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**

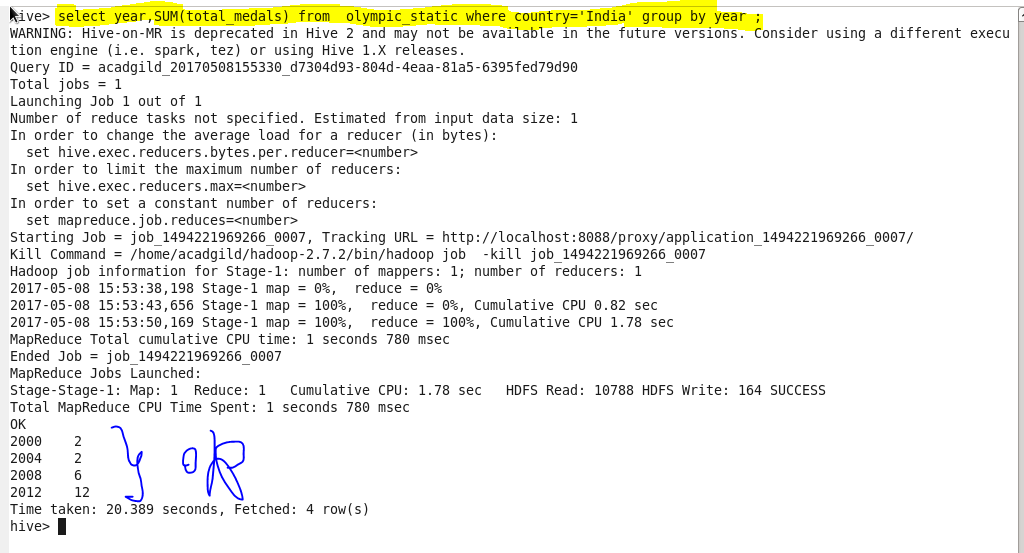
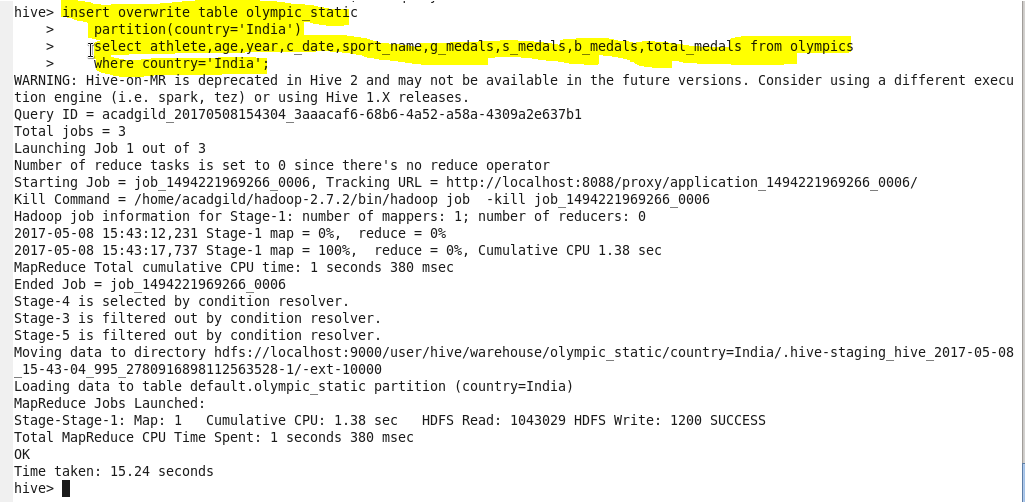
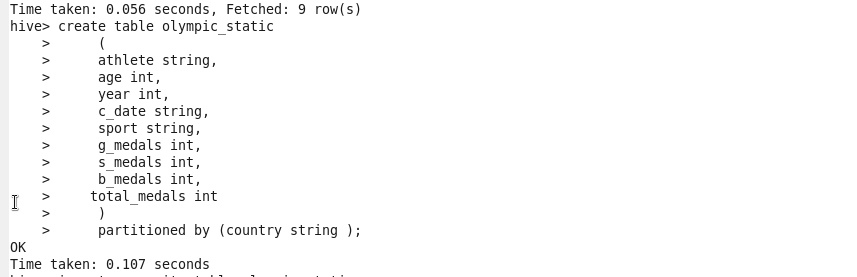
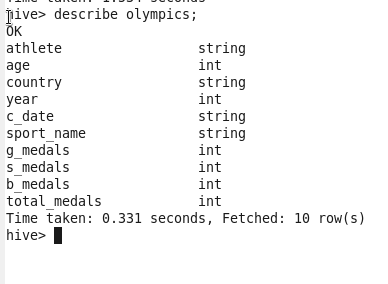
