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

**STATIC PARTITION:**

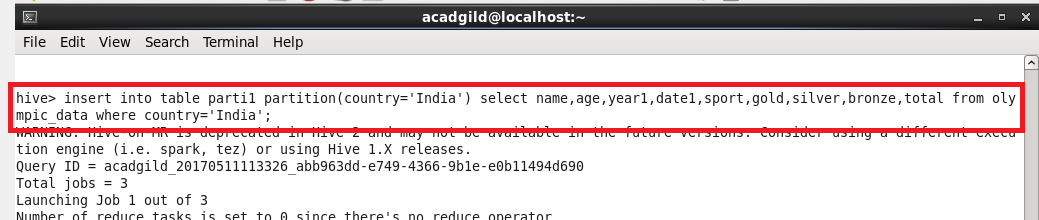
Static partitioning needs to be applied when we know data(supposed to be inserted) belongs to which partition.

In static partitioning we need to specify the partition column value in each and every LOAD statement.

If we know the data in the columns very well, then we can go for static partitioning.

* Insert input data files individually into a partition table is Static Partition
* Usually when loading files (big files) into [Hive tables](http://www.hadooptpoint.com/hive-create-table-examples/) static partitions are preferred
* Static Partition saves your time in loading data compared to dynamic partition
* You “statically” add a partition in table and move the file into the partition of the table.
* We can alter the partition in static partition.
* You can get the partition column value form the filename, day of date etc., without reading the whole big file.
* If you want to use Static partition in hive you should set property **set hive.mapred.mode = strict** This property set by default in hive-site.xml.
* Static partition is in Strict Mode.
* You should use where clause to use limit in static partition.
* You can perform Static partition on Hive Manage table or external table.
* 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.

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



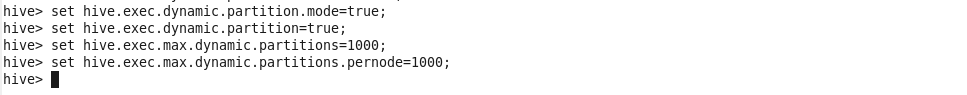
**DYNAMIC PARTITION:**

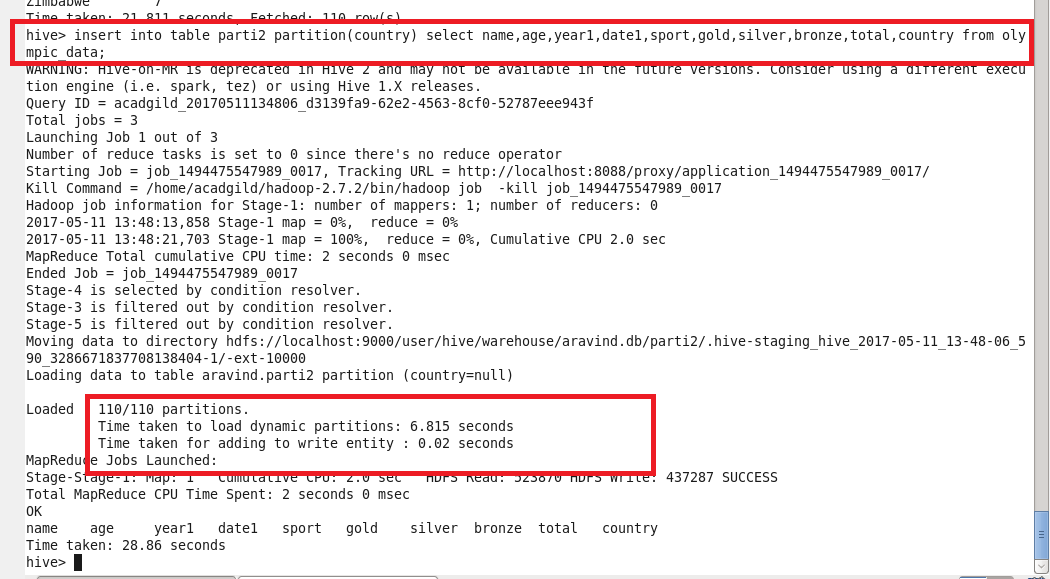
**In Dynamic partitioning** every row of the data is read and data is partitioned through a MR job into the destination tables depending on certain field in file.

