Assignment 7.2 ( Hive Introduction)

First, create a table olympics\_details with the given fields. The query is as below:

CREATE TABLE IF NOT EXISTS olympics\_details

(

Athlete string,

Age int,

Country string,

Year int,

ClosingDate string,

Sport string,

GoldMedals int,

SilverMedals int,

BronzeMedals int,

TotalMedals int

)

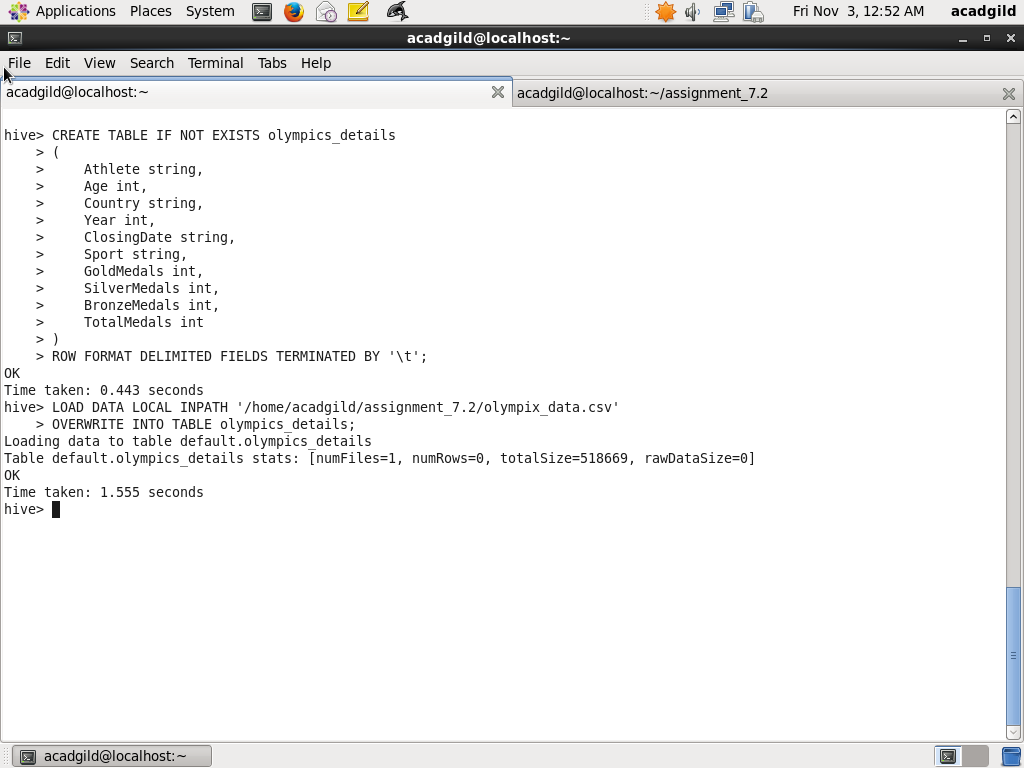
ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t';

Next Load data from the given dataset using query

LOAD DATA LOCAL INPATH ‘/home/acadgild/assignment\_7.2/olympics\_data.csv’

OVERWRITE INTO TABLE olympics\_details;

The screenshot of both the queries are as below:



Task1: Write a hive program to number of medals won by each country in swimming

Here we select the medals won group by country and Sports is Swimming. The query is as below:

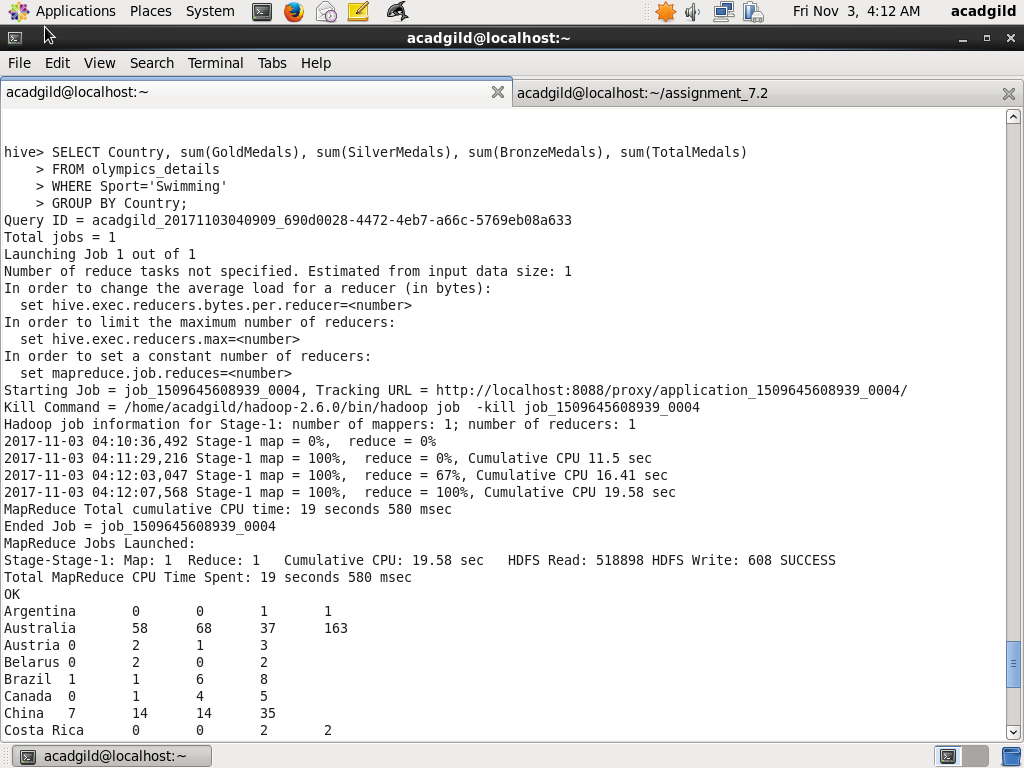
SELECT Country, sum(GoldMedals), sum(SilverMedals), sum(BronzeMedals), sum(TotalMedals)

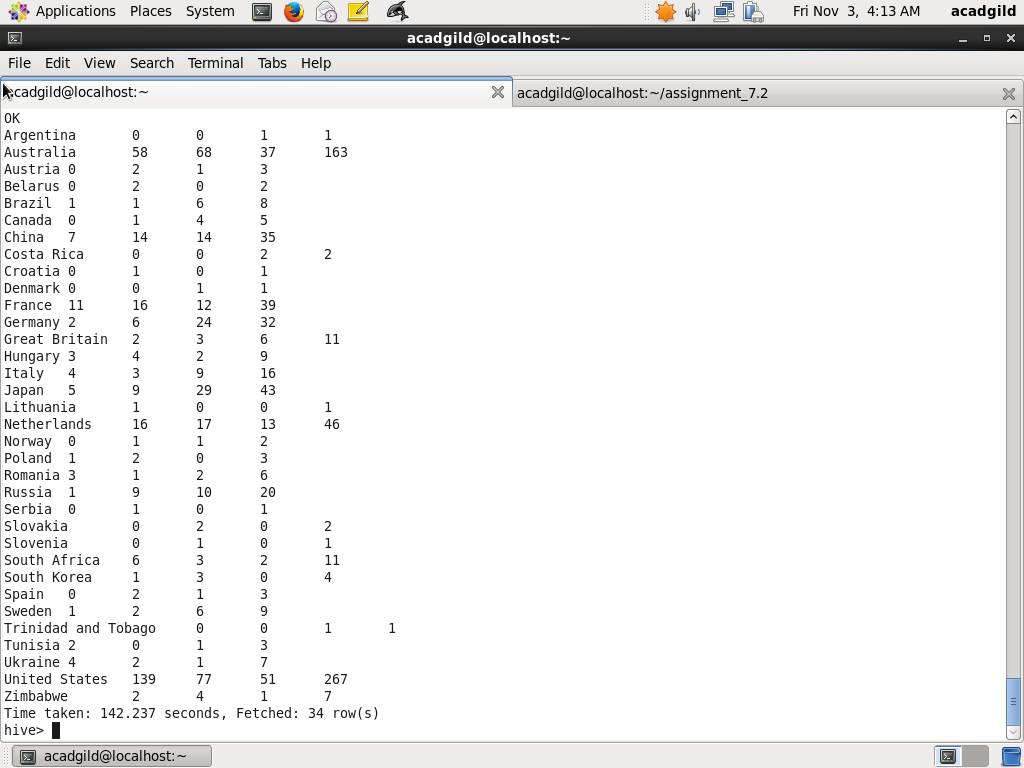
FROM olympics\_details

WHERE Sport='Swimming'

GROUP BY Country;

The screenshots of query and output is as below:





Task2: Write a hive program to find the numbers of medals won by India year wise

The following query will do the task, where sum function is used to sum medals(Gold, Silver< Bronze. Total) won by Country India Group By Year

