1. Explain the differences between static and dynamic partitioning in hive and their working procedures.

Explain the working of Partitioning in brief

1) When we are querying large dataset that time if we use the where clause then whole of the dataset has to searched and this will take more time. Suppose if the same query was done on small part of the data then the time taken will be very less.

2) This method of partitioning the dataset into small parts according to the particular field it is called as the partitioning.

3) Partitioning develops temporary files that are fed as input to the map reduce program. But if there are large number of small files then it would overload the namenode and would cause a problem. Hence the number of partitions has to be kept limited. By default the number of partitions are kept 100. This number can be modified.

4) Partitioning can be used in case of real time logging of data to segregate the data based on the time stamp so as to get the output as fast as possible.

5) Also if we are dealing with the customer data then we can segregate the data according to the area and find output related to particular area as fast as possible.

SET hive.exec.dynamic.partition = true;

SET hive.exec.dynamic.partition.mode = nonstrict;

SET hive.exec.max.dynamic.partitions.pernode=10000;

Explain the difference between Static and Dynamic Partitioning in Hive with an example.

Static Partitioning

Partitioning done with known fields is called as the static partitioning. We statically mention what partition has to be entered in to that partition. Static Partition saves your time in loading data compared to dynamic partition. “Statically” adding a partition in table and move the file into the partition of the table. This takes part of the database and stores into a file.

