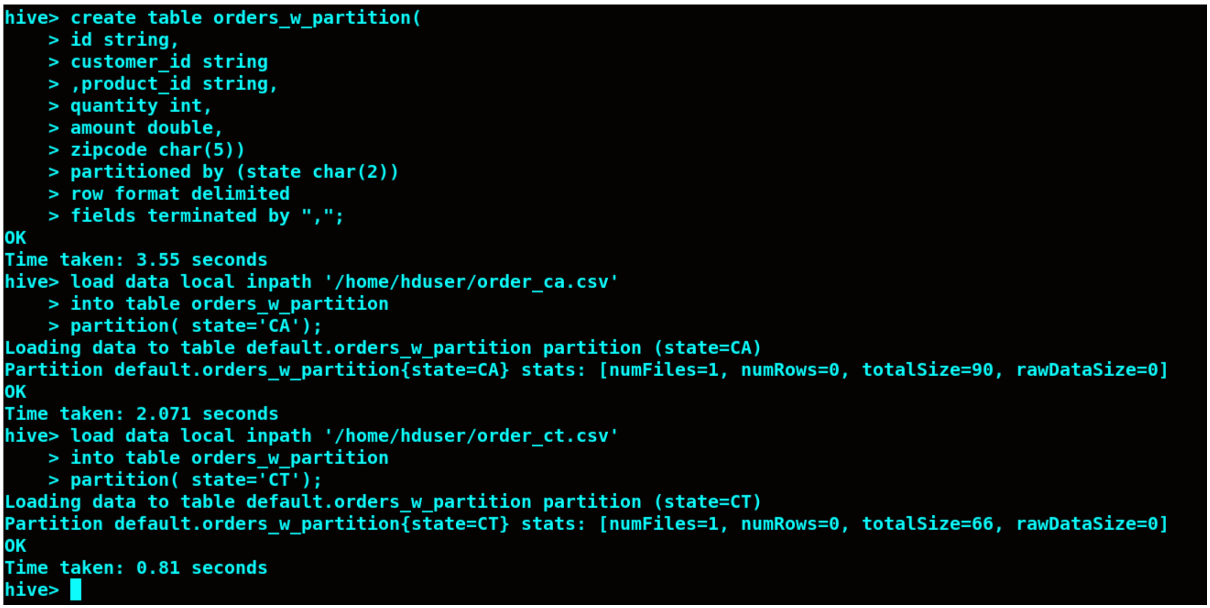
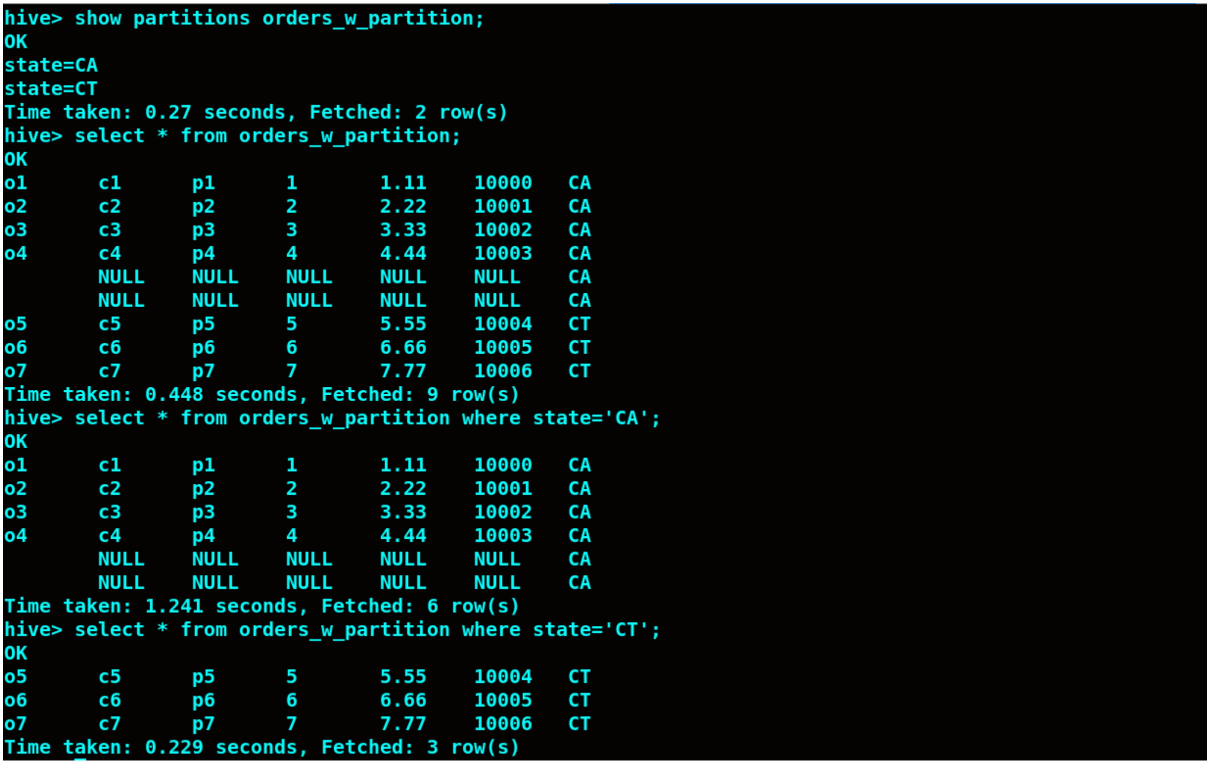
04-07-2022

1. Partitioning:

- ### Static Partitioning:

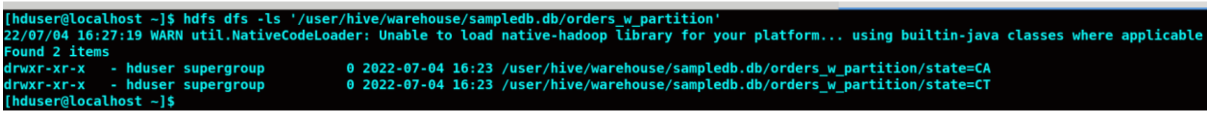


```
hive> create table orders_w_partition(
    > id string,
    > customer_id string
    > ,product_id string,
    > quantity int,
    > amount double,
    > zipcode char(5))
    > partitioned by (state char(2))
    > row format delimited
    > fields terminated by ",";
OK
Time taken: 3.55 seconds
hive> load data local inpath '/home/hduser/order_ca.csv'
    > into table orders_w_partition
    > partition( state='CA');
Loading data to table default.orders_w_partition partition (state=CA)
Partition default.orders_w_partition{state=CA} stats: [numFiles=1, numRows=0, totalSize=90, rawDataSize=0]
OK
Time taken: 2.071 seconds
hive> load data local inpath '/home/hduser/order_ct.csv'
    > into table orders_w_partition
    > partition( state='CT');
Loading data to table default.orders_w_partition partition (state=CT)
Partition default.orders_w_partition{state=CT} stats: [numFiles=1, numRows=0, totalSize=66, rawDataSize=0]
OK
Time taken: 0.81 seconds
hive> 
```



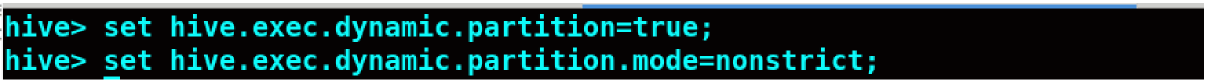
```
hive> show partitions orders_w_partition;
OK
state=CA
state=CT
Time taken: 0.27 seconds, Fetched: 2 row(s)
hive> select * from orders_w_partition;
OK
o1      c1      p1      1       1.11    10000   CA
o2      c2      p2      2       2.22    10001   CA
o3      c3      p3      3       3.33    10002   CA
o4      c4      p4      4       4.44    10003   CA
        NULL    NULL    NULL    NULL    NULL    CA
        NULL    NULL    NULL    NULL    NULL    CA
o5      c5      p5      5       5.55    10004   CT
o6      c6      p6      6       6.66    10005   CT
o7      c7      p7      7       7.77    10006   CT
Time taken: 0.448 seconds, Fetched: 9 row(s)
hive> select * from orders_w_partition where state='CA';
OK
o1      c1      p1      1       1.11    10000   CA
o2      c2      p2      2       2.22    10001   CA
o3      c3      p3      3       3.33    10002   CA
o4      c4      p4      4       4.44    10003   CA
        NULL    NULL    NULL    NULL    NULL    CA
        NULL    NULL    NULL    NULL    NULL    CA
Time taken: 1.241 seconds, Fetched: 6 row(s)
hive> select * from orders_w_partition where state='CT';
OK
o5      c5      p5      5       5.55    10004   CT
o6      c6      p6      6       6.66    10005   CT
o7      c7      p7      7       7.77    10006   CT
Time taken: 0.229 seconds, Fetched: 3 row(s)
```

In warehouse:

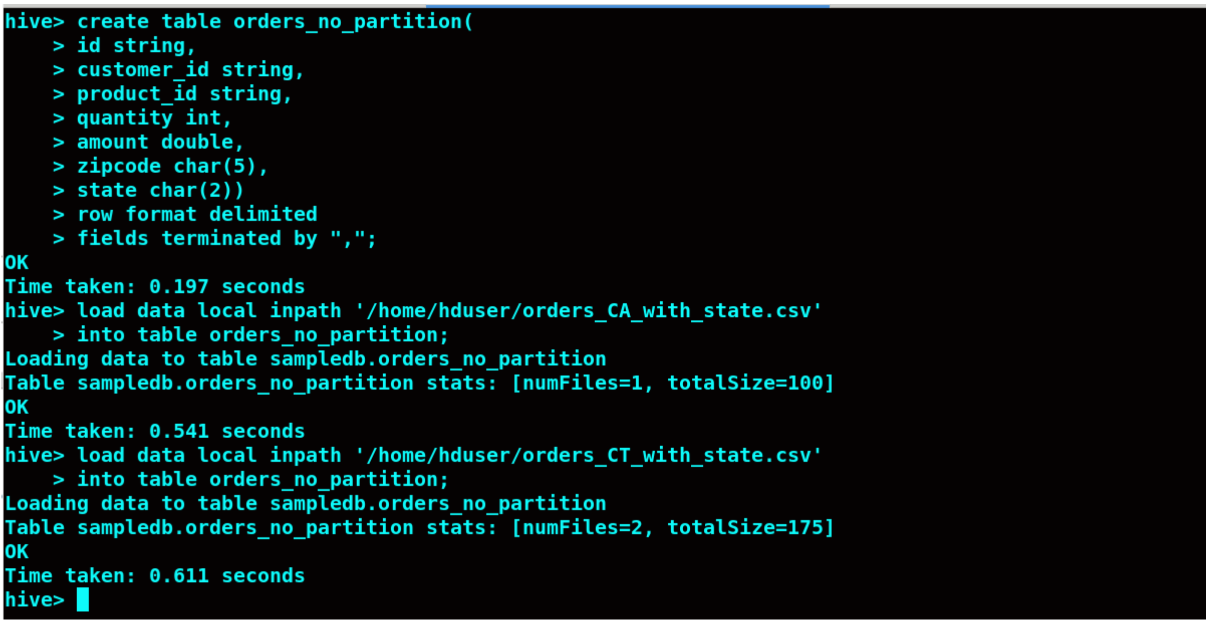


```
[hduser@localhost ~]$ hdfs dfs -ls '/user/hive/warehouse/sampledb.db/orders_w_partition'
22/07/04 16:27:19 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
drwxr-xr-x   - hduser supergroup          0 2022-07-04 16:23 /user/hive/warehouse/sampledb.db/orders_w_partition/state=CA
drwxr-xr-x   - hduser supergroup          0 2022-07-04 16:23 /user/hive/warehouse/sampledb.db/orders_w_partition/state=CT
[hduser@localhost ~]$
```