* 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 you have large data stored in a table then Dynamic partition is suitable.
* If you want to partition number of column but you don’t know how many columns then also dynamic partition is suitable
* Dynamic partition there is no required where clause to use limit.
* we can’t perform alter on Dynamic partition
* You can perform dynamic partition on hive external table and managed table
* If you want to use Dynamic partition in hive then mode is in nonstrict mode

For dynamic partition we need to set some commands in shell

They are,





During inserting in the partitioned table we need not to specify the value of the column based on which we have partitioned .

**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**

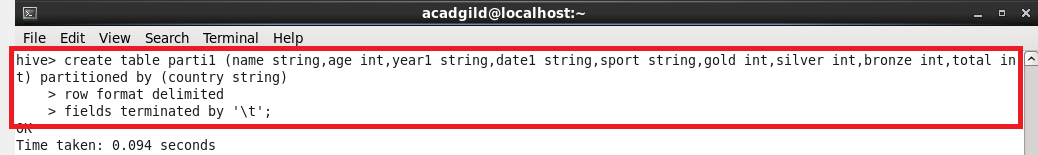
**- Find the number of gold and silver medals india won year wise**

1. **Find the number of medals india won year wise**

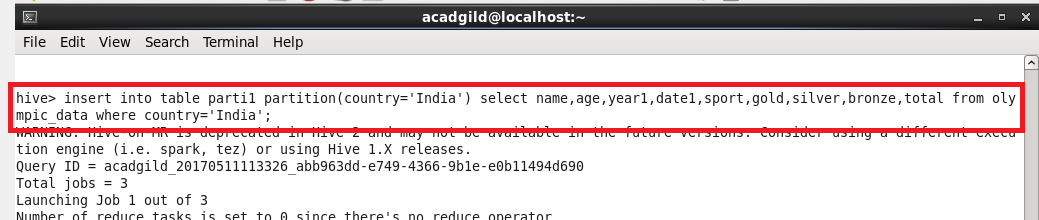
Step 1: creating the table and loading data into it.



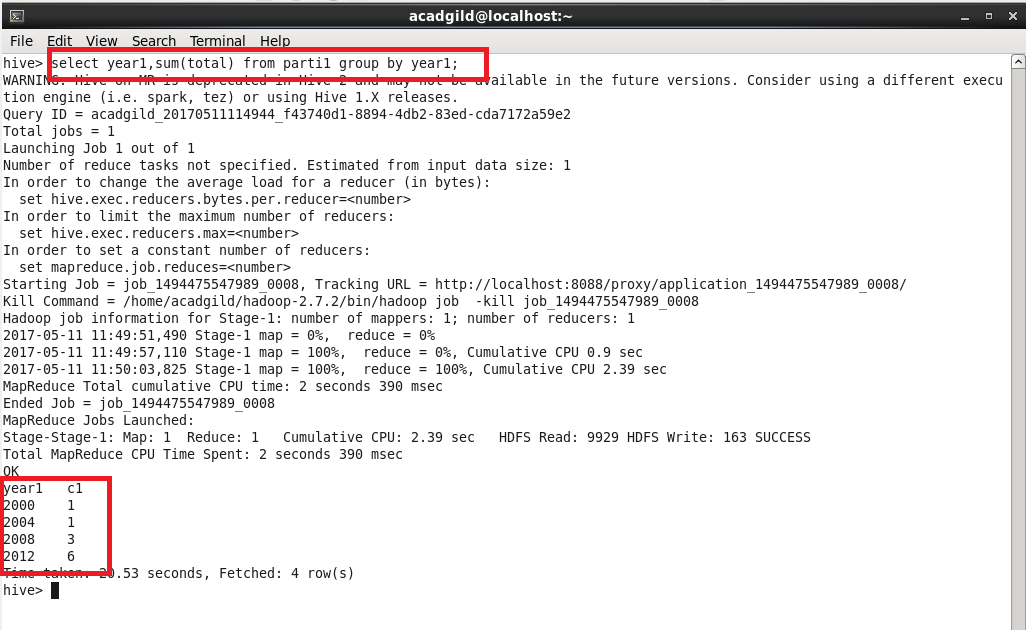
Step 2: creating partitioned table by setting country as partitioning column.



