**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