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

**Partioning:**

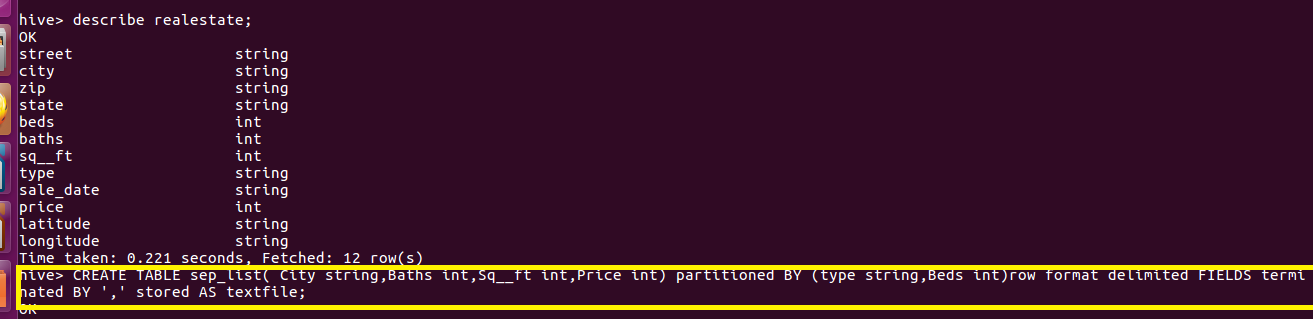
**Working of Partitioning**

Since during querrying in hive for a large dataset,for a simplequerry say we want HOUSES with beds=2 it want to search the entire database and find the result which will take a large time .So in order to optimize querrying time Partition is introduced

In Partitioning,the data is divided into directories based on column specified under Partition so that while querring the hive searches form that directory (eg ) if we use people from beds=2, It will go directly to that directory and the operation will be performed

**Creating Partitioned Table**:It is similar to ordinary table but just adding Partitioned By() where column based on which partitioning needs to be done

Here we will create table with partition for which we will add data from a large realestate data based on bedrooms and flat\_type

