Problem Statement:

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

In Hive Partition concept there is two different type of partitions one is *Static Partition* and another one is Dynamic Partition. Here we will discussed about **Static partition vs dynamic partition in hive** differences.

In Hive There is two types of Partitions

i) Static Partition in Hive

ii) Dynamic Partition in Hive

Static Partition in Hive

* Insert input data files individually into a partition table is Static Partition
* Usually when loading files (big files) into Hivetables 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

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/mahesh/hive-related/hyderabad.log’ INTO TABLE cityreport partition (city = ‘hyderabad’);**

Dynamic Partition in Hive

* 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
* Here is hive dynamic partition properties you should allow

SET hive.exec.dynamic.partition = true;

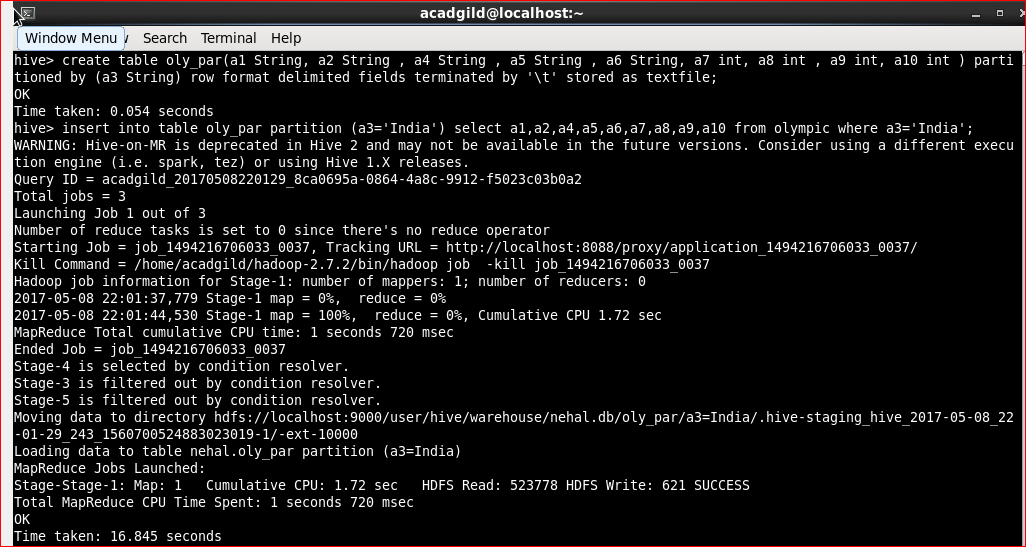
SET hive.exec.dynamic.partition.mode = nonstrict;

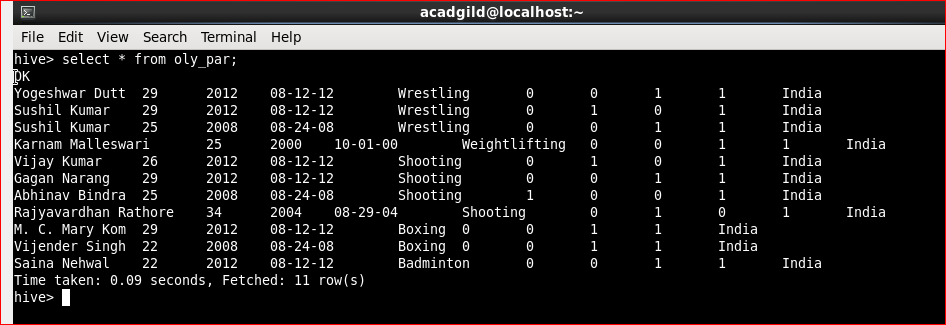
Example:

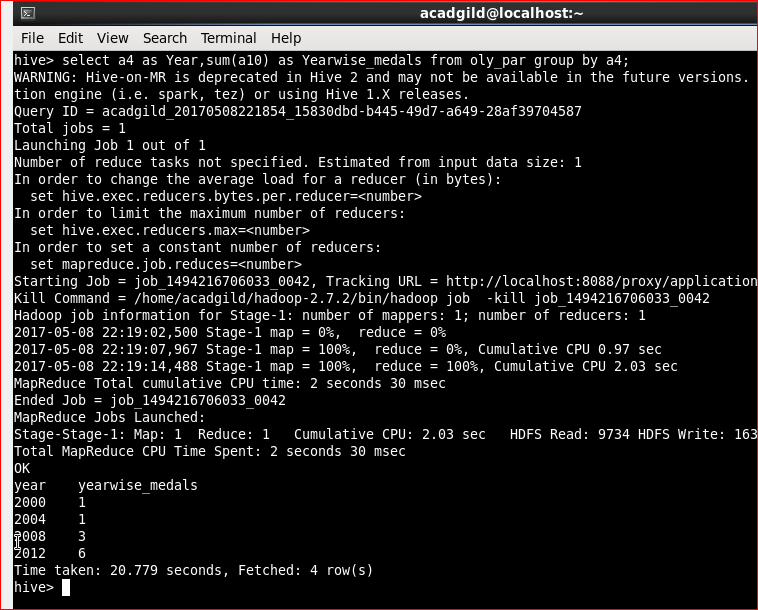
hive> **INSERT** **INTO** **TABLE** t2 PARTITION(country) **SELECT** \* **FROM** T1;

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

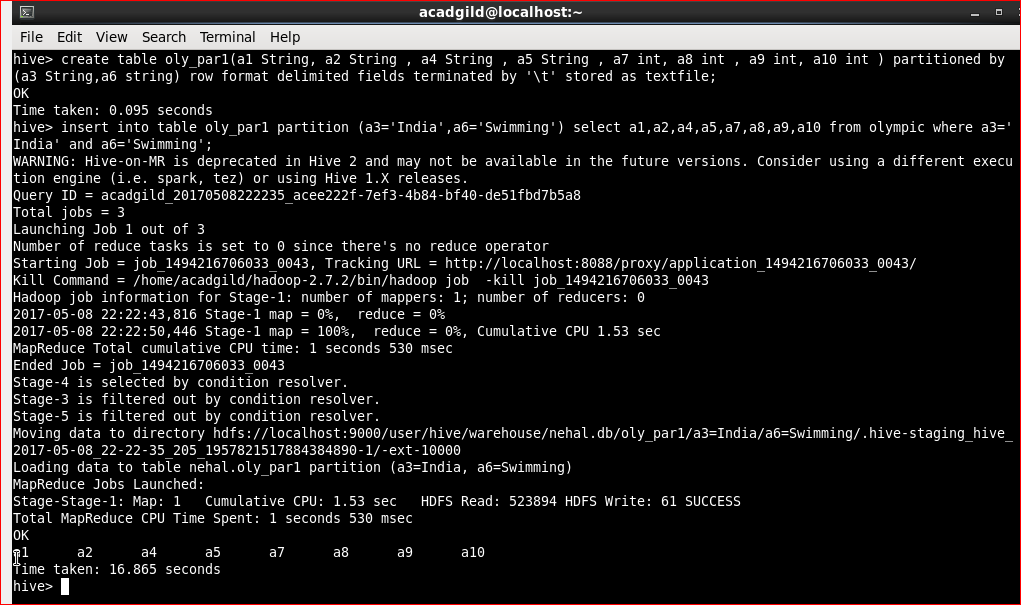
- Find the number of medals India won year wise