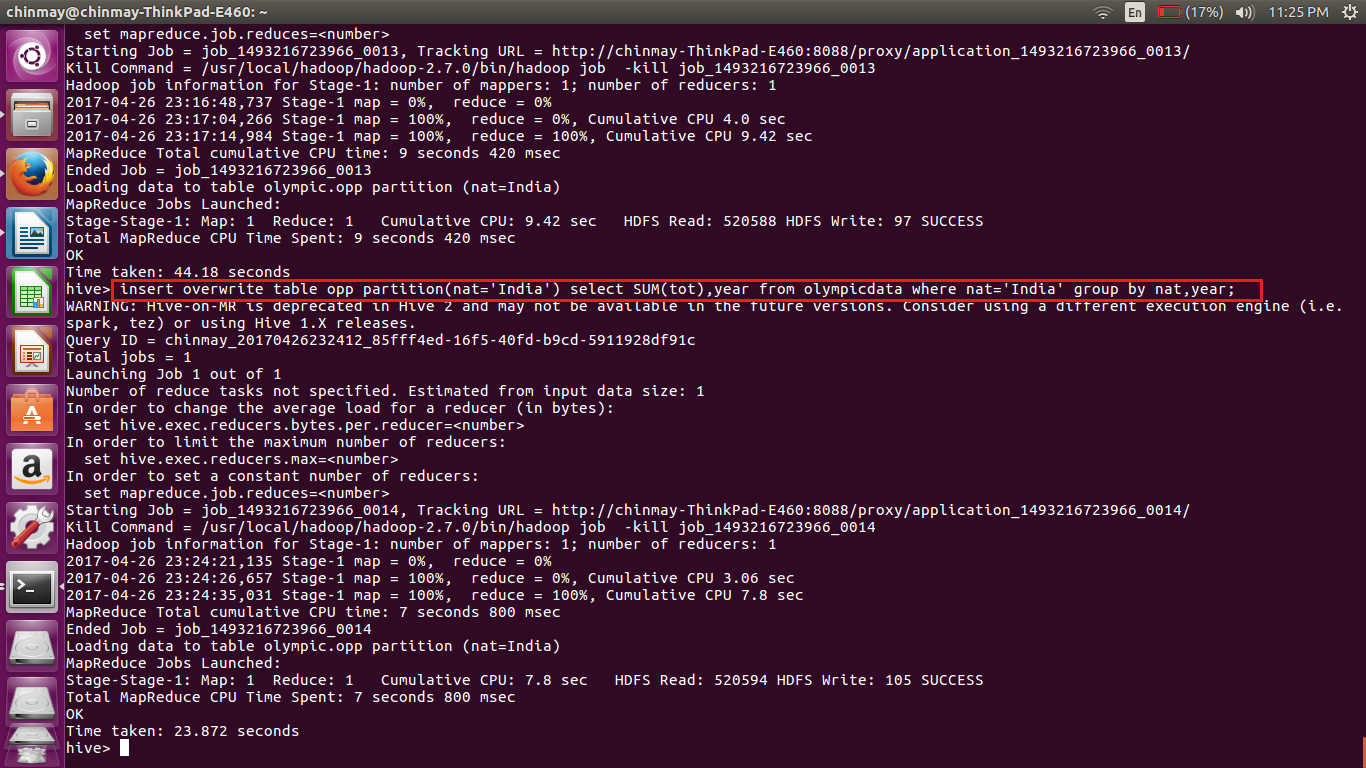
Dynamic Partition

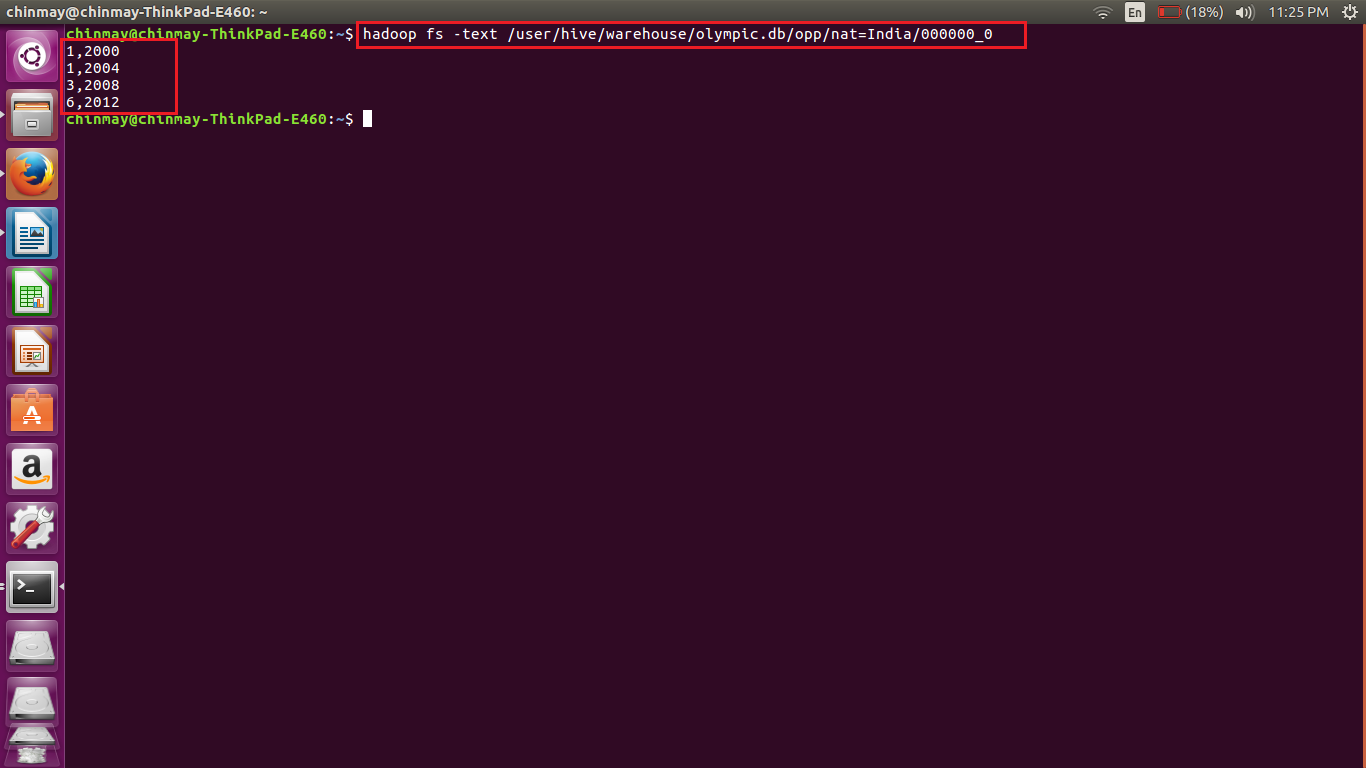
Here no specific condition is given. Only the field is mentioned and the partition and the files are formed for each of the distinct entries.

For Dynamic partition we have to we have to set dynamic partition as nonstrict.

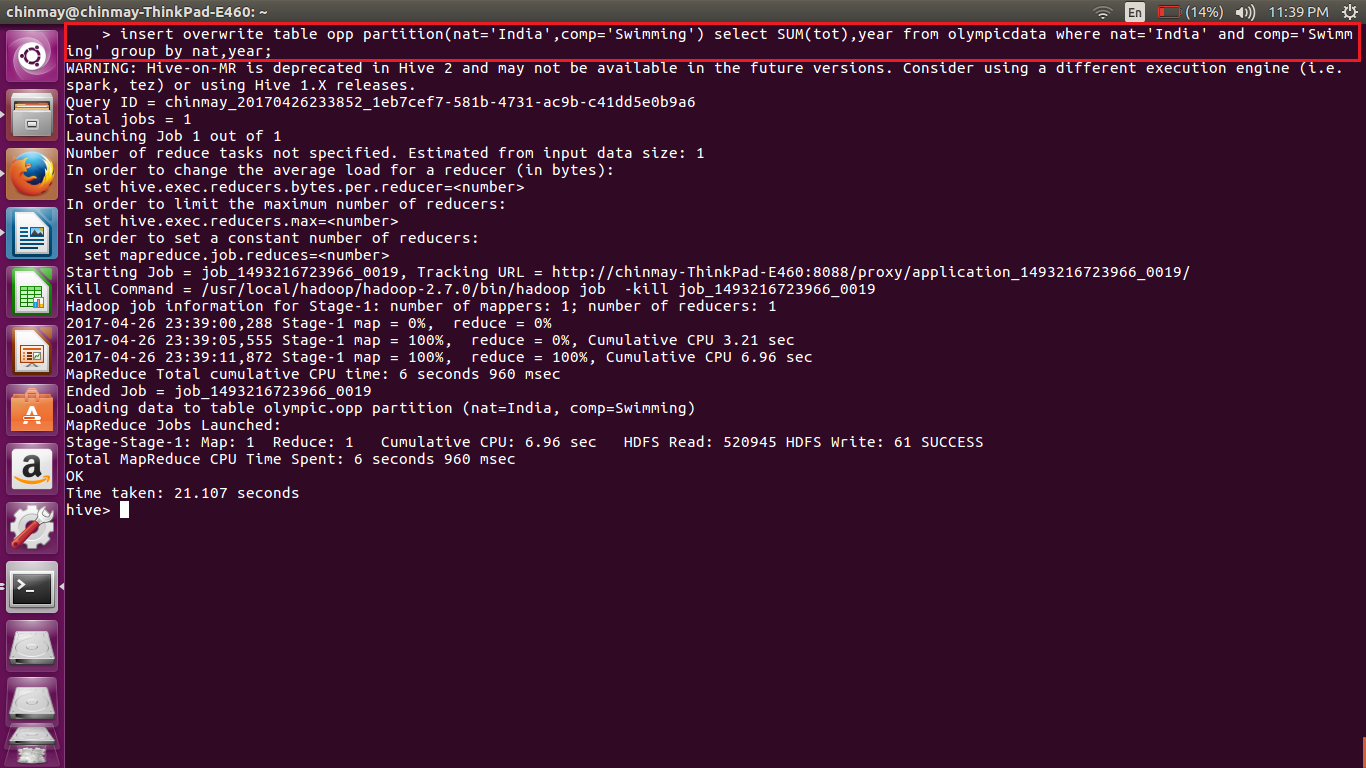
2. Use static partitioning in hive and evaluate the below problem statements