****

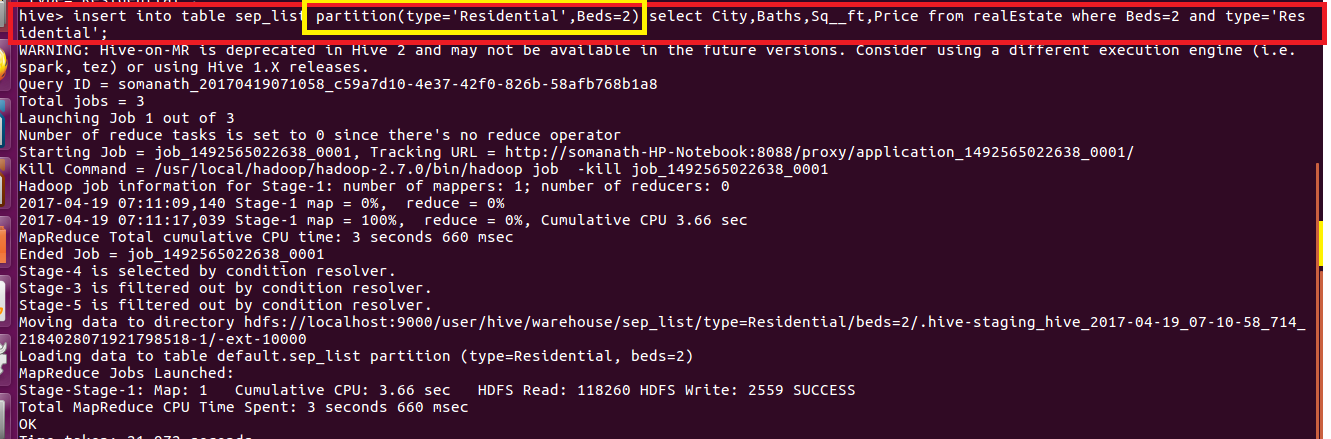
**Static Partitioning**

**Scenario:**

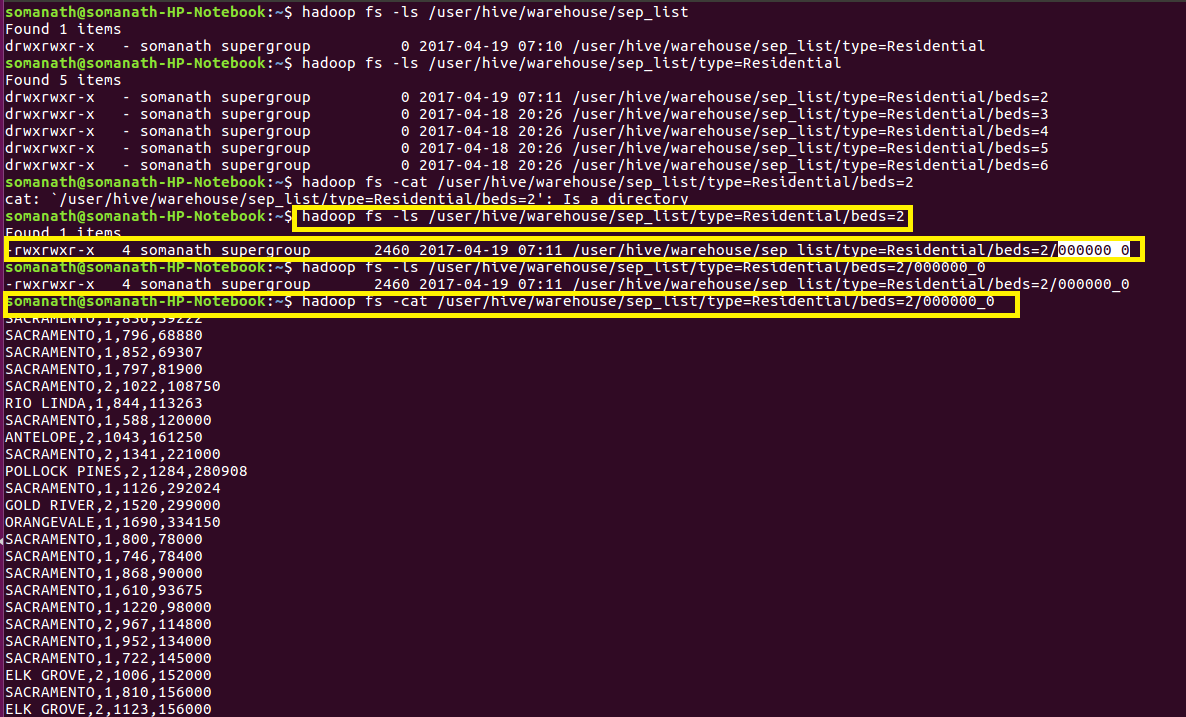
**WE have a real estate database for which we want a separate list of only data with bedroom=2**

**So we will create a static partitioning of bed=2**

**Thus in static partitioning,we know the type of data for example if we know the data is of bedroom with 2 we will create a partitioning into the partitioned table by giving the following command by giving type as residential and beds as2**

****

**Output:we can see that a separate directory for residential as type and bed=2 is created which contains all the data of beds=2**

