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

In Hive Partition - there are 2 different type of partitions

*1.)Static Partition*

2.) Dynamic Partition.

* There are certain difference between **Static partition and dynamic partition** in hive differences.

In Hive there are 2 types of Partitions available:

***i) Static Partition in Hive***

***ii) Dynamic Partition in Hive***

Static Partition in Hive are:

* Inserting the input data files individually into a partition table is called as Static Partition.
* While loading the files (eg-big files) into the Hivetables usually static partitions will be preferred.
* Static Partition will save your time for data loading when compared to dynamic partition.
* While “statically” adding a partition in table and when we move the file into that partition of the table. One can alter those partition in the Static partition.
* One can the partition column value form from *filename, day of date etc.* without reading the whole file.
* In HIVE when one needs to use Static partition one should set the property as **set hive.mapred.mode = strict**
* hive-site.xml will be set as default.
* Static partition will be usually set in Strict Mode.
* One should use where clause in order to use limit in the static partition.
* One can perform Static partition on Hive Manage table or external table.

Example:

**CREATE** **TABLE** cityreport (cityid string, creport string, ctover string)  
partitioned **BY** (city string)  
row format delimited  
**FIELDS** terminated **BY** ‘|’  
stored **AS** textfile;

*Loading data using static partitioning:*

**LOAD DATA LOCAL inpath ‘/home/mathurri/hive-related/hyderabad.log’ INTO TABLE cityreport partition (city = ‘chennai’);**

*Dynamic Partition in Hive*

* single insert into partition table is known as dynamic partition
* Usually dynamic partition will load the data from non-partitioned table
* Dynamic Partition will take more time for loading data when compared to the static partition.
* When a large data is stored in a table in such case Dynamic partition will be suitable.
* In case if one needs to partition number of column but don’t know how many columns were present then dynamic partition can be used.
* Dynamic partition will not be required where clause to use limit.
* One can’t perform any alter on those Dynamic partition
* On HIVE one can perform dynamic partition using the external table as well the managed table.
* In HIVE when one need to use the Dynamic partition in such case the mode will be in nonstrict mode
* Here is hive dynamic partition properties you should allow

SET hive.exec.dynamic.partition = true;

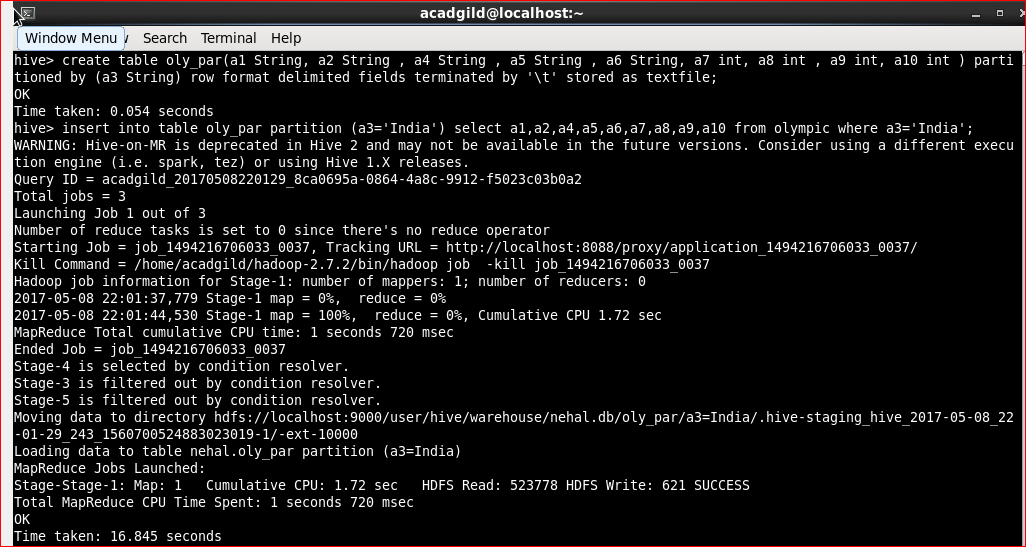
SET hive.exec.dynamic.partition.mode = nonstrict;