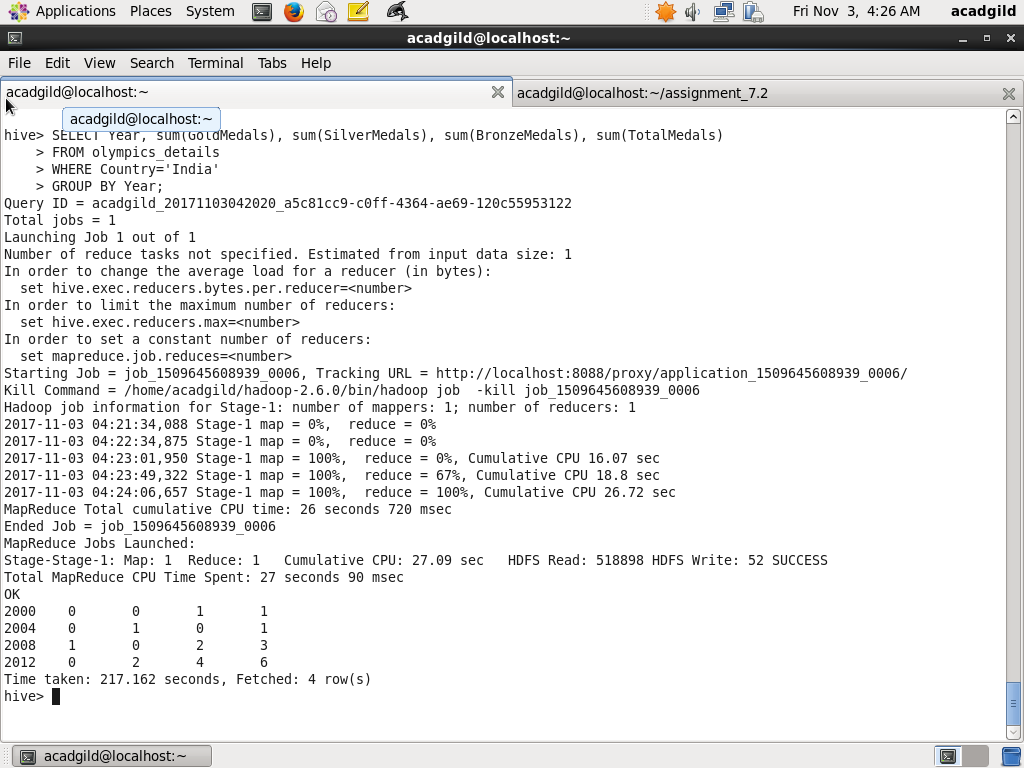
SELECT Year, sum(GoldMedals), sum(SilverMedals), sum(BronzeMedals), sum(TotalMedals)

FROM olympics\_details

WHERE Country='India'

GROUP BY Year;

Screenshot of Hive query and output is as below:



Task3: Write a hive program to find total number of medals each country won

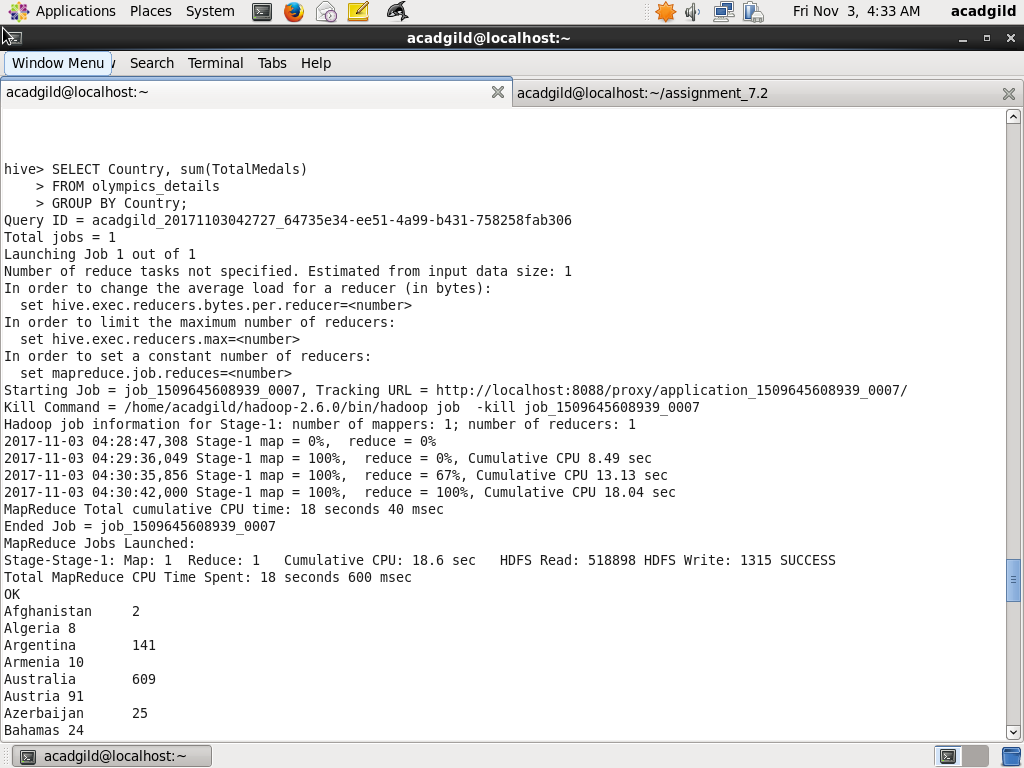
The following query will do the task, sum function is used to total medals group by Country

