**PROBLEM STATEMENT:**

**Explain the key concepts of Bucketing and perform bucketing operations using our attached Blog. Share and explain the commands used with the final result.**

**SOLUTION:**

1. To overcome the problem of partitioning, Hive provides Bucketing concept, which allows user to divide table data sets into more manageable parts. Thus, Bucketing helps user to maintain parts that are more manageable and user can set the size of the manageable parts or Buckets too.

2. Hive partition divides table into number of partitions and these partitions can be further subdivided into more manageable parts known as Buckets or Clusters.

3. The Bucketing concept is based on Hash function, which depends on the type of the bucketing column. Records which are bucketed by the same column will always be saved in the same bucket.

4. Here, CLUSTERED BY clause is used to divide the table into buckets.

5. In Hive Partition, each partition will be created as directory. But in Hive Buckets, each bucket will be created as file.

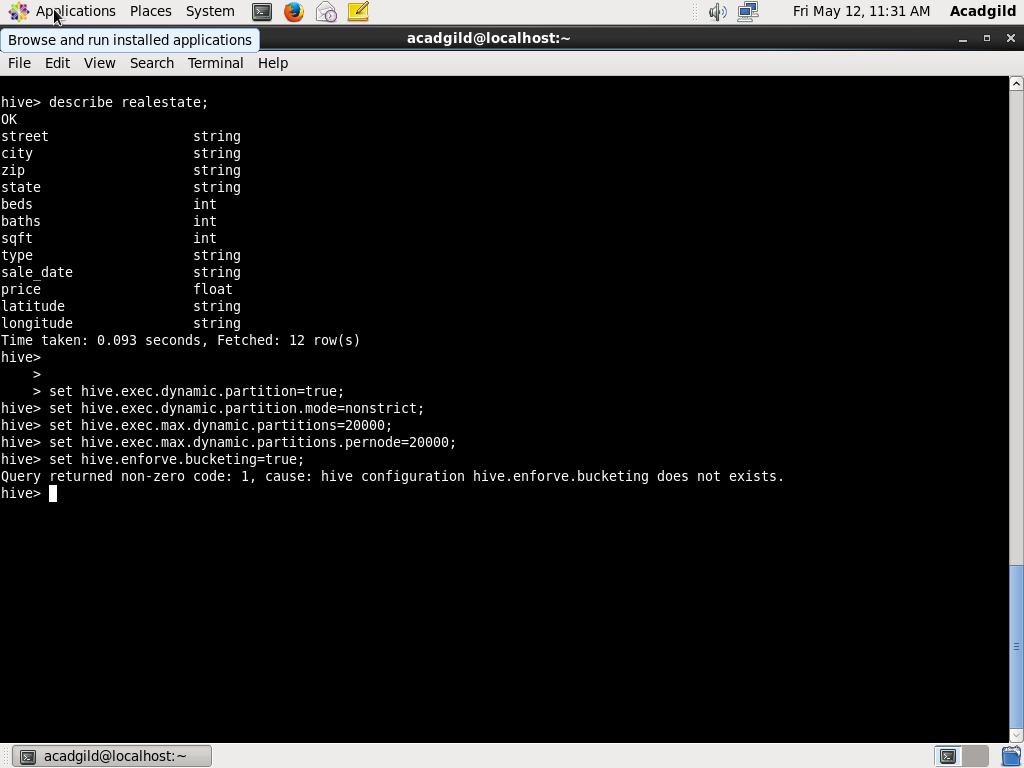
6. Bucketing can also be done even without partitioning on Hive tables.

7. Bucketed tables allows much more efficient sampling than the non-bucketed tables. With sampling, we can try out queries on a section of data for testing and debugging purpose when the original data sets are very huge. Here, the user can fix the size of buckets according to the need.