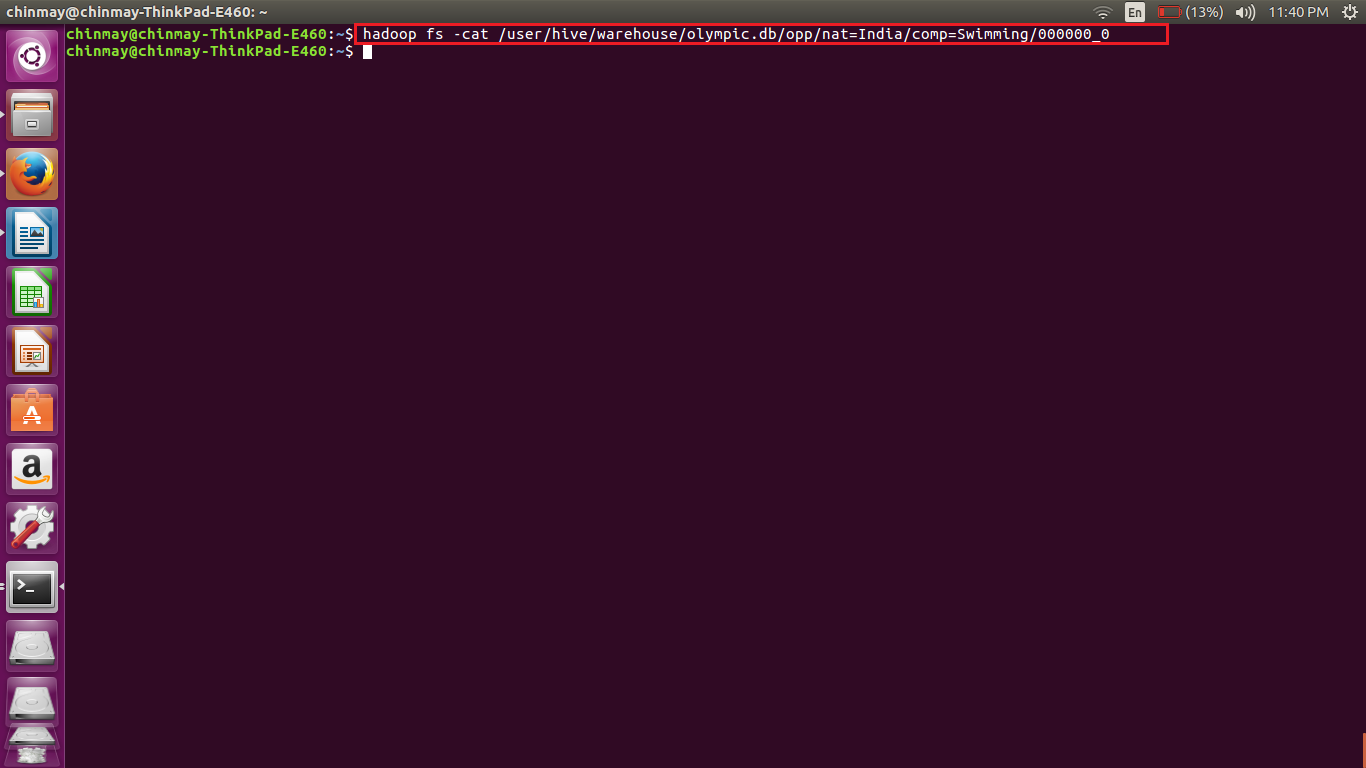
- Find the number of medals india won year wise