****

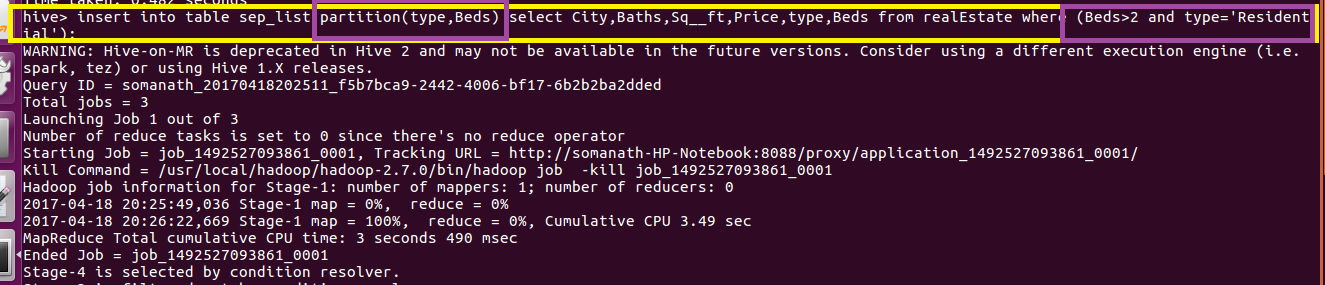
**Drawback: the major problem with static partitioning is that suppose the dataset is large and contains house with bedroom from 2 to 6 .**

**So we want to type the same SQL querry 5 times with beds=3, beds=4, beds=5, beds=6 if we use static partitioning**

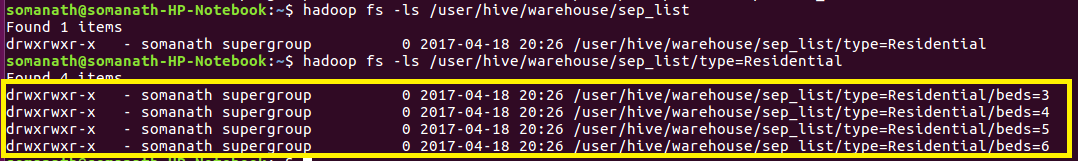
**So for this we use dynamic partitioning in which we will not specify the beds as 2 but we will just mention the columns on which partition to be done and the columns will be added as last two column in select statement so that hive will automatically do partitioning**

**Dynamic Partitioning**

**So if I want the partitioning to be done on houses on bedrooms greater than 2 I will just specify the columns based on which partitioning need to be done and the hive will automatically do partitioning with beds=3, beds=4, beds=5, beds=6 as shown**

****

**Output**

****

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

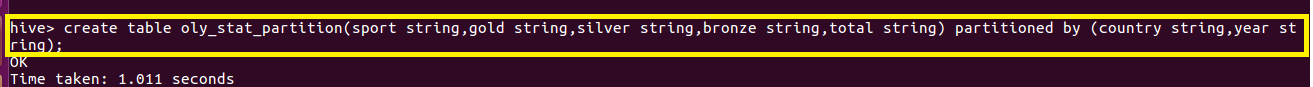
**- Find the number of medals india won year wise**

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

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

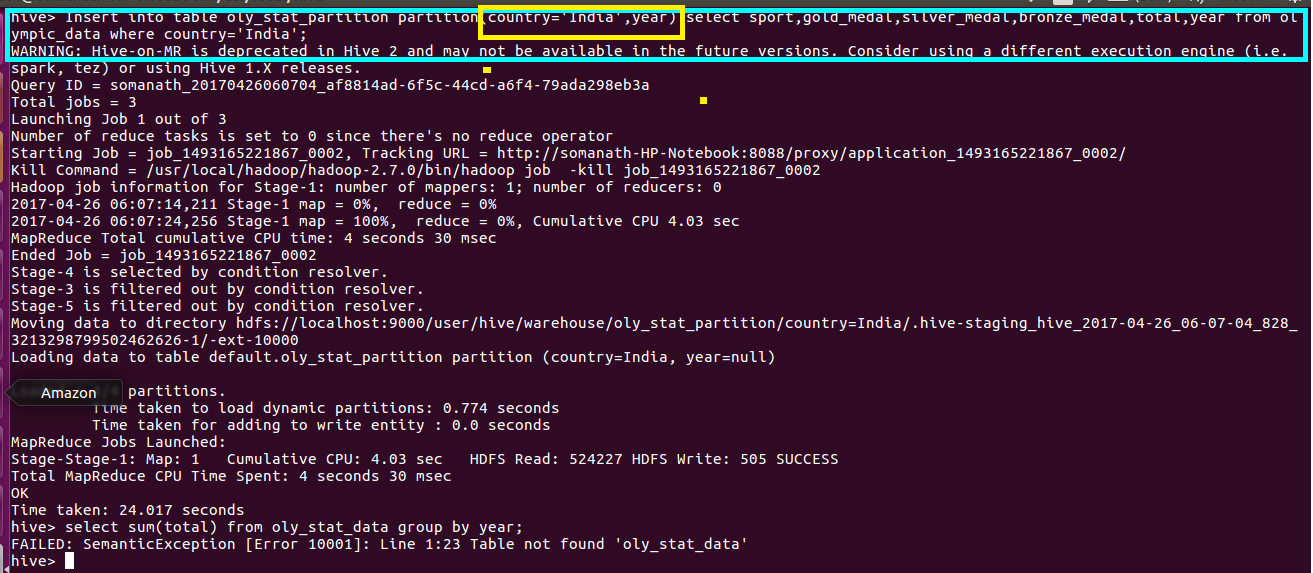
**Creating Table**

**Create a partion table where partitioning is based on country and year**

****

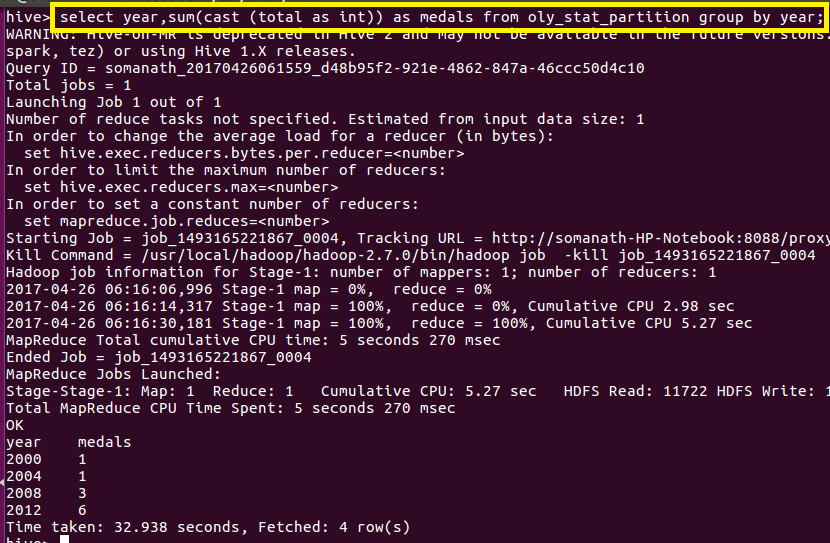
**Loading Data**

**Since it is static partitioning and we want country India I am inserting statically country=’India’ and but the same we don’t have have the years in which India won the medal so I am loading year dynamically as shown**

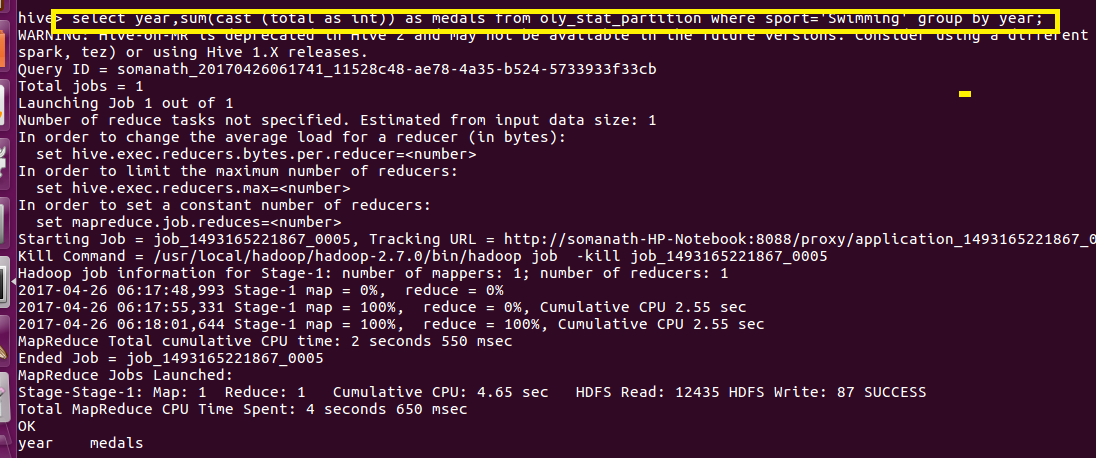
****

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

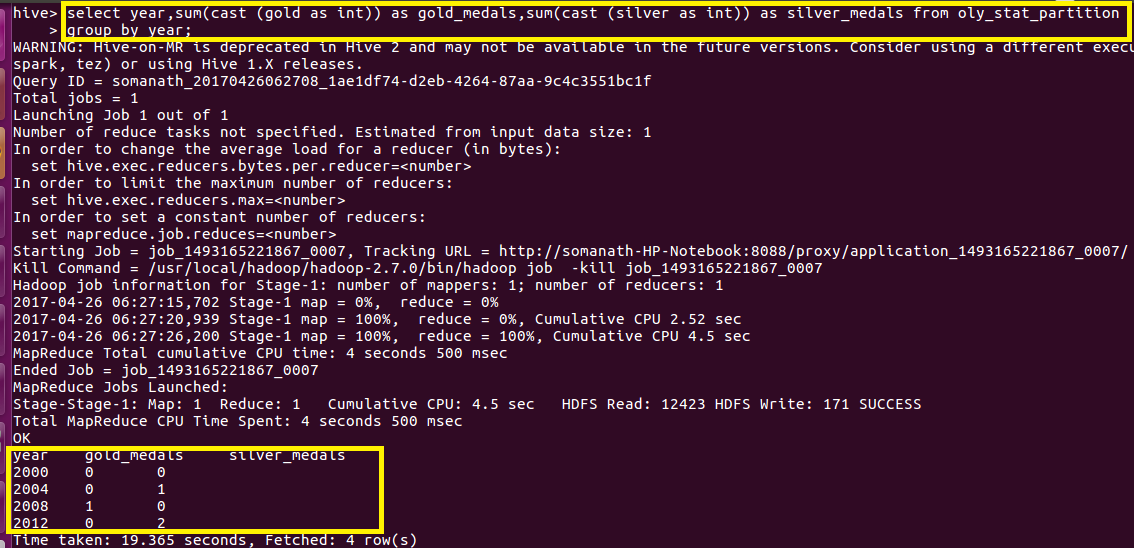
**Since data contains only India just grouping by year and sum total medals**

