PARTITION:

**Hive** organizes tables into **Partitions**. It is a way of dividing a table into related parts based on the values of **partitioned** columns such as date, city, and department.

Using **partition**, it is easy to query a portion of the data.

LIMITATIONS:

##### Partitioning gives effective results when,

##### 1.There are limited number of partitions,

##### 2.Comparatively equal sized partitions.

          But this may not possible in all scenarios, like when are partitioning our tables based geographic locations like country, some bigger countries will have large partitions where as small countries data will create small partitions . So, In these cases Partitioning will not be ideal.

  To overcome the problem of over partitioning, Hive provides Bucketing concept, another technique for decomposing table data sets into more manageable parts.

#### Bucketing:

##### Hive Partition can be further subdivided into Clusters or Buckets

Hive Buckets is nothing but another technique of decomposing data or decreasing the data into more manageable parts or equal parts. we can’t create number of Hive Buckets the reason is we should declare the number of buckets for a table in the time of table creation.

EXAMPLE:

1.create table patient1(patient\_id int, patient\_name string, gender string, total\_amount int, drug string) row format delimited fields terminated by ',' stored as textfile;

load data local inpath '/home/geouser/Documents/patient1' into table patient1;

**Create a  table in hive with partition and bucketing**

create table partition\_bucket (patient\_id int, patient\_name string, gender string, total\_amount int) partitioned by (drug string) clustered by (gender) into 4 buckets;

### ****Set the properties for partition and bucketing****

SET hive.exec.dynamic.partition=true;

SET hive.exec.dynamic.partition.mode=non-strict;

SET hive.enforce.bucketing =true;

### ****Insert value into the table****

insert overwrite table partition\_bucket partition(drug) select \* from patient1;

### View the bucket value in the table partition\_bucket

select \* from partition\_bucket TABLESAMPLE(BUCKET 1 OUT OF 4 ON gender);

select \* from partition\_bucket TABLESAMPLE(BUCKET 2 OUT OF 4 ON gender);