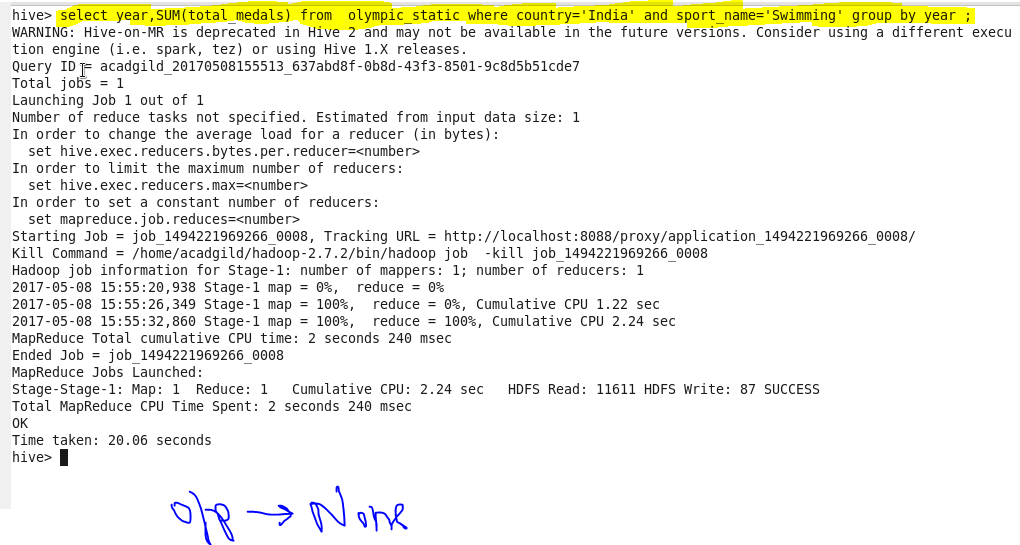
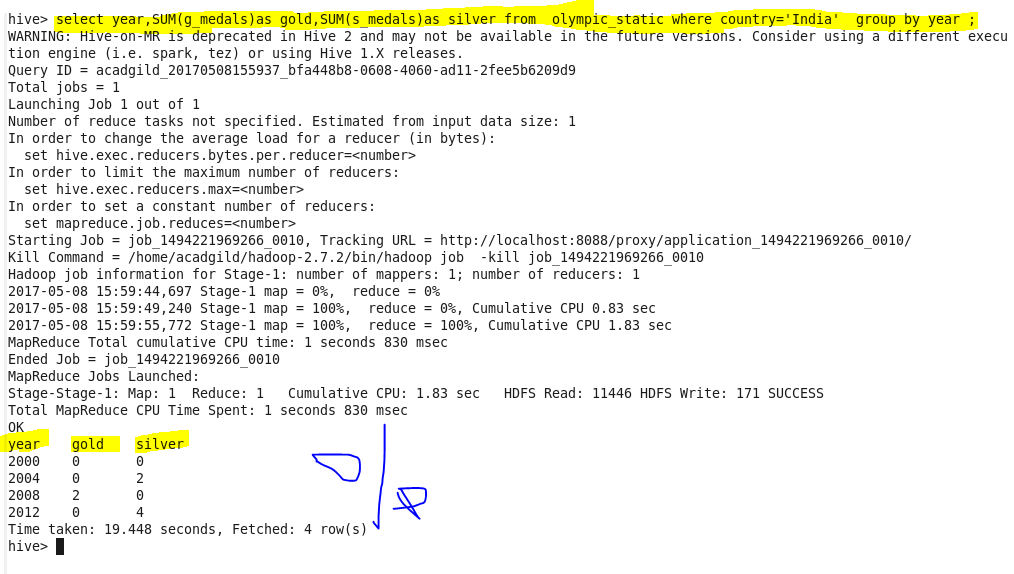
Static partitioning needs to be applied when we know data (supposed to be inserted) belongs to which partition.

**When to use dynamic partitioning**

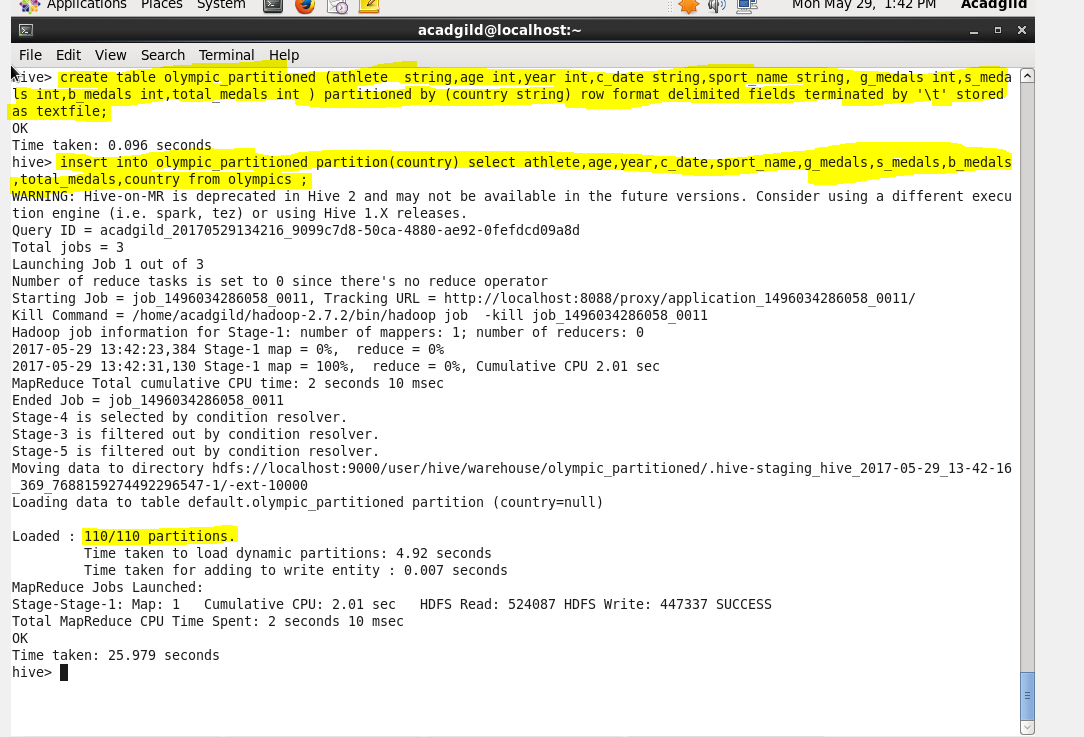
In static partitioning, every partitioning needs to be backed with individual hive statement which is not feasible for large number of partitions as it will require writing of lot of hive statements. In that scenario dynamic partitioning is suggested as we can create as many number of partitions

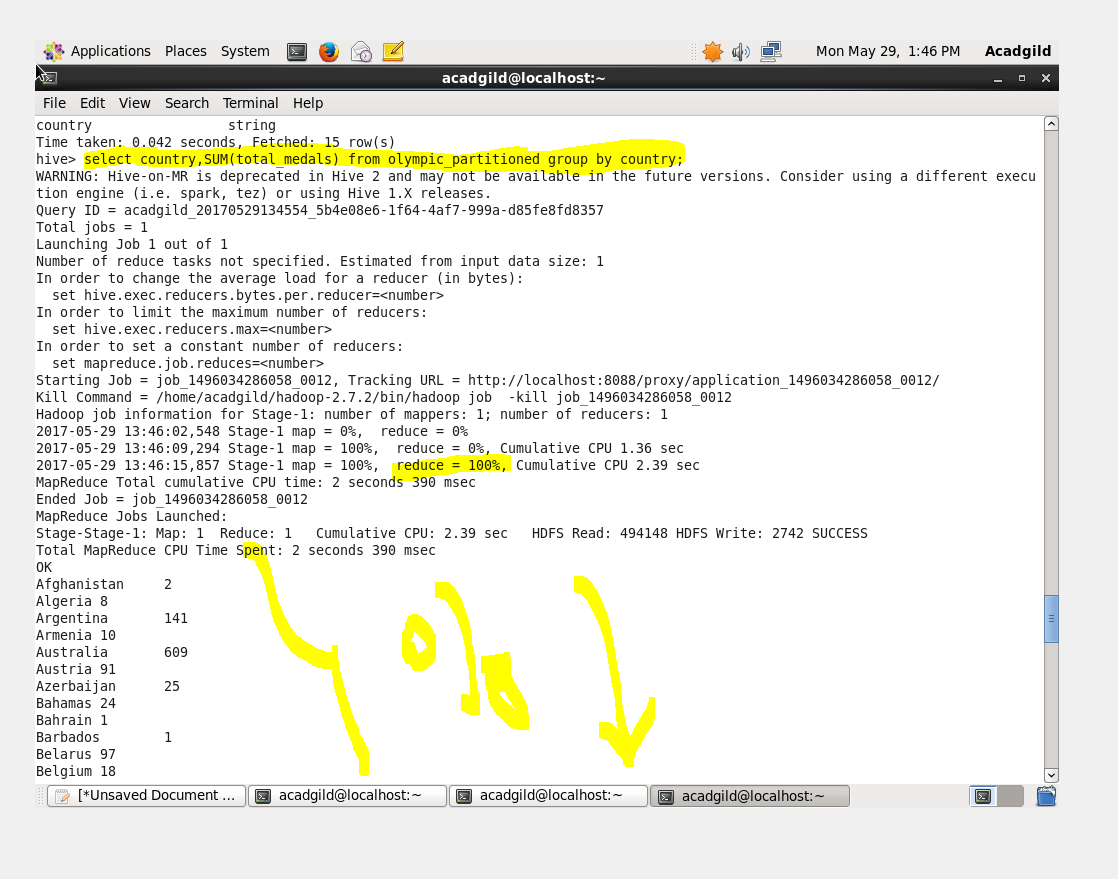
**2. Use static partitioning in hive and evaluate the below problem statements**

**- Find the number of medals india won year