**ASSIGNMENT 27.3:**

**Explain the key concepts of Bucketing:**

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.

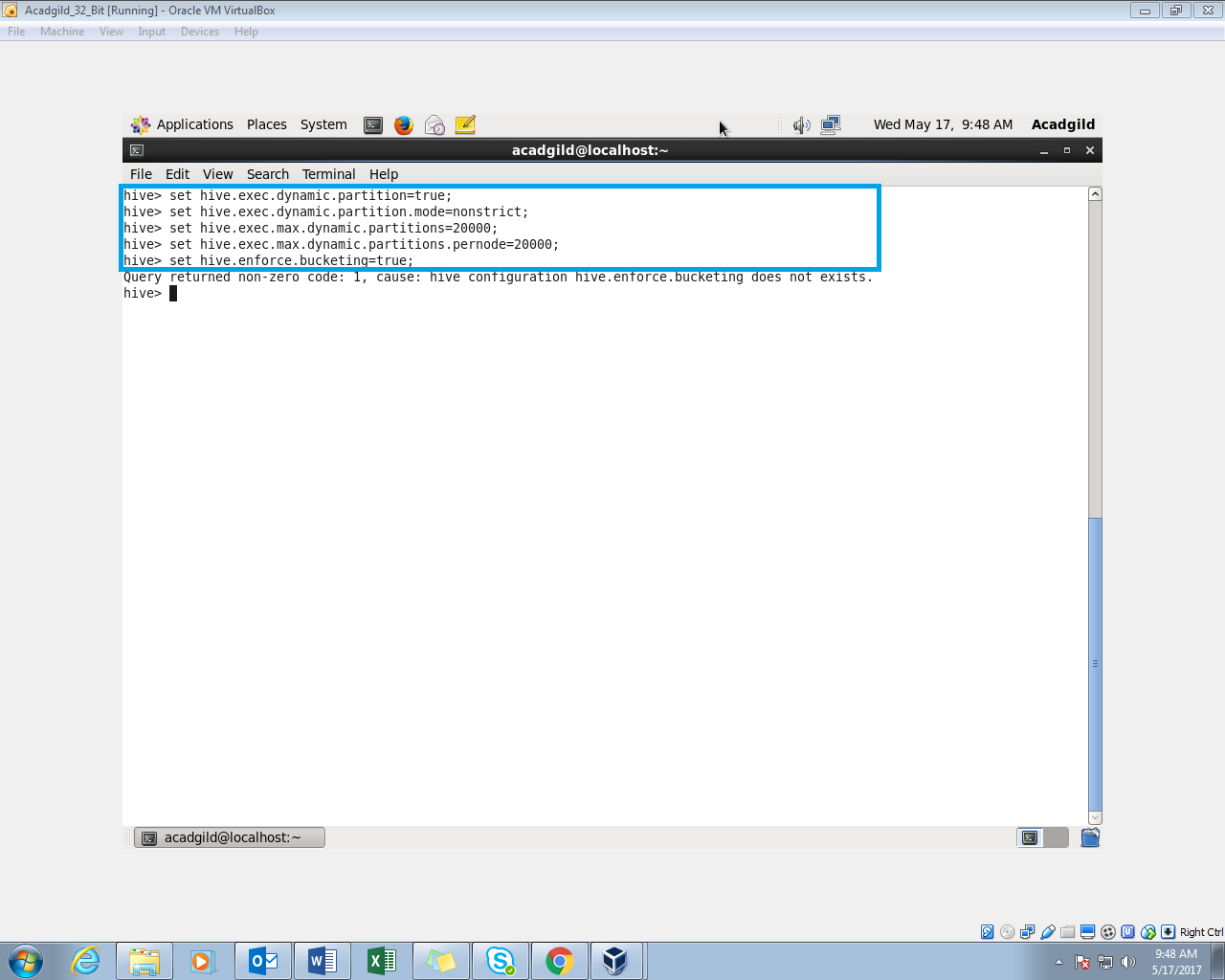
4. Here, CLUSTERED BY clause is used to divide the table into buckets.