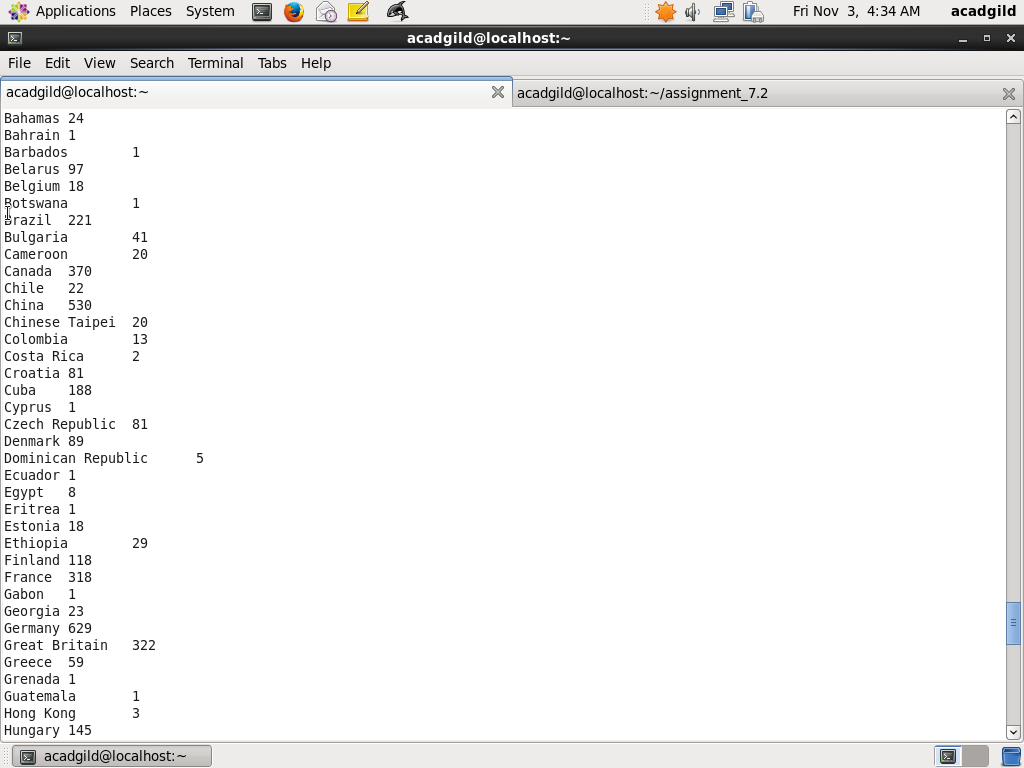
SELECT Country, sum(TotalMedals)