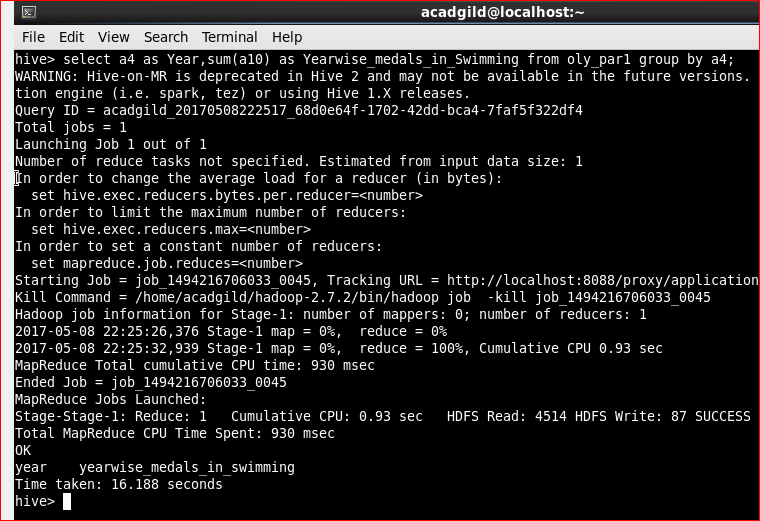




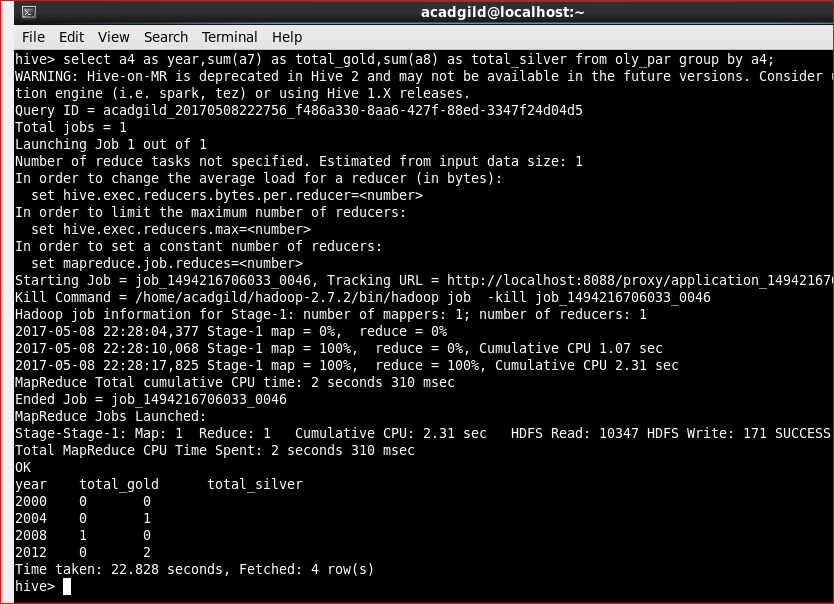


- Find the number of medals India won in swimming year wise