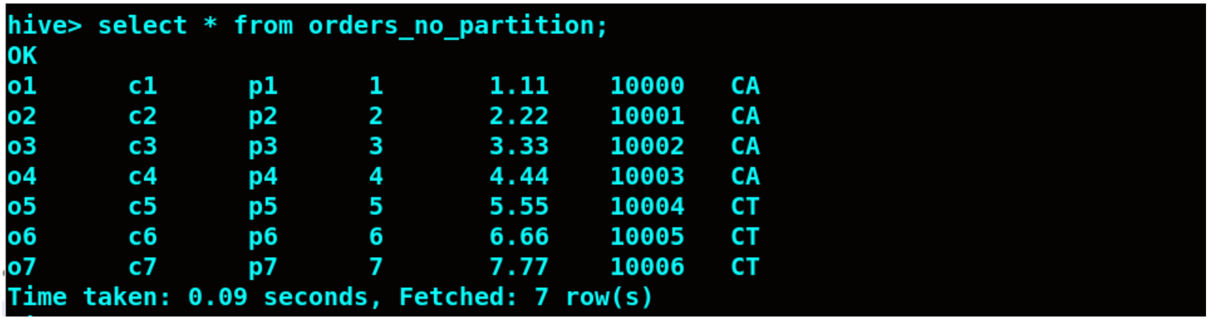
- **Dynamic Partitioning:**



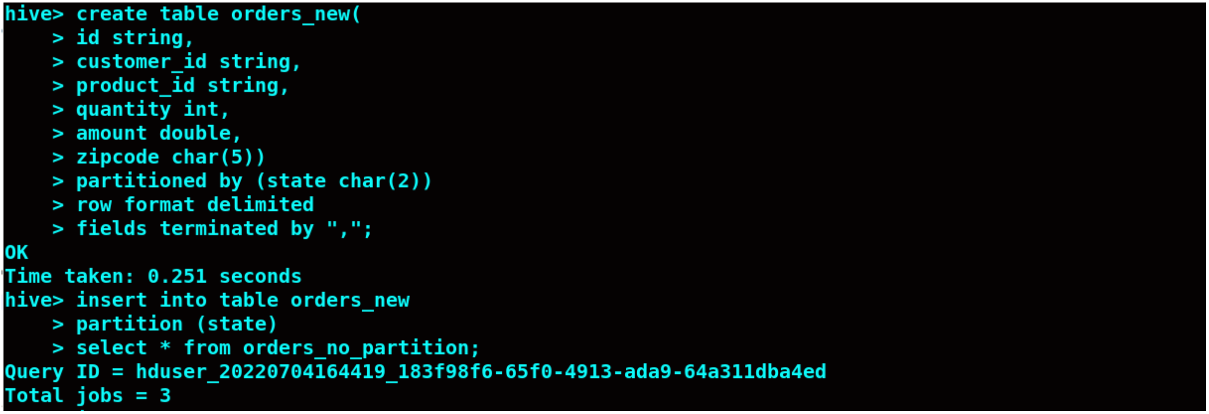
```
hive> set hive.exec.dynamic.partition=true;
hive> set hive.exec.dynamic.partition.mode=nonstrict;
```



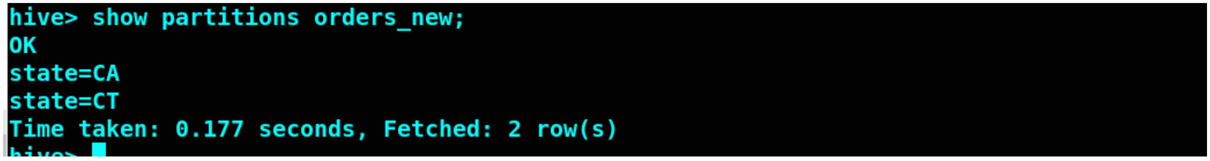
```
hive> create table orders_no_partition(
    > id string,
    > customer_id string,
    > product_id string,
    > quantity int,
    > amount double,
    > zipcode char(5),
    > state char(2))
    > row format delimited
    > fields terminated by ",";
OK
Time taken: 0.197 seconds
hive> load data local inpath '/home/hduser/orders_CA_with_state.csv'
    > into table orders_no_partition;
Loading data to table sampledb.orders_no_partition
Table sampledb.orders_no_partition stats: [numFiles=1, totalSize=100]
OK
Time taken: 0.541 seconds
hive> load data local inpath '/home/hduser/orders_CT_with_state.csv'
    > into table orders_no_partition;
Loading data to table sampledb.orders_no_partition
Table sampledb.orders_no_partition stats: [numFiles=2, totalSize=175]
OK
Time taken: 0.611 seconds
hive> 
```



```
hive> select * from orders_no_partition;
OK
o1      c1      p1      1       1.11    10000   CA
o2      c2      p2      2       2.22    10001   CA
o3      c3      p3      3       3.33    10002   CA
o4      c4      p4      4       4.44    10003   CA
o5      c5      p5      5       5.55    10004   CT
o6      c6      p6      6       6.66    10005   CT
o7      c7      p7      7       7.77    10006   CT
Time taken: 0.09 seconds, Fetched: 7 row(s)
```