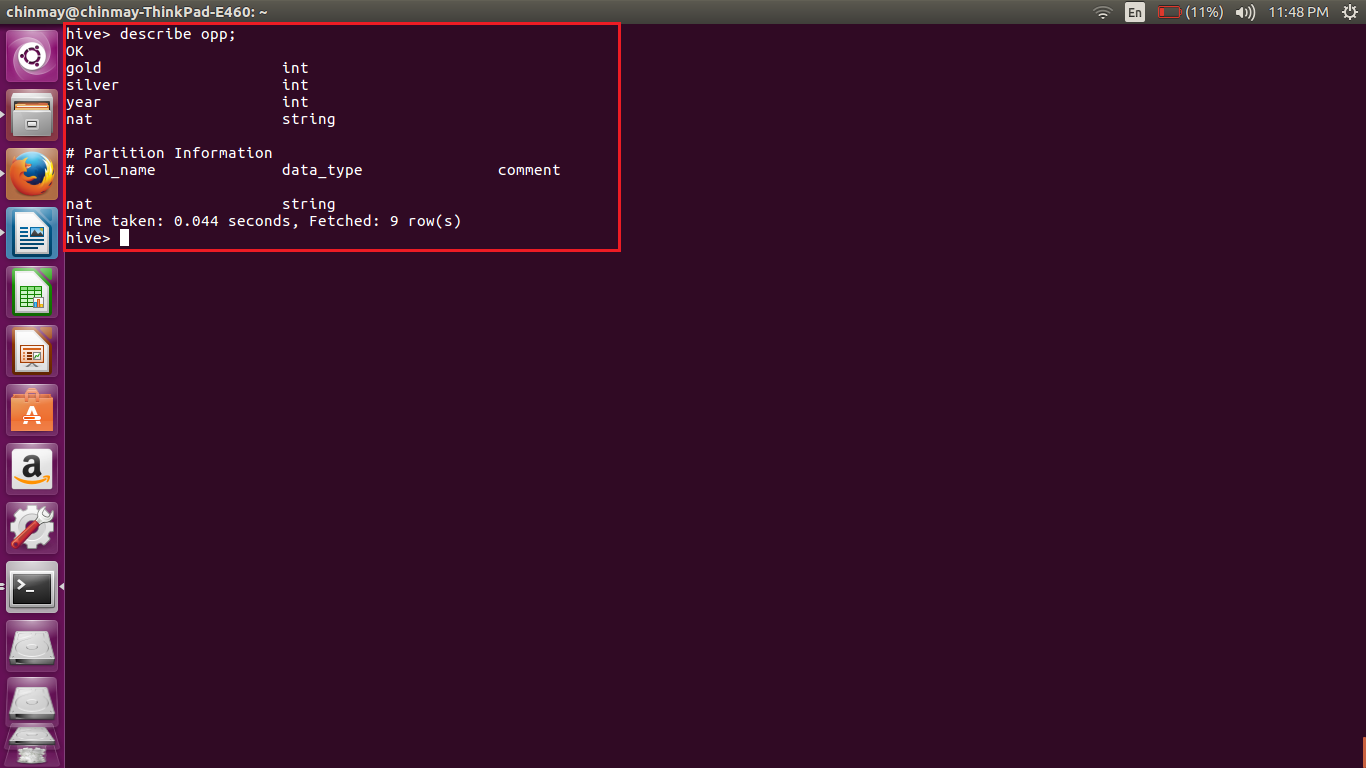


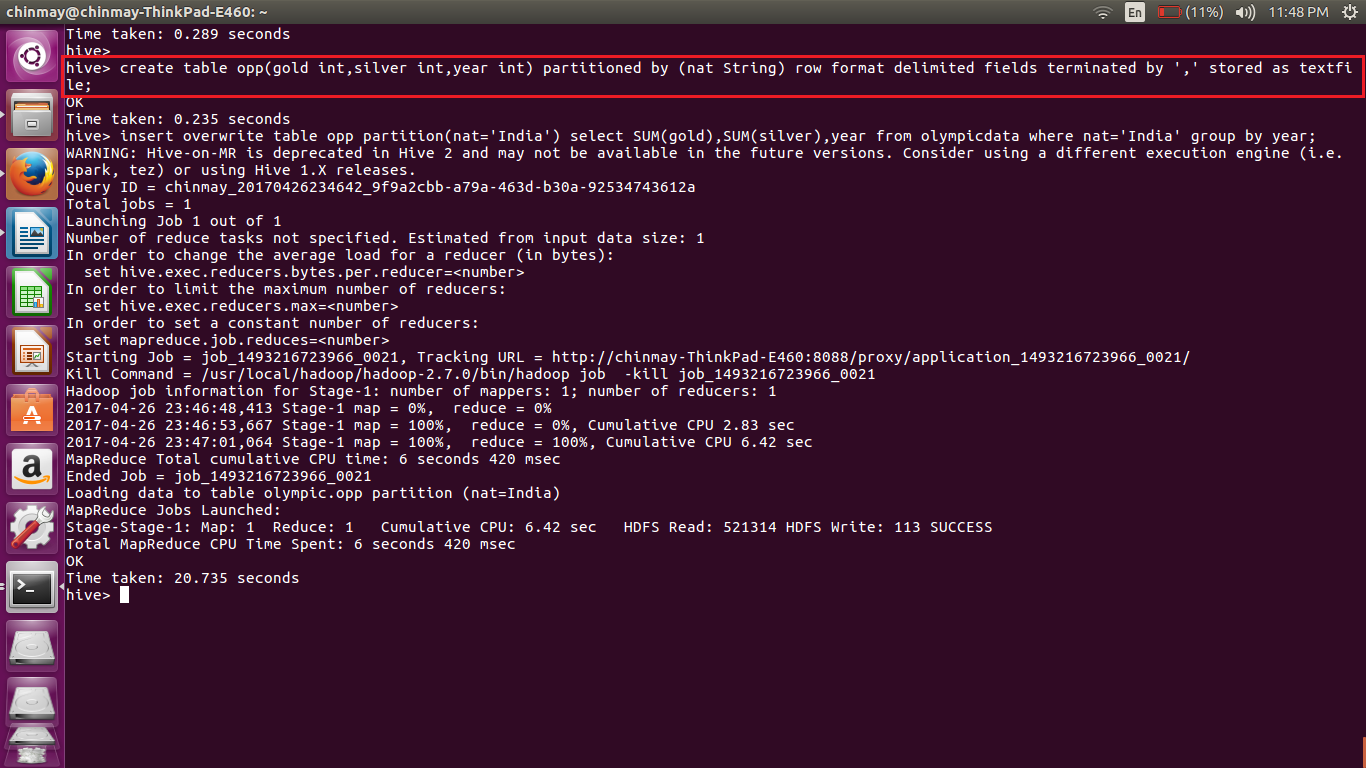
- Find the number of medals india won in swimming year wise