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

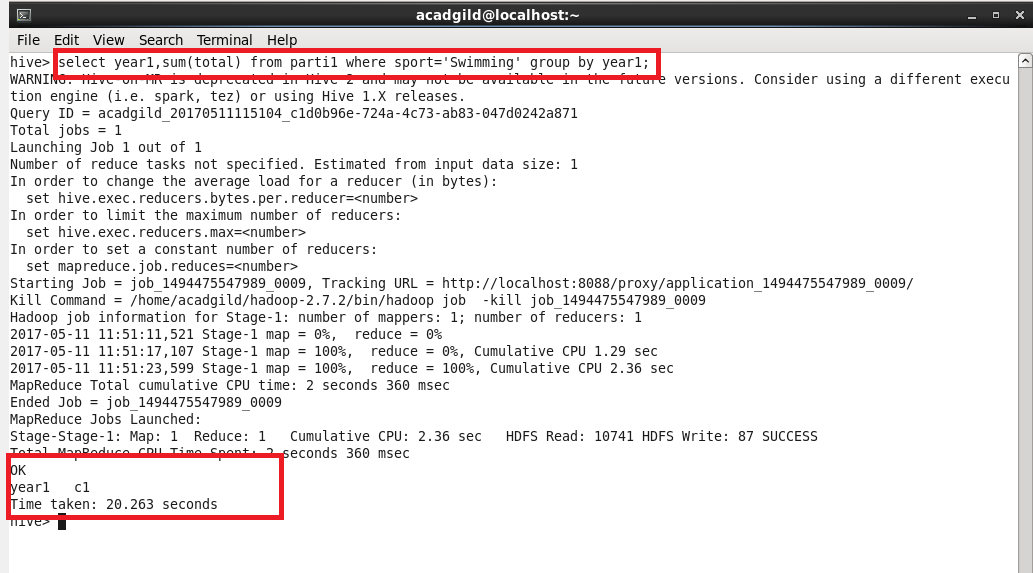


Step 4: QUERY



1. **Find the number of medals india won in swimming year wise**

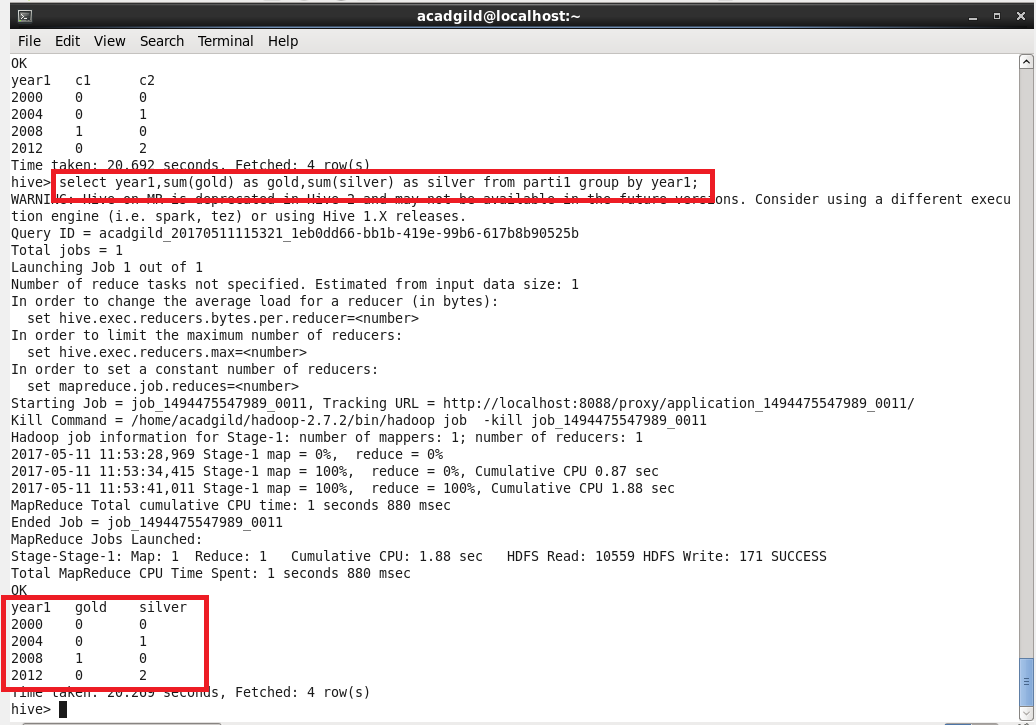
***Sum(total) and sport = swimming will give the sum of total medals won in swimming***



No one has got medal in swiming from india so the answer is null

**-3)** **Find the number of gold and silver medals india won year wise**

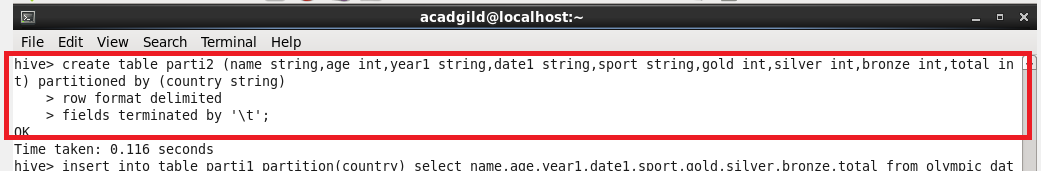
***Sum(gold), sum(silver) will give the sum of gold and silver medals won so far***



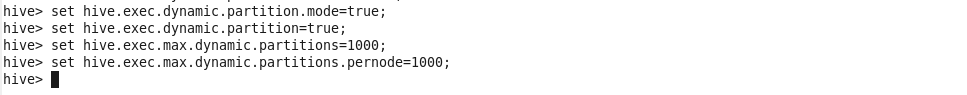
**DYNAMIC PARTITIONING:**

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

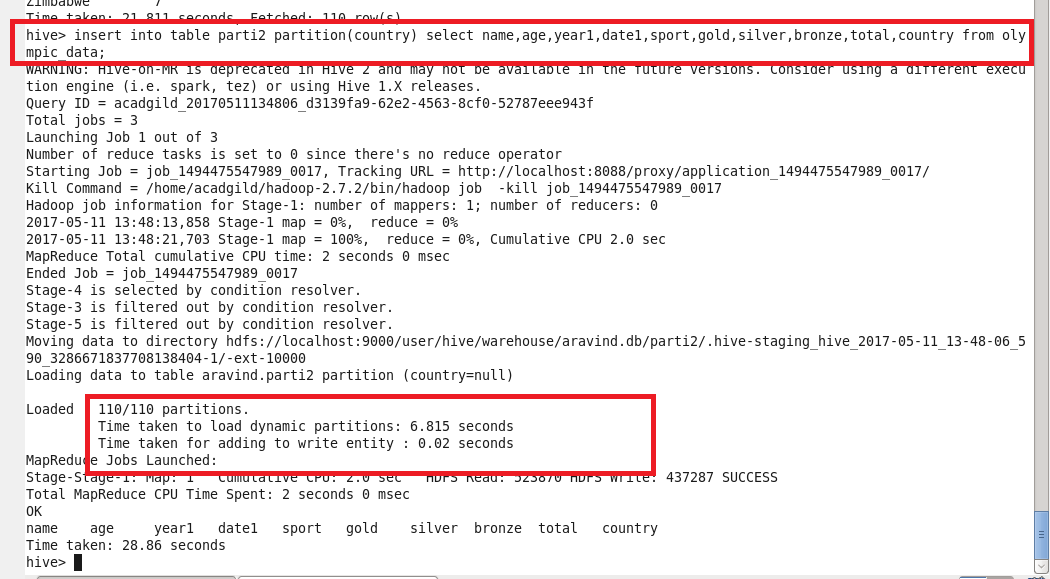
STEP 1: creating a table and partitioning by country.

****

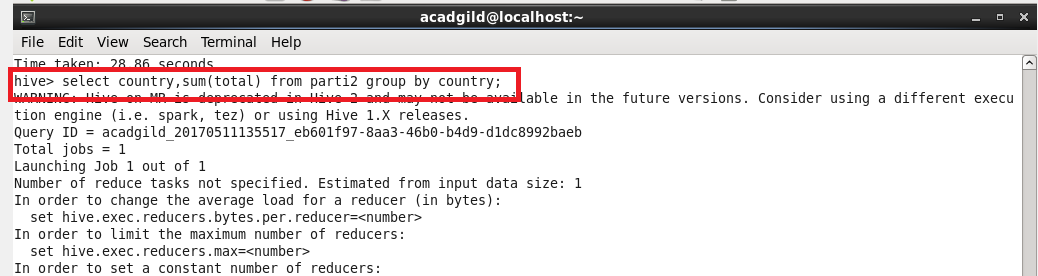
Step 2:setting some configurations in shell for running dynamic configurations

****

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

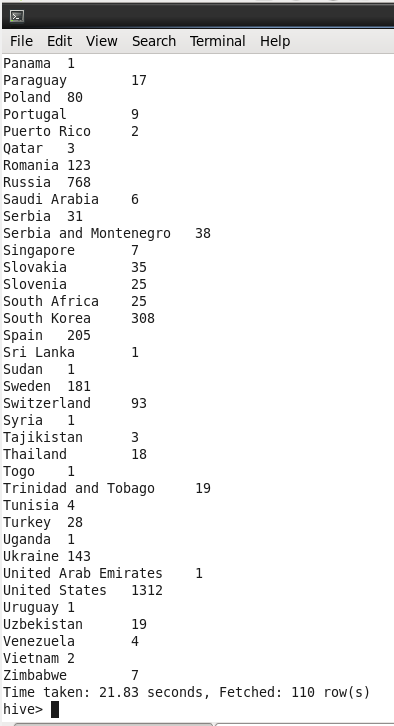
****

Step 4: querying

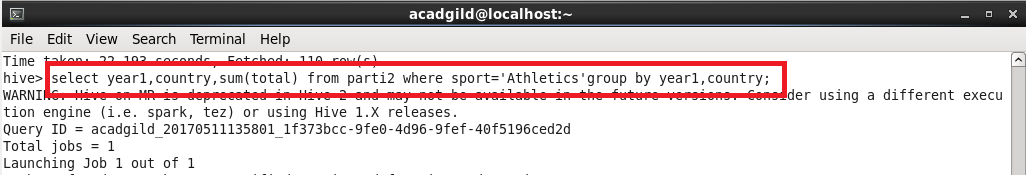
****

**Output:**

**I m attaching only random screenshots of the o/p since I cant include everything**

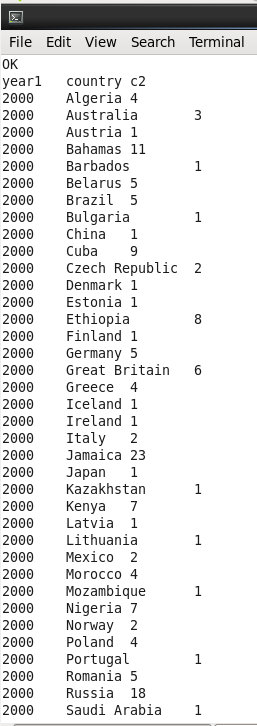
** **

1. **Find the number ot medals each country won in Athletics year wise**

****

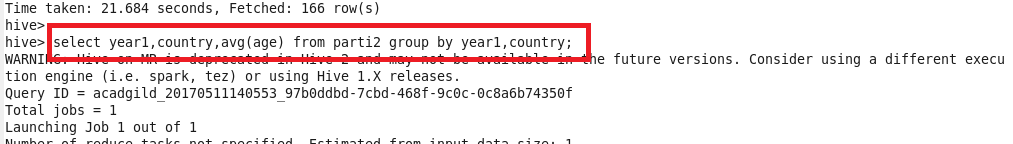
***Sum(total) and sport=’athletics’ and grouping by year and country will give number of medals each country won year wise***

**Output:**

****

**3)** **Find the average age of atheltes participated from each country in olympics year wise**

***Avg(age) and group by country and year will give the average age of players from each country year wise***

****

**o/p is in the next page**

****