HIVE BUCKETING

BUCKETING - There are times when there are a lot of distinct values in columns and partitioning is not a good option.

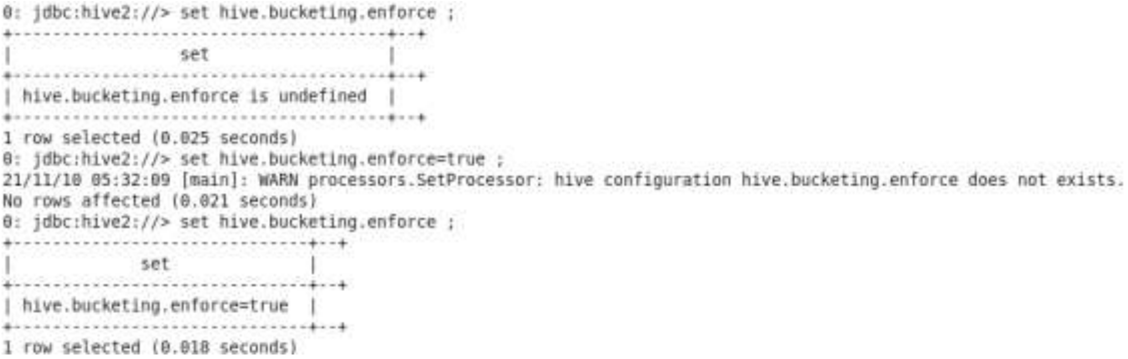
Hence we can select a column which is most frequently accessed using “WHERE” clause and then perform bucketing on that column.

When you perform bucketing on a column, you have to specify the number of buckets which you wish to create. These number of buckets are actually the no of files into which your table’s data will be divided and stored in the form of files.

The number of buckets will depend on the data size as well as you have to do trial and error to find out the best number suited for the optimization.

To create a bucket, you have to enable bucketing:

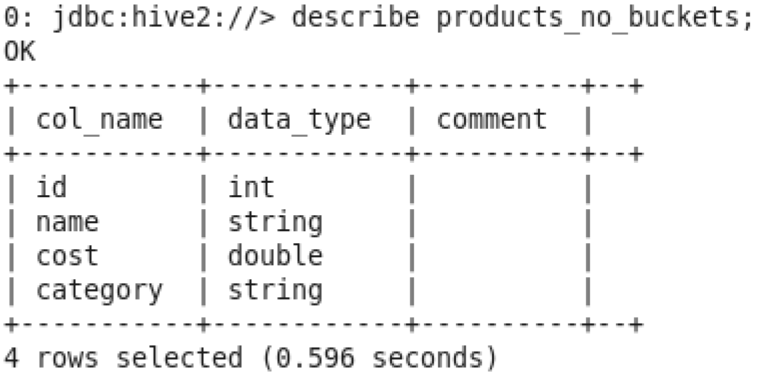
```
0: jdbc:hive2://> SET hive.bucketing.enforce=true ;
```



```
0: jdbc:hive2://> set hive.bucketing.enforce ;
+--------------------------------------+--+
|                 set                  |
+--------------------------------------+--+
| hive.bucketing.enforce is undefined  |
+--------------------------------------+--+
1 row selected (0.025 seconds)
0: jdbc:hive2://> set hive.bucketing.enforce=true ;
21/11/10 05:32:09 [main]: WARN processors.SetProcessor: hive configuration hive.bucketing.enforce does not exists.
No rows affected (0.021 seconds)
0: jdbc:hive2://> set hive.bucketing.enforce ;
+------------------------------+--+
|             set              |
+------------------------------+--+
| hive.bucketing.enforce=true  |
+------------------------------+--+
1 row selected (0.018 seconds)
```

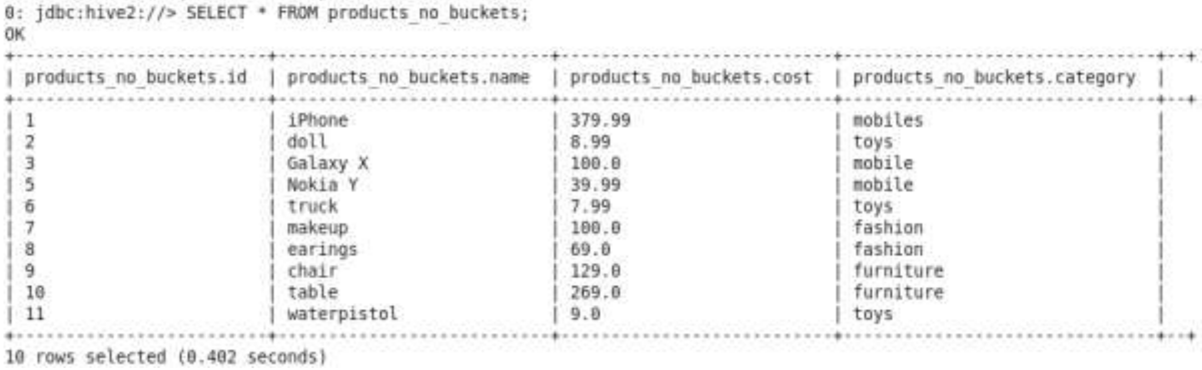
```
0: jdbc:hive2://> CREATE TABLE products_no_buckets(
. . . . . . . . > id int,
. . . . . . . . > name string,
. . . . . . . . > cost double,
. . . . . . . . > category string)
. . . . . . . . > ROW FORMAT DELIMITED FIELDS TERMINATED BY ',';
OK
No rows affected (0.325 seconds)
```

```
0: jdbc:hive2://> DESCRIBE products_no_buckets;
```



```
0: jdbc:hive2://> describe products_no_buckets;
OK
+-----------+-----------+----------+--+
| col_name  | data_type | comment  |
+-----------+-----------+----------+--+
| id        | int       |          |
| name      | string    |          |
| cost      | double    |          |
| category  | string    |          |
+-----------+-----------+----------+--+
4 rows selected (0.596 seconds)
```

0: jdbc:hive2://> SELECT * FROM products_no_buckets;



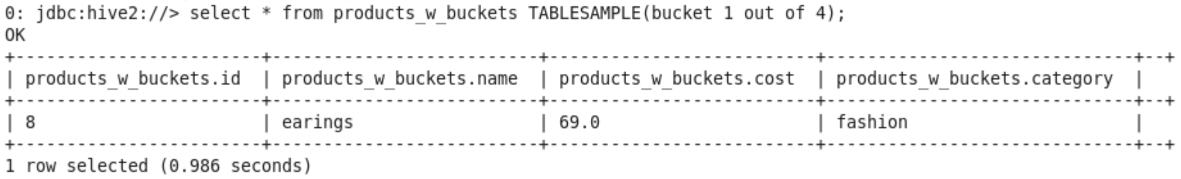
```
0: jdbc:hive2://> SELECT * FROM products_no_buckets;
OK
+-------------------------+---------------------------+---------------------------+-------------------------------+--+
| products_no_buckets.id  | products_no_buckets.name  | products_no_buckets.cost  | products_no_buckets.category  |
+-------------------------+---------------------------+---------------------------+-------------------------------+--+
| 1                       | iPhone                    | 379.99                    | mobiles                       |
| 2                       | doll                      | 8.99                      | toys                          |
| 3                       | Galaxy X                  | 100.0                     | mobile                        |
| 5                       | Nokia Y                   | 39.99                     | mobile                        |
| 6                       | truck                     | 7.99                      | toys                          |
| 7                       | makeup                    | 100.0                     | fashion                       |
| 8                       | earings                   | 69.0                      | fashion                       |
| 9                       | chair                     | 129.0                     | furniture                     |
| 10                      | table                     | 269.0                     | furniture                     |
| 11                      | waterpistol               | 9.0                       | toys                          |
+-------------------------+---------------------------+---------------------------+-------------------------------+--+
10 rows selected (0.402 seconds)
```

```
0: jdbc:hive2://> CREATE TABLE products_w_buckets(
. . . . . . . . > id int,
. . . . . . . . > name string,
. . . . . . . . > cost double,
. . . . . . . . > category string)
. . . . . . . . > CLUSTERED BY (id) INTO 4 BUCKETS;
OK
No rows affected (0.325 seconds)
```

```
0: jdbc:hive2://> INSERT INTO products_w_buckets
. . . . . . . . > select * from products_no_buckets;
```

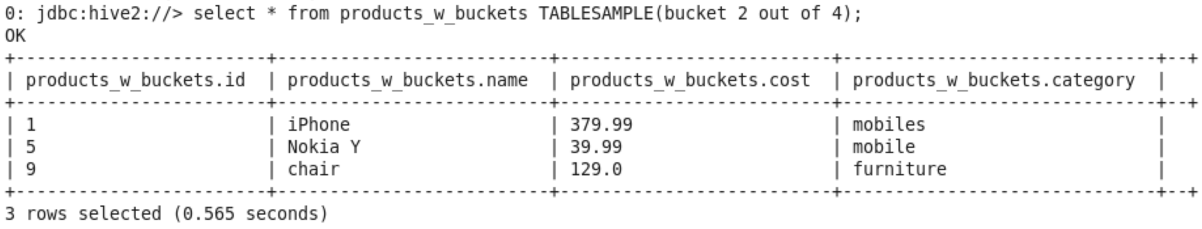
Verify that the records are inserted in the table products_w_buckets and check the data present in each bucket :

0: jdbc:hive2://> select * from products_w_buckets TABLESAMPLE(bucket 1 out of 4);