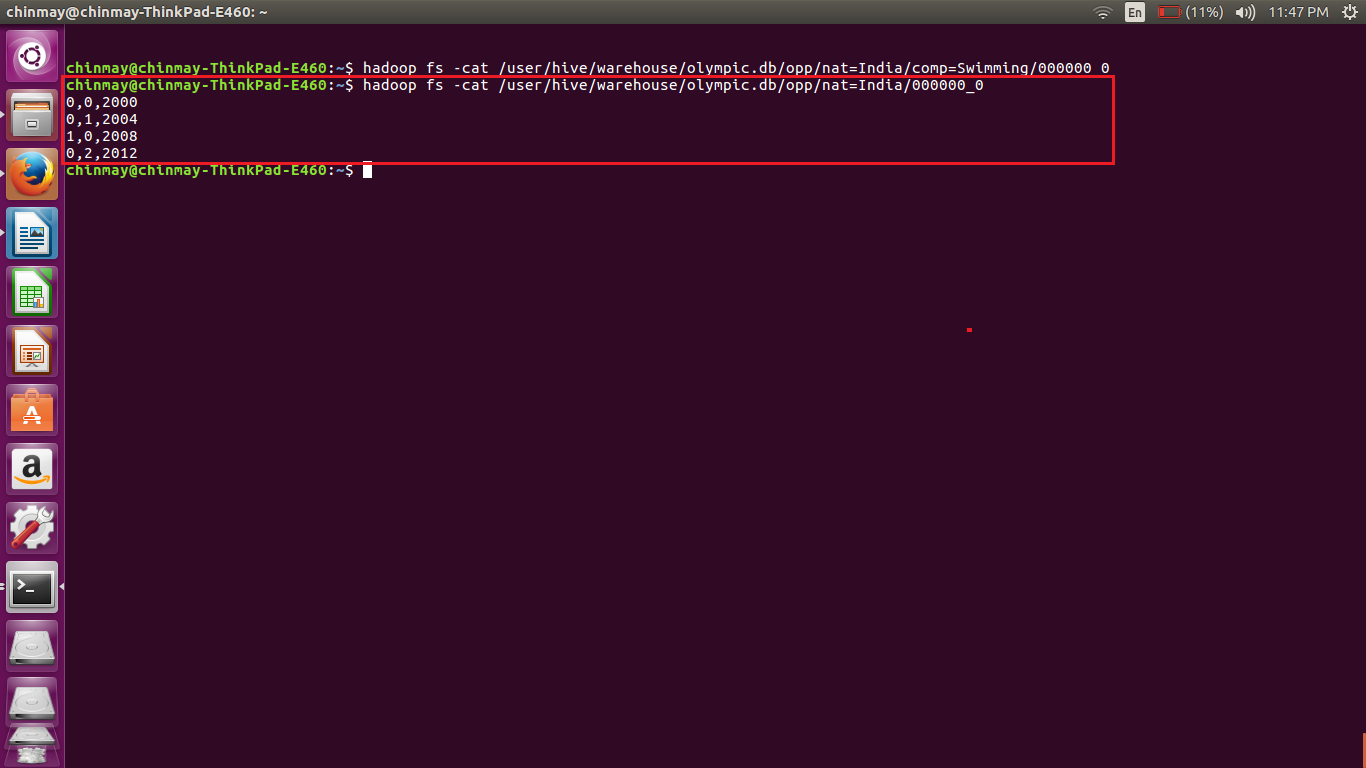




- Find the number of gold and silver medals india won year wise

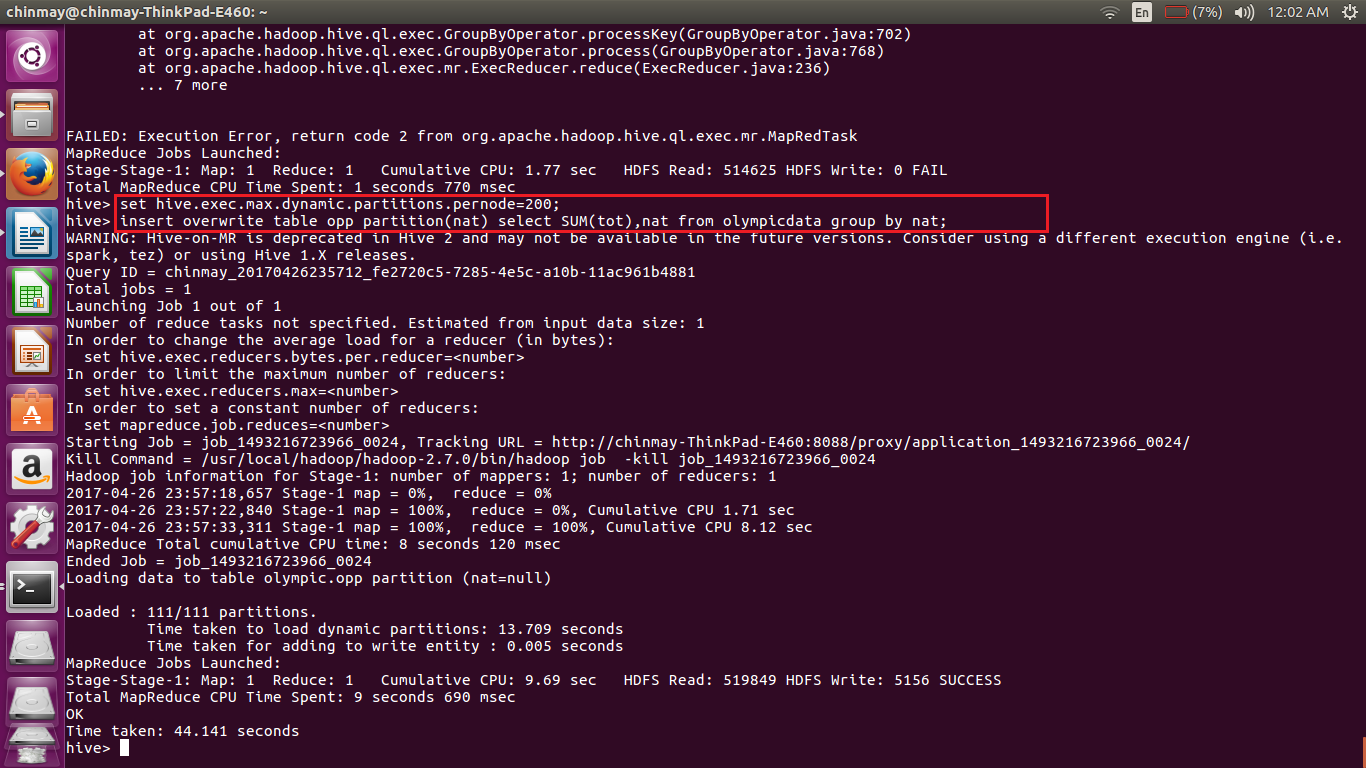


