

**Solution-**

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

**Advantages of Bucketing-**

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.

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 tables.

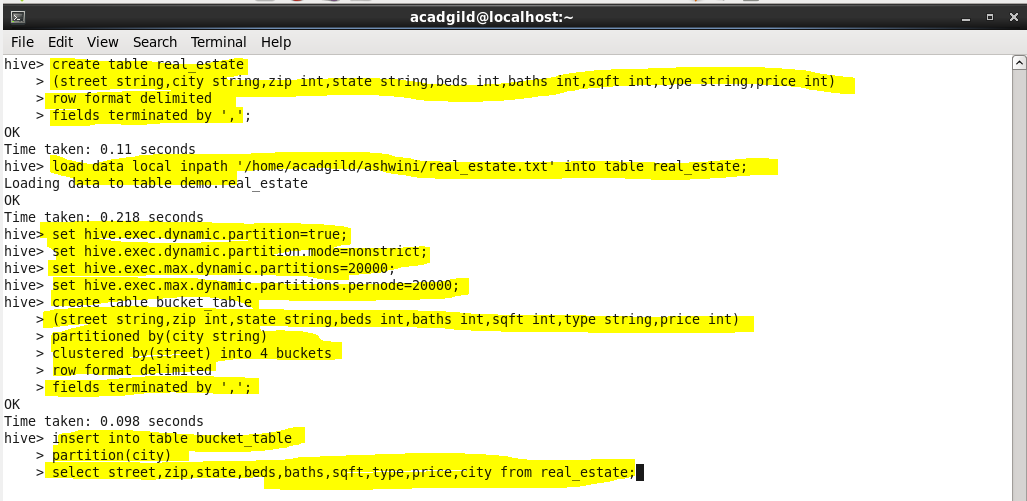
**Example-**

**Input file-**

****

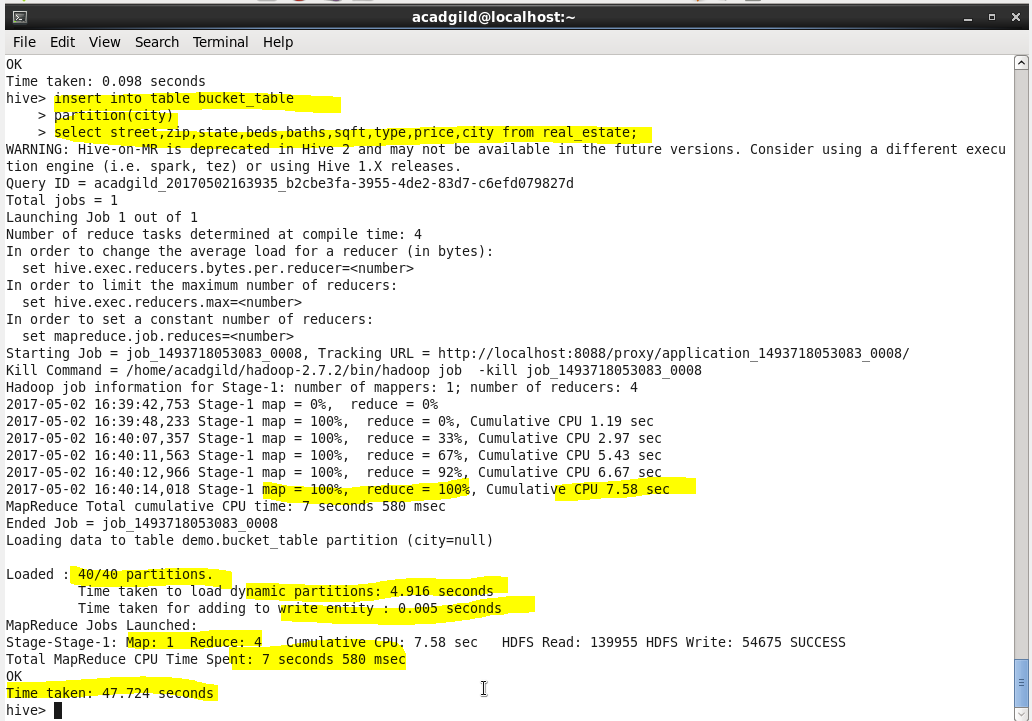
**Step 1-**

* Creation of table named real\_estate.
* Inserting data into the table.
* Setting hive properties for bucketing.
* Creating bucket table which is partitioned by the city.



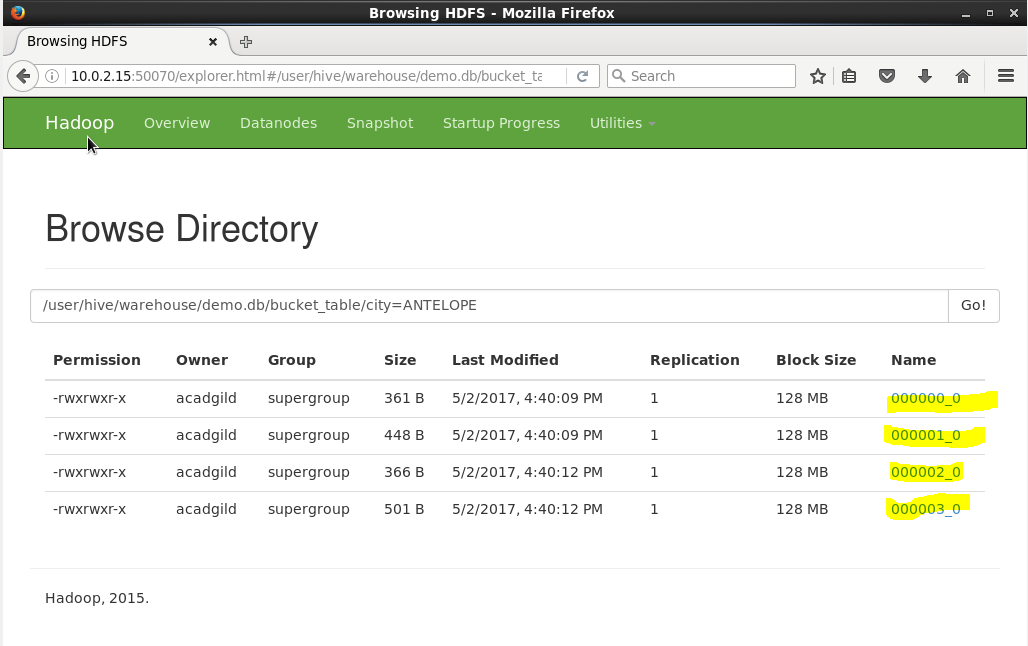
**Step 2-**

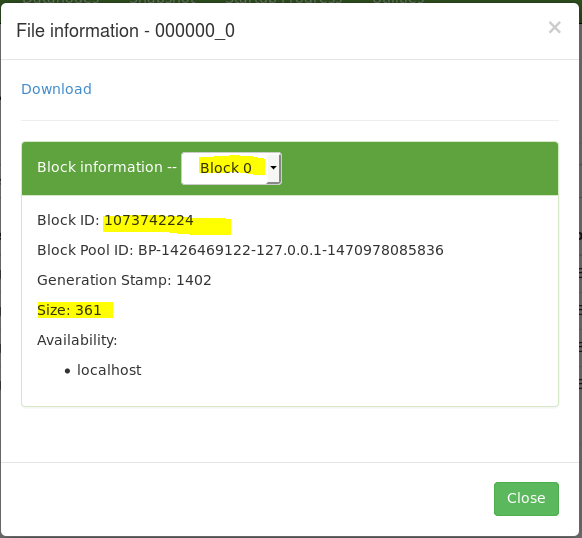
* Inserting data into bucket table, which is partitioned by city.
* Select query to display data from bucket table.



**Output-**

* As we can see 4 buckets have been created, because we have specified the number of buckets.





* If we download the partition and display, it will give the data.

