**Session 26**

**Assignment 6**

**Problem Statement:**

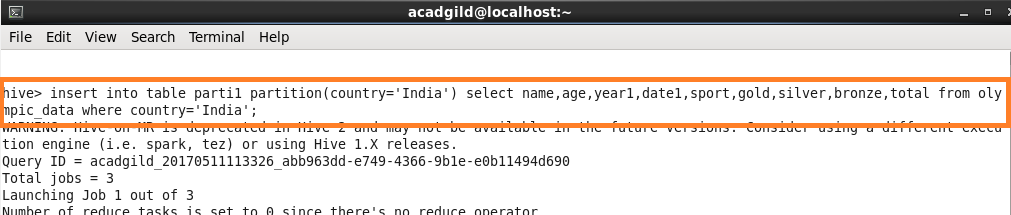
**Part 1:**

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

**Static Partition:**

* Static partitioning needs to be applied when we know data (supposed to be inserted) belongs to which partition.
* In static partitioning we need to specify the partition column value in each and every LOAD statement.
* If the data in the columns is known very well, then static partitioning can be used.
* Insert input data files individually into a partition table is Static Partition.
* Usually when loading files (big files) into Hive tables static partitions are preferred.
* Static Partition saves your time in loading data compared to dynamic partition.
* The user can “statically” add a partition in table and move the file into the partition of the table.
* The partition in static partition can be altered.
* The partition column value can get to form the filename, day of date etc. without reading the whole big file.
* If you want to use Static partition in hive the property should be set as,
* set hive.mapred.mode = strict;
* This property set by default in hive-site.xml.
* Static partition is in Strict Mode.
* The where clause could be used to use limit in static partition.
* Static partition can be performed on Hive Manage table or external table.

During inserting we should specify the value of the column based on which we have partitioned.

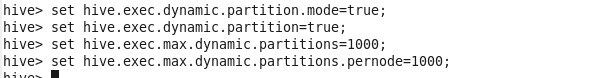


**Dynamic Partition:**

In Dynamic partitioning every row of the data is read and data is partitioned through a MR Job into the destination tables depending on certain field in file.

* Single insert to partition table is known as dynamic partition.
* Usually dynamic partition load the data from non-partitioned table.
* Dynamic Partition takes more time in loading data compared to Static Partition.
* When the data stored in a table is large then Dynamic partition is suitable.
* If number of column should be partitioned but the number of columns is not known then also dynamic partition is suitable.
* Dynamic partition there is no required where clause to use limit.
* The alter option cannot be performed on Dynamic partition.
* The dynamic partition can be performed on hive external table and managed table
* If you want to use Dynamic partition in hive then mode is in non-strict mode.
* Here is hive dynamic partition properties you should allow
* SET hive.exec.dynamic.partition = true;
* SET hive.exec.dynamic.partition.mode = nonstrict;

For dynamic partition we need to set some commands in shell. They are,





During inserting in the partitioned table, the value of the column can be or not be specified based on which the table should be partitioned.

**Part 2:**

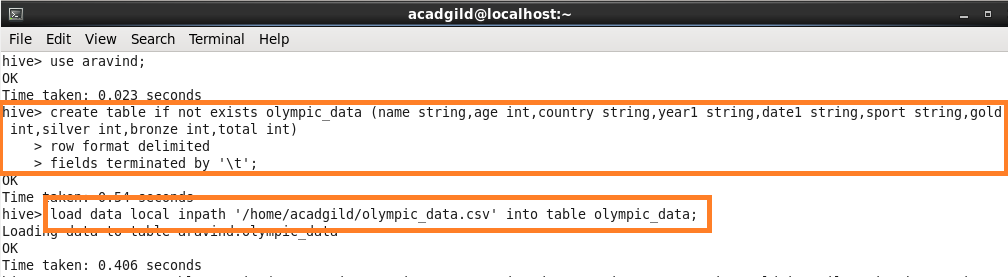
Use static partitioning in hive and evaluate the below problem statements

