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

Partitioning in Hive is very useful to purne data during query to reduce query times.

Partitions are created when data is inserted into table. Depending on how you load data you would need partitions. Usually when loading files (big files) into Hive tables static partitions are preferred. That 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. Since the files are big they are usually generated in HDFS. You can get the partition column value form the filename, day of date etc without reading the whole big file.

Incase of dynamic partition whole big file i.e. 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. So usually dynamic partition are useful when you are doing sort of a ETL flow in your data pipeline. e.g. you load a huge file through a move command into a Table X. then you run a inert query into a Table Y and partition data based on field in table X say day , country. You may want to further run a ETL step to partition the data in country partition in Table Y into a Table Z where data is partitioned based on cities for a particular country only. etc.

Static Partition in Hive

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 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.

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

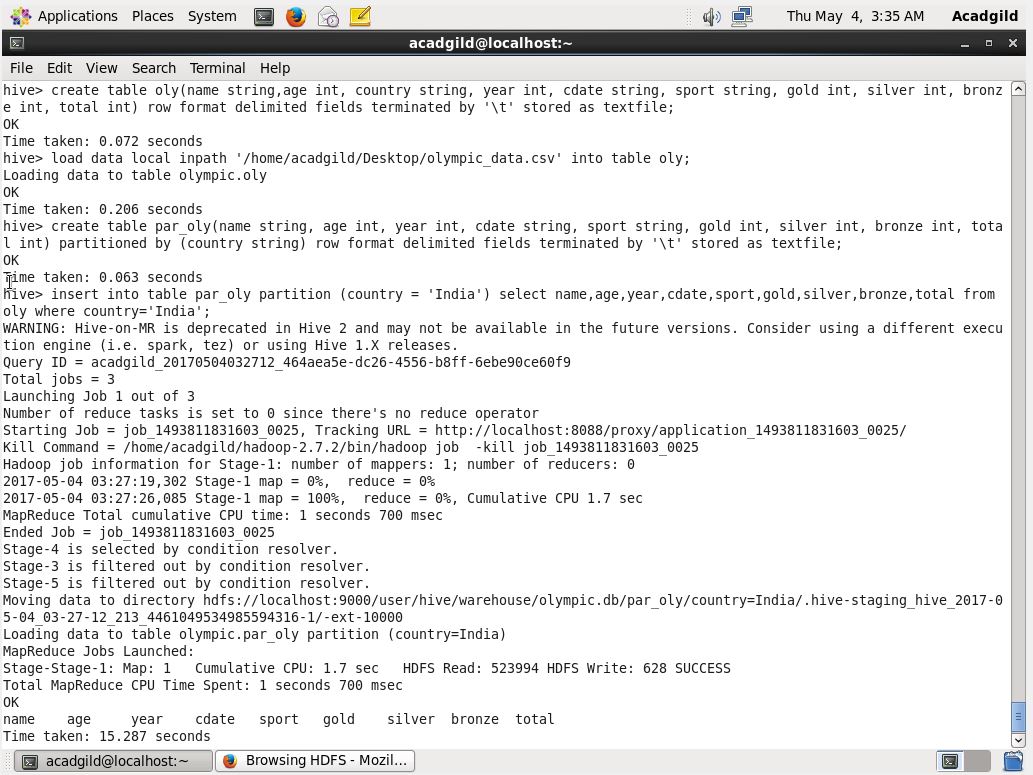
SET hive.exec.dynamic.partition = true;

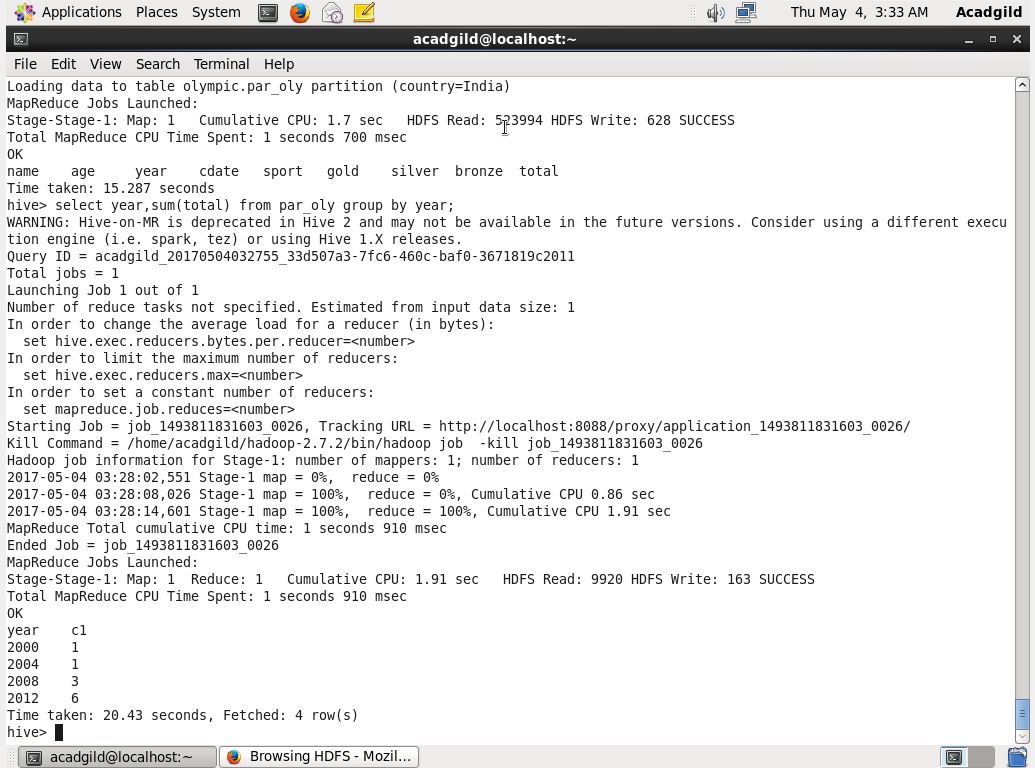
SET hive.exec.dynamic.partition.mode = nonstrict;

Thus depending on your end table or requirements for data and in what form data is produced at source you may choose static or dynamic partition.

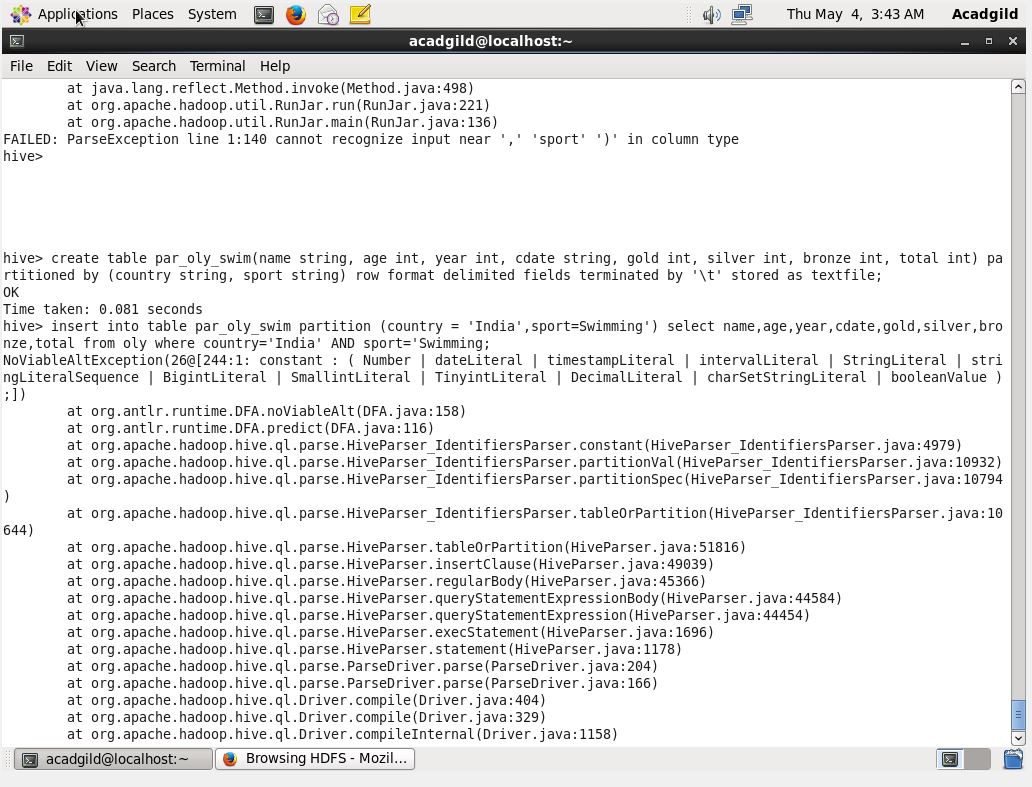
1. Use static partitioning in hive and evaluate the below problem statements

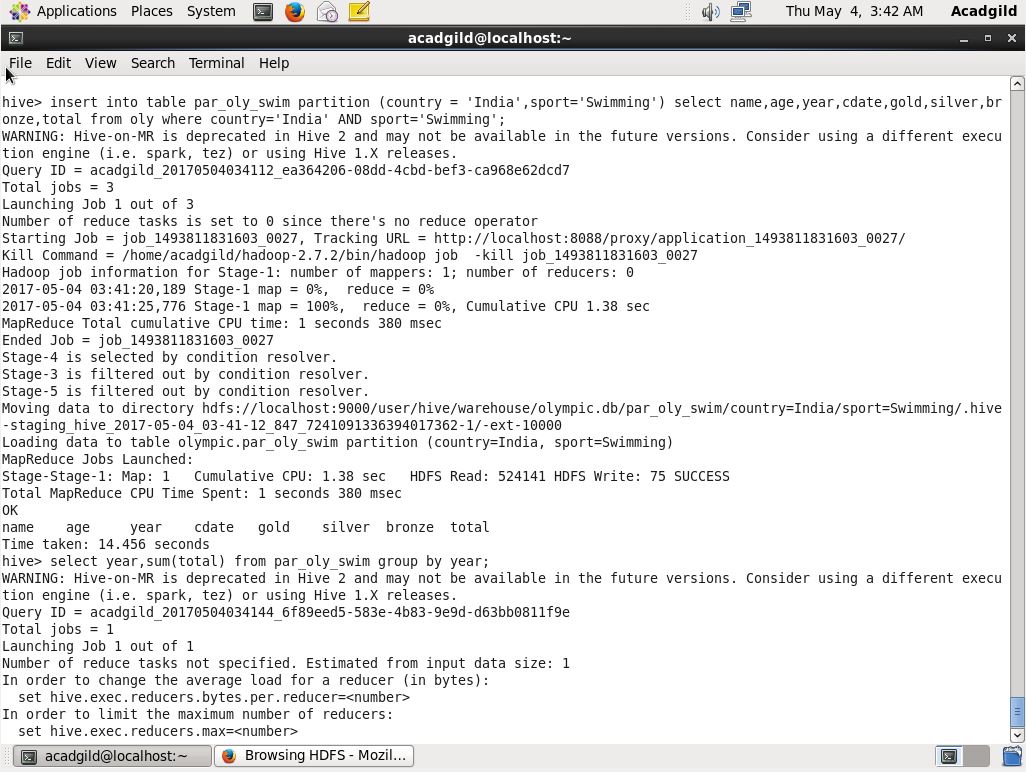
- Find the number of medals india won year wise

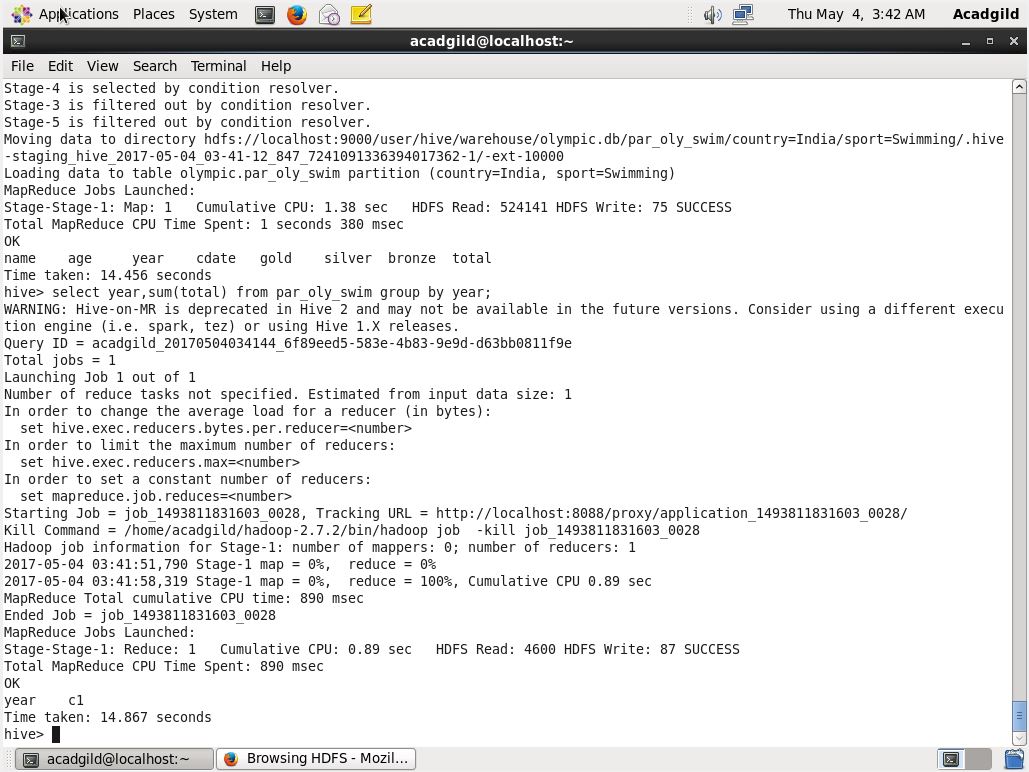




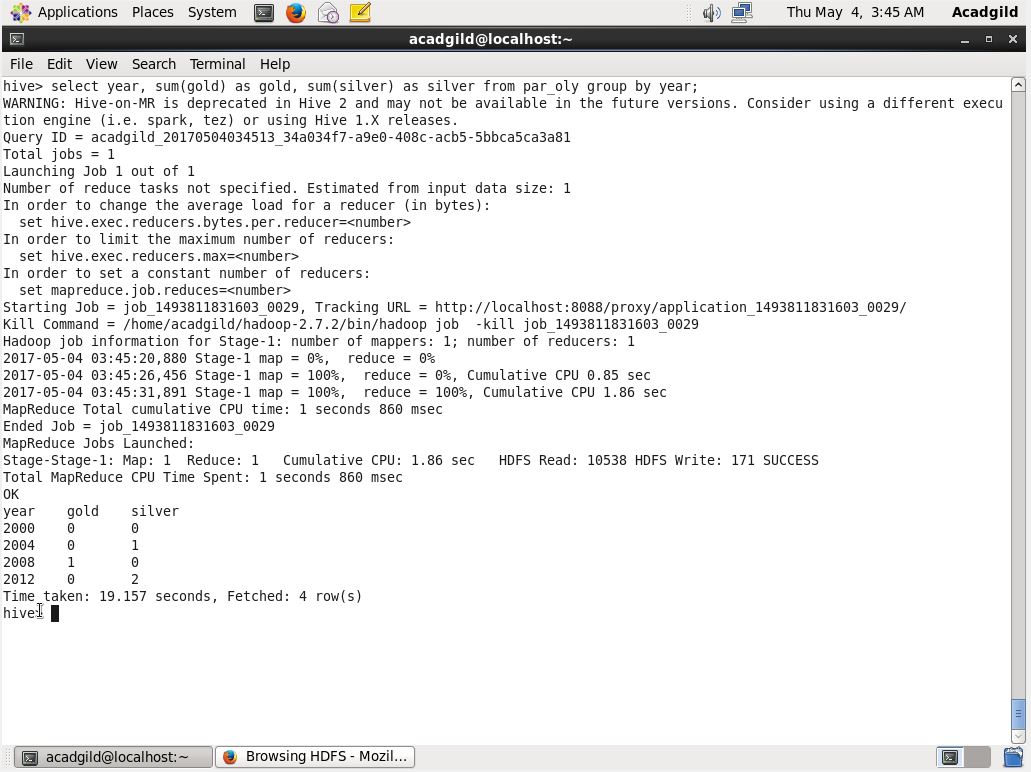
- Find the number of medals india won in swimming year wise







- 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 ot medals each country won in Athletics year wise

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

