**Assignment 26.6**

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

Hive has been one of the preferred tool for performing queries on large datasets, especially when full table scan is done on the datasets.

In the case of tables which are not partitioned, all the files in a table’s data directory is read and then filters are applied on it as a subsequent phase. This becomes a slow and expensive affair especially in cases of large tables.

Without partitioning Hive reads all the data in the directory and applies the query filters on it. This is slow and expensive since all data has to be read.

Classification of partitioning:

* Static partitioning
* Dynamic Partitioning

**When to use static partitioning**

Static partitioning needs to be applied when we know data (supposed to be inserted) belongs to which partition.

**When to use dynamic partitioning**

* In static partitioning, every partitioning needs to be backed with individual hive statement which is not feasible for large number of partitions as it will require writing of lot of hive statements. In that scenario dynamic partitioning is suggested as we can create as many number of partitions with single hive statement.

We can understand the procedure of static and dynamic partitioning by the below queries.

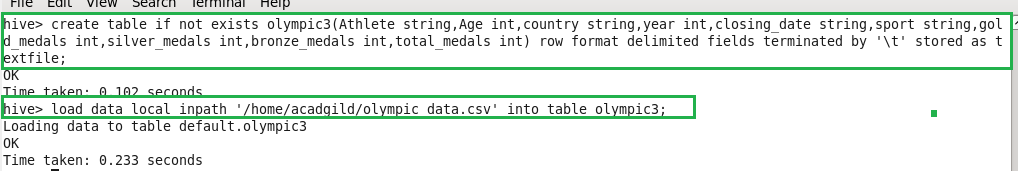
Static partitioning needs to be applied when we know data (supposed to be inserted) belongs to which partition.

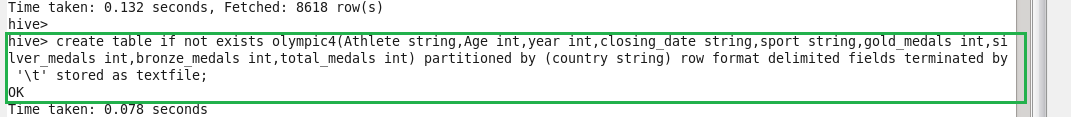
1. Use static partitioning in hive and evaluate the below problem statements –

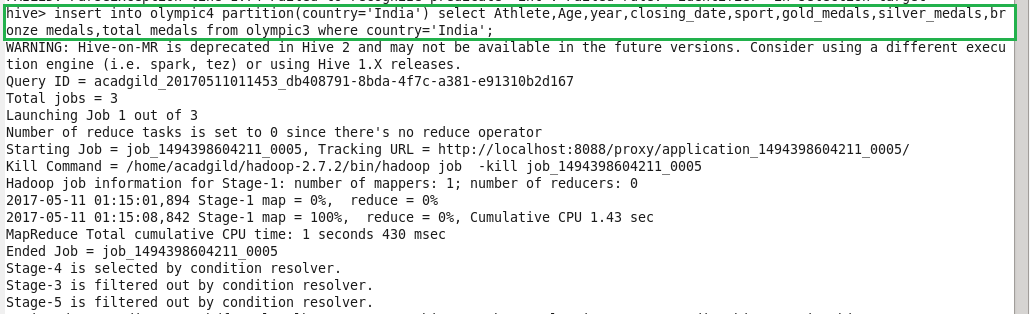
**Find the number of medals India won year wise –**

