

# Bucketing in hive

We will use below dataset to perform the operations:

- real\_estate.xls

## Understanding Data

real\_estate.xls file has following columns.

street

city

zip

state

beds

baths

sq\_\_ft

type

price

## create the table

```
hive> create table realestate(  
  > street string,city string,zip string,state string,beds string,baths string,sqft string,type string,price  
double)  
  > row format delimited fields terminated by ',';
```

## Load the input dataset

```
hive> load data local inpath 'file:///home/cloudera/khasimbabu/HIVE_Excercise/real_estate.csv'  
into table realestate;
```

## Display the content of the table

```
hive> select * from realestate limit 10;
```

```
hive> set hive.enforce.bucketing=true;
```

It will set the number of reduce tasks to be equal to the number of buckets

```
hive> create table bucket_table(  
  > street string,zip string,state string,beds string,baths string,sqft string,type string,price double)  
  > partitioned by (city string)  
  > clustered by (street) into 4 buckets  
  > row format delimited fields terminated by ',';
```

```
hive> insert into table bucket_table partition(city)  
  > select street,zip,state,beds,baths,sqft,type,price,city from realestate;
```

## Browse the directory

```
[cloudera@quickstart /]$ hdfs dfs -ls /user/hive/warehouse/retail.db/bucket_table/
```

Found 40 items

```
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=ANTELOPE
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=AUBURN
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=CAMERON PARK
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=CARMICHAEL
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=CITRUS HEIGHTS
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=COOL
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=DIAMOND SPRINGS
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=EL DORADO
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=EL DORADO HILLS
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=ELK GROVE
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=ELVERTA
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=FAIR OAKS
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=FOLSOM
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=FORESTHILL
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=GALT
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=GARDEN VALLEY
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=GOLD RIVER
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=GRANITE BAY
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=GREENWOOD
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=LINCOLN
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=LOOMIS
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=MATHER
```

```

drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=MEADOW VISTA
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=NORTH HIGHLANDS
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=ORANGEVALE
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=PENRYN
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=PLACERVILLE
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=POLLOCK PINES
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=RANCHO CORDOVA
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=RANCHO MURIETA
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=RIO LINDA
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=ROCKLIN
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=ROSEVILLE
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=SACRAMENTO
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=SHINGLE SPRINGS
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=SLOUGHHOUSE
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=WALNUT GROVE
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=WEST SACRAMENTO
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=WILTON
drwxrwxrwx - cloudera supergroup      0 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=city

```

```

[cloudera@quickstart /]$ hdfs dfs -ls /user/hive/warehouse/retail.db/bucket_table/city=ANTELOPE
Found 4 items

```

```

-rwxrwxrwx 1 cloudera supergroup      403 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=ANTELOPE/000000_0
-rwxrwxrwx 1 cloudera supergroup      502 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=ANTELOPE/000001_0
-rwxrwxrwx 1 cloudera supergroup      408 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=ANTELOPE/000002_0
-rwxrwxrwx 1 cloudera supergroup      561 2021-01-27 09:30
/user/hive/warehouse/retail.db/bucket_table/city=ANTELOPE/000003_0

```

```
[cloudera@quickstart /]$ hdfs dfs -ls  
/user/hive/warehouse/retail.db/bucket_table/city=ANTELOPE/000000_0  
-rwxrwxrwx  1 cloudera supergroup    403 2021-01-27 09:30  
/user/hive/warehouse/retail.db/bucket_table/city=ANTELOPE/000000_0
```