First we have to create the database and load the dataset into it

Commands: 1- creating database

1:- Create database Olympic\_data;

Then use this database by using given command

2: - use olympic\_data;

Then we create table Olympic by given command

3:-Create table Olympic(Athlete string,Age int,Country string,Year string,Closing string,Spot string,Gold int,Silver int,Bronze int,Total int)

Now load the csv file present locally at /home/acadgild/Downloads/olympix\_data.csv

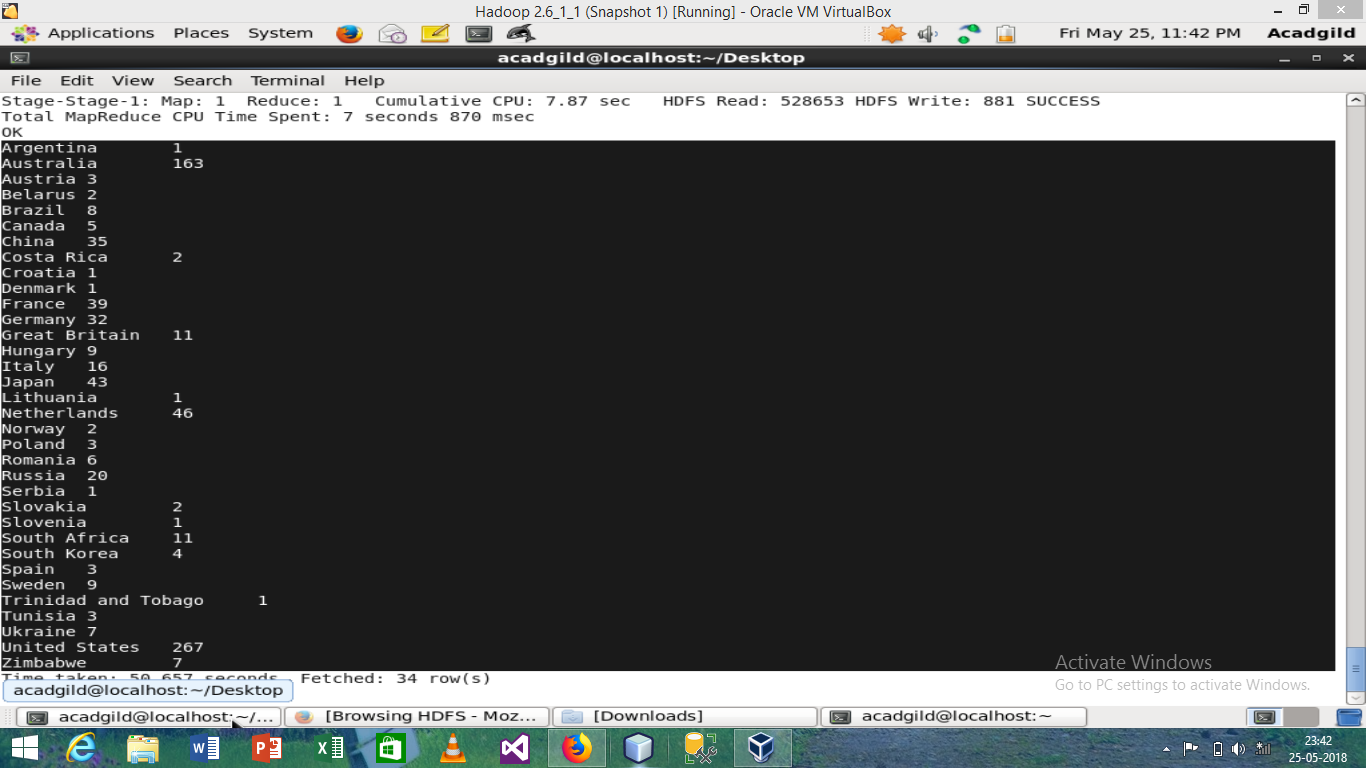
4:-Load data local inpath ‘/home/acadgild/Downloads/olympix\_data.csv’ into table Olympic

**Task1:- Write a Hive program to find the number of medals won by each country in swimming.**

**Solution:** select country, sum(total) from Olympic where spot = “Swimming” group by country

Is the command through which we can achieve this task.

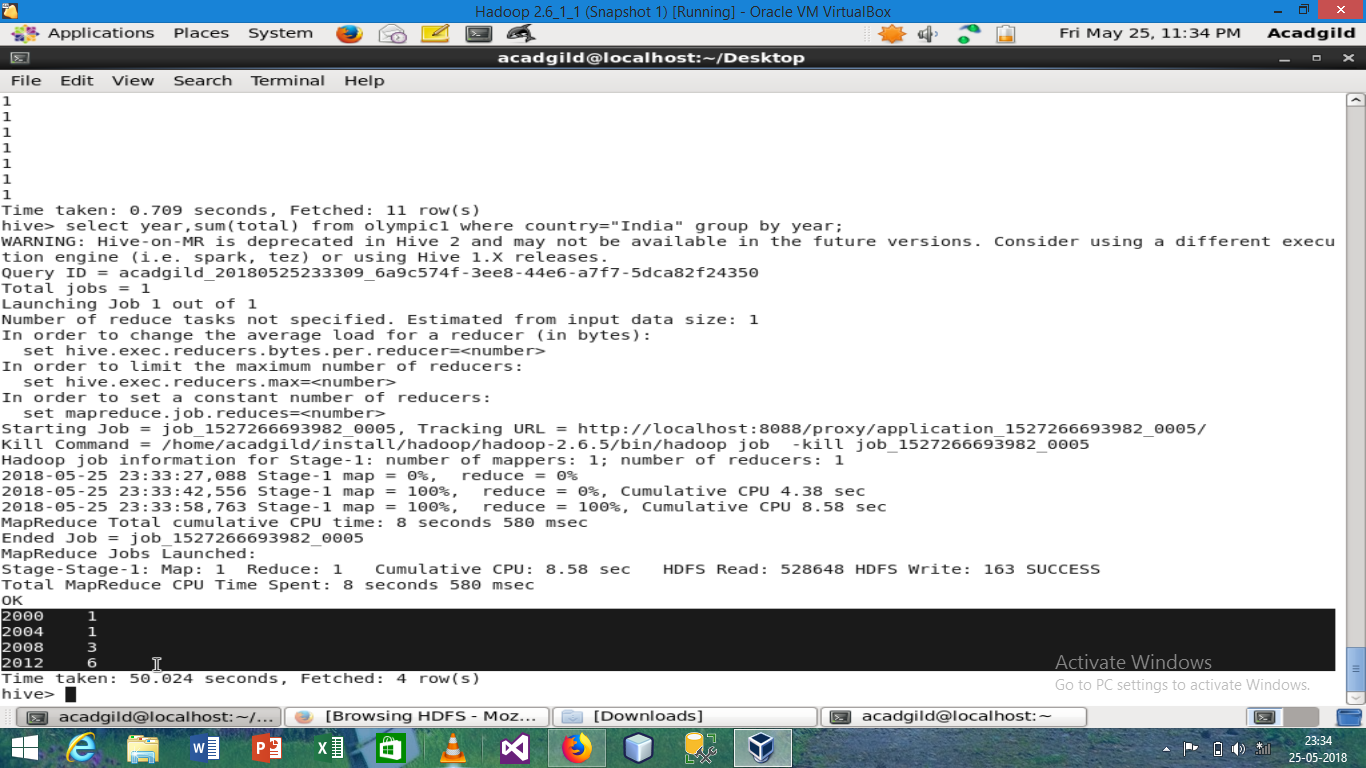
Screenshot of output is given below



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

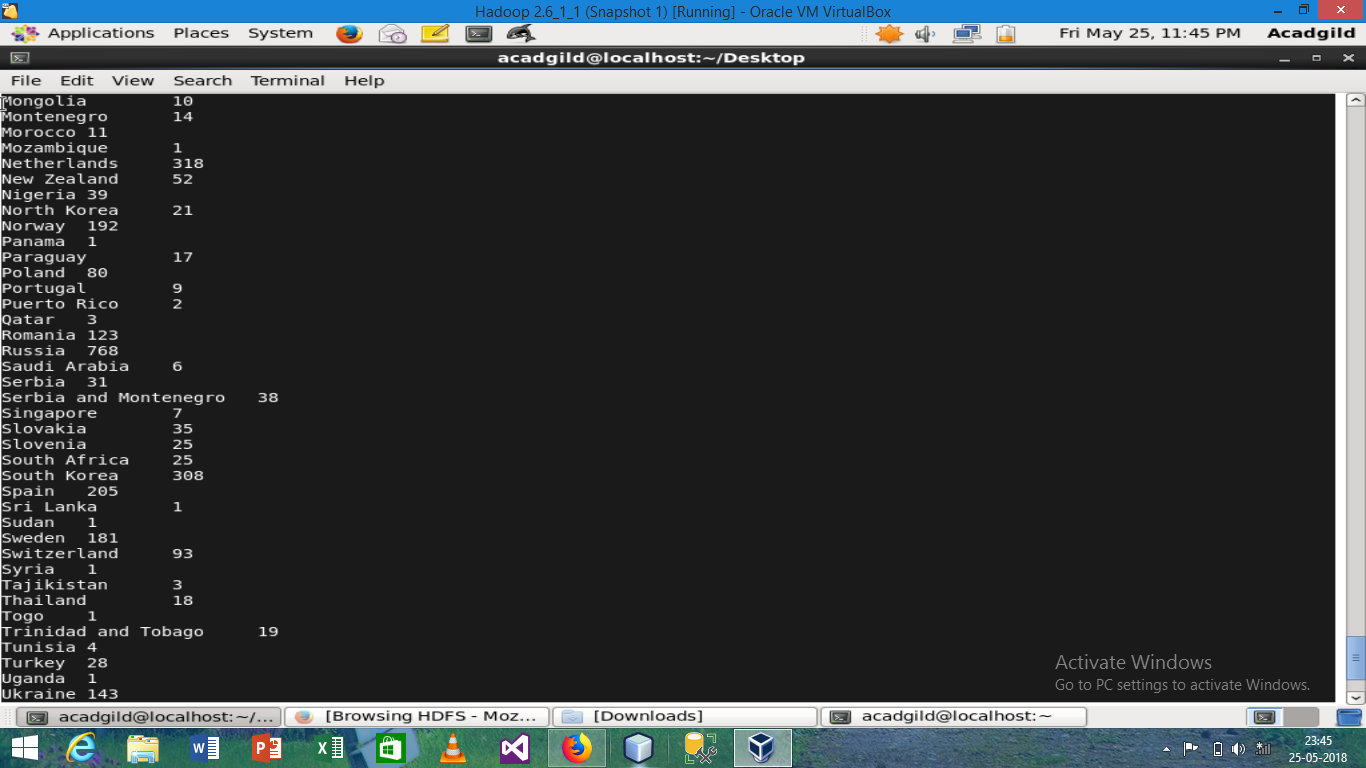
**Solution:** select year,sum(total) from Olympic where country = “India” group by year

Output



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

Solution: select country, sum(total) from Olympic group by country;

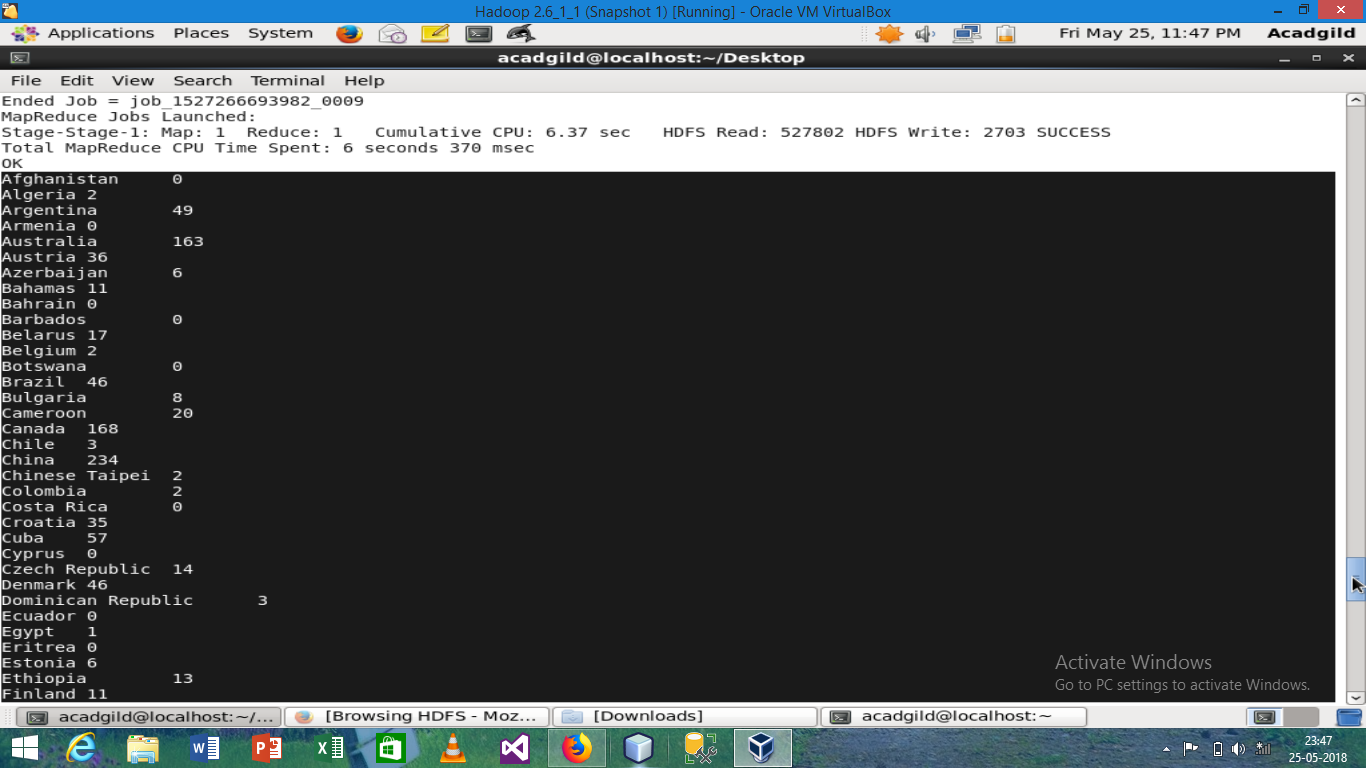


And many more………

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

**Solution:** Select country, sum(gold) from Olympic group by country;

Output:



And many more…….

**Task 2 :- Write a hive UDF that implements functionality of string concat\_ws(string SEP, array<string>).**

**This UDF will accept two arguments, one string and one array of string.**

**It will return a single string where all the elements of the array are separated by the SEP.**

**Solution:** for this we have to write UDF means we will create a class who extends UDF and this UDF class is present under org.apache.hadoop.hive.ql.exec.UDF package so we have to import this package after importing this package we will create a class who will extends UDF class and override evaluvate() method and under this we will write the logic.

Import java.util.ArayList;

Import org.apache.hadoop.hive.ql.exec.UDF;

Public class Append extends UDF

{

Public String evaluvate(String saperator, ArrayList<String> array)

{

If(array==null)

{return null;}

String a = “ ”;

a = a+array.get(0);

for(int i=1;i<array.size();i++)

{

a +=saperator;

a += array.get(i);

}

return a.toString();

}

}

Now we will export it to jar and include this jar to hive by following command

Add jar Append ‘/home/acadgild/Desktop/’

Create temporary function concatenation as Append

Now the UDF function is ready to use

The output is