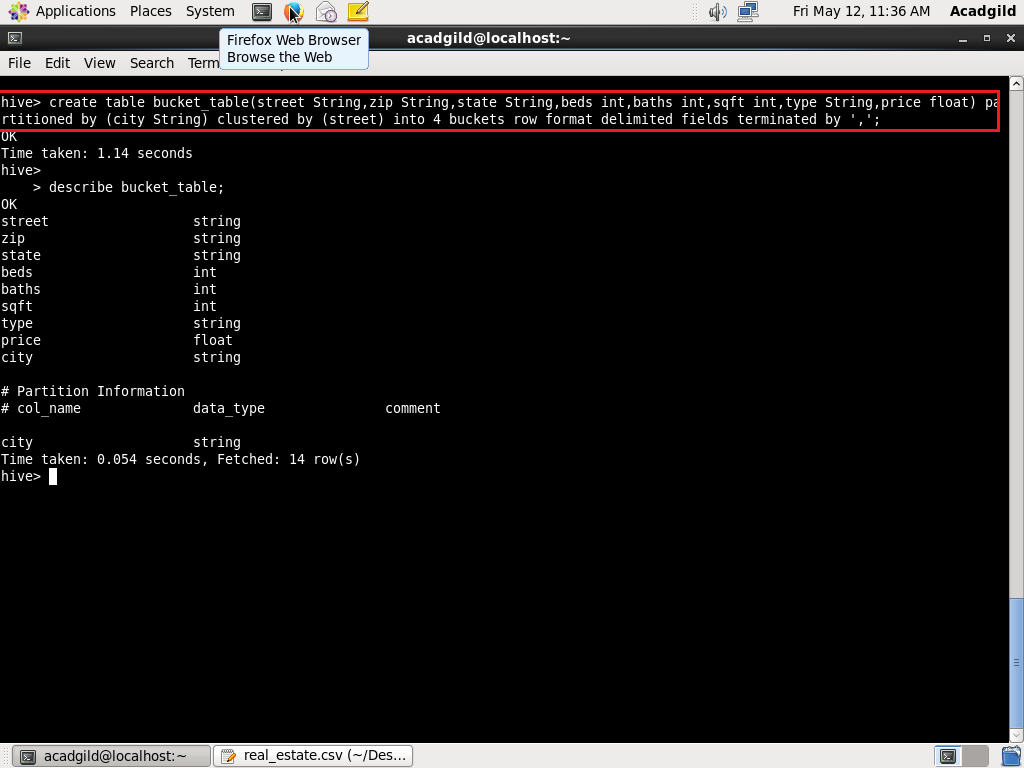
8. Bucketing concept also provides the flexibility to keep the records in each bucket to be sorted by one or more columns. Since the data files are equal sized parts, map-side joins will be faster on the bucketed table.

//The table ‘realestate’ contains entire dataset.

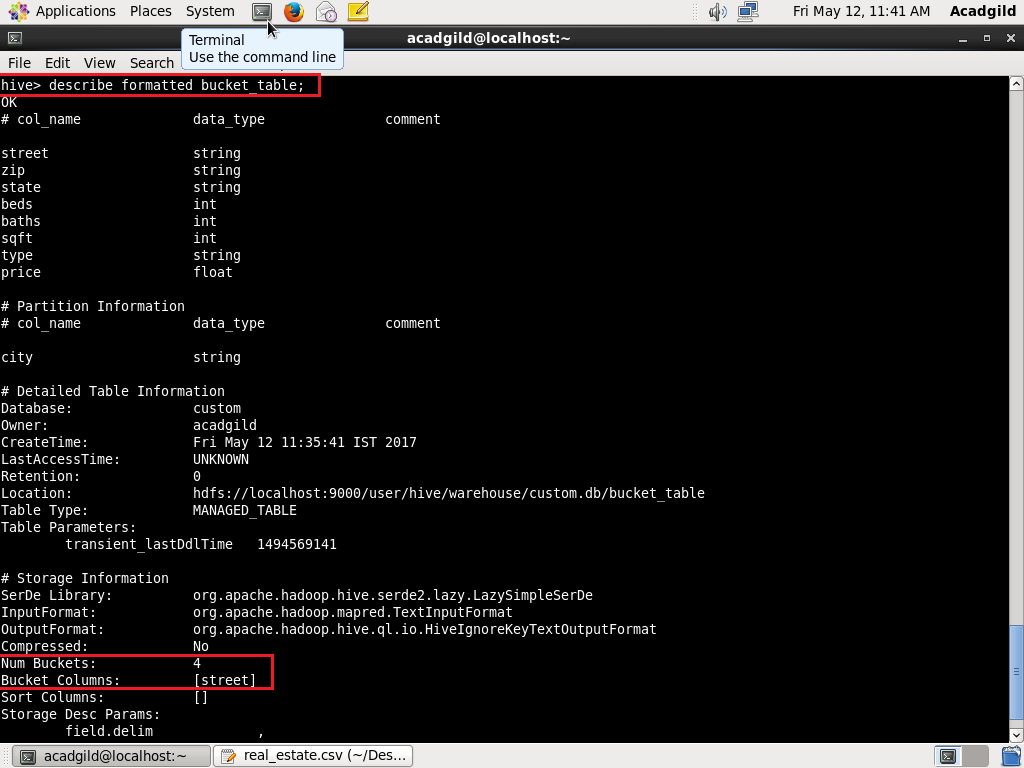
// The properties are set to enable partitions and bucketing.



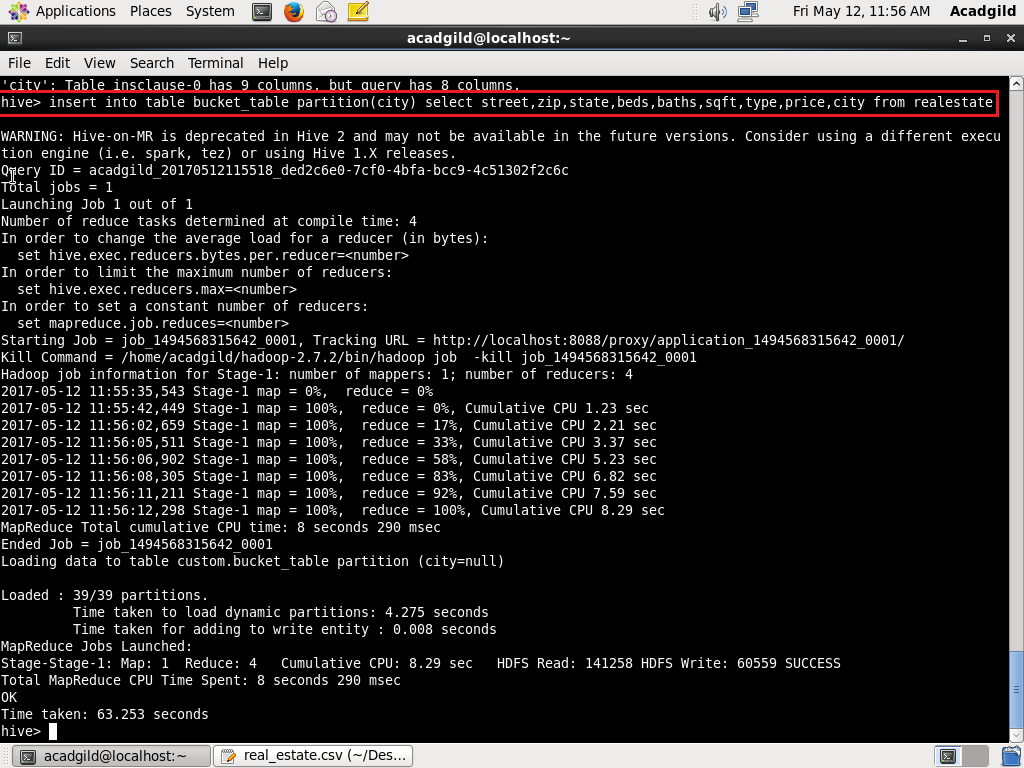
// creating bucket table

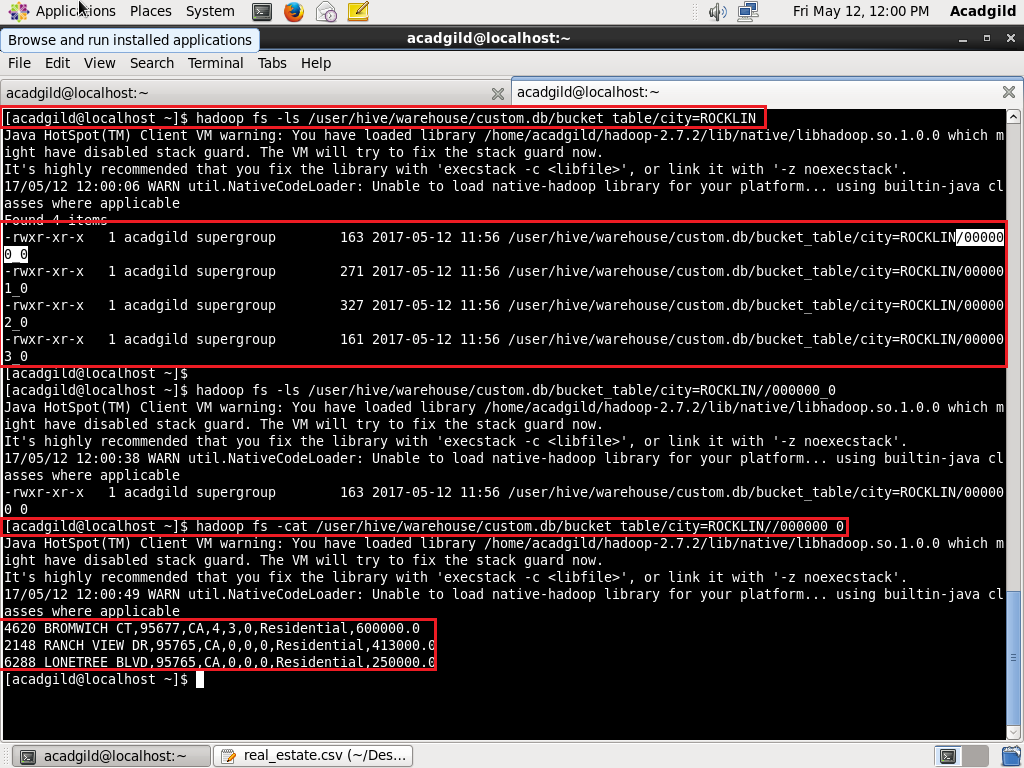
****

//displaying table details



//inserting data into bucket table





// The directory of the partition city=ROCKLIN contains four files because the number of buckets is set to 4.

//contents one file are displayed