**ASSIGNMENT 27.3**

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

Let’s assume a condition that there is a huge [dataset](https://acadgild.com/big-data/big-data-development-training-certification). At times, even after partitioning on a particular field or fields, the partitioned file size doesn’t match with the actual expectation and remains huge and we want to manage the partition results into different parts. To overcome this problem of partitioning, Hive provides Bucketing concept, which allows user to divide table data sets into more manageable parts.

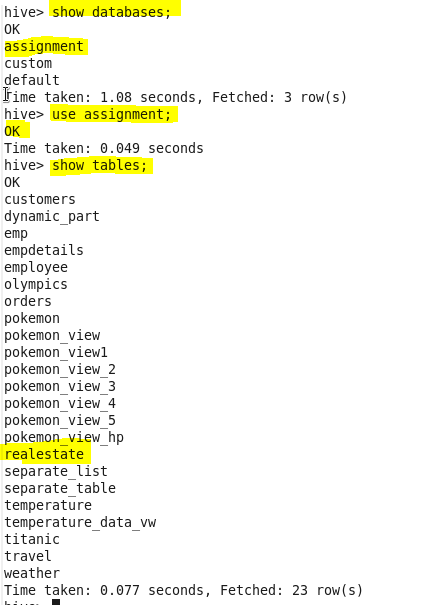
Thus, Bucketing helps user to maintain parts that are more manageable and user can set the size of the manageable parts or Buckets too.

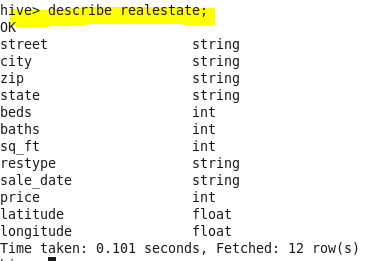
**Bucketing features in hive:**

* 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.
* 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.

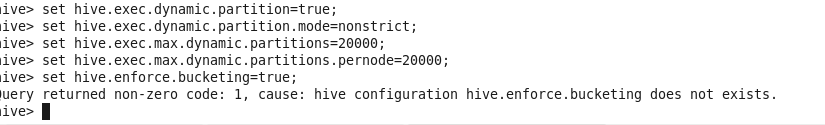
**Example:**

We start by selecting a database and performing queries on the bucketed table.

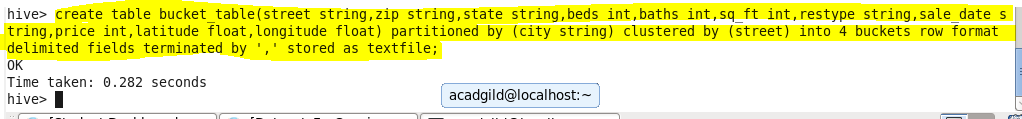




Set the Below Properties in Hive Command Line Before Proceeding Further for Bucketing Scripts

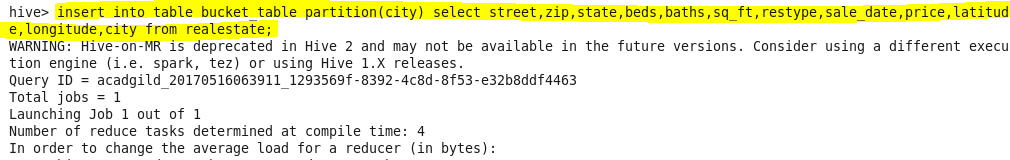


Creating Bucket Table:

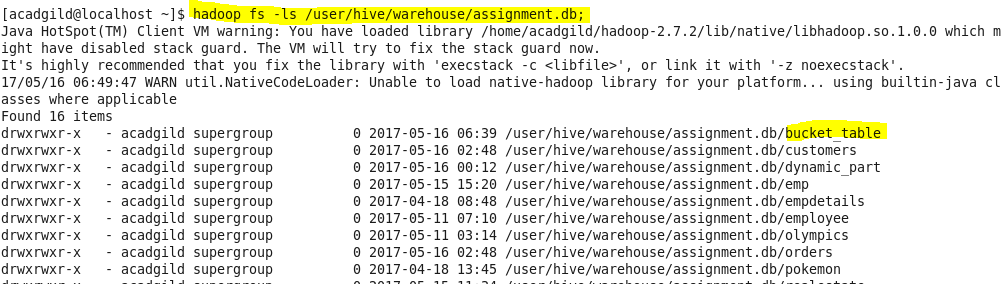


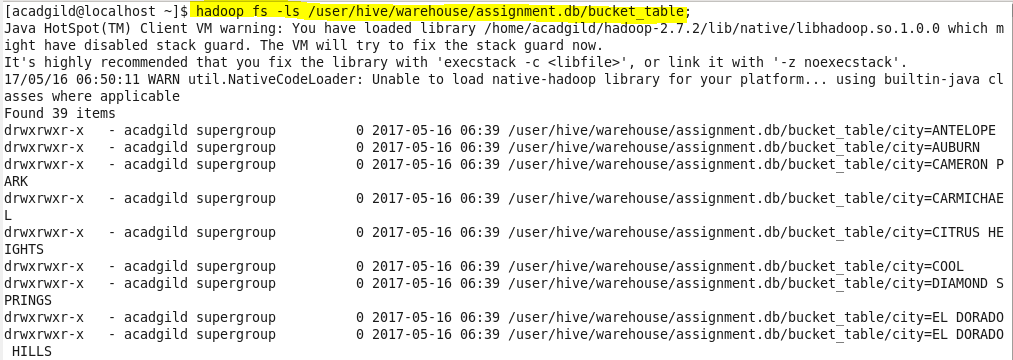
 we have decomposed Hive Buckets into ***‘4’***parts.

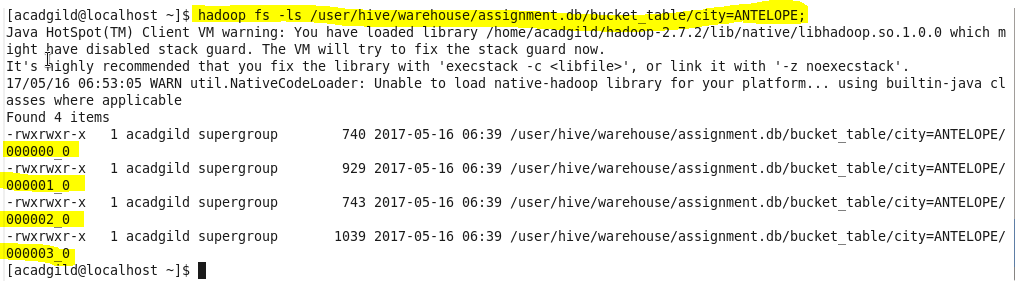
Inserting data in the bucket table



Here the bucketed table has been created







Hence for every city 4 files are created

**Output:**