****

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

****

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

****

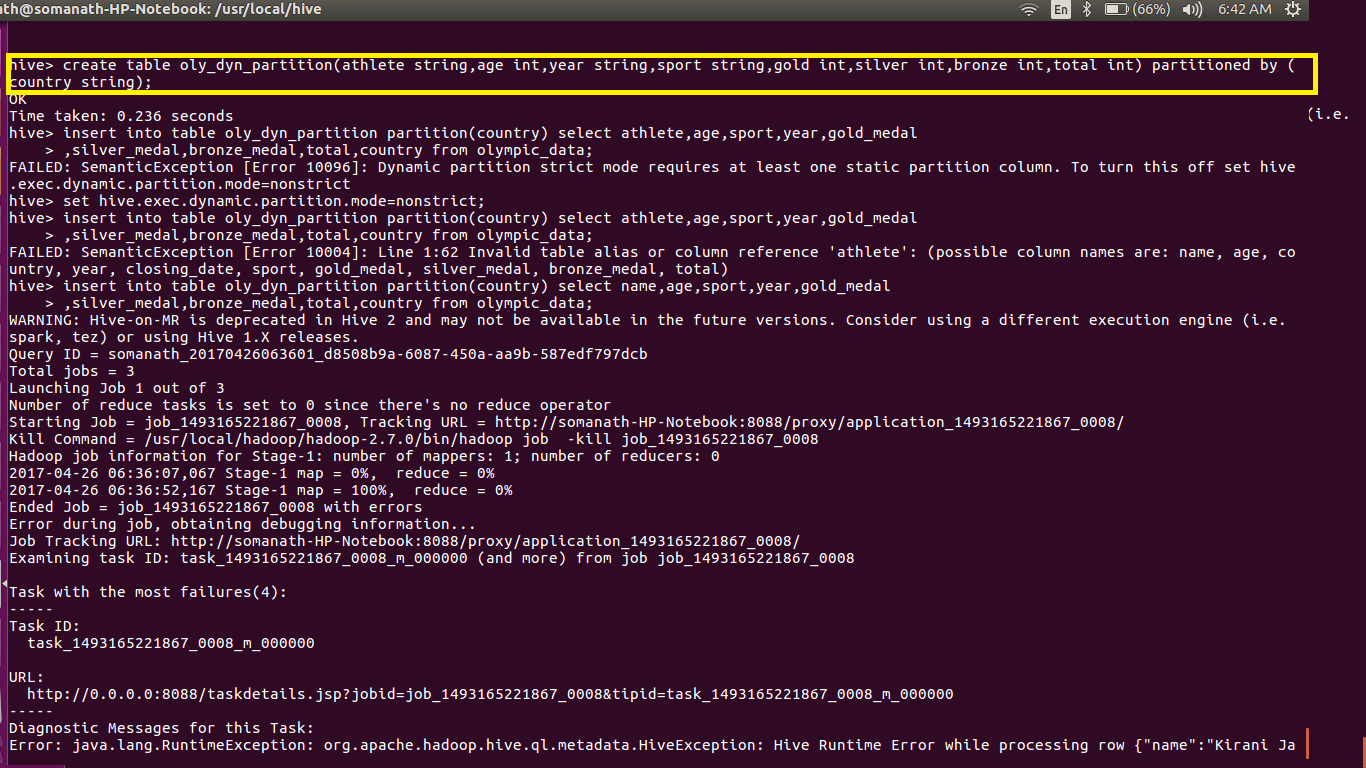
**3. Use dynamic partitioning in hive and evaluate the below problem statements**

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

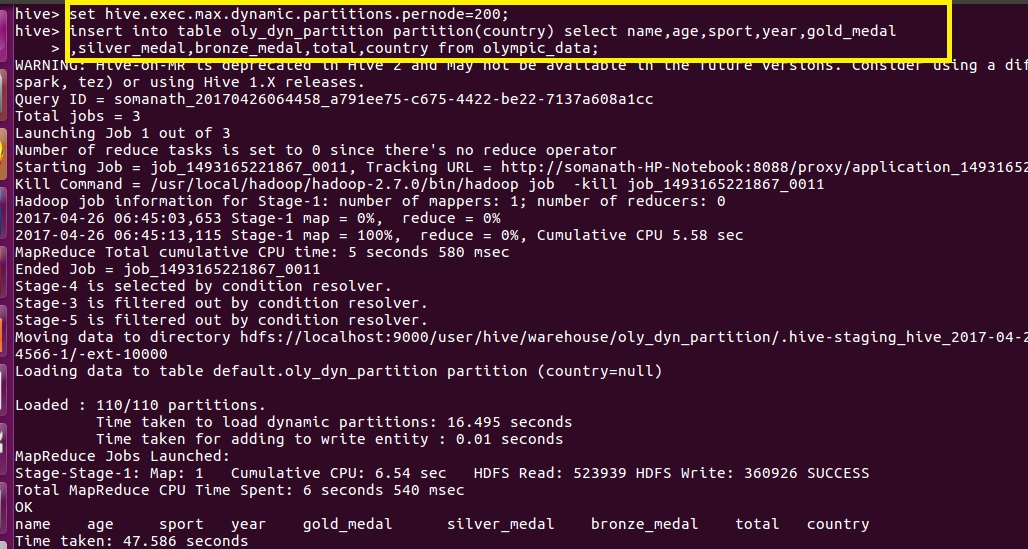
**- Find the number of medals each country won in Athletics year wise**

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

**Since It is a dynamic partitioning and all querrying is based on country I am creating a dynamic table partitioned by country to ensure faster querying**

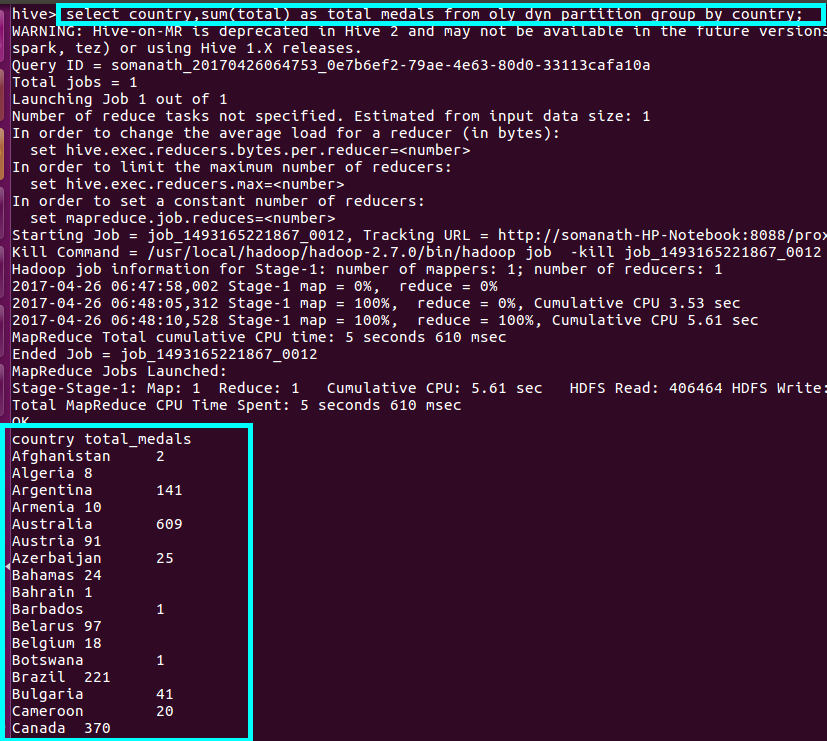
****

**Loading data dynamically based on country**

****

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

**Since we want total medals won by each country,group by country and added the total medals using sum()**

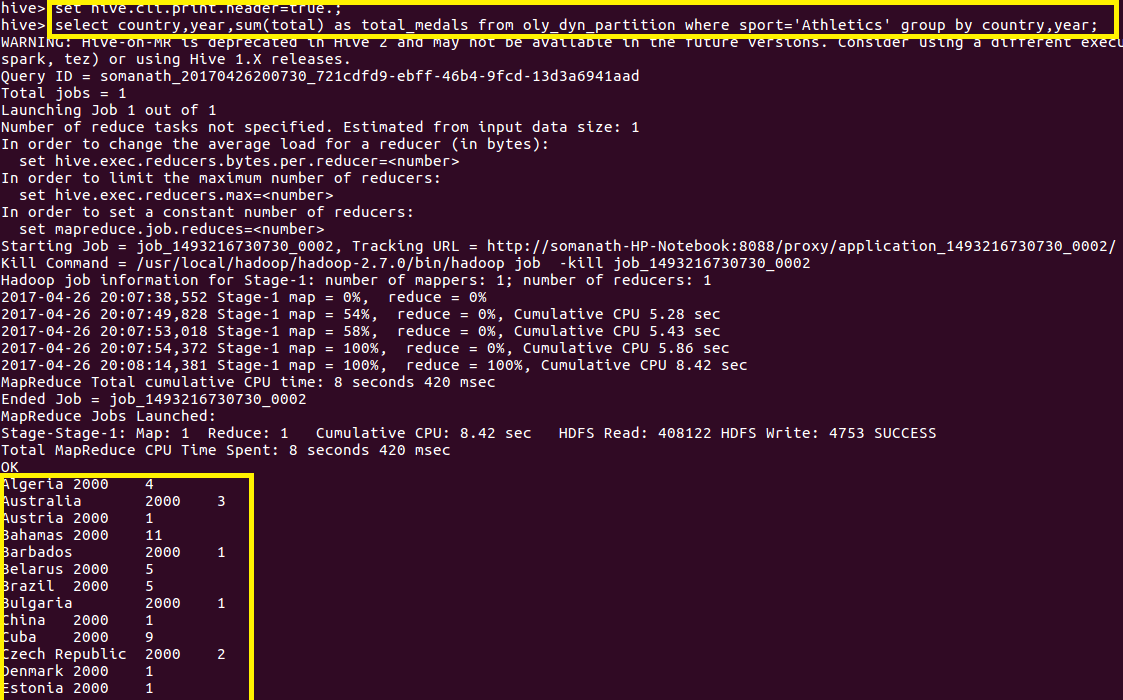
****

**Output**

**Full output attached(total)**

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

**Since we want medals won by athletics group by country and year and filter by sport=athletics**

****

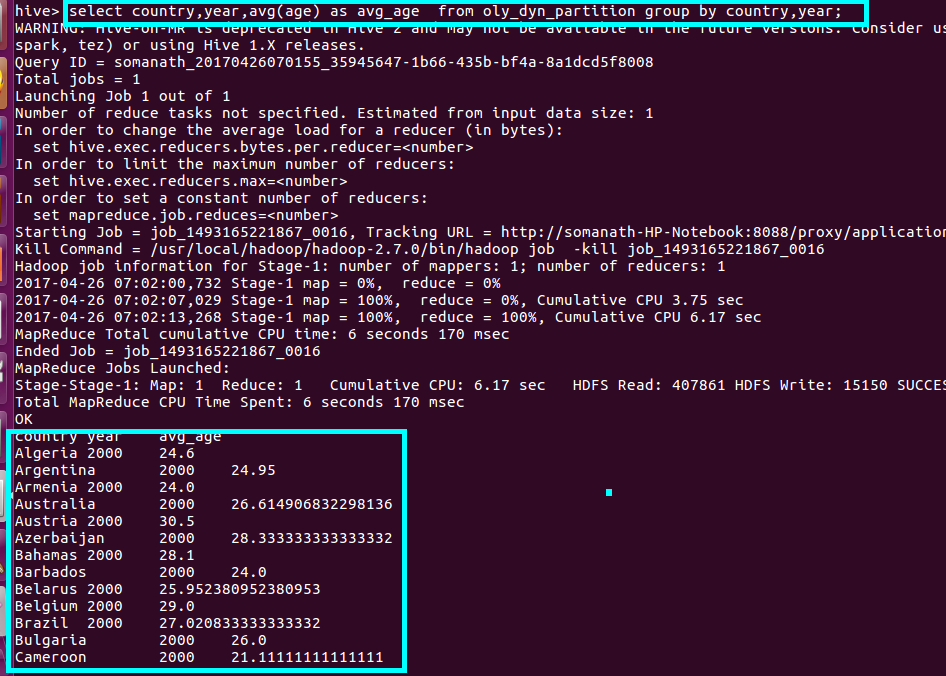
**Output**

**Full output attached(athlete medal year wise)**

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

**Since we want avg age of athletes in Olympics for each country yearwise,**

**I grouped by country and year and used avg() on age**

****

**Output**

**Full output attached(avg age)**