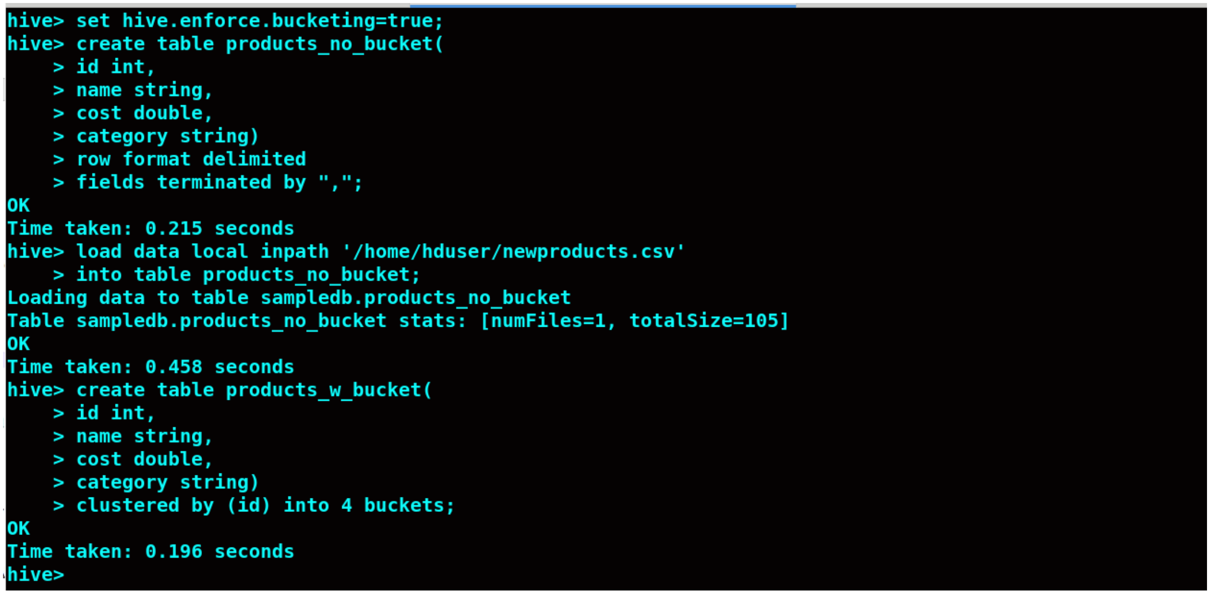


```
hive> create table orders_new(
    > id string,
    > customer_id string,
    > product_id string,
    > quantity int,
    > amount double,
    > zipcode char(5))
    > partitioned by (state char(2))
    > row format delimited
    > fields terminated by ",";
OK
Time taken: 0.251 seconds
hive> insert into table orders_new
    > partition (state)
    > select * from orders_no_partition;
Query ID = hduser_20220704164419_183f98f6-65f0-4913-ada9-64a311dba4ed
Total jobs = 3
```

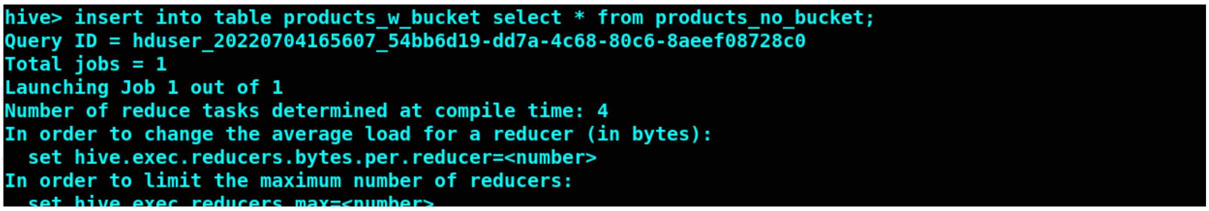


```
hive> show partitions orders_new;
OK
state=CA
state=CT
Time taken: 0.177 seconds, Fetched: 2 row(s)
hive> 
```

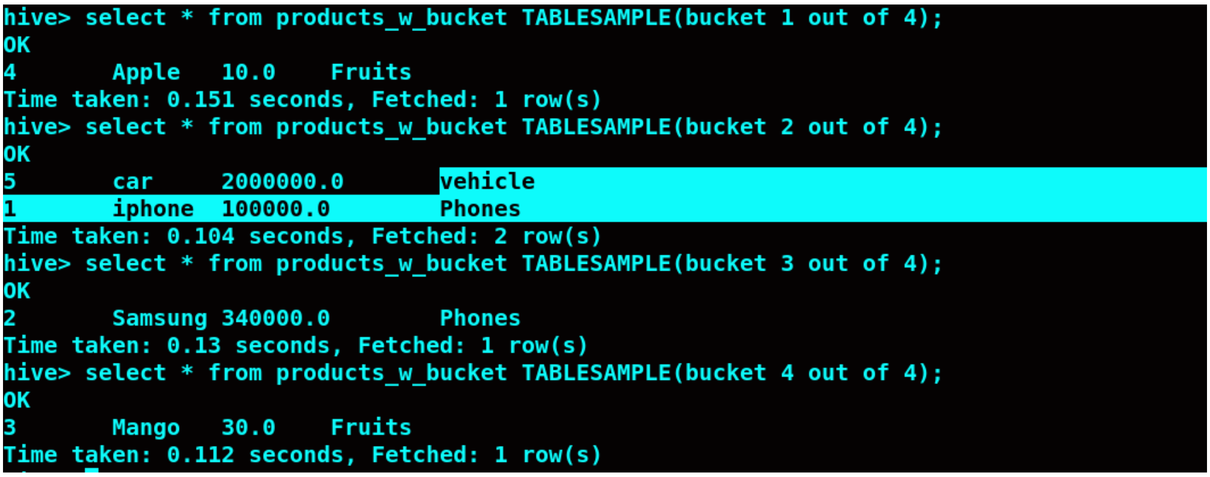
2. Bucketing:



```
hive> set hive.enforce.bucketing=true;
hive> create table products_no_bucket(
    > id int,
    > name string,
    > cost double,
    > category string)
    > row format delimited
    > fields terminated by ",";
OK
Time taken: 0.215 seconds
hive> load data local inpath '/home/hduser/newproducts.csv'
    > into table products_no_bucket;
Loading data to table sampledb.products_no_bucket
Table sampledb.products_no_bucket stats: [numFiles=1, totalSize=105]
OK
Time taken: 0.458 seconds
hive> create table products_w_bucket(
    > id int,
    > name string,
    > cost double,
    > category string)
    > clustered by (id) into 4 buckets;
OK
Time taken: 0.196 seconds
hive>
```