- Find the number of medals India won year wise

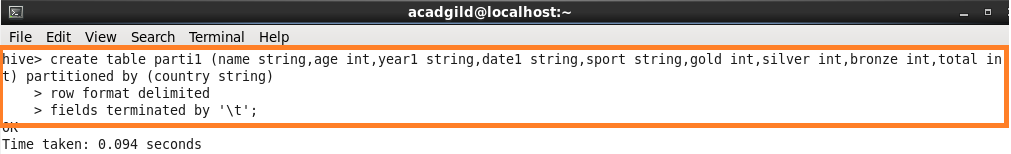
- Find the number of medals India won in swimming year wise

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

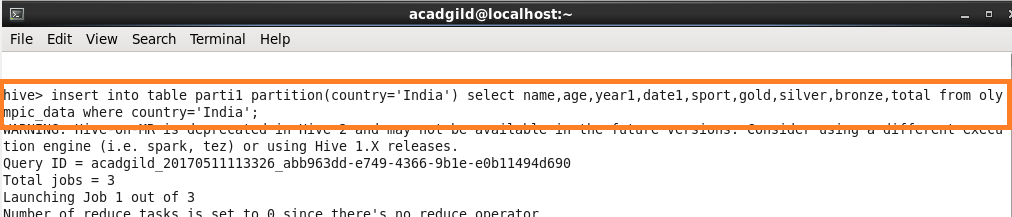
**Step 1:** Creating the table and loading data into it.



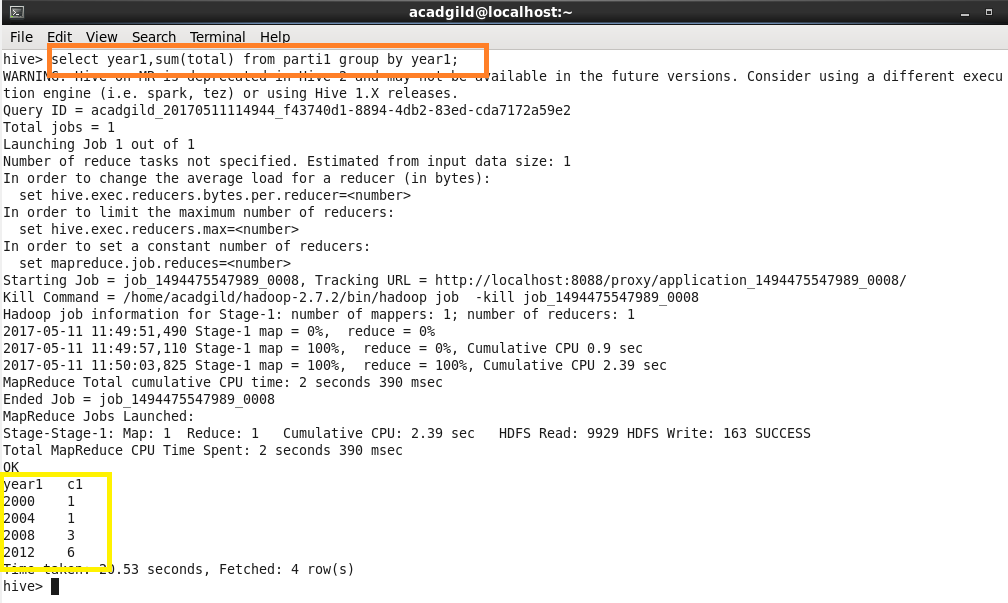
**Step 2:** Creating partitioned table by setting country as partitioning column.