wise**

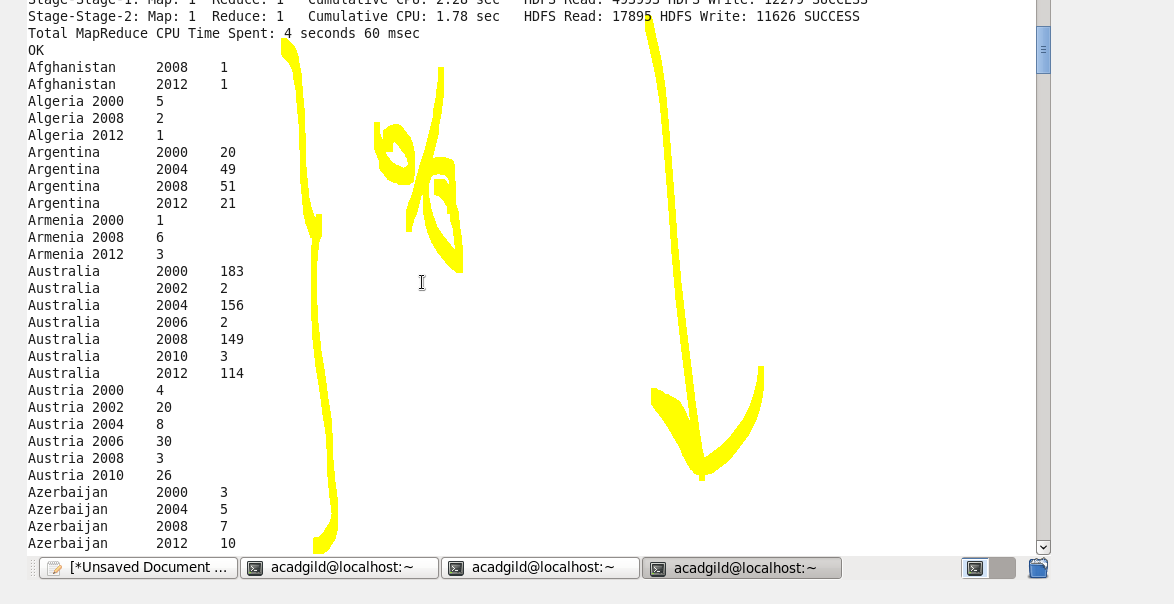
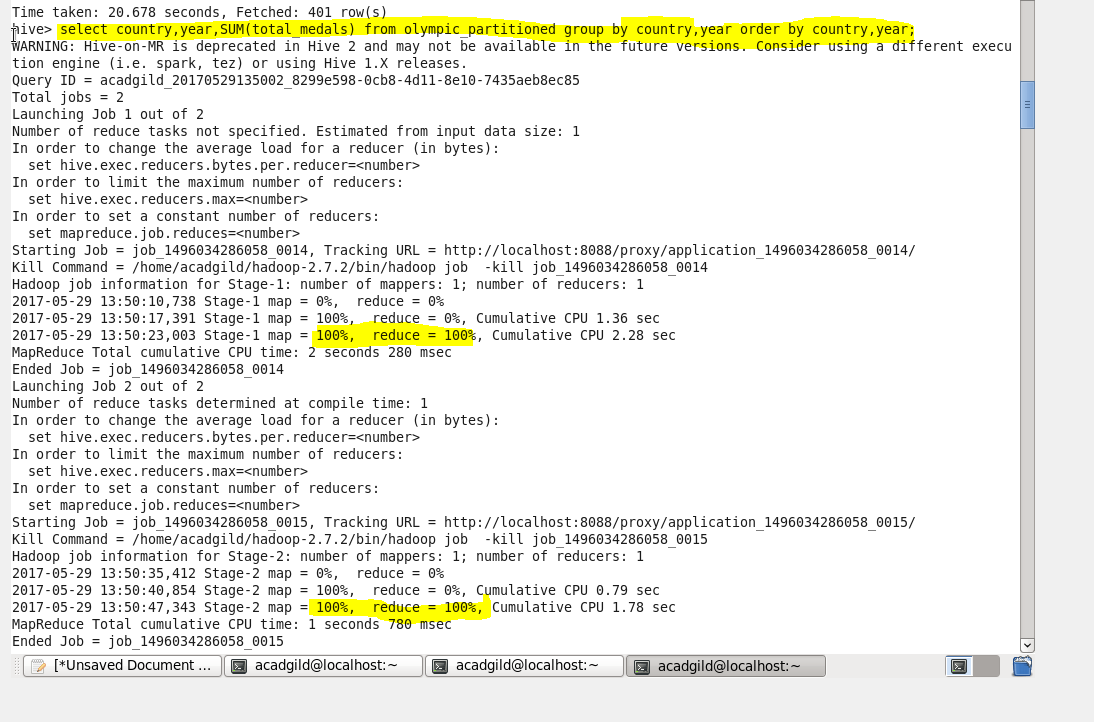
* 
* - Find the number of medals india won in swimming year wise  Find the number of gold and silver medals india won year wise
* 

1. 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 