**Problem Statement**

CREATE DATABASE IF NOT EXISTS GAMES;

USE GAMES;

**Creating a staging table with closing\_data as string so that this value should not be skipped while loading**

CREATE DATABASE IF NOT EXISTS GAMES;

USE GAMES;

CREATE TABLE IF NOT EXISTS olympics\_stg

(

Athlete\_Name 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 SERDE 'org.apache.hadoop.hive.serde2.OpenCSVSerde'

WITH SERDEPROPERTIES (

"separatorChar" = "\t",

"quoteChar" = "\"",

"escapeChar" = "\\"

)

STORED AS TEXTFILE;

**Loading staging table:**

LOAD DATA LOCAL INPATH '/home/acadgild/sridhar/hive/assignment7/olympix\_data.csv' INTO TABLE olympics\_stg

**Final table with changing closing\_date to Data type;**

CREATE TABLE IF NOT EXISTS olympics

(

Athlete\_Name string,

Age int,

Country string,

Year int,

Closing\_Date date,

Sport string,

Gold\_Medals int,

Silver\_Medals int,

Bronze\_Medals int,

Total\_Medals int

);

**Loading into table**

INSERT INTO TABLE olympics

SELECT

Athlete\_Name,

Age,

Country,

Year,

to\_date(from\_unixtime(unix\_timestamp(Closing\_Date,'MM-dd-yy'))) ,

Sport,

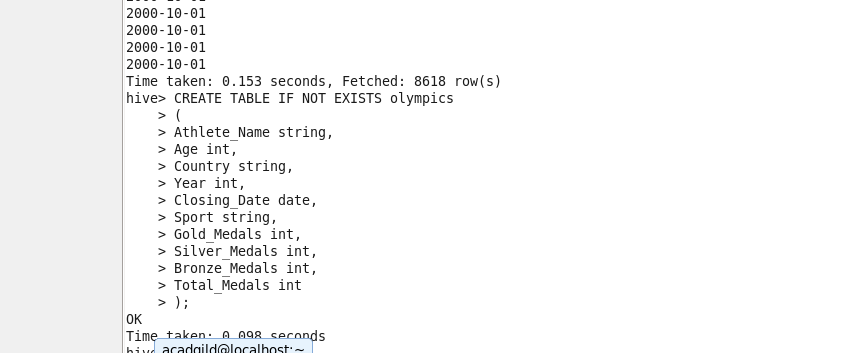
Gold\_Medals,

Silver\_Medals,

Bronze\_Medals,

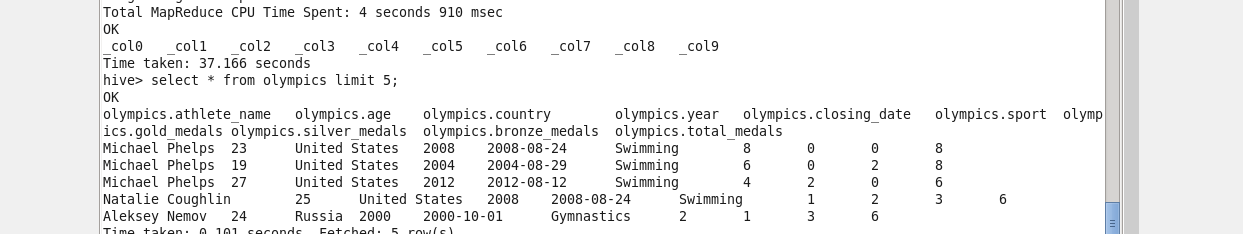
Total\_Medals

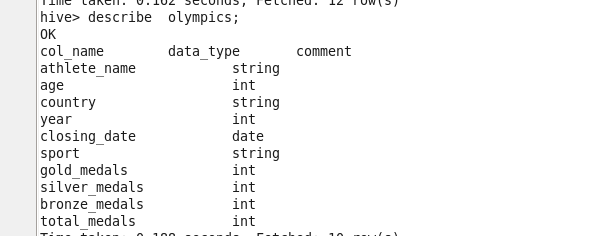
FROM olympics\_stg;





Sample records:





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

select DISTINCT Country, No\_of\_medals FROM

(

SELECT

Country,

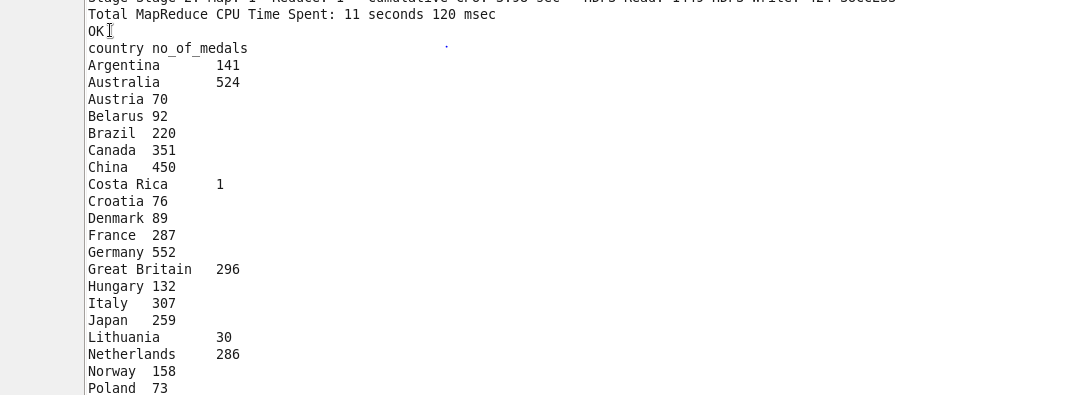
COUNT(Total\_Medals) OVER (PARTITION by(Country)) as No\_of\_medals,

sport

FROM olympics

)temp

WHERE Sport='Swimming';



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

select DISTINCT Year, No\_of\_medals FROM

(

SELECT

Year,

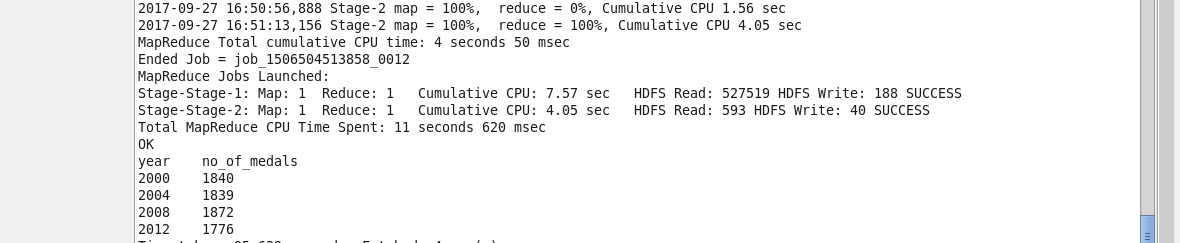
COUNT(Total\_Medals) OVER (PARTITION by(Year)) as No\_of\_medals,

Country

FROM olympics

)temp

WHERE Country='India';



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

select DISTINCT Country, No\_of\_medals FROM

(

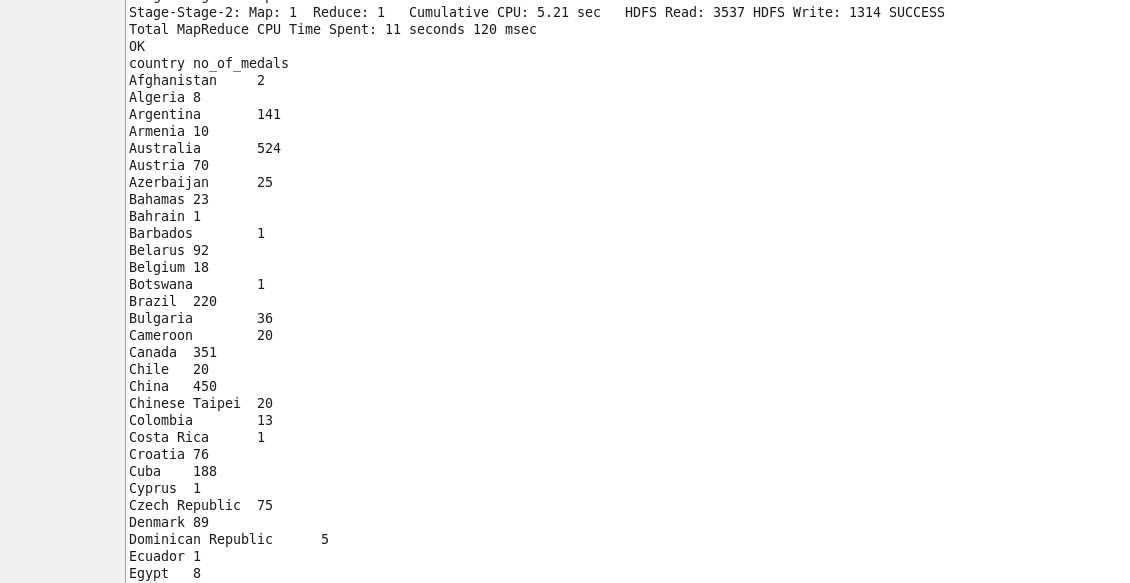
SELECT

Country,

COUNT(Total\_Medals) OVER (PARTITION by(Country)) as No\_of\_medals

FROM olympics

)temp;



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

select DISTINCT Country, No\_of\_medals FROM

(

SELECT

Country,

COUNT(Gold\_Medals) OVER (PARTITION by(Country)) as No\_of\_medals

FROM olympics

)temp;

