**Hive 7.2 Assignment:**

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

**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.

Solution:

**Step 1:**

Use custom;

// Creating table on CSVSerDe

------------------------------------------------------------------

Create table Olympic (Athlete string, Age int, Country string, Year string, ClosingDate string, Sport string, Gold\_Medals int, Silver\_Medals int, Bronze\_Medals int, Total\_Medals int) ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.OpenCSVSerde'

WITH SERDEPROPERTIES (

"separatorChar" = "\t",

"quoteChar" = "\"",

"escapeChar" = "\\"

)

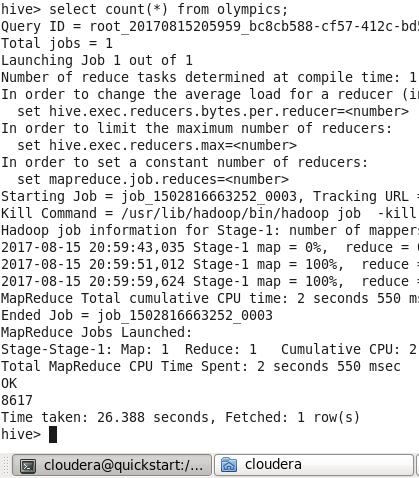
STORED AS TEXTFILE TBLPROPERTIES ("skip.header.line.count"="1");

Describe Olympic;

Step 2:

---------------------------Loading in the table-----------------------------------------------------

LOAD DATA LOCAL INPATH '/home/cloudera/olympix\_data.csv' OVERWRITE INTO TABLE Olympic;



view table in hdfs

--> hadoop fs -cat /user/hive/warehouse/custom.db/olympic/olympix\_data.csv

**Problem Statement :1**

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

SELECT COUNT(Total\_Medals), COUNTRY FROM OLYMPIC WHERE SPORT= 'Swimming'

GROUP BY COUNTRY;

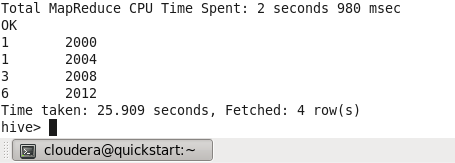


**Problem Statement :2**

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

SELECT COUNT(Total\_Medals),Year FROM OLYMPIC WHERE country= 'India'

Group by Year ;



**Problem Statement :3**

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

SELECT COUNT(Total\_Medals),Country FROM OLYMPIC Group by Country;



**Problem Statement :4**

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

SELECT COUNT(Gold\_Medals),Country FROM OLYMPIC where Gold\_Medals=1

Group by Country;