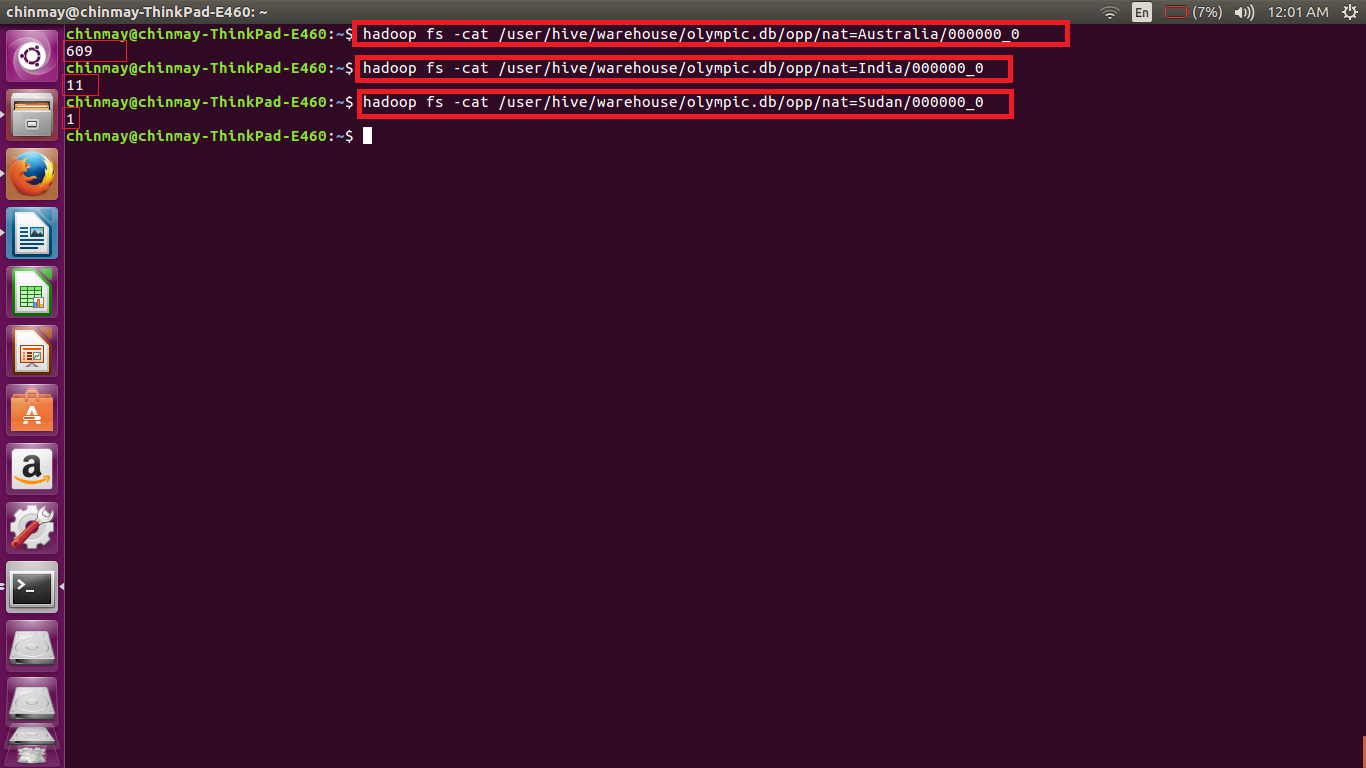


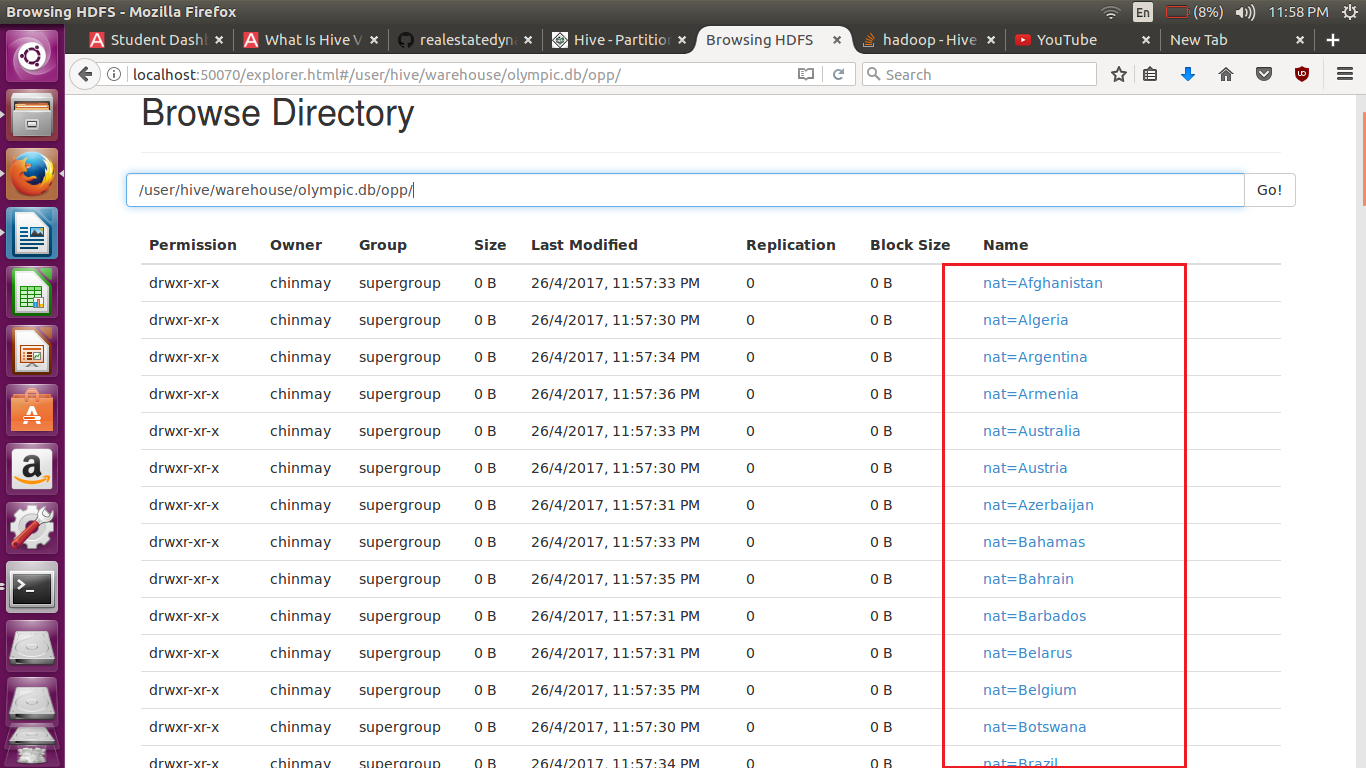


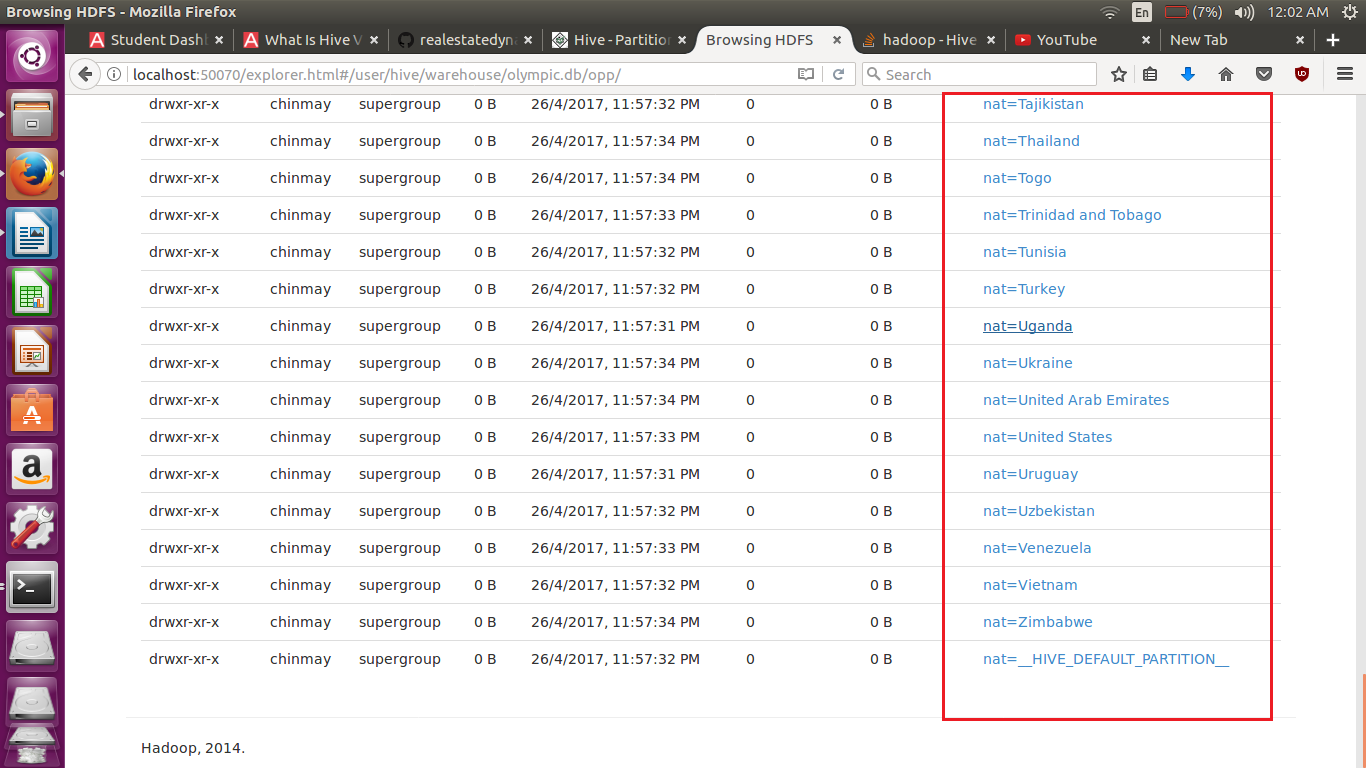
3. Use dynamic partitioning in hive and evaluate the below problem statements

- Find the total number of medals won by each country.