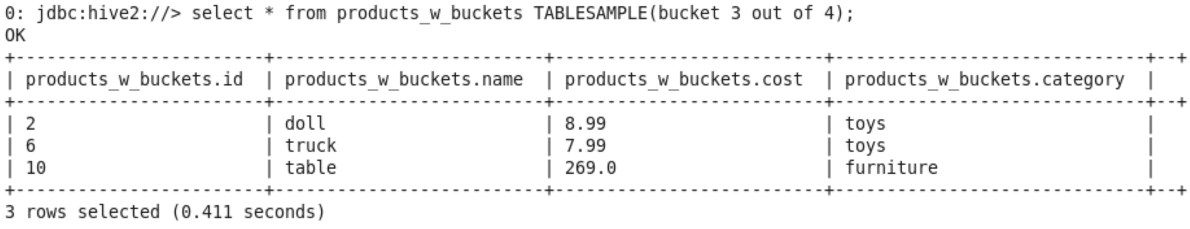
```
0: jdbc:hive2://> select * from products_w_buckets TABLESAMPLE(bucket 1 out of 4);
OK
+------------------------+--------------------------+--------------------------+------------------------------+--+
| products_w_buckets.id  | products_w_buckets.name  | products_w_buckets.cost  | products_w_buckets.category  |
+------------------------+--------------------------+--------------------------+------------------------------+--+
| 8                      | earings                  | 69.0                     | fashion                      |
+------------------------+--------------------------+--------------------------+------------------------------+--+
1 row selected (0.986 seconds)
```

0: jdbc:hive2://> select * from products_w_buckets TABLESAMPLE(bucket 2 out of 4);



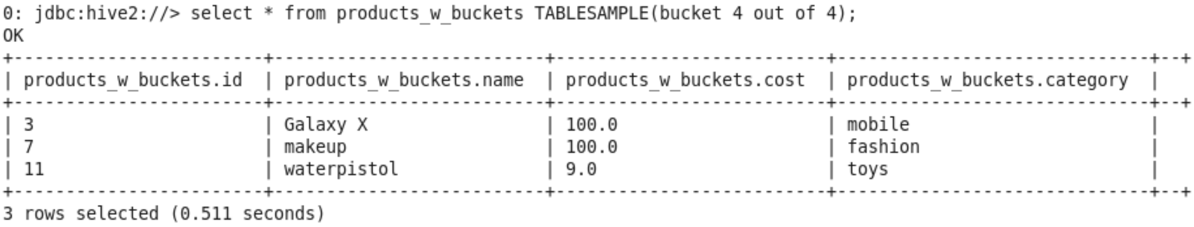
```
0: jdbc:hive2://> select * from products_w_buckets TABLESAMPLE(bucket 2 out of 4);
OK
+------------------------+--------------------------+--------------------------+------------------------------+--+
| products_w_buckets.id  | products_w_buckets.name  | products_w_buckets.cost  | products_w_buckets.category  |
+------------------------+--------------------------+--------------------------+------------------------------+--+
| 1                      | iPhone                   | 379.99                   | mobiles                      |
| 5                      | Nokia Y                  | 39.99                    | mobile                       |
| 9                      | chair                    | 129.0                    | furniture                    |
+------------------------+--------------------------+--------------------------+------------------------------+--+
3 rows selected (0.565 seconds)
```

0: jdbc:hive2://> select * from products_w_buckets TABLESAMPLE(bucket 3 out of 4);



```
0: jdbc:hive2://> select * from products_w_buckets TABLESAMPLE(bucket 3 out of 4);
OK
+------------------------+--------------------------+--------------------------+------------------------------+--+
| products_w_buckets.id  | products_w_buckets.name  | products_w_buckets.cost  | products_w_buckets.category  |
+------------------------+--------------------------+--------------------------+------------------------------+--+
| 2                      | doll                     | 8.99                     | toys                         |
| 6                      | truck                    | 7.99                     | toys                         |
| 10                     | table                    | 269.0                    | furniture                    |
+------------------------+--------------------------+--------------------------+------------------------------+--+
3 rows selected (0.411 seconds)
```

0: jdbc:hive2://> select * from products_w_buckets TABLESAMPLE(bucket 4 out of 4);



```
0: jdbc:hive2://> select * from products_w_buckets TABLESAMPLE(bucket 4 out of 4);
OK
+-------------------------+--------------------------+---------------------------+------------------------------+--+
| products_w_buckets.id   | products_w_buckets.name  | products_w_buckets.cost   | products_w_buckets.category  |
+-------------------------+--------------------------+---------------------------+------------------------------+--+
| 3                       | Galaxy X                 | 100.0                     | mobile                       |
| 7                       | makeup                   | 100.0                     | fashion                      |
| 11                      | waterpistol              | 9.0                       | toys                         |
+-------------------------+--------------------------+---------------------------+------------------------------+--+
3 rows selected (0.511 seconds)
```