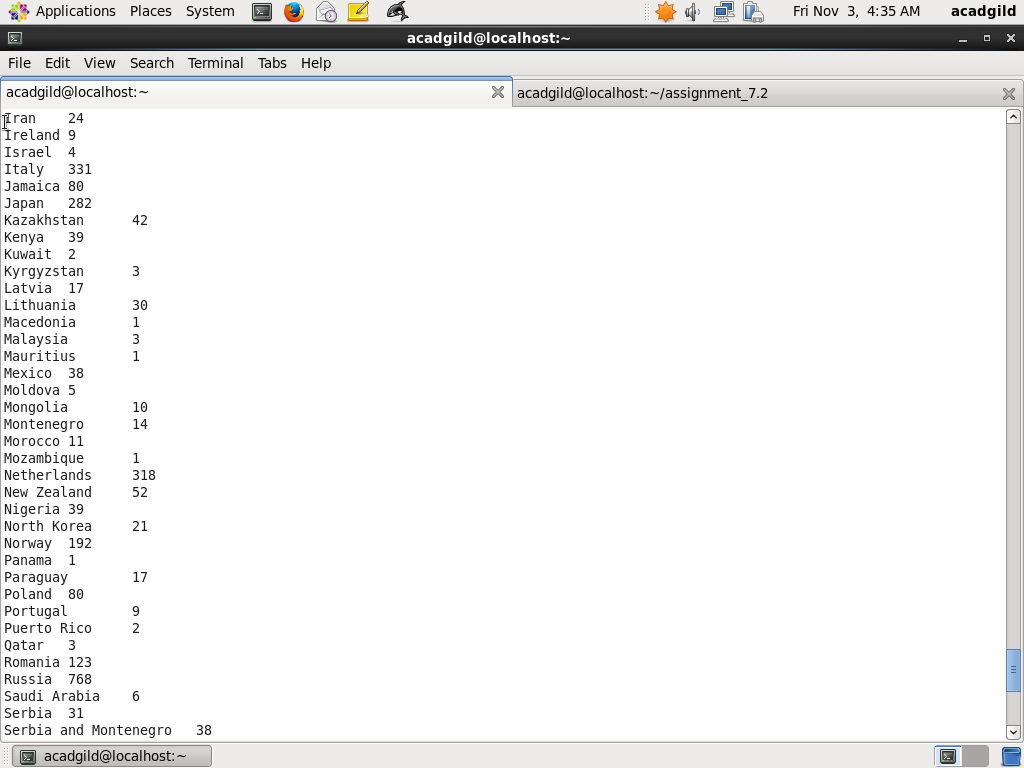
FROM olympics\_details

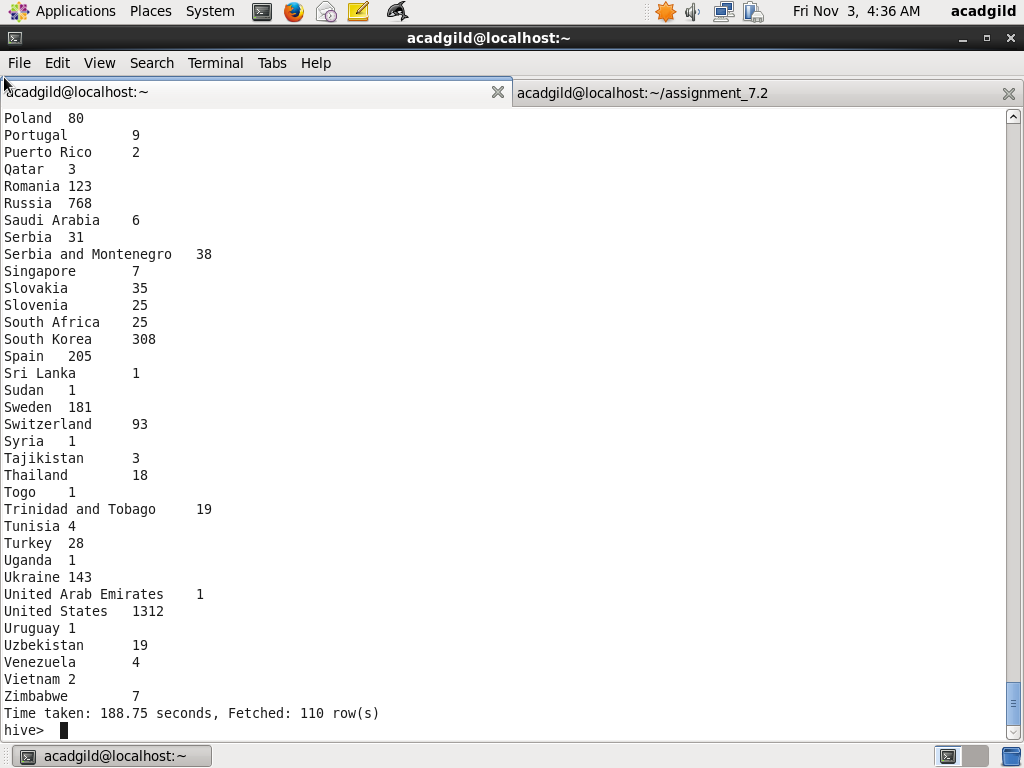
GROUP BY Country;

Screenshot of Hive query and output is as below:









Task4: Write a hive program to find total number of gold medals each country won

The following query will do the task, sum function is used to total Gold medals group by Country

SELECT Country, sum(GoldMedals)

FROM olympics\_details

GROUP BY Country;

Screenshots of Hive query and output is as below:

