static and dynamic partitioning

Partition manages the underlying structures of the table’s data directory. In case of partitioned tables, subdirectories are created under the table’s data directory for each unique value of a partition column. When a partition table is queried with one or both partition columns in criteria or in the where clause, what hive effectively does is partition elimination by scanning only those data directories that are needed.

Types of partition

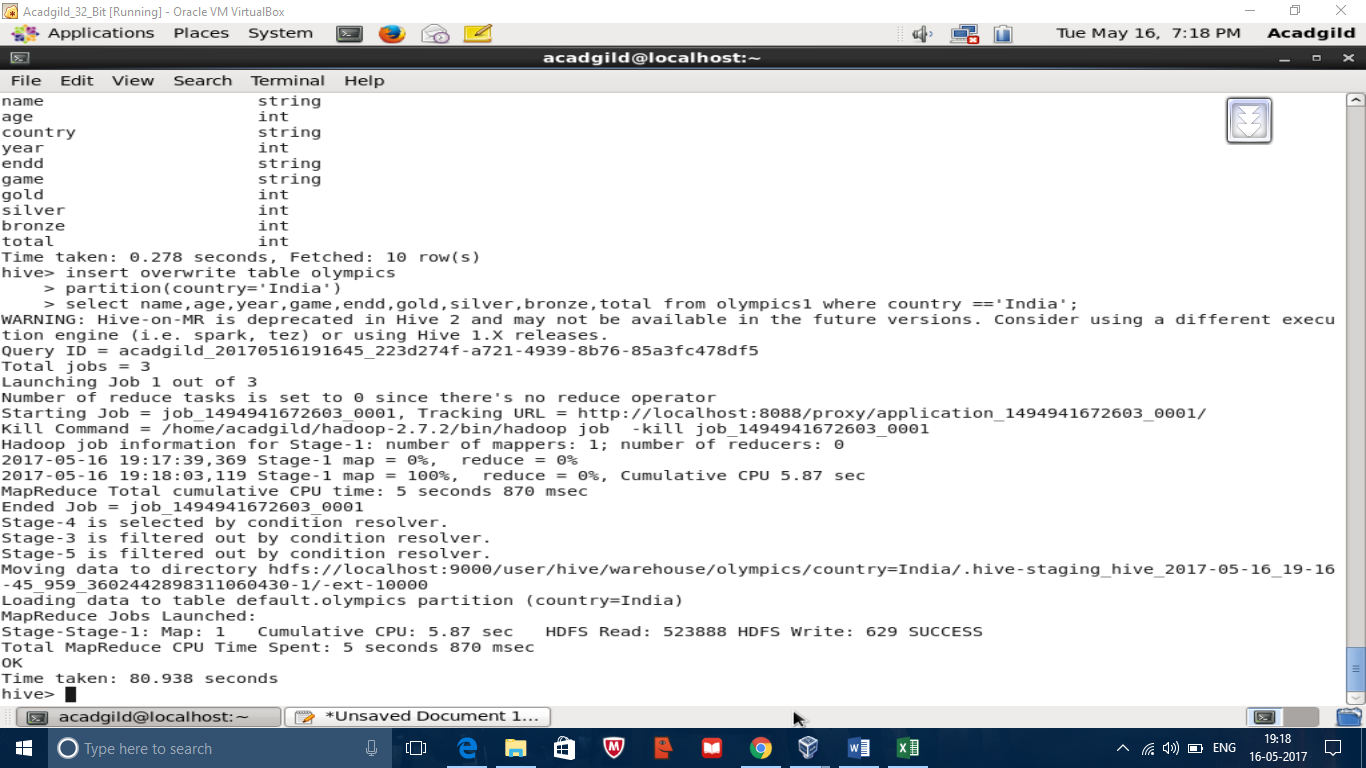
Static partitioning

Dynamic partitioning

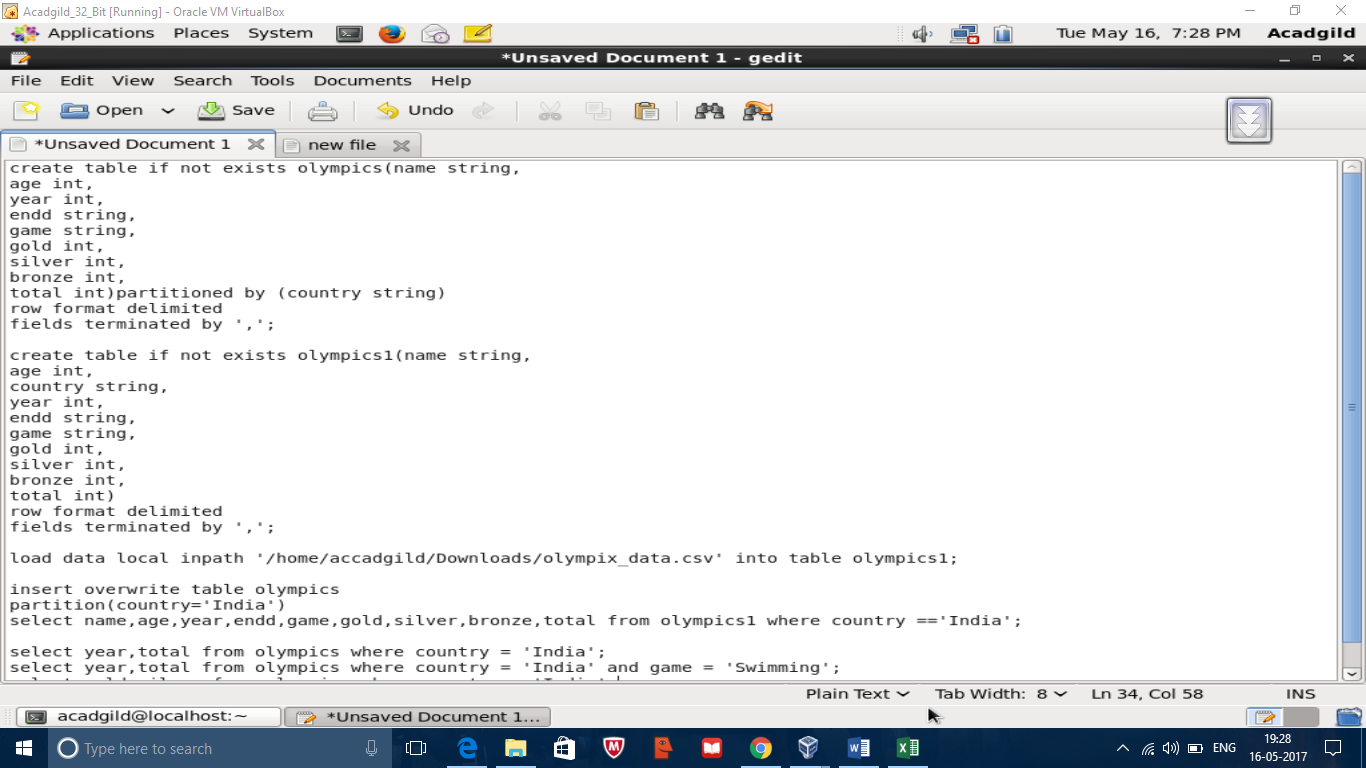
Static partitioning needs to be applied when we know data belongs to which partition In static partitioning, every partition needs to be backed with individual hive statement which is not feasible for large number of partitions as it will require writing of lot of hive statements. In that scenario dynamic partitioning is suggested s we can create as many number of partitions with single hive statement. Dynamic partition are the partition column which are known only at the execution time

Example for static partition

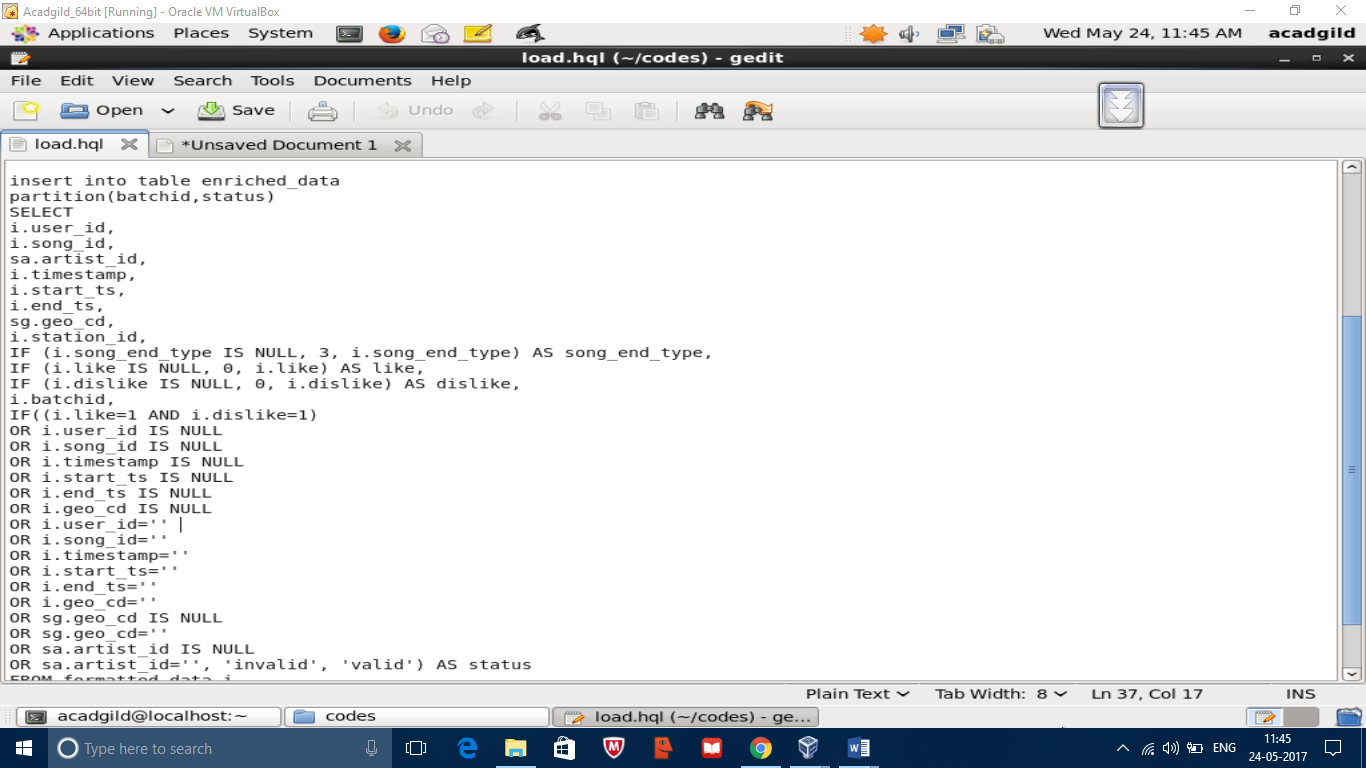
Loading data in static partition

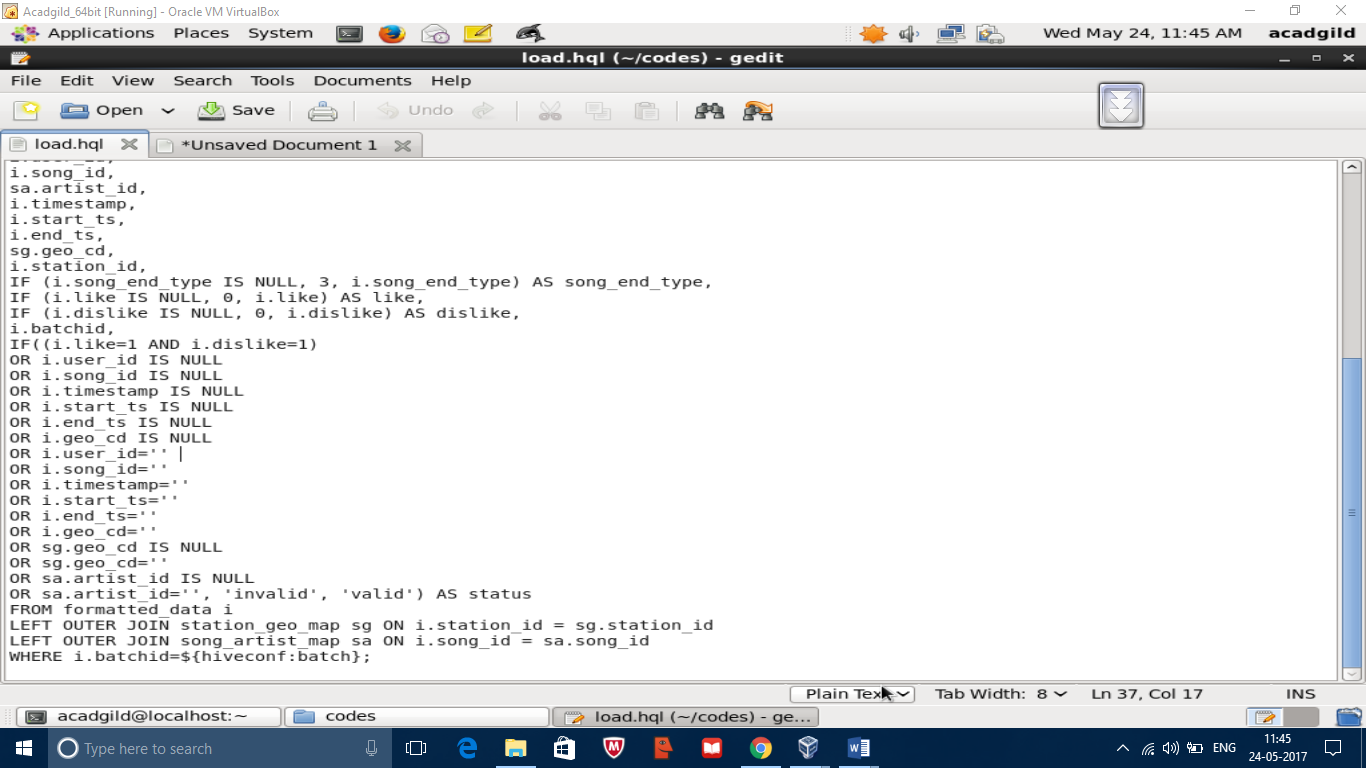


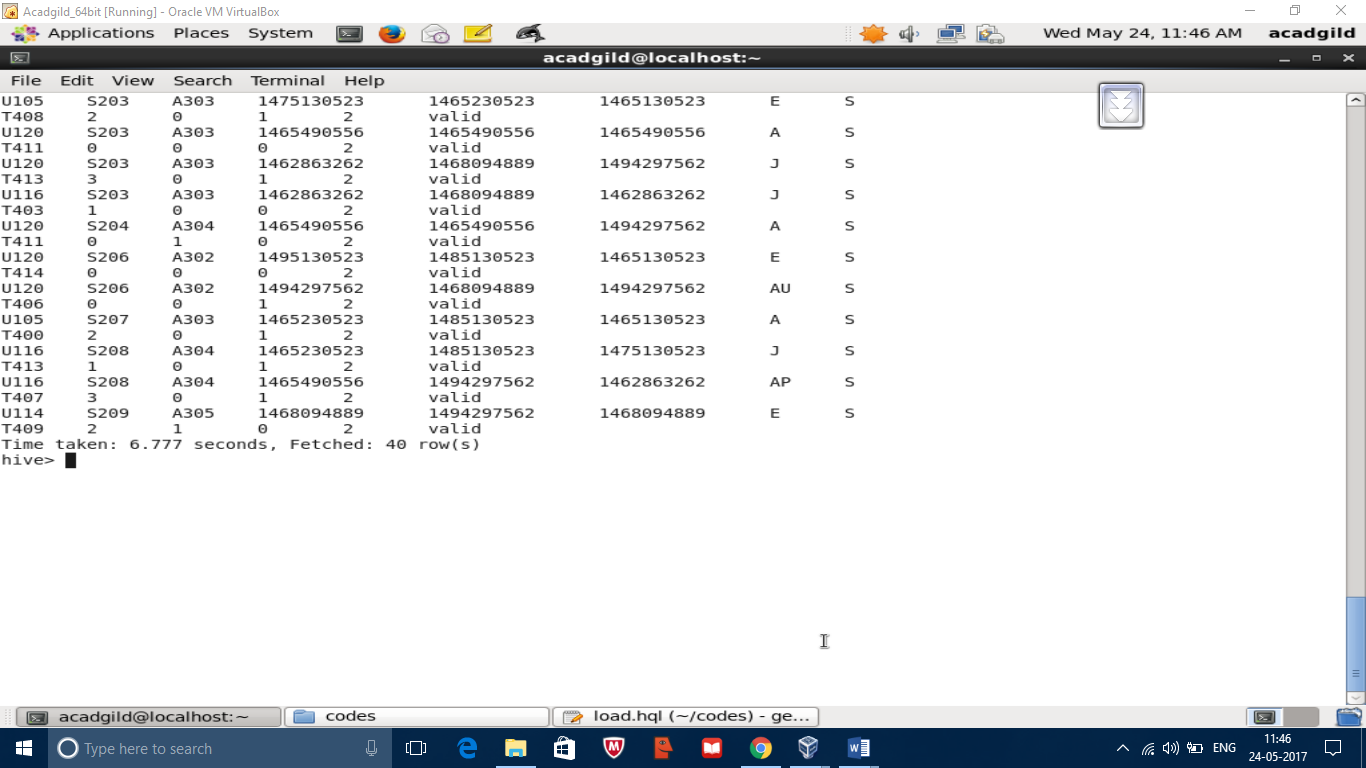
Code



Loadaing data in dynamic partition





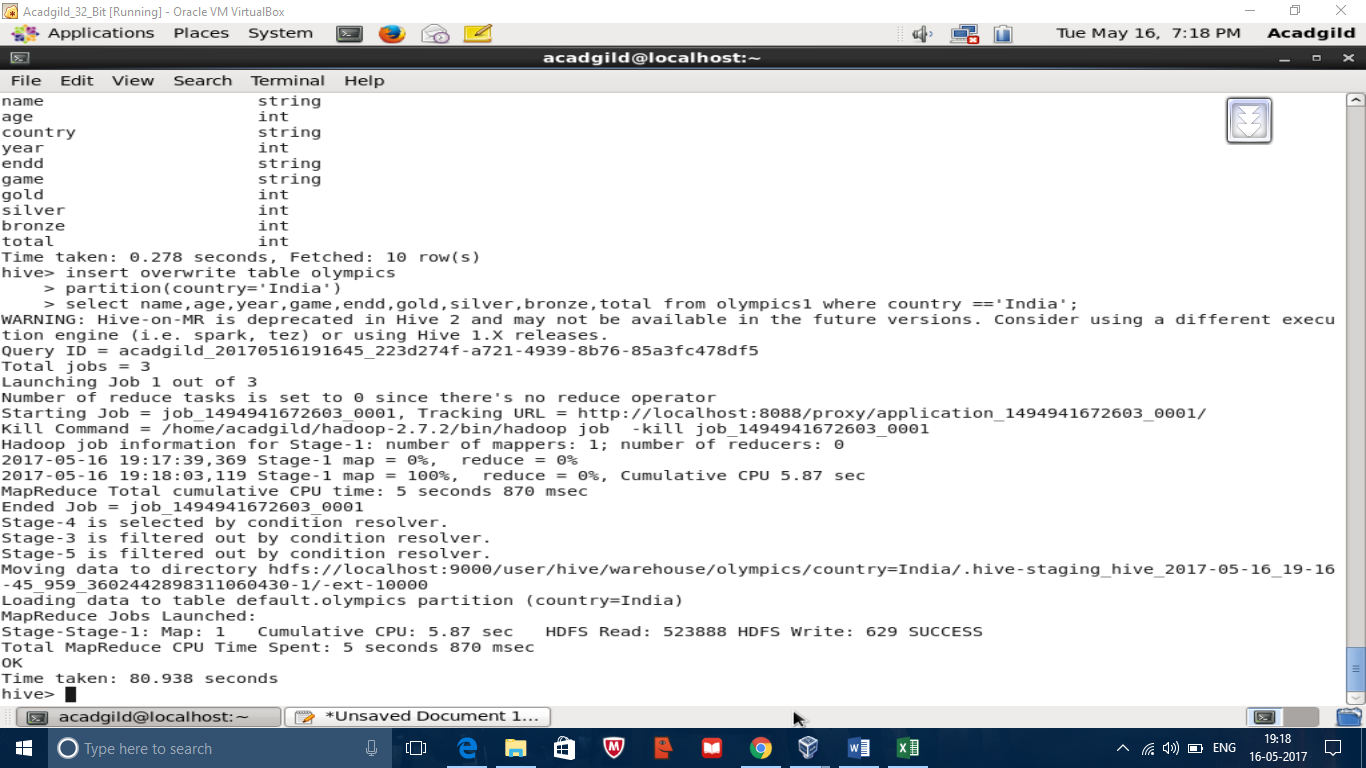


partitioning and bucketing

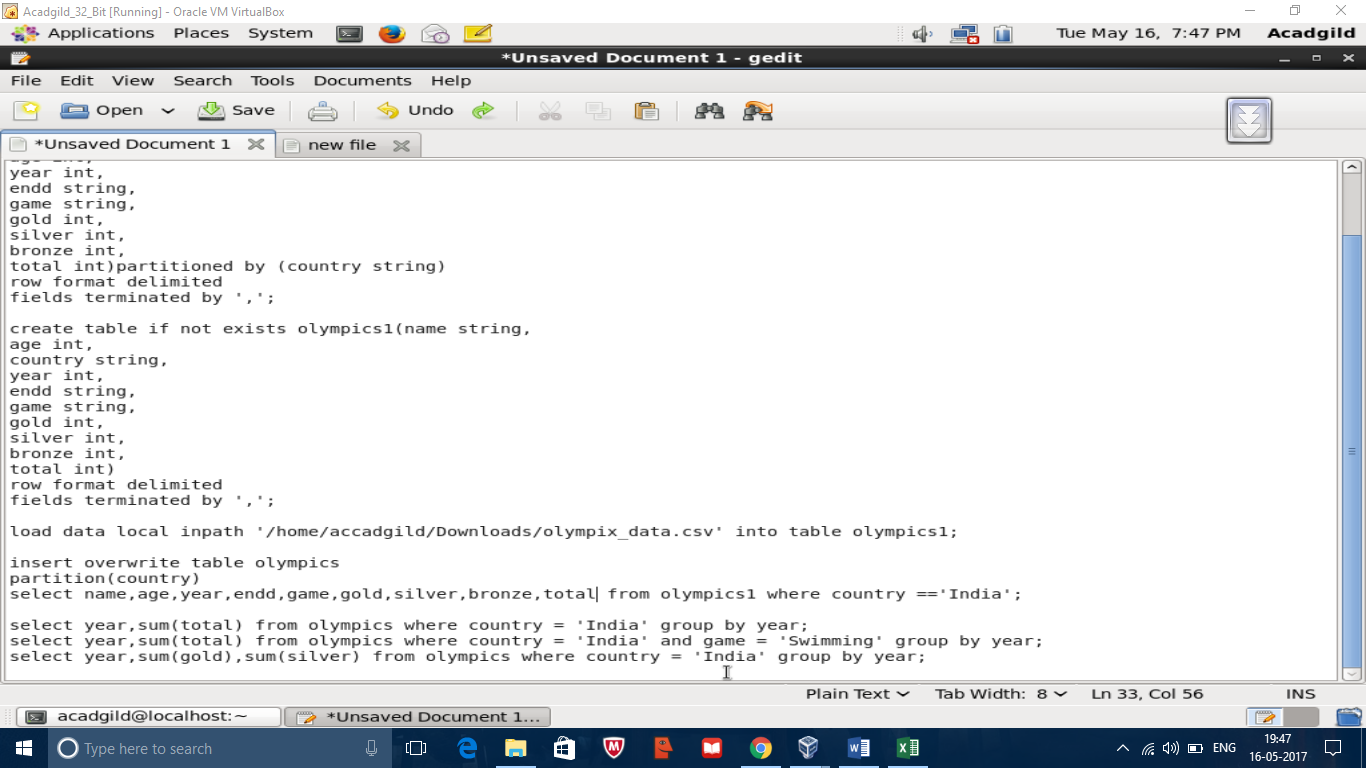
Partitioning data is often used for dristibuting load horizontally, this has performance benefit and helps in organizing data in a logical fashion. Hive table can be partitioned by using partitioned by clause with the column to partition

Bucketing is another technique for decomposing data sets into more manageable parts. For example suppose a table using data as the top level partition and employee\_id as the second level partition leads to too man small partitions. Instead if we bucket the employee table and use employee\_id as the bucketing column the value of this column will be hashed by user-defined number into buckets. Records with same employee\_id will always same in same bucket.

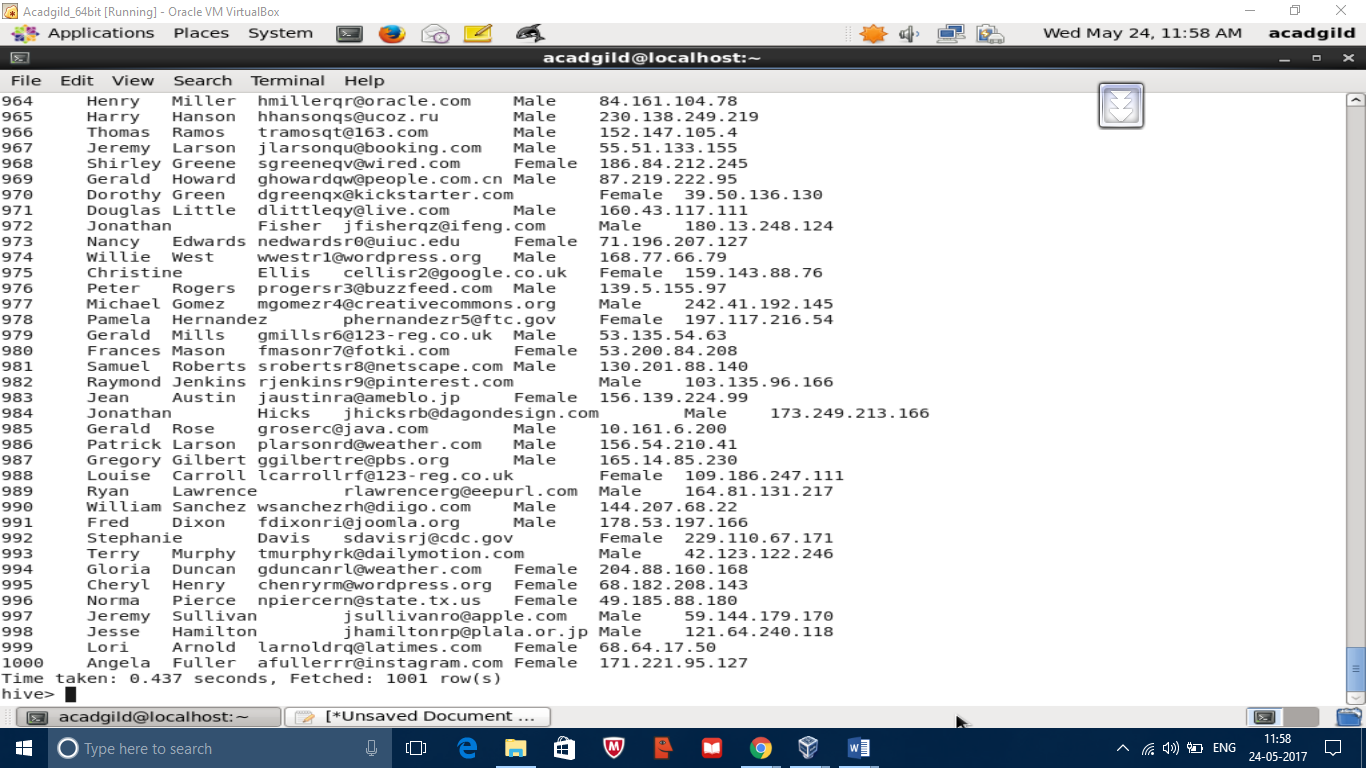
Example for partition



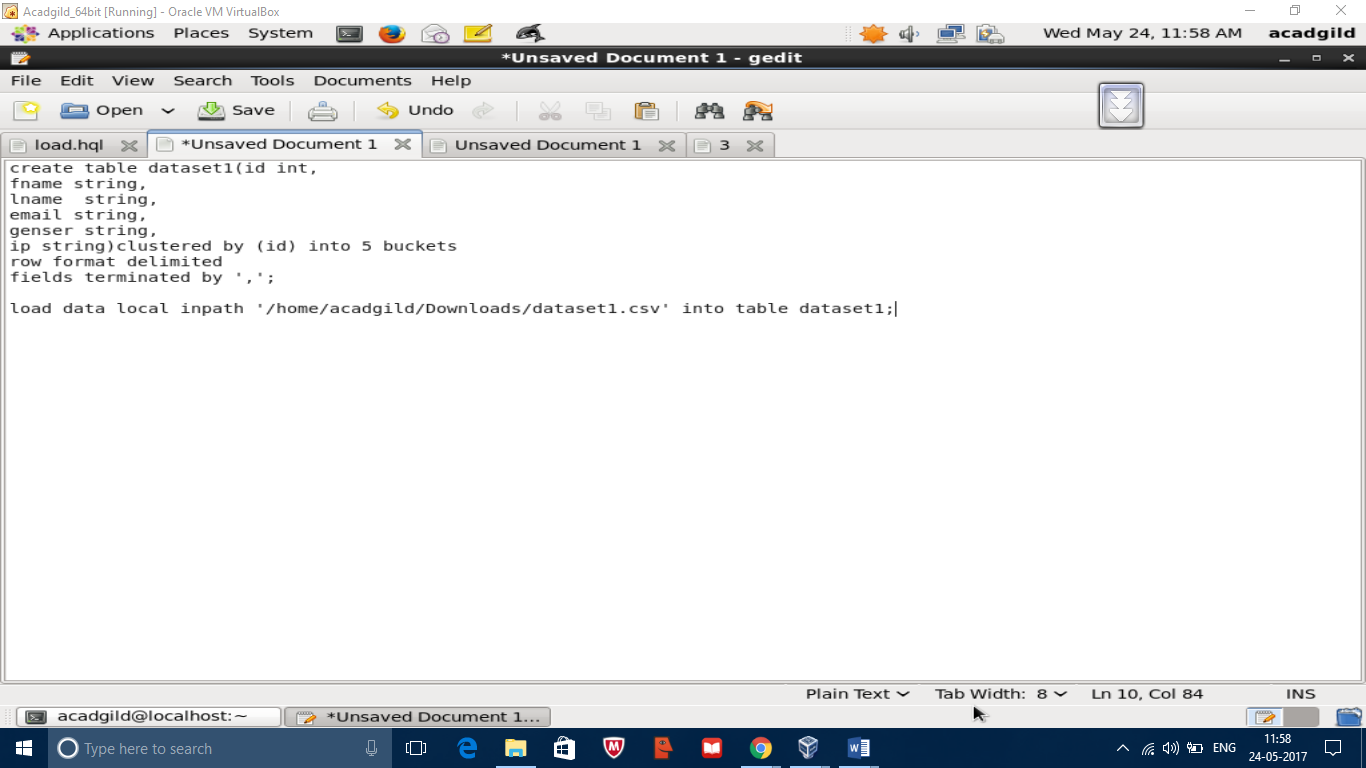
Code



Example for bucketing



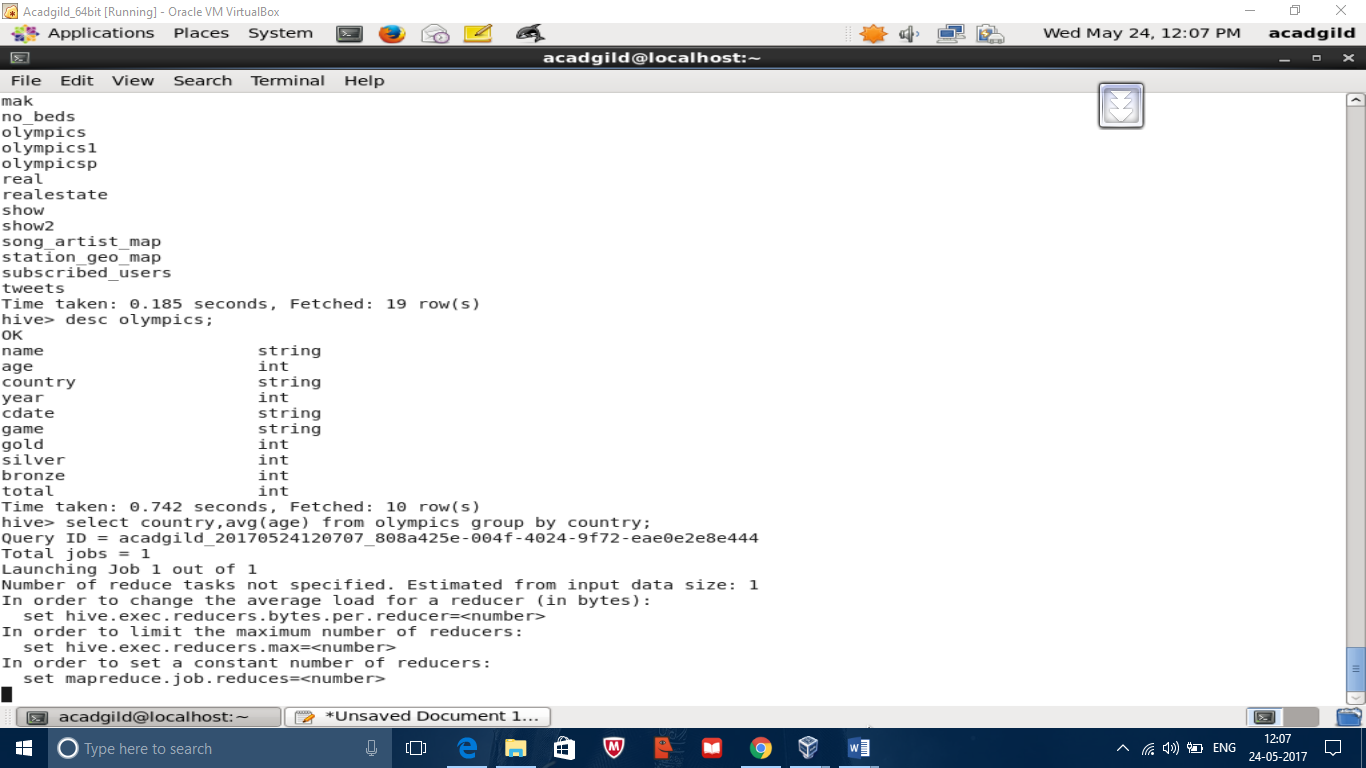
Code

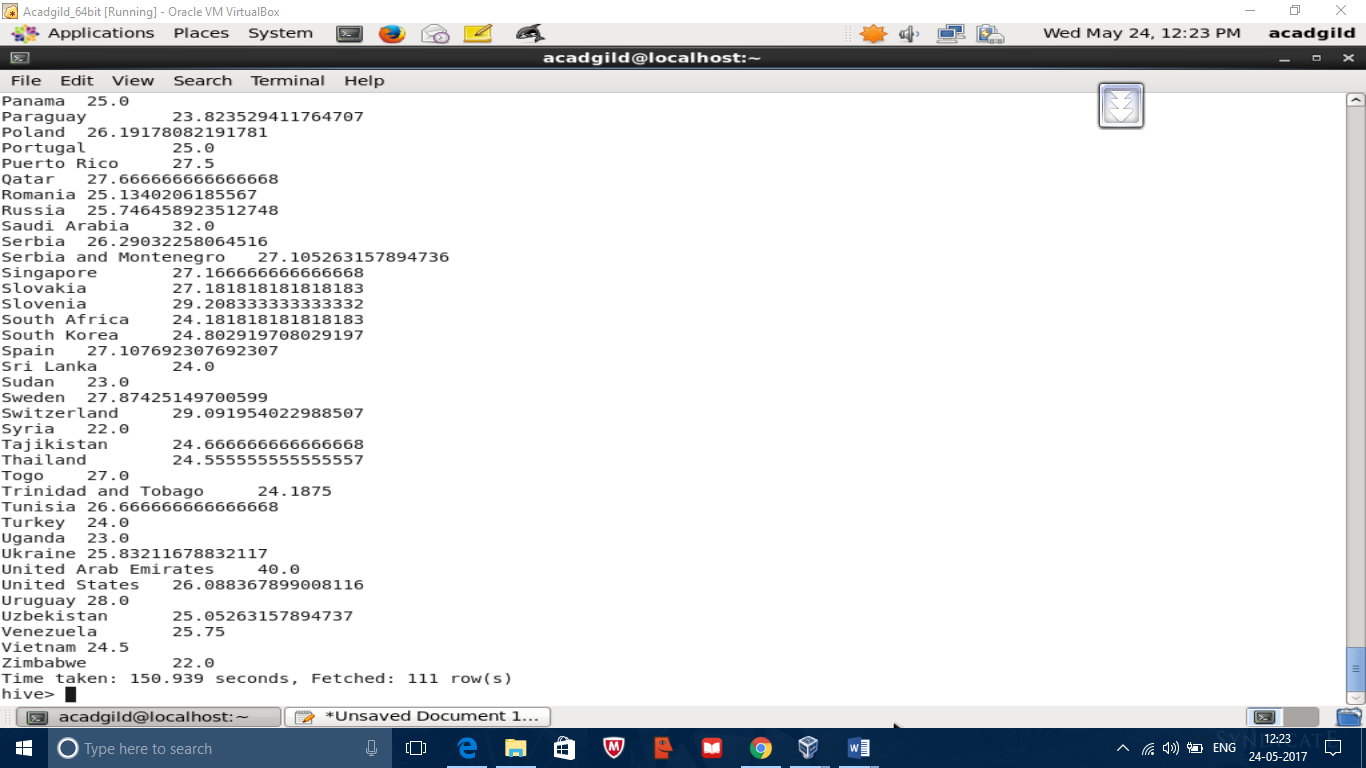


Different types of groups in hive

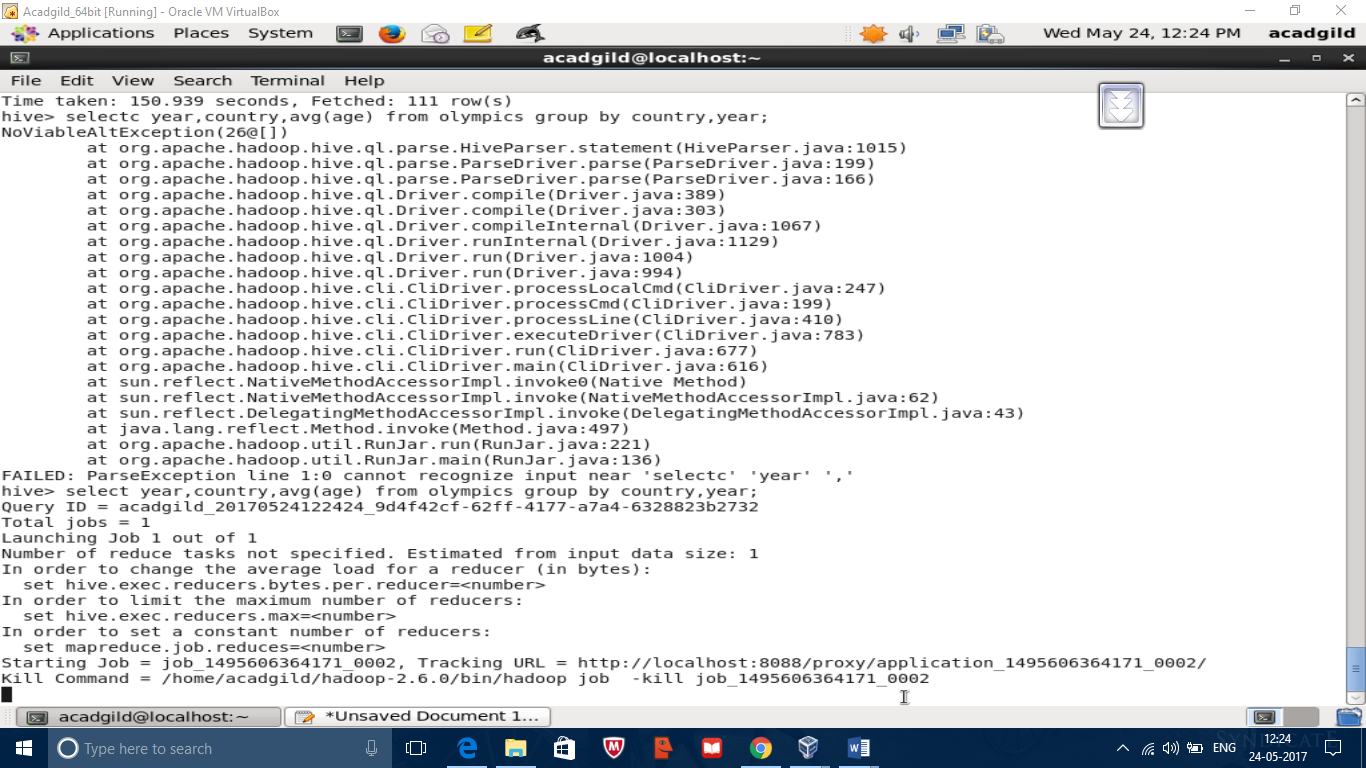
In hive grouping can be done on single column, multiple column, group by having

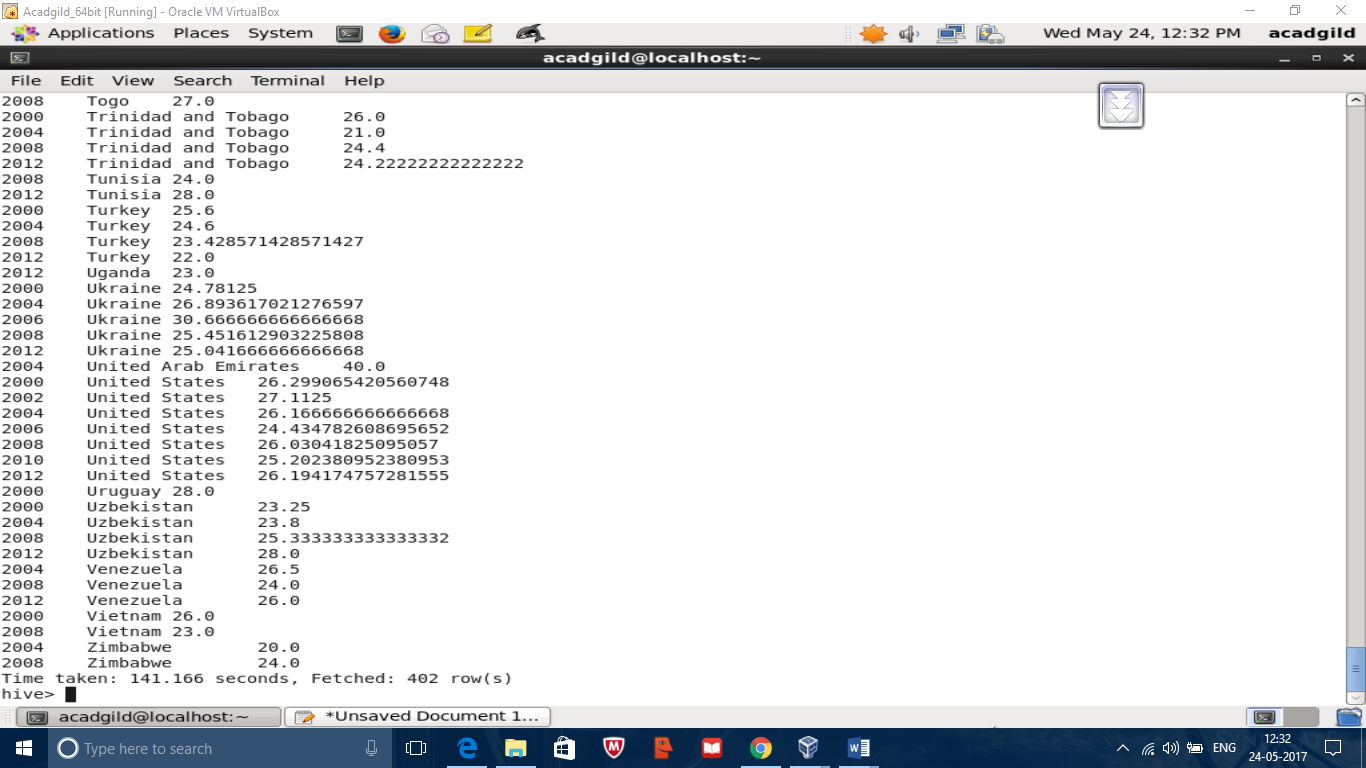
Group by single column





Group with multiple column





Group by having

