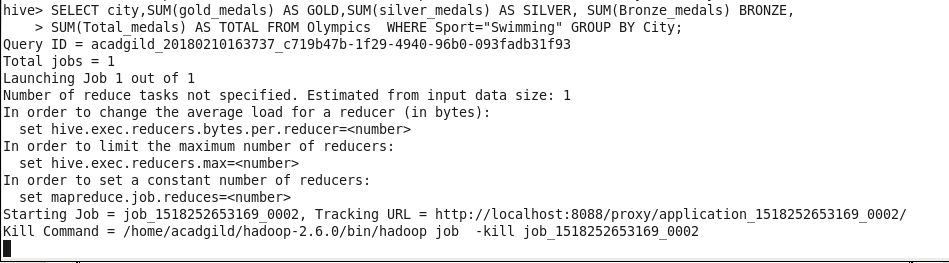
**Page 3**

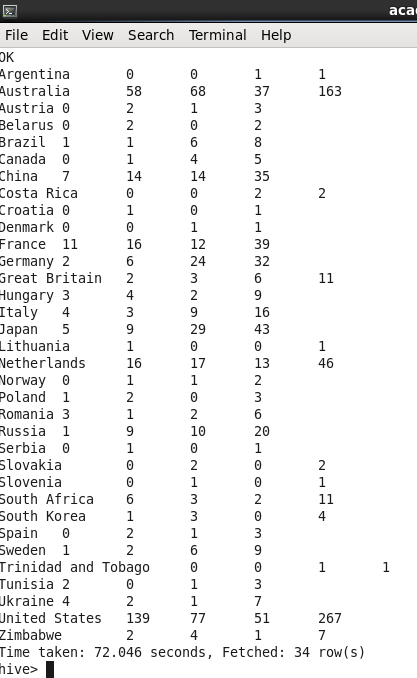
*Big Data and Hadoop Development*

**A C A D G I L D**

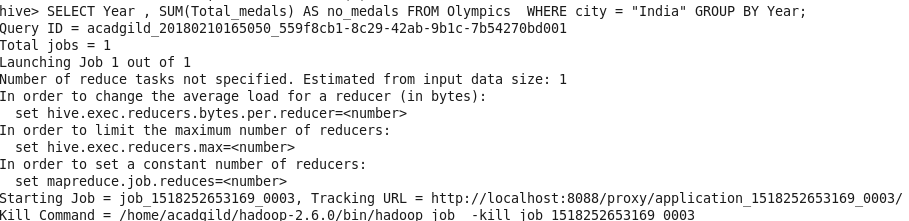
**5. Problem Statement**

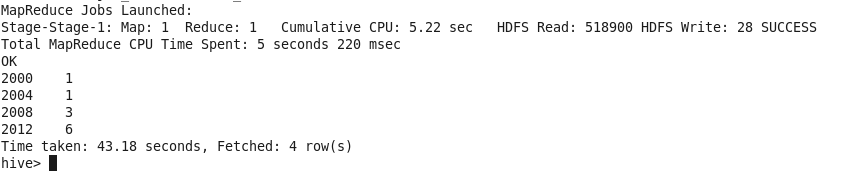
1. Write a Hive program to find the number of medals won by each country in swimming.



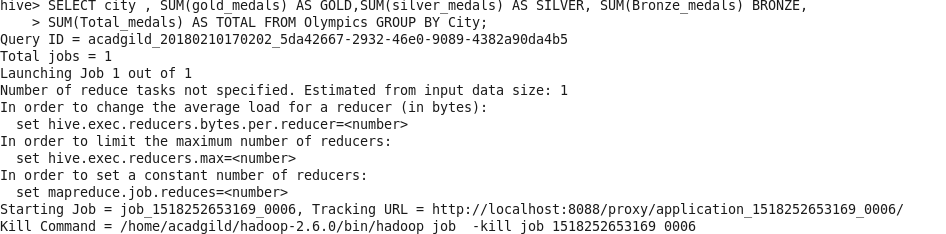


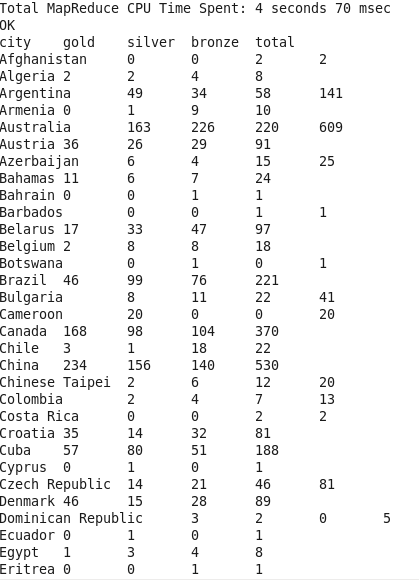
2. Write a Hive program to find the number of medals that India won year wise.

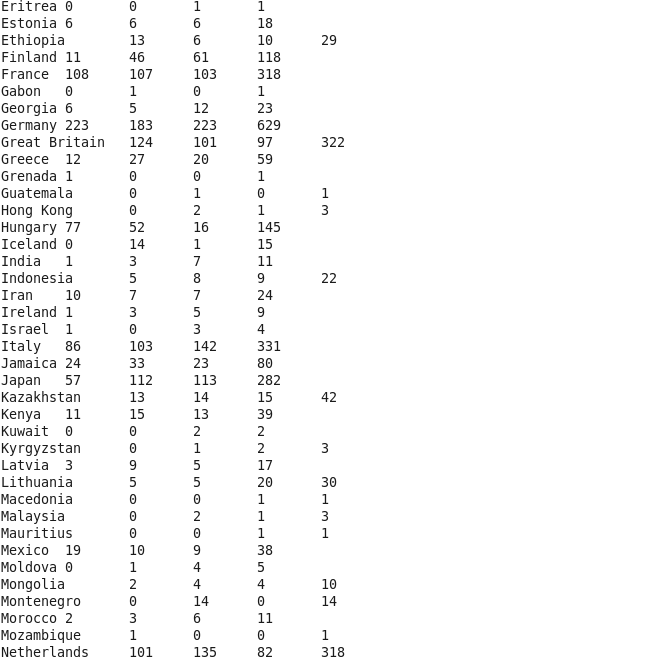


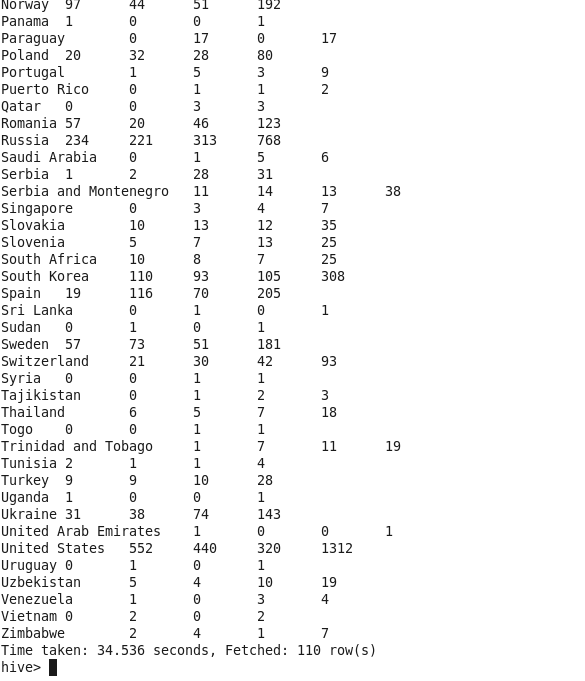


3. Write a Hive Program to find the total number of medals each country won.

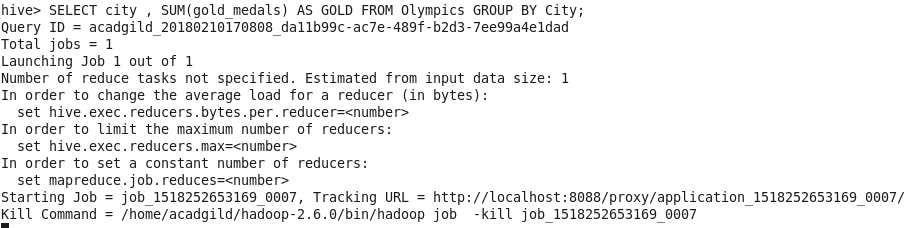


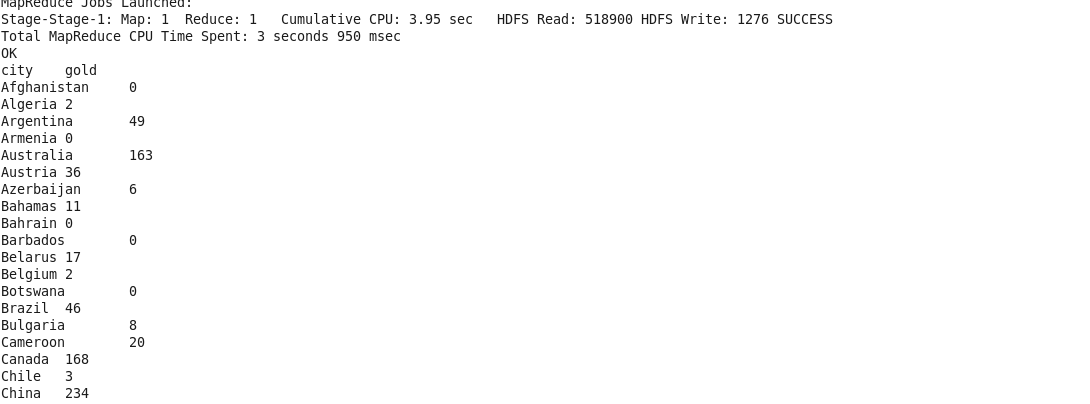


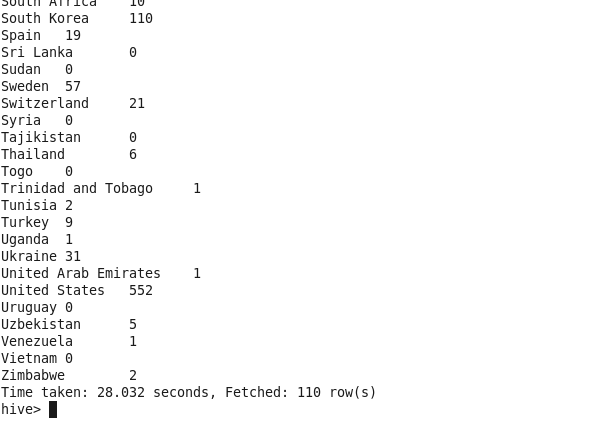




4. Write a Hive program to find the number of gold medals each country won.







**ge 4**