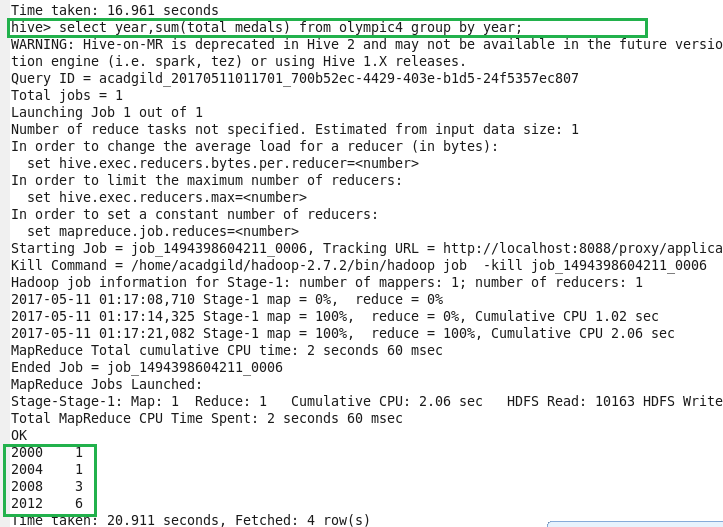
First, creating table and loading data into the table from the given dataset.

Then, creating table for static partitioning as olympic4 which is partitioned by country and then inserting the data into the partitioned table where country=’India’ and then according to the question giving command for finding the number of medals india won year wise using sum function with output.



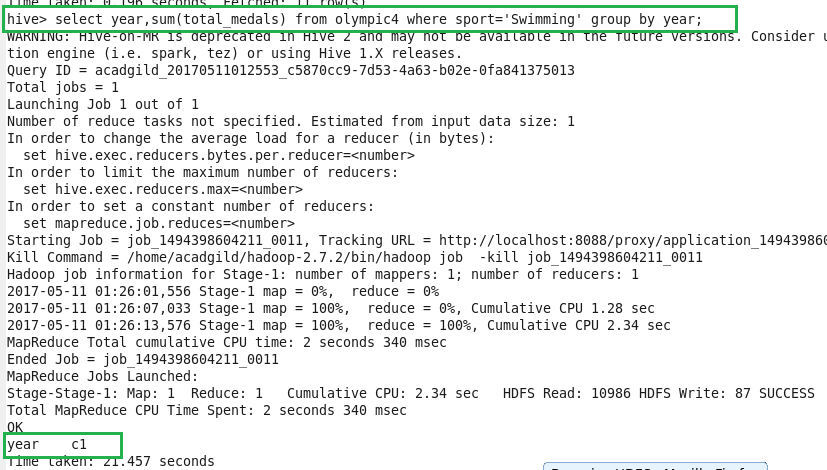






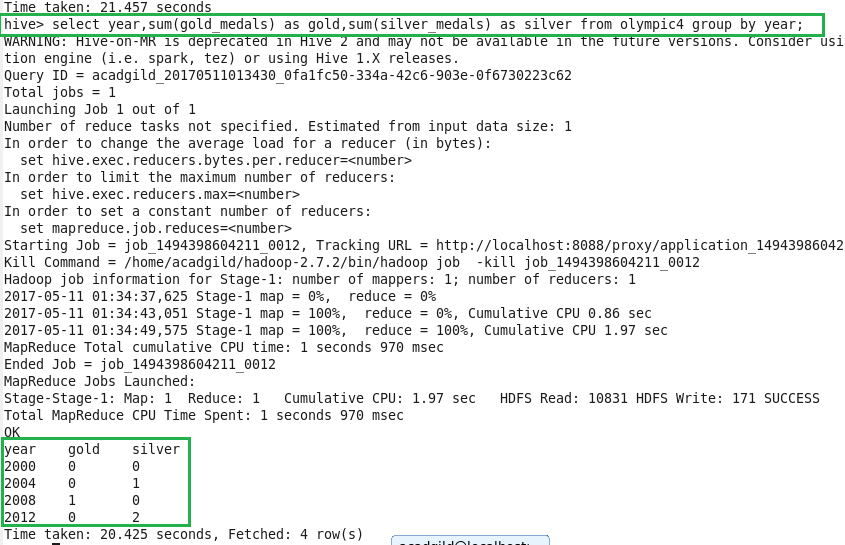
-**Find the number of medals india won in swimming year wise –**

Here, according to the question giving command for finding number of medals India won in swimming year wise using sum function and in where condition sport is giving as swimming with output.



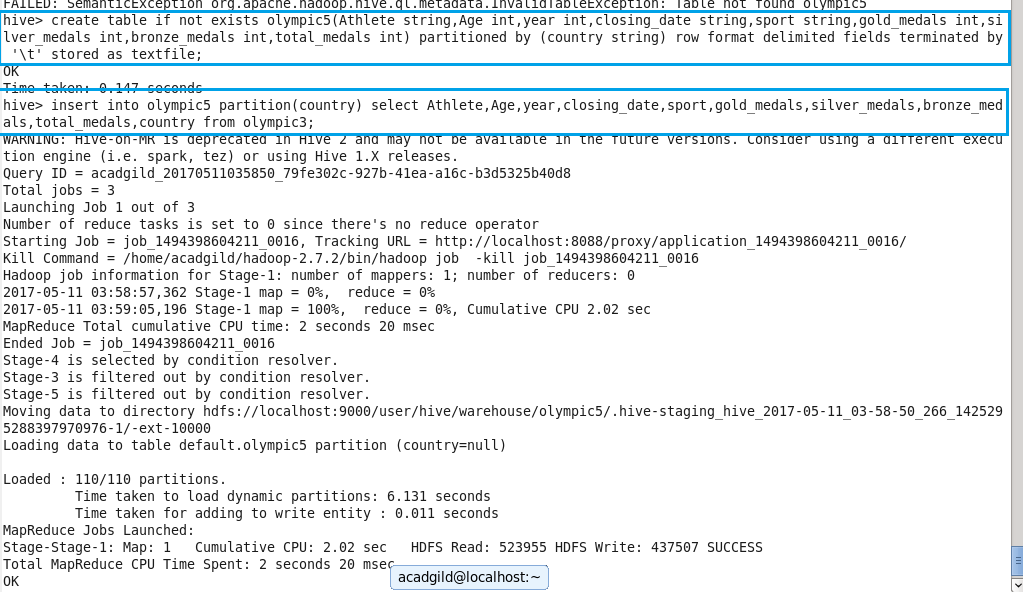
**Find the number of gold and silver medals india won year wise.**

Here, according to the question giving command for finding number of gold and silver medals India won year wise using sum function with output..