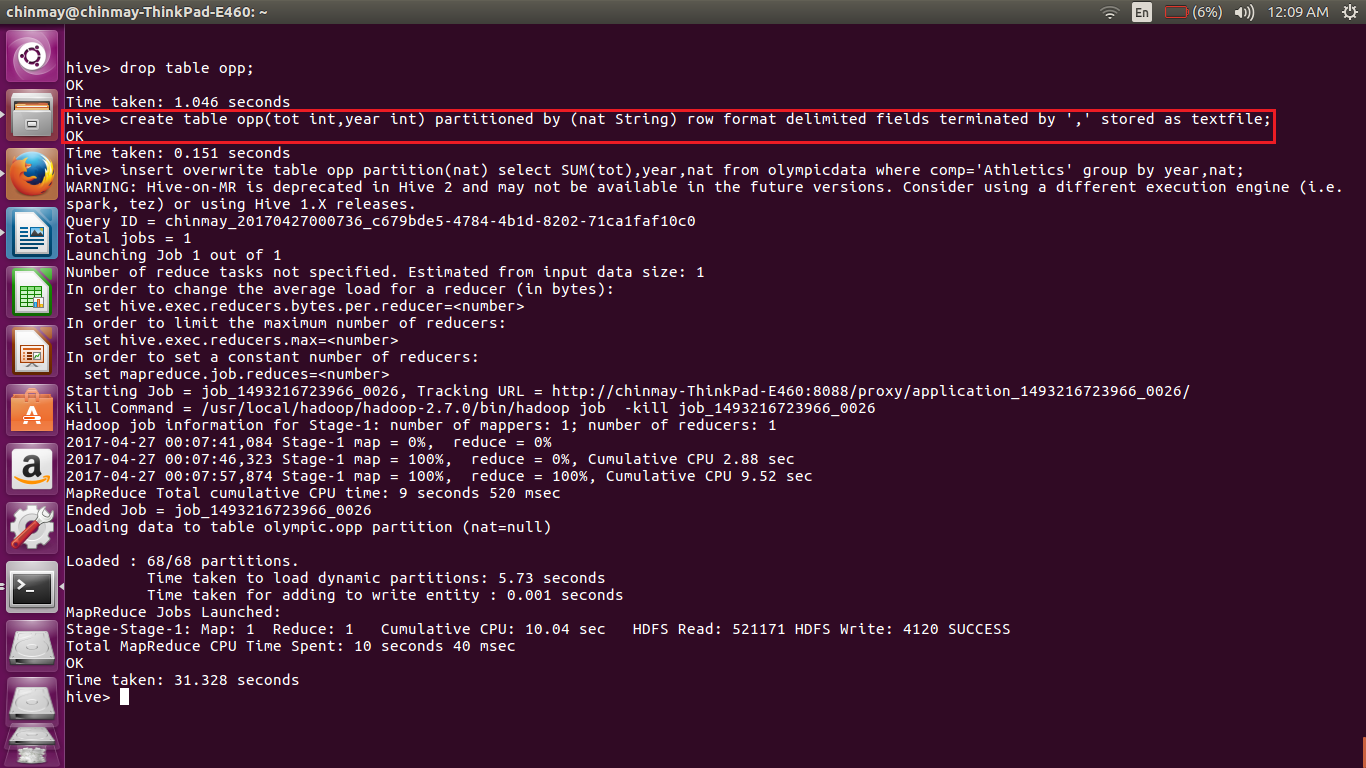


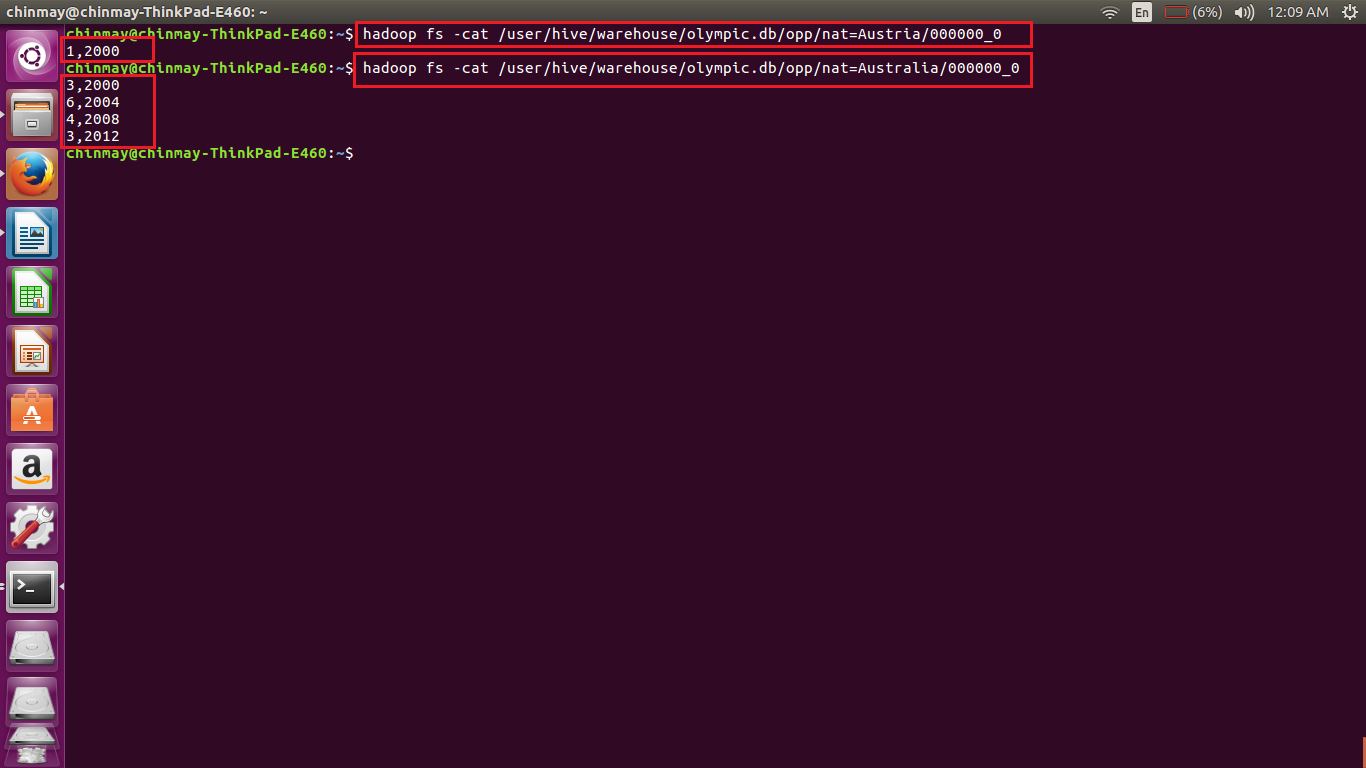


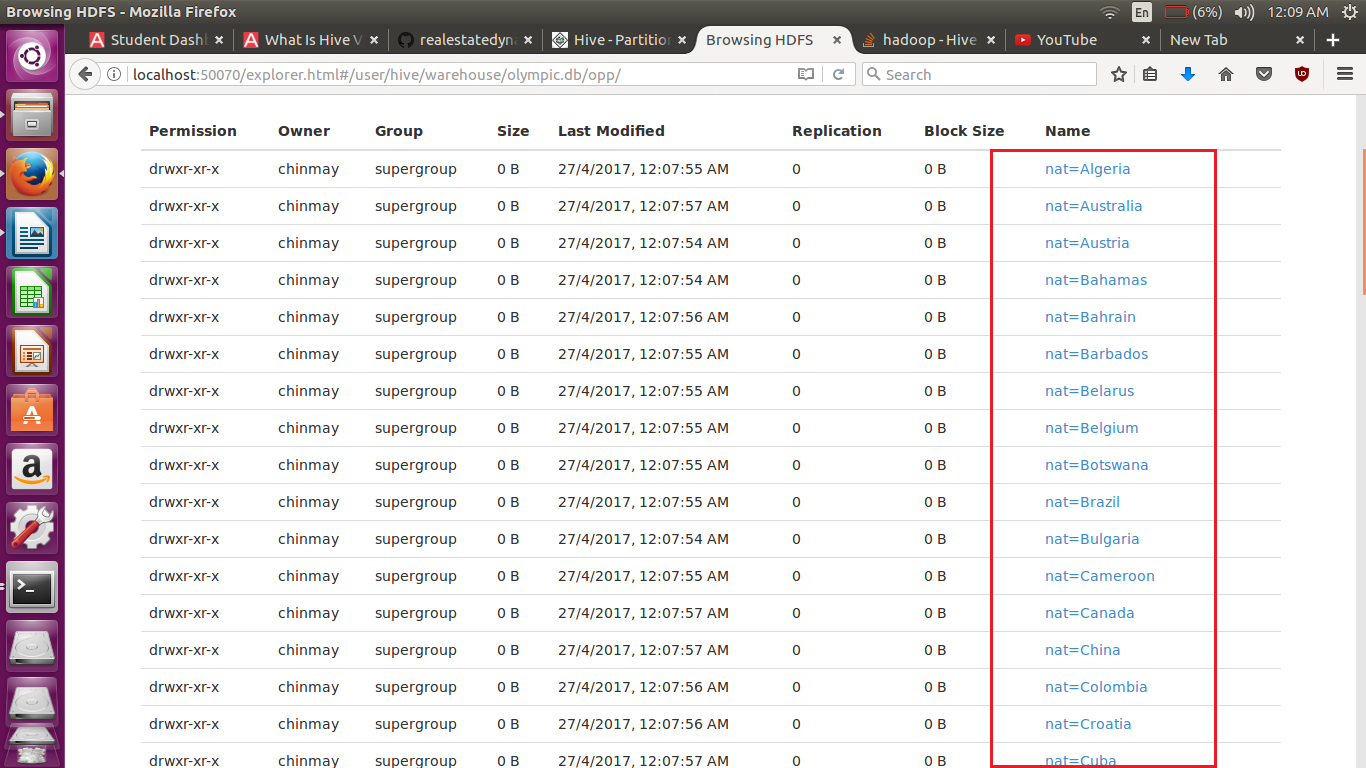




- Find the number of medals each country won in Athletics year wise







- Find the average age of atheltes participated from each country in olympics year wise