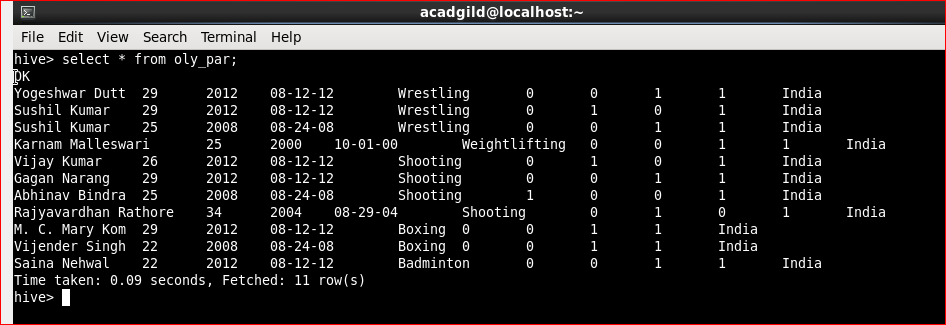
Example:

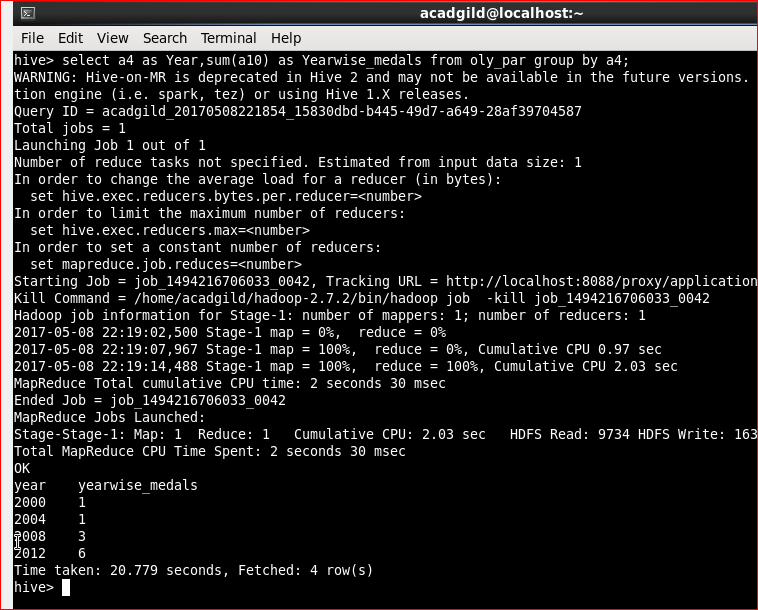
hive> **INSERT** **INTO** **TABLE** t2 PARTITION(country) **SELECT** \* **FROM** T1;

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

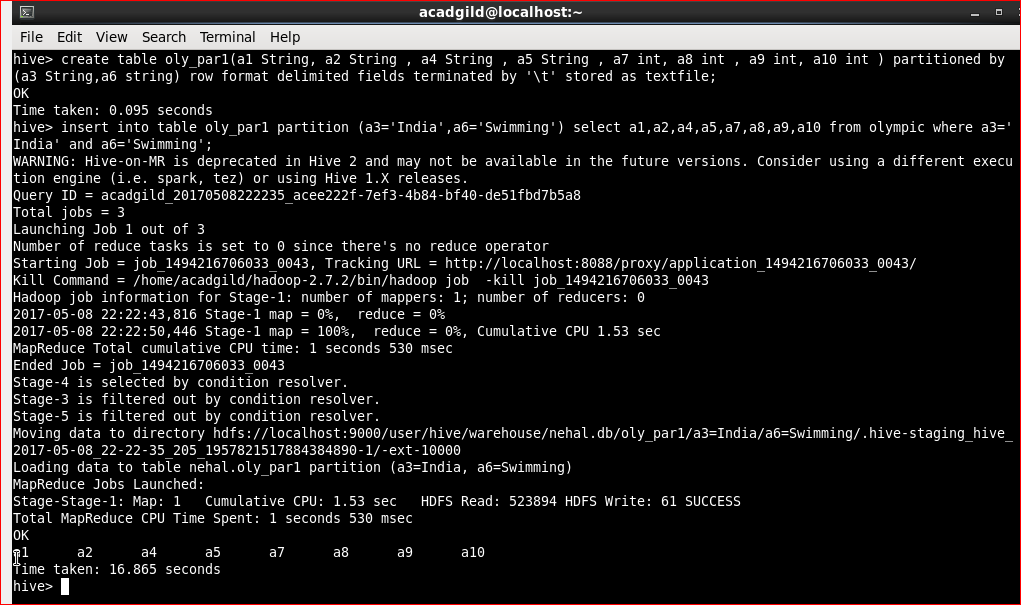
- Find the number of medals India won year wise

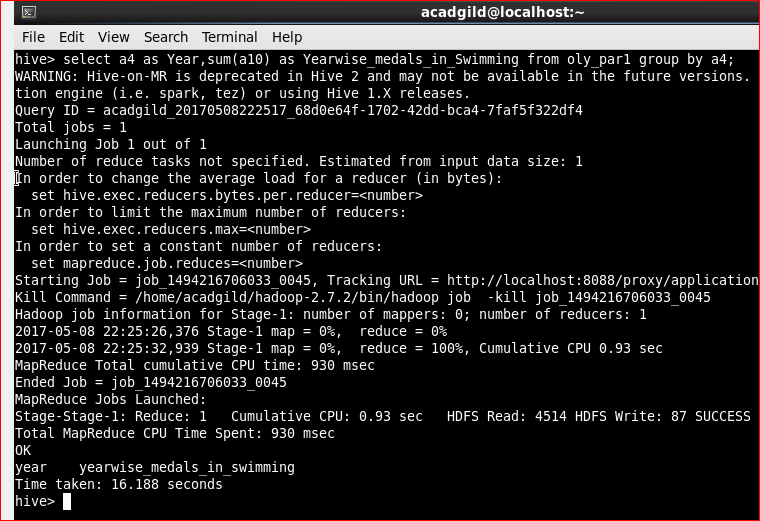




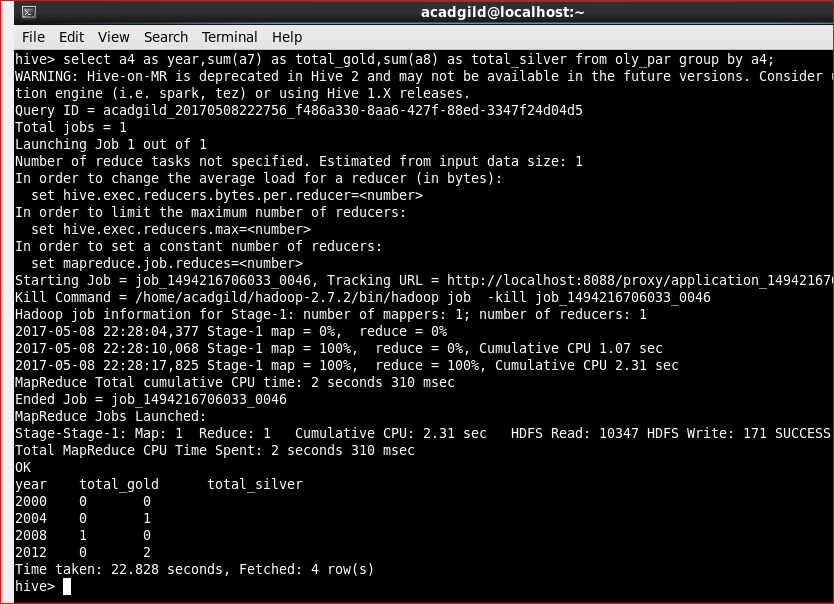


- Find the number of medals India won in swimming year wise



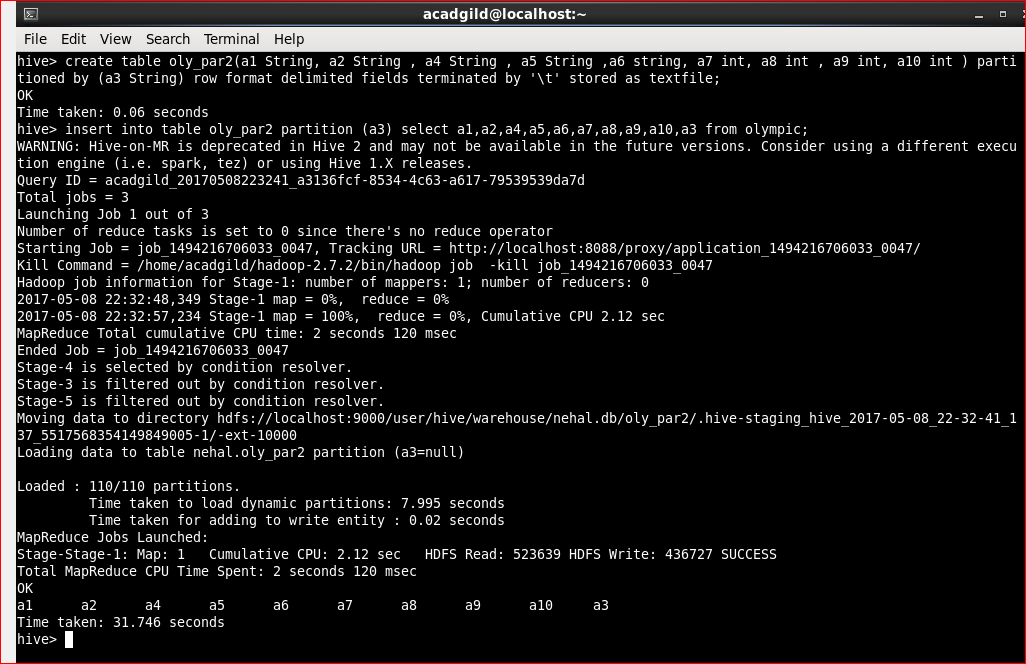


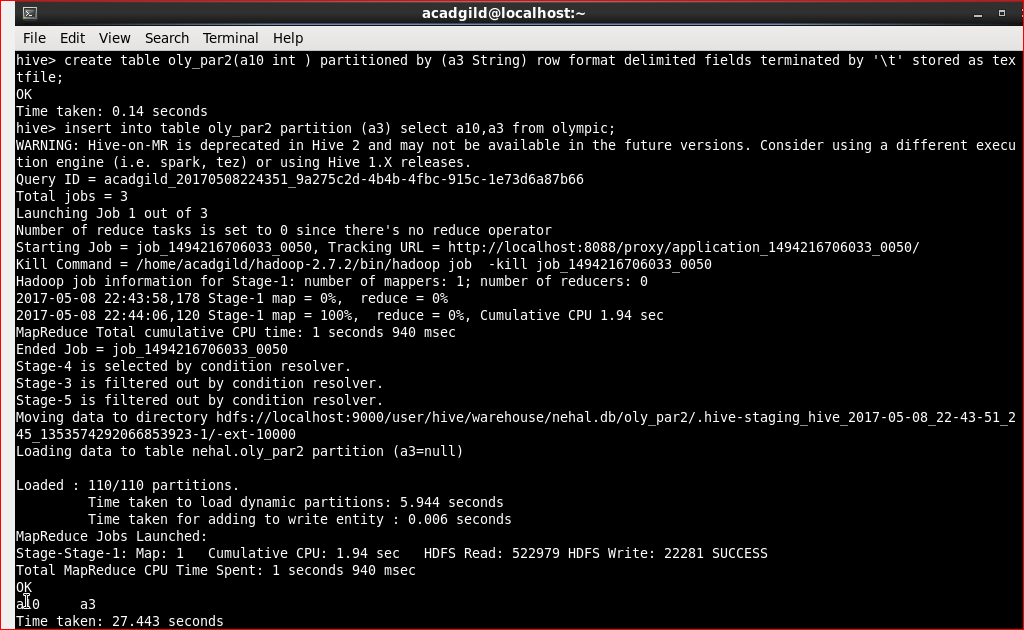
- Find the number of gold and silver medals India won year wise

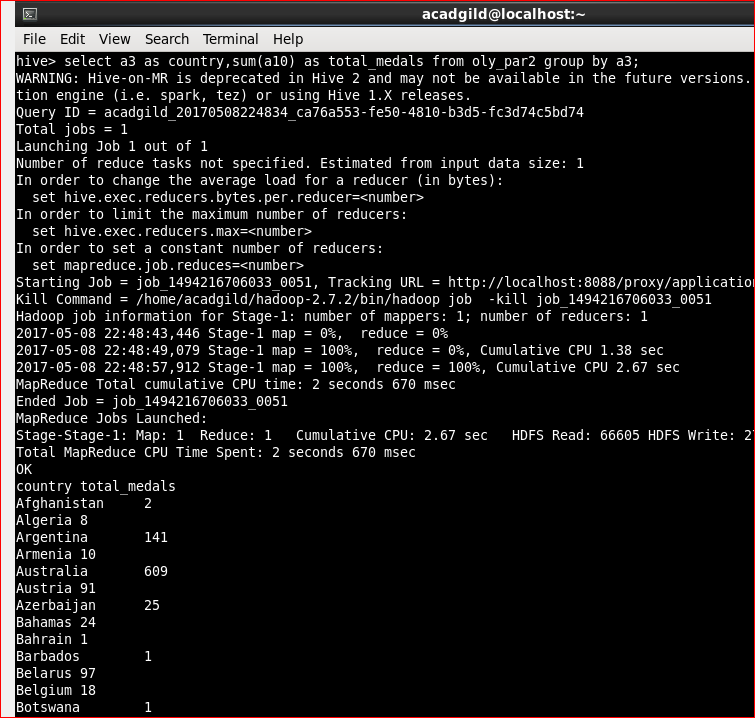


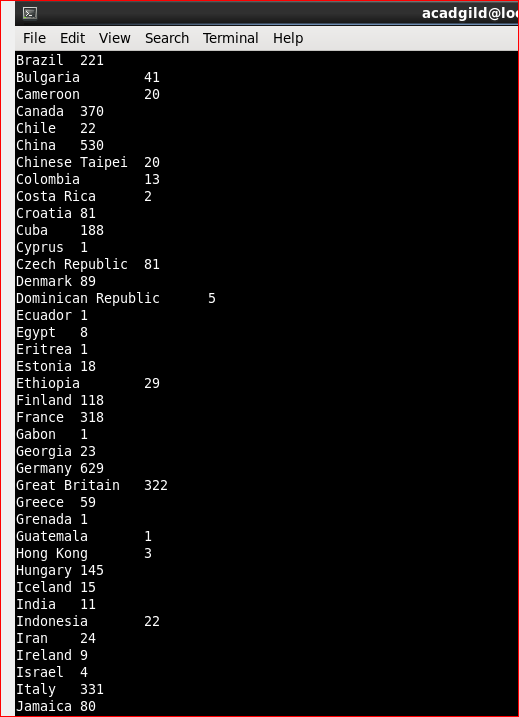
**3. Use dynamic partitioning in hive and evaluate the below problem statements**

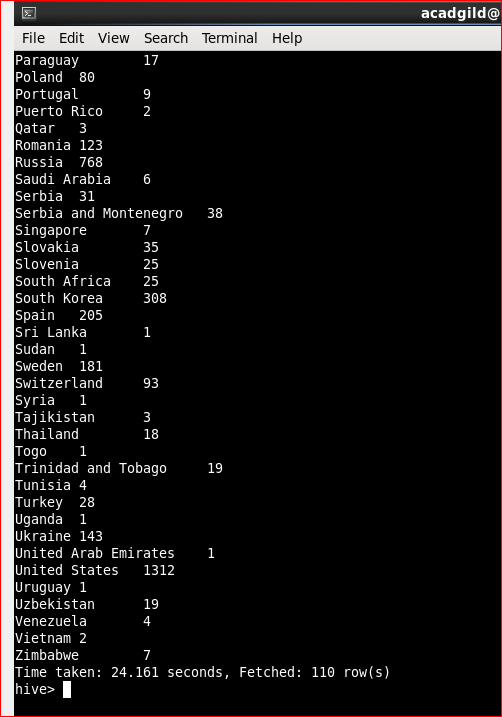
- Find the total number of medals won by each country.



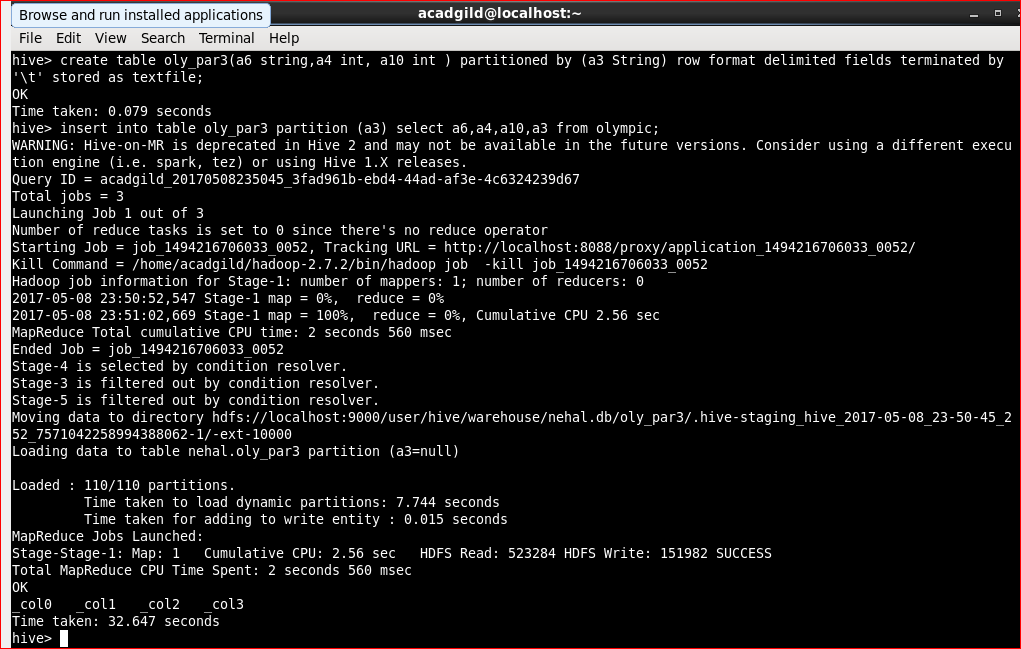


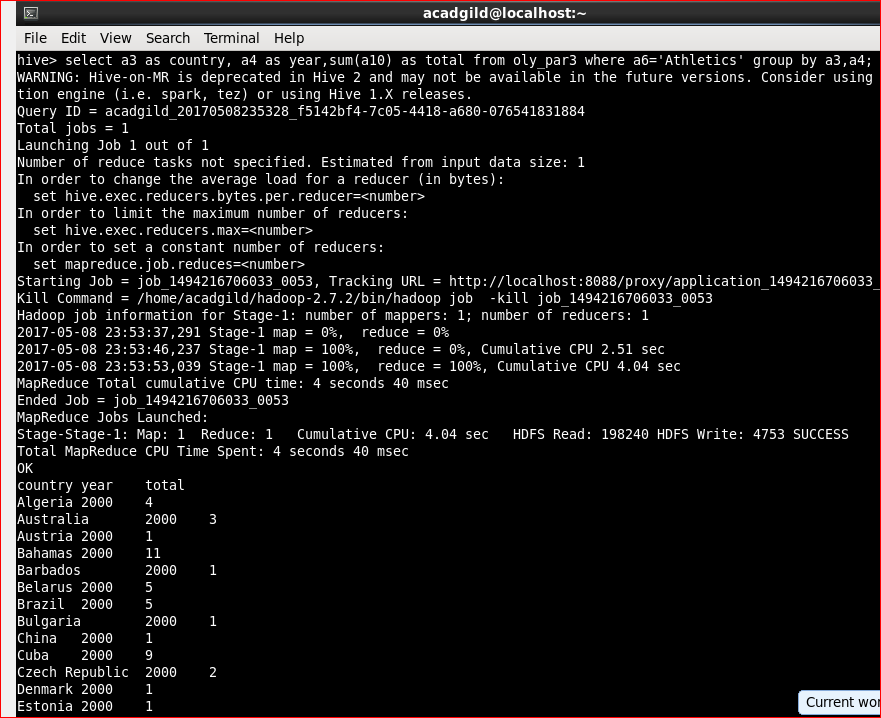


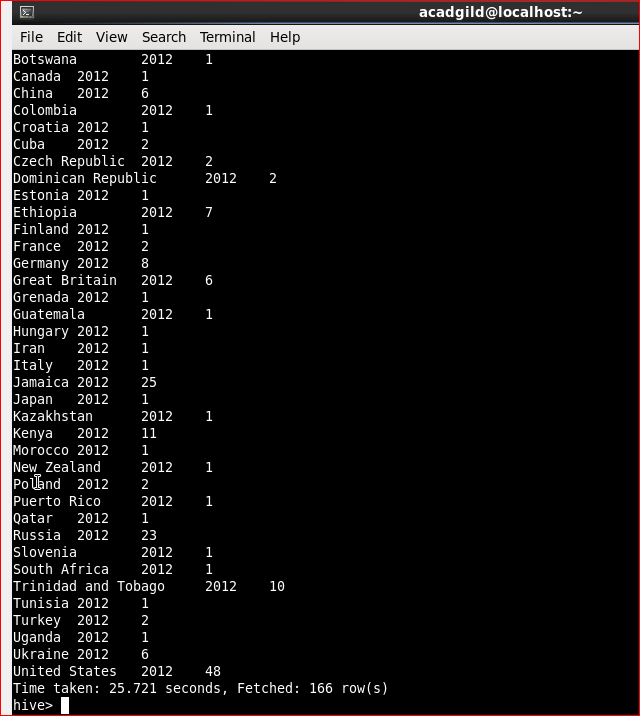




- Find the number of medals each country won in Athletics year wise







- Find the average age of athletes participated from each country in Olympics year wise