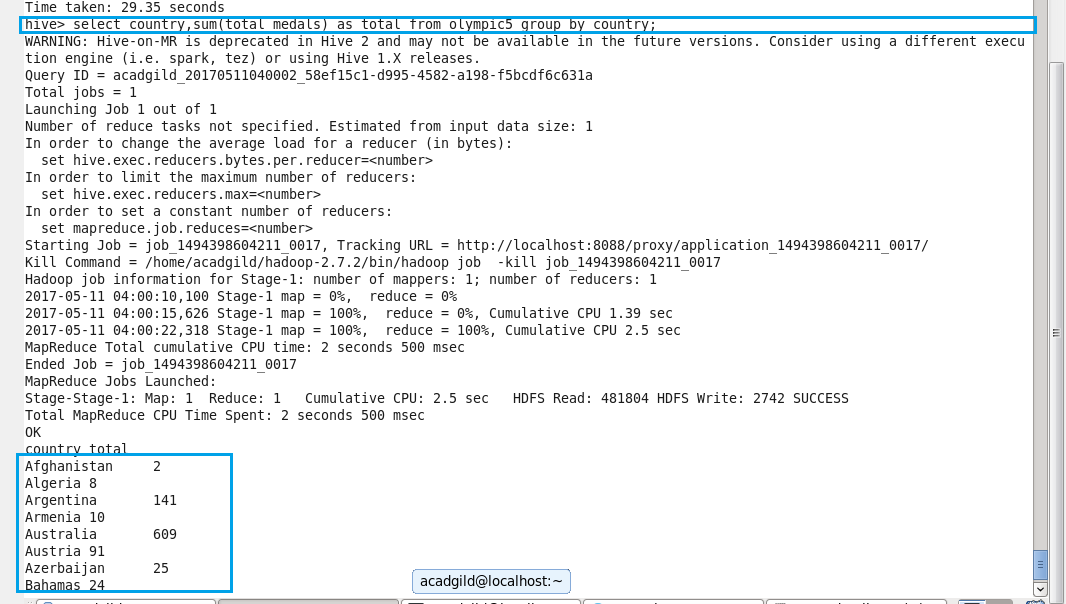


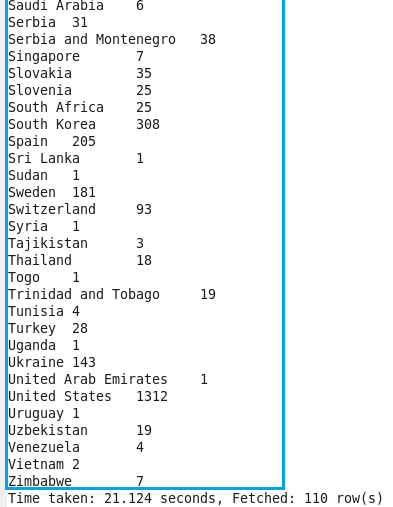
1. Use dynamic partitioning in hive and evaluate the below problem statements –

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

Here, table needs to be created with all the columns including country as olympic5 which is partitioned by country and then inserting the data into the partitioned table olympic5.

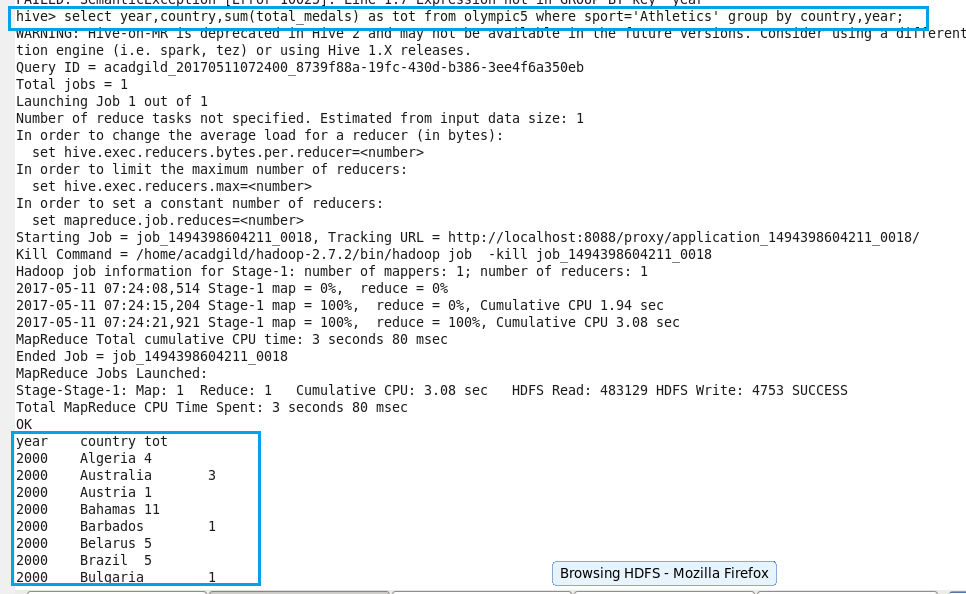
Then, according to the question giving command for finding the number of medals won by each country using sum function along with output.





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

According to the question giving command for finding the number of medals each country won in Athletics year wise using sum function and sport name is given as athletics with output.





**- Find the average age of athletes participated from each country in olympics year wise.**

According to the question giving command for finding the average age of athletes participated from each country in olympics year wise using avg function with output.