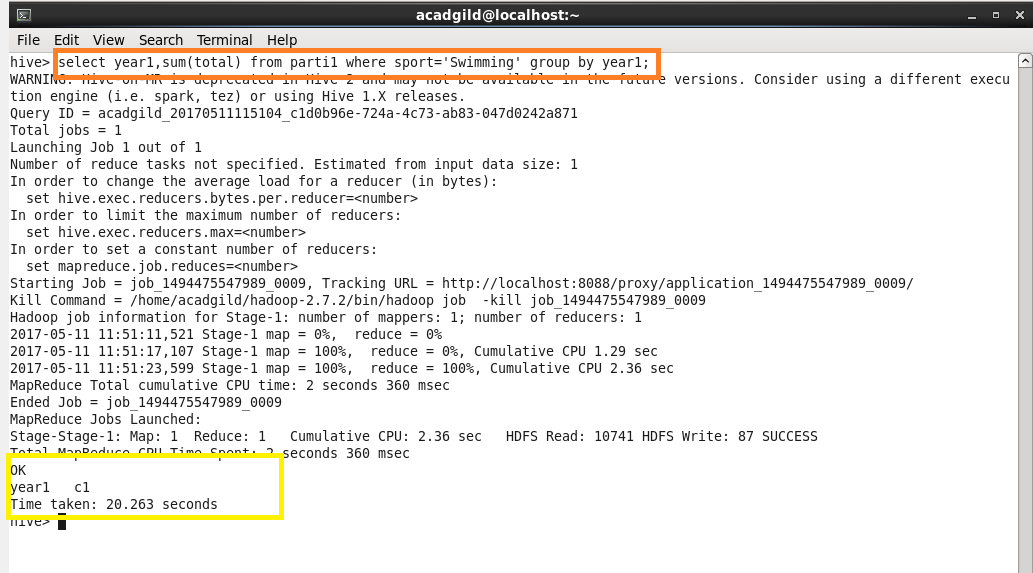
**Step 3:** Inserting into the partitioned table by copying the data from olympic\_data table and setting country = ’India’.