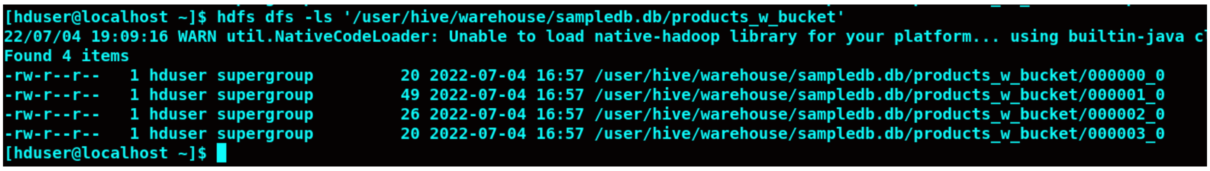


```
hive> insert into table products_w_bucket select * from products_no_bucket;
Query ID = hduser_20220704165607_54bb6d19-dd7a-4c68-80c6-8aeef08728c0
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 4
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
```



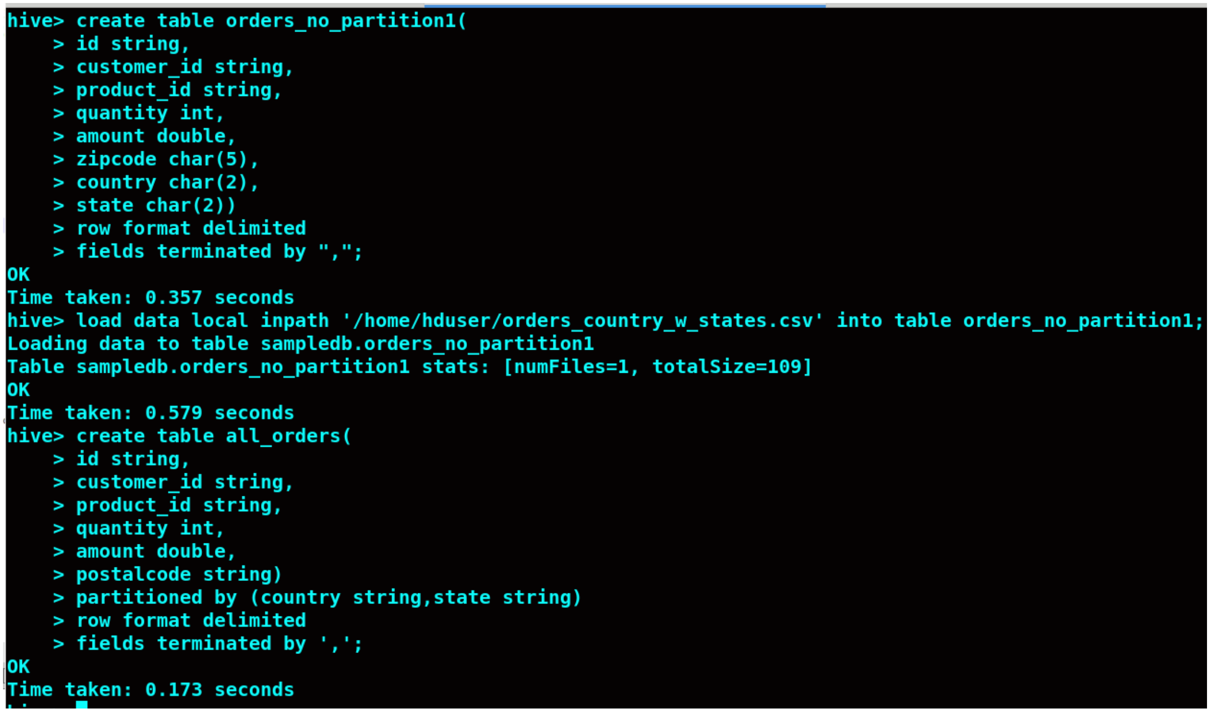
```
hive> select * from products_w_bucket TABLESAMPLE(bucket 1 out of 4);
OK
4       Apple   10.0    Fruits
Time taken: 0.151 seconds, Fetched: 1 row(s)
hive> select * from products_w_bucket TABLESAMPLE(bucket 2 out of 4);
OK
5       car     2000000.0       vehicle
1       iphone  100000.0        Phones
Time taken: 0.104 seconds, Fetched: 2 row(s)
hive> select * from products_w_bucket TABLESAMPLE(bucket 3 out of 4);
OK
2       Samsung 340000.0        Phones
Time taken: 0.13 seconds, Fetched: 1 row(s)
hive> select * from products_w_bucket TABLESAMPLE(bucket 4 out of 4);
OK
3       Mango   30.0    Fruits
Time taken: 0.112 seconds, Fetched: 1 row(s)
```

In warehouse:

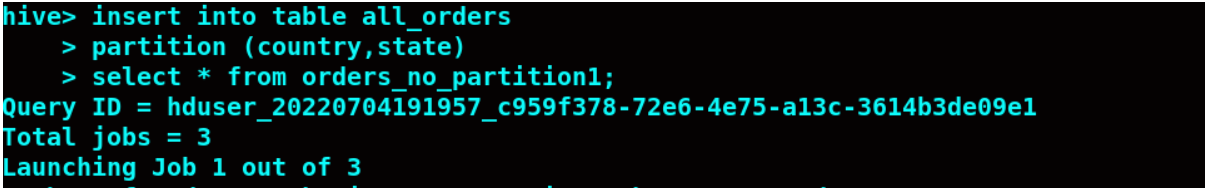


```
[hduser@localhost ~]$ hdfs dfs -ls '/user/hive/warehouse/sampledb.db/products_w_bucket'
22/07/04 19:09:16 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java c
Found 4 items
-rw-r--r--   1 hduser supergroup         20 2022-07-04 16:57 /user/hive/warehouse/sampledb.db/products_w_bucket/000000_0
-rw-r--r--   1 hduser supergroup         49 2022-07-04 16:57 /user/hive/warehouse/sampledb.db/products_w_bucket/000001_0
-rw-r--r--   1 hduser supergroup         26 2022-07-04 16:57 /user/hive/warehouse/sampledb.db/products_w_bucket/000002_0
-rw-r--r--   1 hduser supergroup         20 2022-07-04 16:57 /user/hive/warehouse/sampledb.db/products_w_bucket/000003_0
[hduser@localhost ~]$ 
```

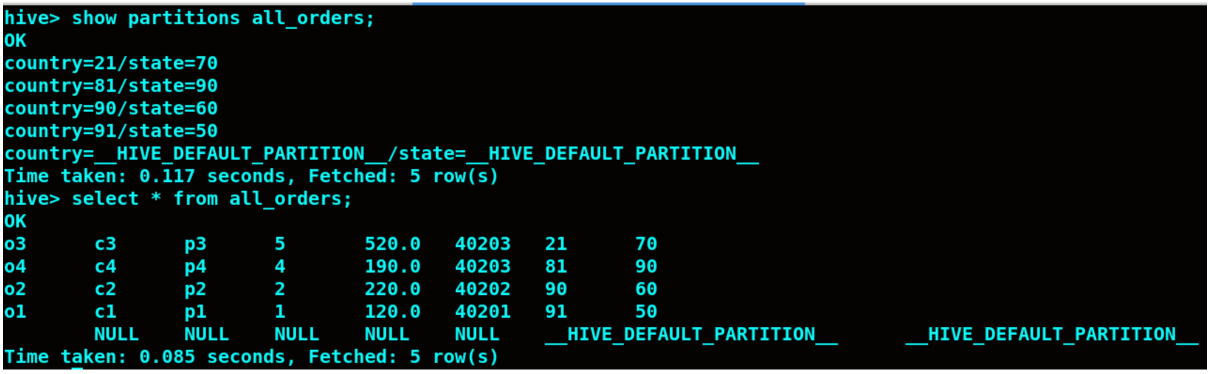
3. Partitioning with 2 columns:



```
hive> create table orders_no_partition1(
    > id string,
    > customer_id string,
    > product_id string,
    > quantity int,
    > amount double,
    > zipcode char(5),
    > country char(2),
    > state char(2))
    > row format delimited
    > fields terminated by ",";
OK
Time taken: 0.357 seconds
hive> load data local inpath '/home/hduser/orders_country_w_states.csv' into table orders_no_partition1;
Loading data to table sampledb.orders_no_partition1
Table sampledb.orders_no_partition1 stats: [numFiles=1, totalSize=109]
OK
Time taken: 0.579 seconds
hive> create table all_orders(
    > id string,
    > customer_id string,
    > product_id string,
    > quantity int,
    > amount double,
    > postalcode string)
    > partitioned by (country string,state string)
    > row format delimited
    > fields terminated by ',';
OK
Time taken: 0.173 seconds
```

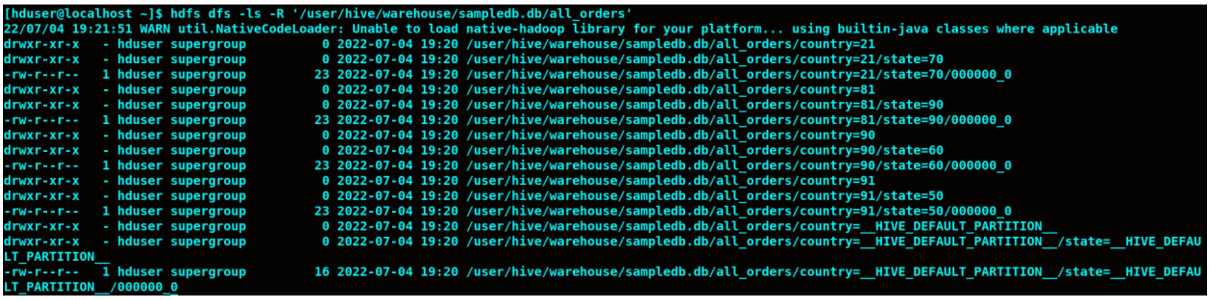


```
hive> insert into table all_orders
    > partition (country,state)
    > select * from orders_no_partition1;
Query ID = hduser_20220704191957_c959f378-72e6-4e75-a13c-3614b3de09e1
Total jobs = 3
Launching Job 1 out of 3
```



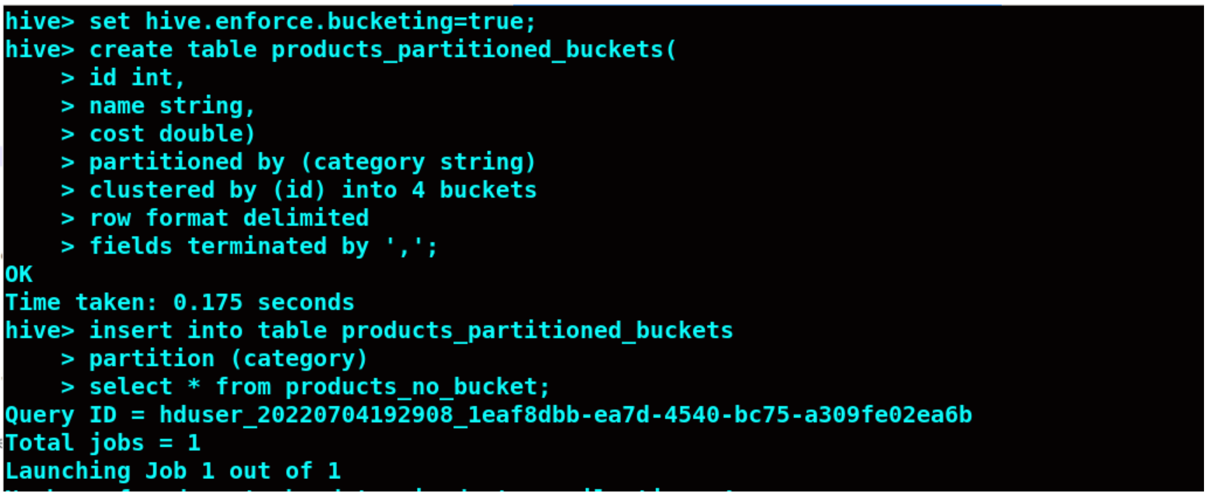
```
hive> show partitions all_orders;
OK
country=21/state=70
country=81/state=90
country=90/state=60
country=91/state=50
country=__HIVE_DEFAULT_PARTITION__/state=__HIVE_DEFAULT_PARTITION__
Time taken: 0.117 seconds, Fetched: 5 row(s)
hive> select * from all_orders;
OK
o3      c3      p3      5       520.0   40203   21      70
o4      c4      p4      4       190.0   40203   81      90
o2      c2      p2      2       220.0   40202   90      60
o1      c1      p1      1       120.0   40201   91      50
        NULL    NULL    NULL    NULL    NULL    __HIVE_DEFAULT_PARTITION__      __HIVE_DEFAULT_PARTITION__
Time taken: 0.085 seconds, Fetched: 5 row(s)
```

In warehouse:



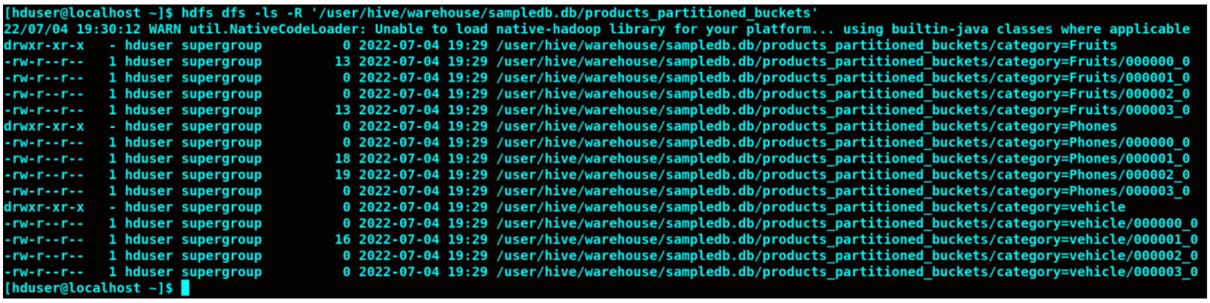
```
[hduser@localhost ~]$ hdfs dfs -ls -R '/user/hive/warehouse/sampledb.db/all_orders'
22/07/04 19:21:51 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
drwxr-xr-x   - hduser supergroup          0 2022-07-04 19:20 /user/hive/warehouse/sampledb.db/all_orders/country=21
drwxr-xr-x   - hduser supergroup          0 2022-07-04 19:20 /user/hive/warehouse/sampledb.db/all_orders/country=21/state=70
-rw-r--r--   1 hduser supergroup         23 2022-07-04 19:20 /user/hive/warehouse/sampledb.db/all_orders/country=21/state=70/000000_0
drwxr-xr-x   - hduser supergroup          0 2022-07-04 19:20 /user/hive/warehouse/sampledb.db/all_orders/country=81
drwxr-xr-x   - hduser supergroup          0 2022-07-04 19:20 /user/hive/warehouse/sampledb.db/all_orders/country=81/state=90
-rw-r--r--   1 hduser supergroup         23 2022-07-04 19:20 /user/hive/warehouse/sampledb.db/all_orders/country=81/state=90/000000_0
drwxr-xr-x   - hduser supergroup          0 2022-07-04 19:20 /user/hive/warehouse/sampledb.db/all_orders/country=90
drwxr-xr-x   - hduser supergroup          0 2022-07-04 19:20 /user/hive/warehouse/sampledb.db/all_orders/country=90/state=60
-rw-r--r--   1 hduser supergroup         23 2022-07-04 19:20 /user/hive/warehouse/sampledb.db/all_orders/country=90/state=60/000000_0
drwxr-xr-x   - hduser supergroup          0 2022-07-04 19:20 /user/hive/warehouse/sampledb.db/all_orders/country=91
drwxr-xr-x   - hduser supergroup          0 2022-07-04 19:20 /user/hive/warehouse/sampledb.db/all_orders/country=91/state=50
-rw-r--r--   1 hduser supergroup         23 2022-07-04 19:20 /user/hive/warehouse/sampledb.db/all_orders/country=91/state=50/000000_0
drwxr-xr-x   - hduser supergroup          0 2022-07-04 19:20 /user/hive/warehouse/sampledb.db/all_orders/country=__HIVE_DEFAULT_PARTITION__
drwxr-xr-x   - hduser supergroup          0 2022-07-04 19:20 /user/hive/warehouse/sampledb.db/all_orders/country=__HIVE_DEFAULT_PARTITION__/state=__HIVE_DEFAU
LT_PARTITION__
-rw-r--r--   1 hduser supergroup         16 2022-07-04 19:20 /user/hive/warehouse/sampledb.db/all_orders/country=__HIVE_DEFAULT_PARTITION__/state=__HIVE_DEFAU
LT_PARTITION__/000000_0
```

4. Partitioning with bucketing:



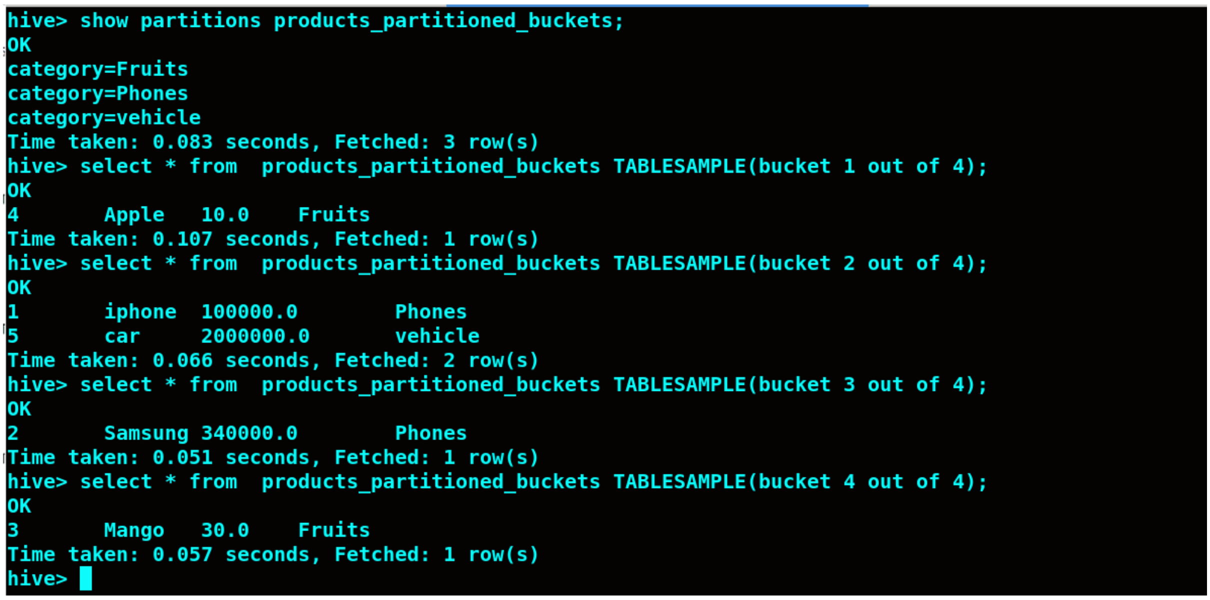
```
hive> set hive.enforce.bucketing=true;
hive> create table products_partitioned_buckets(
    > id int,
    > name string,
    > cost double)
    > partitioned by (category string)
    > clustered by (id) into 4 buckets
    > row format delimited
    > fields terminated by ',';
OK
Time taken: 0.175 seconds
hive> insert into table products_partitioned_buckets
    > partition (category)
    > select * from products_no_bucket;
Query ID = hduser_20220704192908_1eaf8dbb-ea7d-4540-bc75-a309fe02ea6b
Total jobs = 1
Launching Job 1 out of 1
```

In warehouse:



```
[hduser@localhost ~]$ hdfs dfs -ls -R '/user/hive/warehouse/sampledb.db/products_partitioned_buckets'
22/07/04 19:30:12 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
drwxr-xr-x   - hduser supergroup          0 2022-07-04 19:29 /user/hive/warehouse/sampledb.db/products_partitioned_buckets/category=Fruits
-rw-r--r--   1 hduser supergroup         13 2022-07-04 19:29 /user/hive/warehouse/sampledb.db/products_partitioned_buckets/category=Fruits/000000_0
-rw-r--r--   1 hduser supergroup          0 2022-07-04 19:29 /user/hive/warehouse/sampledb.db/products_partitioned_buckets/category=Fruits/000001_0
-rw-r--r--   1 hduser supergroup          0 2022-07-04 19:29 /user/hive/warehouse/sampledb.db/products_partitioned_buckets/category=Fruits/000002_0
-rw-r--r--   1 hduser supergroup         13 2022-07-04 19:29 /user/hive/warehouse/sampledb.db/products_partitioned_buckets/category=Fruits/000003_0
drwxr-xr-x   - hduser supergroup          0 2022-07-04 19:29 /user/hive/warehouse/sampledb.db/products_partitioned_buckets/category=Phones
-rw-r--r--   1 hduser supergroup          0 2022-07-04 19:29 /user/hive/warehouse/sampledb.db/products_partitioned_buckets/category=Phones/000000_0
-rw-r--r--   1 hduser supergroup         18 2022-07-04 19:29 /user/hive/warehouse/sampledb.db/products_partitioned_buckets/category=Phones/000001_0
-rw-r--r--   1 hduser supergroup         19 2022-07-04 19:29 /user/hive/warehouse/sampledb.db/products_partitioned_buckets/category=Phones/000002_0
-rw-r--r--   1 hduser supergroup          0 2022-07-04 19:29 /user/hive/warehouse/sampledb.db/products_partitioned_buckets/category=Phones/000003_0
drwxr-xr-x   - hduser supergroup          0 2022-07-04 19:29 /user/hive/warehouse/sampledb.db/products_partitioned_buckets/category=vehicle
-rw-r--r--   1 hduser supergroup          0 2022-07-04 19:29 /user/hive/warehouse/sampledb.db/products_partitioned_buckets/category=vehicle/000000_0
-rw-r--r--   1 hduser supergroup         16 2022-07-04 19:29 /user/hive/warehouse/sampledb.db/products_partitioned_buckets/category=vehicle/000001_0
-rw-r--r--   1 hduser supergroup          0 2022-07-04 19:29 /user/hive/warehouse/sampledb.db/products_partitioned_buckets/category=vehicle/000002_0
-rw-r--r--   1 hduser supergroup          0 2022-07-04 19:29 /user/hive/warehouse/sampledb.db/products_partitioned_buckets/category=vehicle/000003_0
[hduser@localhost ~]$
```

Partitions and buckets:



```
hive> show partitions products_partitioned_buckets;
OK
category=Fruits
category=Phones
category=vehicle
Time taken: 0.083 seconds, Fetched: 3 row(s)
hive> select * from  products_partitioned_buckets TABLESAMPLE(bucket 1 out of 4);
OK
4       Apple   10.0    Fruits
Time taken: 0.107 seconds, Fetched: 1 row(s)
hive> select * from  products_partitioned_buckets TABLESAMPLE(bucket 2 out of 4);
OK
1       iphone  100000.0        Phones
5       car     2000000.0       vehicle
Time taken: 0.066 seconds, Fetched: 2 row(s)
hive> select * from  products_partitioned_buckets TABLESAMPLE(bucket 3 out of 4);
OK
2       Samsung 340000.0        Phones
Time taken: 0.051 seconds, Fetched: 1 row(s)
hive> select * from  products_partitioned_buckets TABLESAMPLE(bucket 4 out of 4);
OK
3       Mango   30.0    Fruits
Time taken: 0.057 seconds, Fetched: 1 row(s)
hive>
```