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

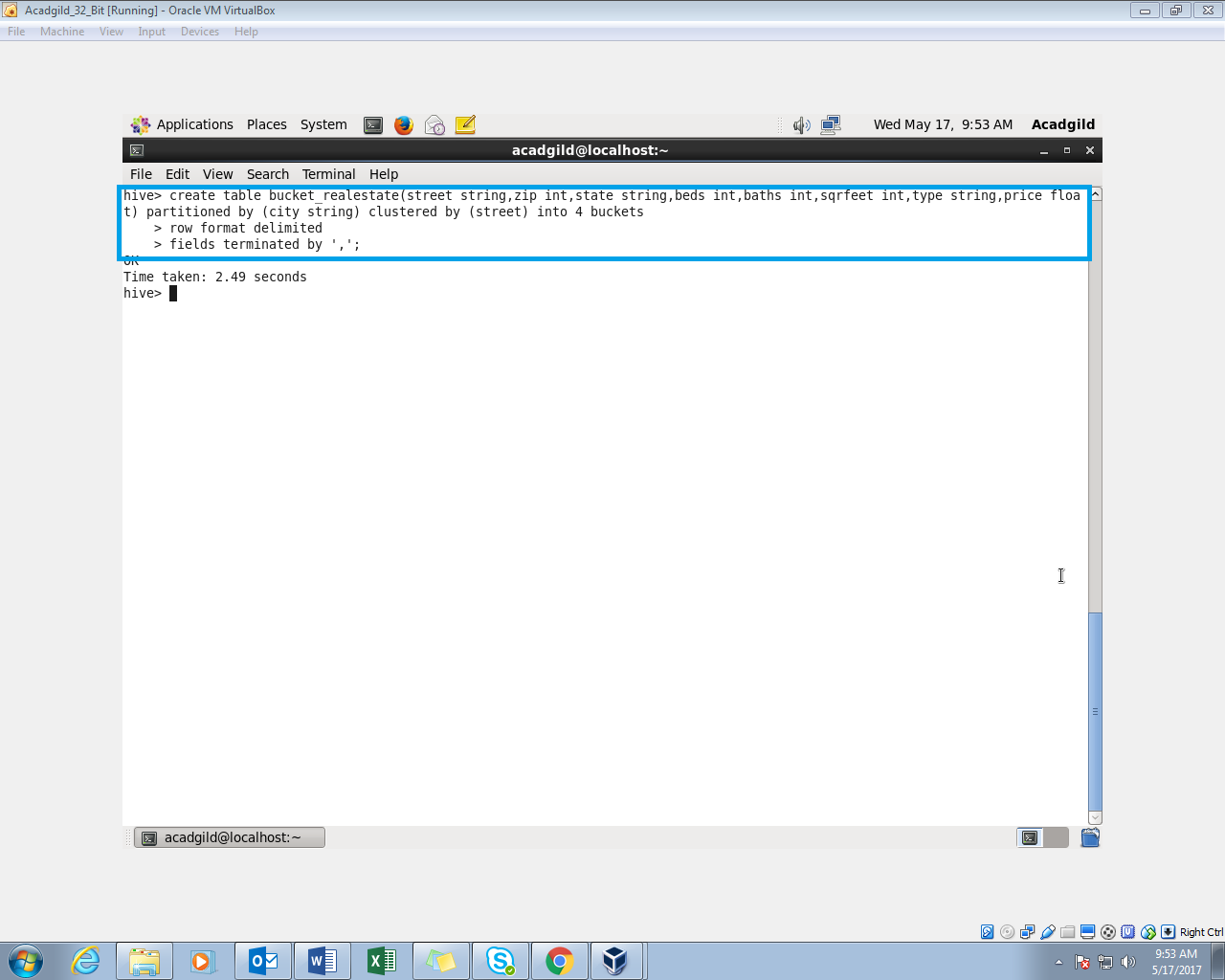
6. Bucketing can also be done even without partitioning on Hive tables

7. 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.