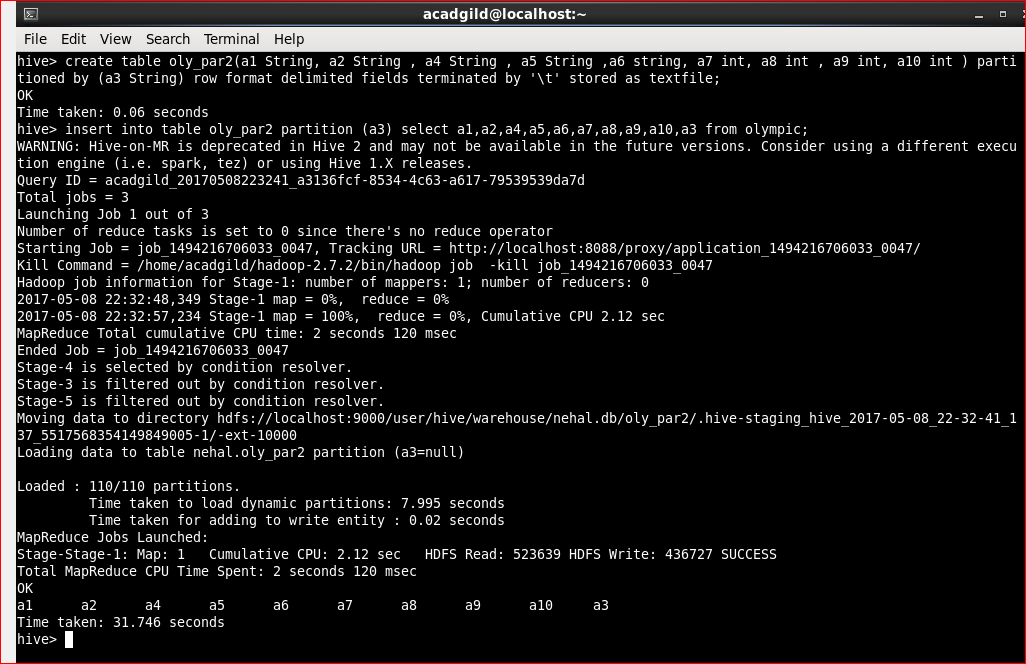


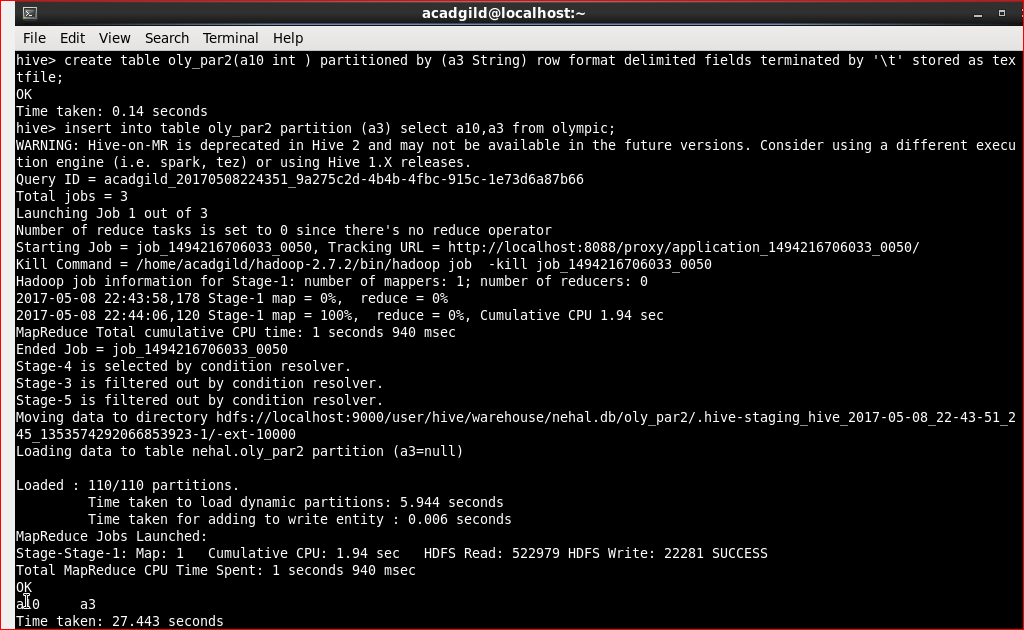
- Find the number of gold and silver medals India won year wise

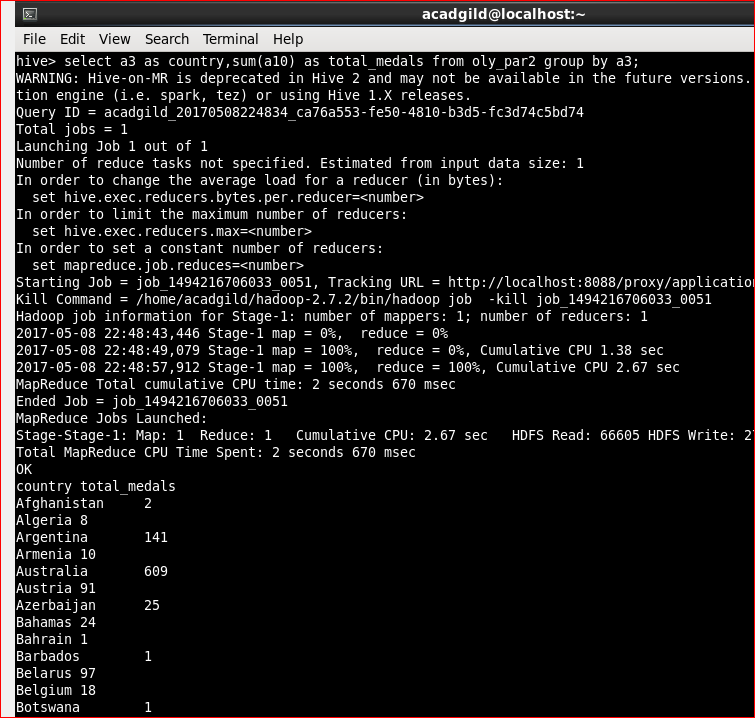


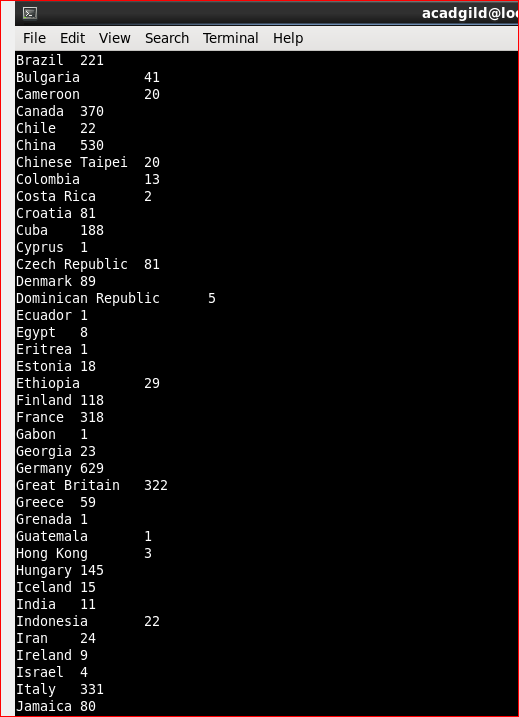
**3. Use dynamic partitioning in hive and evaluate the below problem statements**

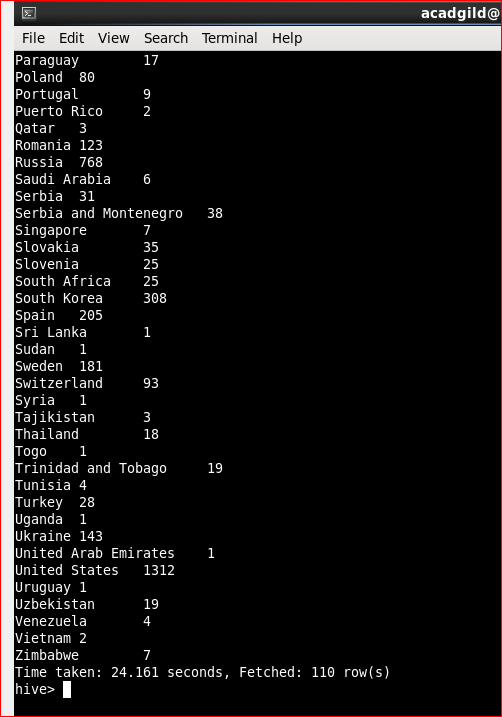
- Find the total number of medals won by each country.



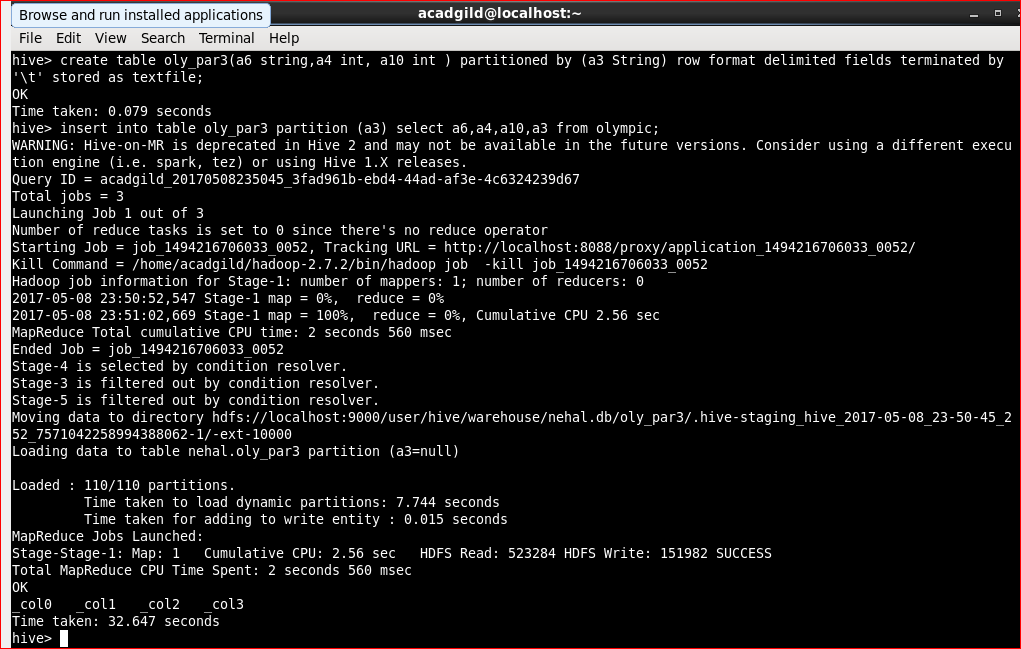


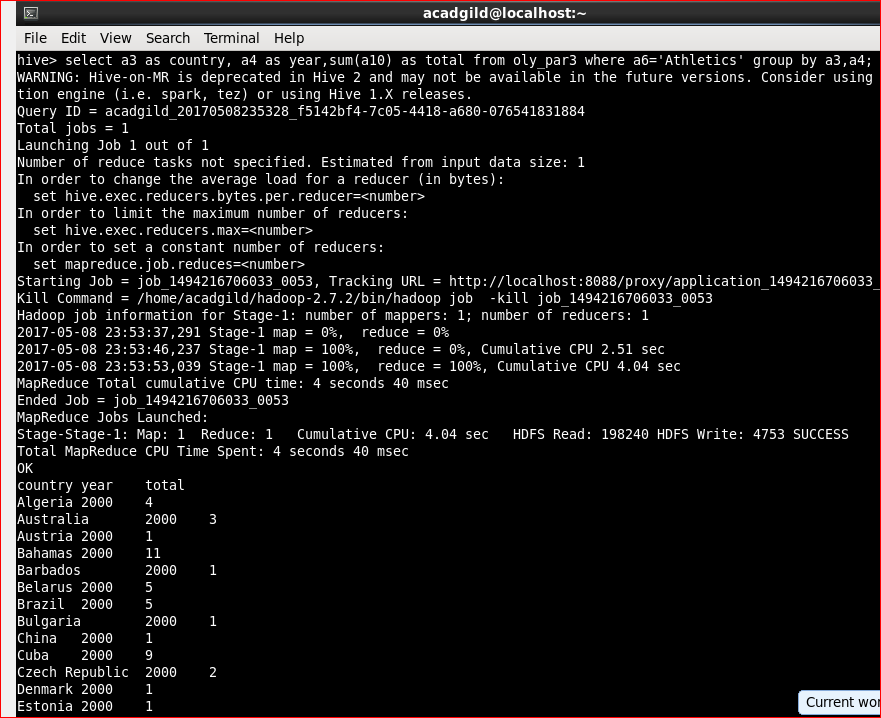


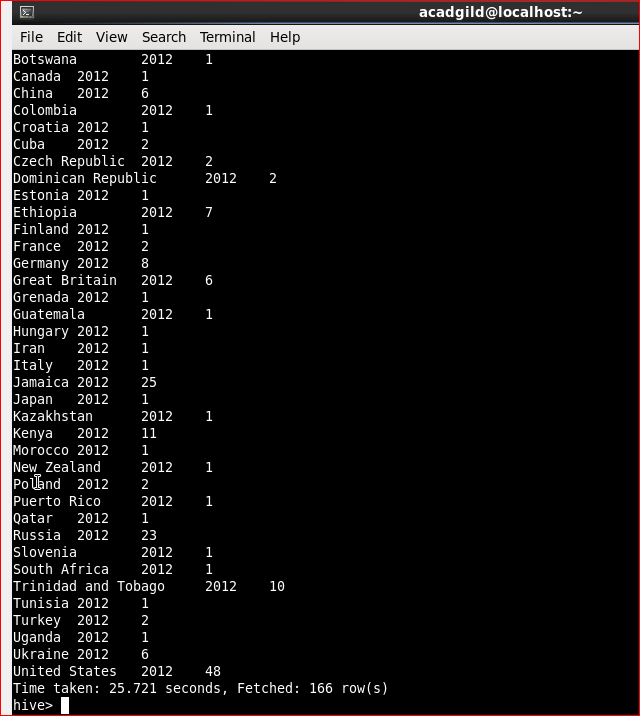




- Find the number of medals each country won in Athletics year wise







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