**Step 4:** Find the number of medals India won year wise.

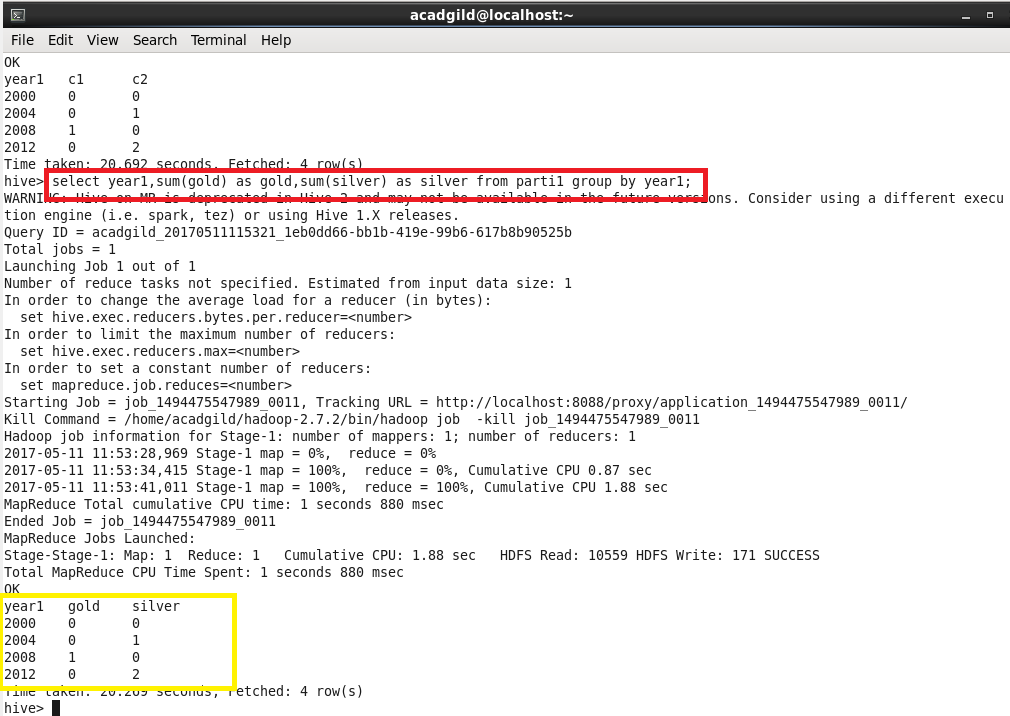


**Step 5:** Find the number of medals India won in swimming year wise.



**NOTE:** No one has got medal in swimming from India so the answer is null.

**Step 6:** Find the number of gold and silver medals India won year wise.

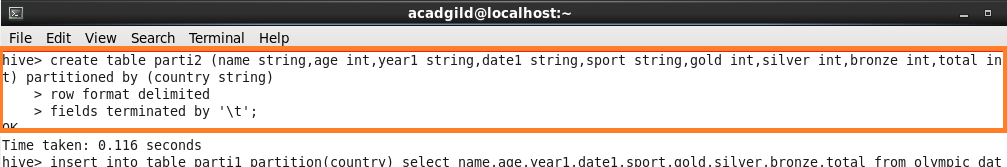


**Part 3:**

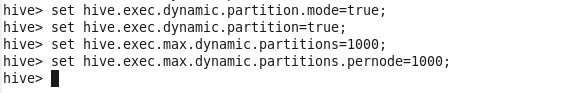
Use dynamic partitioning in hive and evaluate the below problem statements.

* 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 athletes participated from each country in olympics year wise.

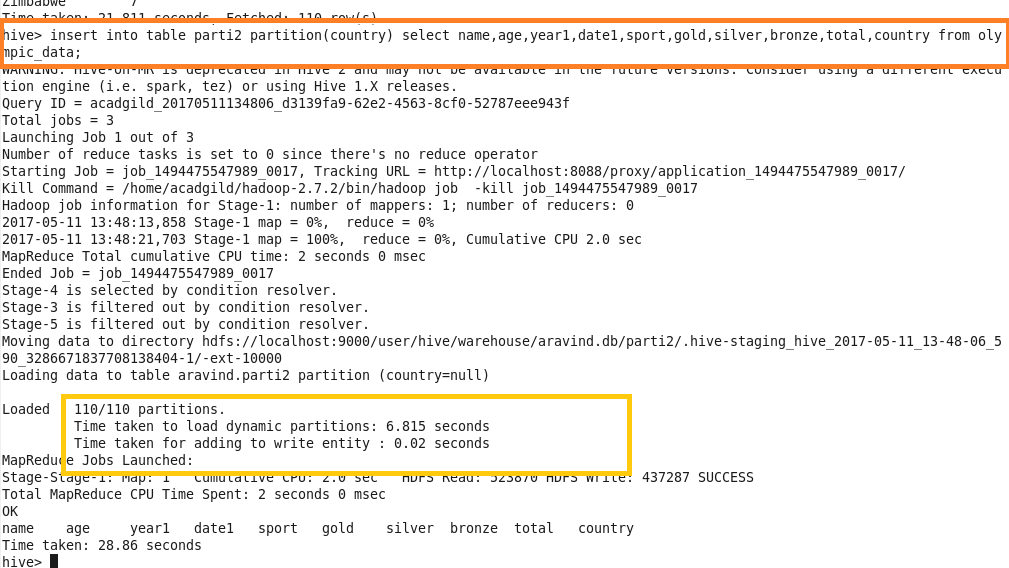
**Step 1:** Creating a table and partitioning by country.



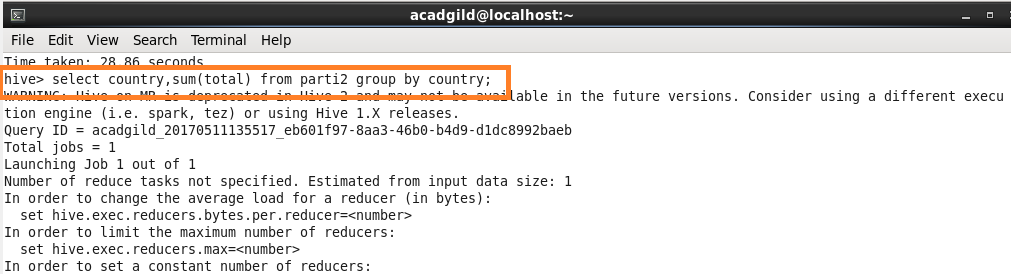
**Step 2:** Setting some configurations in shell for running dynamic configurations.



**Step 3:** Inserting into the partitioned table that I have created already.

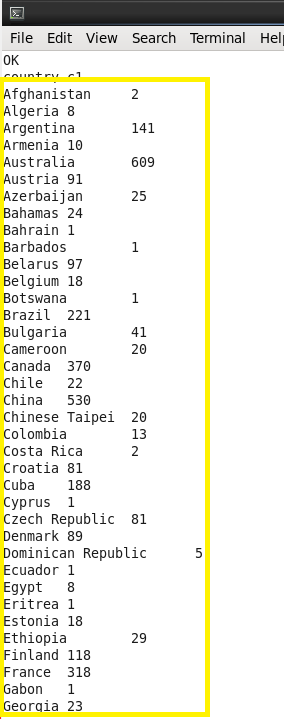
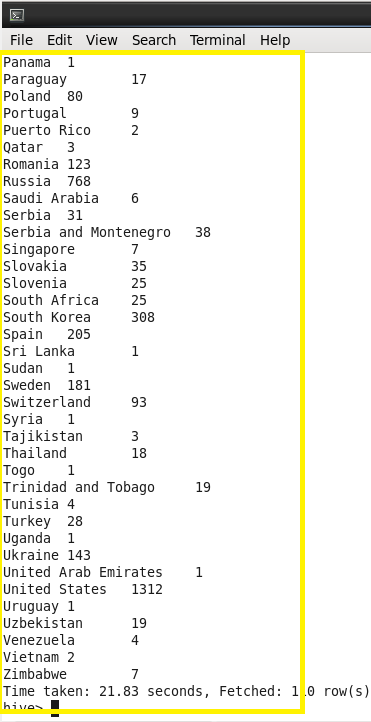


**Step 4:** Find the total number of medals won by each country.

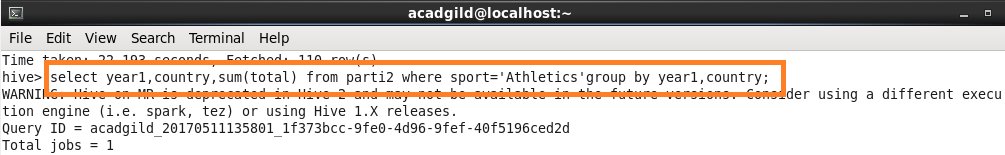


**Output:**

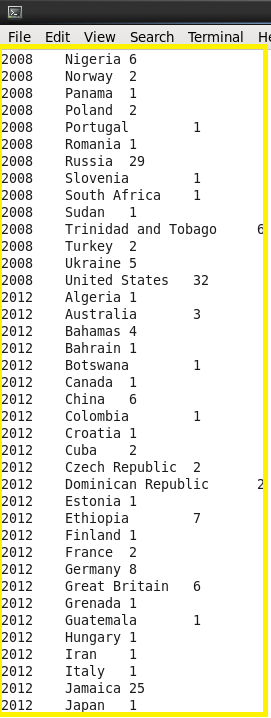
Some random screenshots of the output is attached since complete output cannot be included.

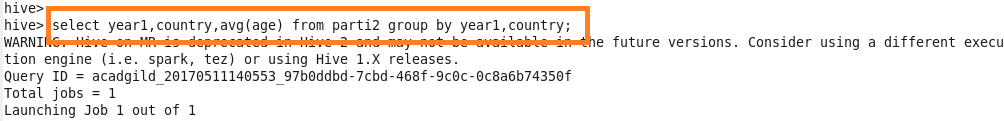
**Step 5:** Find the number of medals each country won in Athletics year wise.



**Output:**

**Step 6:** Find the average age of athletes participated from each country in olympics year wise.



**Output:**