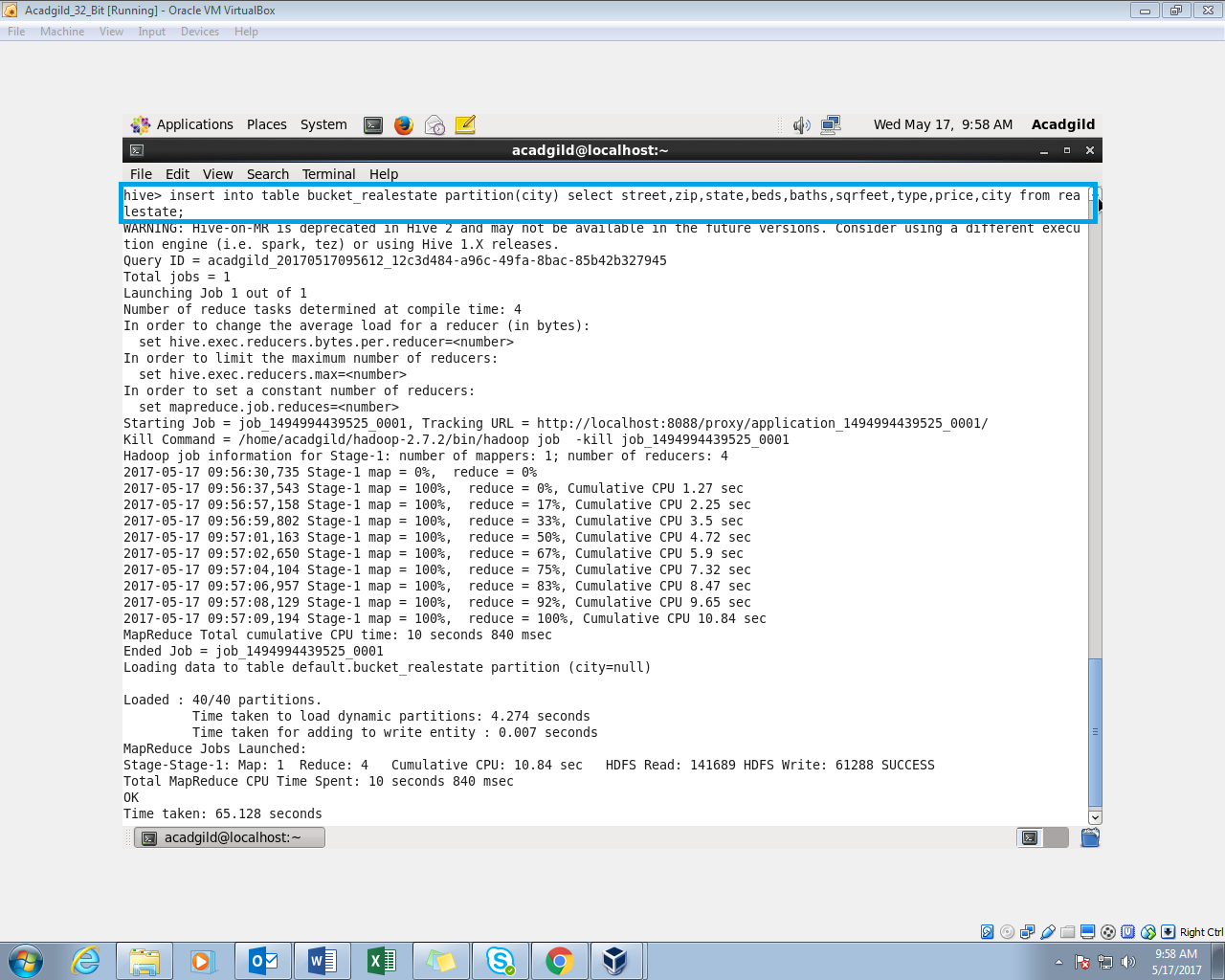
Setting the properties:



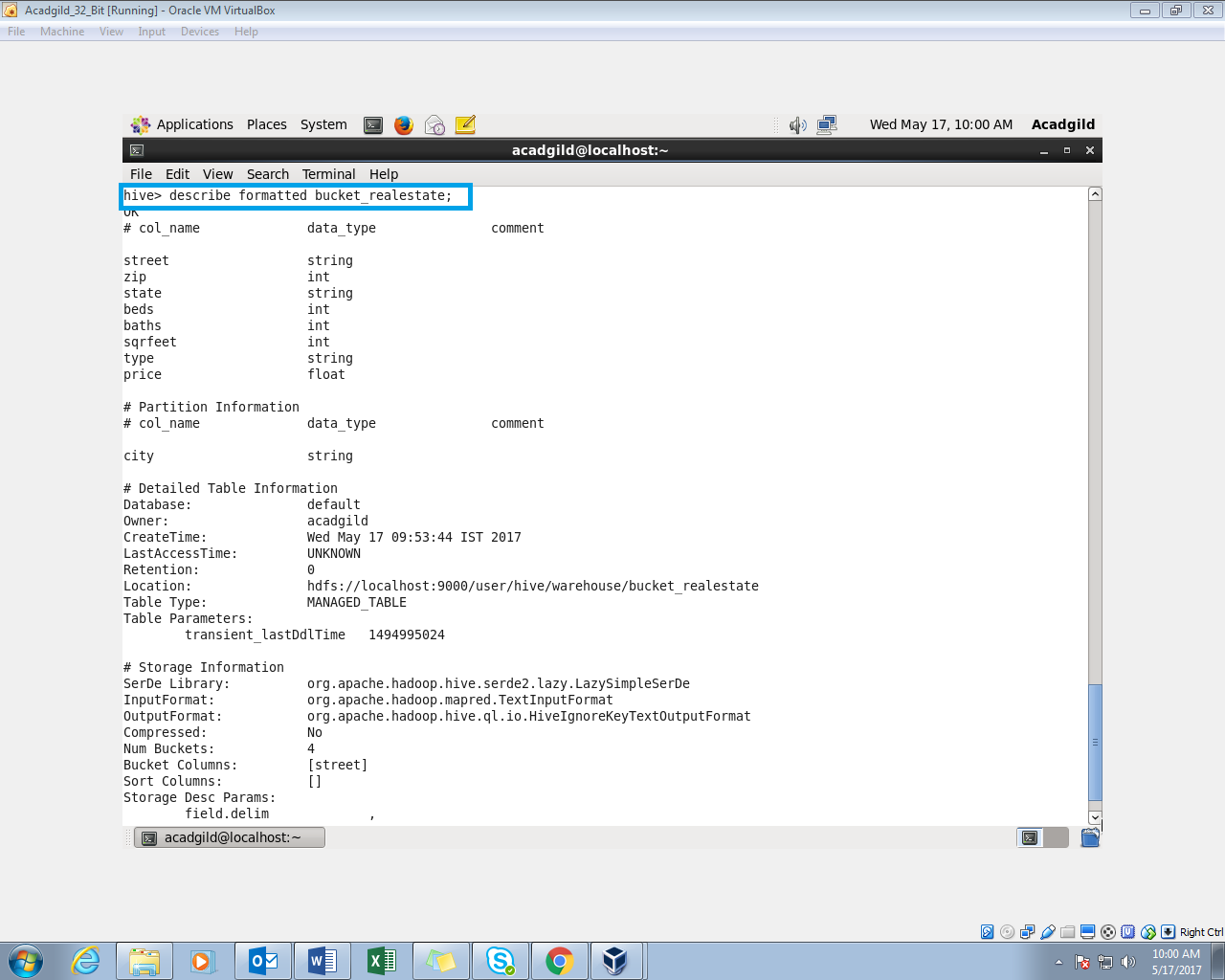
Creating the bucket table:



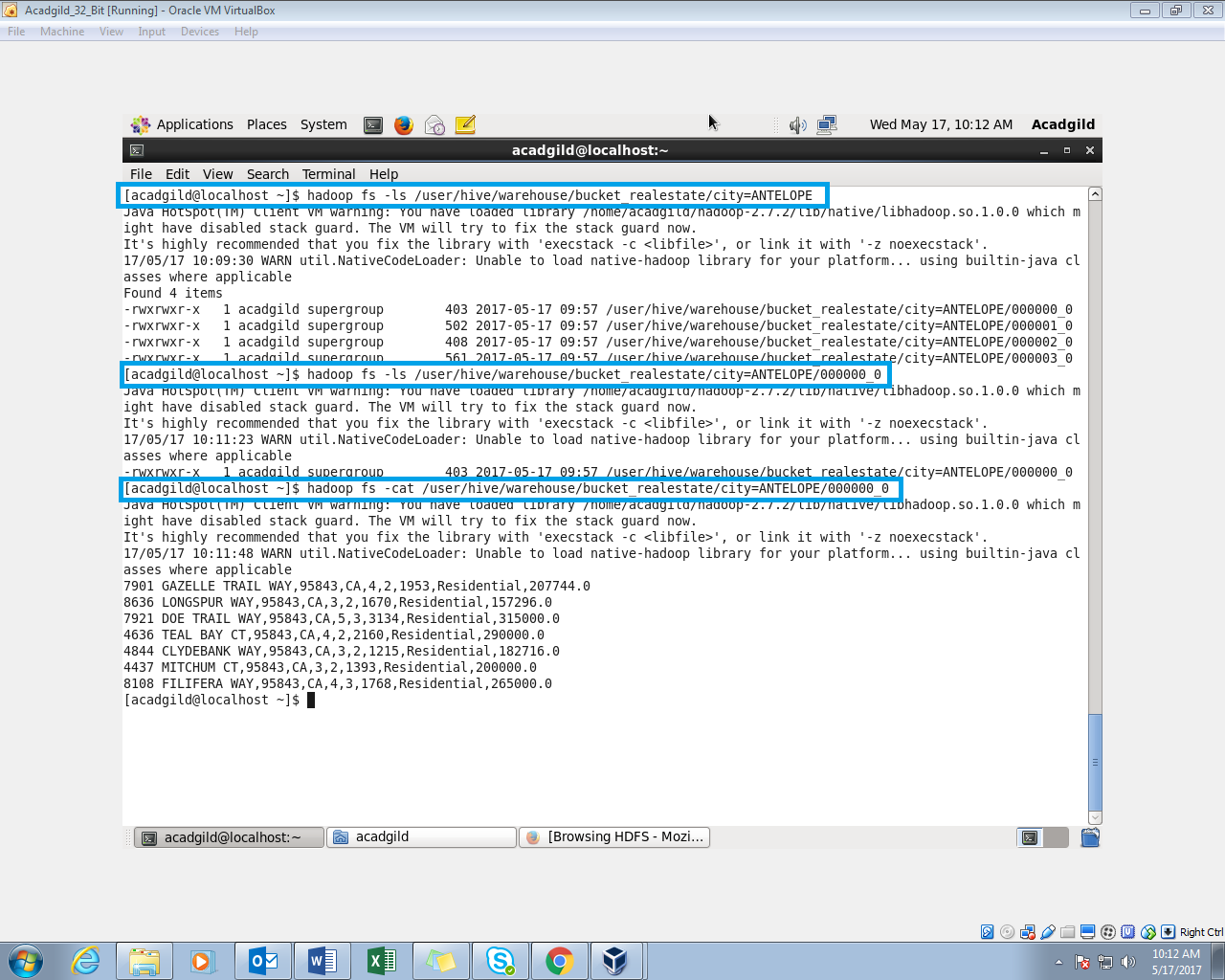
Inserting the data into bucketed table:



Describing the bucketed table:



Buckets are created:



Displaying the contents of the bucket. For each city, four buckets are created.

Browsing the hdfs directory:

