**Bucketing in Hive**

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. In Hive Buckets, each bucket will be created as file. Bucketing can also be done even without partitioning on Hive tables.

**Advantage:**

1) Bucketed tables allows much more efficient sampling than the non-bucketed tables.

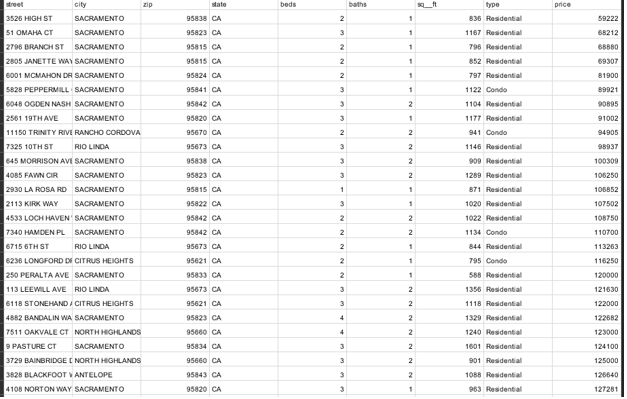
2) With sampling, we can try out queries on a section of data for testing and debugging purpose when the original data sets are very huge. Here, the user can fix the size of buckets according to the need.

3) Bucketing concept also provides the flexibility to keep the records in each bucket to be sorted by one or more columns.

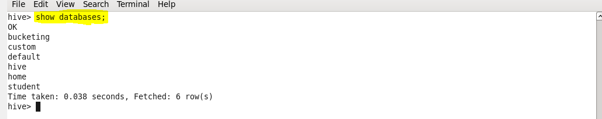
4) Since the data files are equal sized parts, map-side joins will be faster on the bucketed tables.

Ex:

*Input Dataset :*



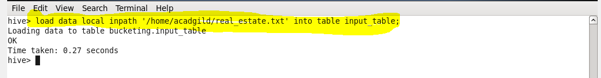
Selection of the data set:



Creation of a new i/P table:

C:\Users\613008\Desktop\a3.PNG

Load the i/P data set:



***Display the Contents of Table I/P table:***

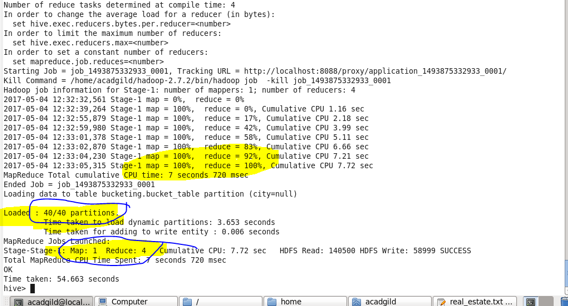
******

***Setting the hive properties:***

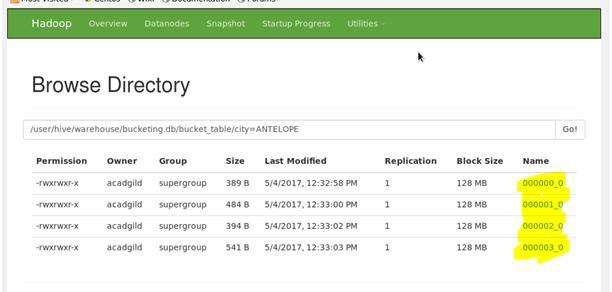
******

***Creating Bucket Table &Query :***

******

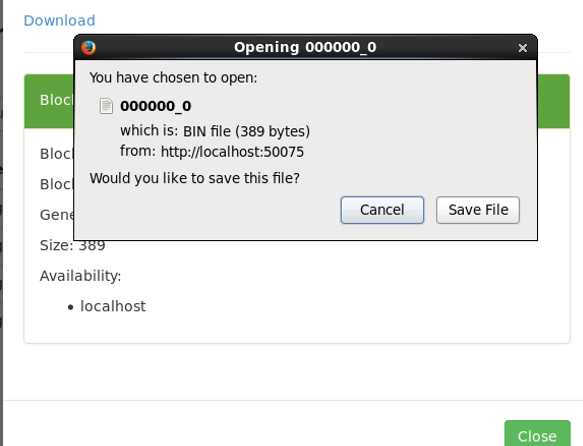
******

***Browse Directory:***

******

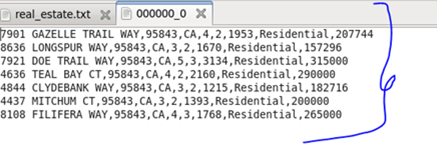
***Browse file:***

******

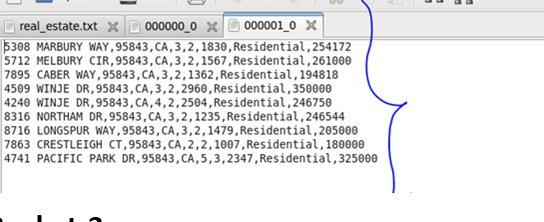
******

***O/P:***

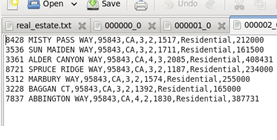
***Bucket1***

******

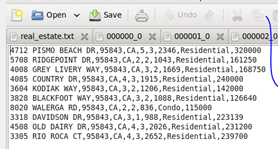
***Bucket2***

******

***Bucket 3***

******

***Bucket 4***

******