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: 1.962 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=88, rawDataSize=0]
OK
Time taken: 2.82 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=92, rawDataSize=0]
OK
Time taken: 1.187 seconds
```

```
hive> show partitions orders_w_partition;
OK
state=CA
state=CT
Time taken: 0.63 seconds, Fetched: 2 row(s)
hive>
```

```
[hduser@localhost ~]$ hdfs dfs -ls "/user/hive/warehouse/demo_db.db/orders_w_partition"
22/07/04 16:27:02 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 16:24 /user/hive/warehouse/demo_db.db/orders_w_partition/state=CA
drwxr-xr-x   - hduser supergroup          0 2022-07-04 16:24 /user/hive/warehouse/demo_db.db/orders_w_partition/state=CT
```

```
hive> select * from orders_w_partition
    > where state="CA";
OK
o1      c1      p1      1       1.11    50001   CA
o2      c2      p2      1       2.12    50002   CA
o3      c3      p3      1       3.31    50003   CA
o4      c4      p4      1       4.44    50004   CA
Time taken: 1.987 seconds, Fetched: 4 row(s)
hive> select * from orders_w_partition
    > where state="CT";
OK
o5      c5      p5      1       10.11   50011   CT
o6      c6      p6      3       22.12   50022   CT
o7      c7      p7      2       33.31   50033   CT
o8      c8      p8      1       48.44   50034   CT
Time taken: 0.267 seconds, Fetched: 4 row(s)
hive> select * from orders_w_partition;
OK
o1      c1      p1      1       1.11    50001   CA
o2      c2      p2      1       2.12    50002   CA
o3      c3      p3      1       3.31    50003   CA
o4      c4      p4      1       4.44    50004   CA
o5      c5      p5      1       10.11   50011   CT
o6      c6      p6      3       22.12   50022   CT
o7      c7      p7      2       33.31   50033   CT
o8      c8      p8      1       48.44   50034   CT
Time taken: 0.269 seconds, Fetched: 8 row(s)
```

Dynamic Programing

```
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: 0.268 seconds
hive> load data local inpath "/home/hduser/order_ca_w_state.csv" into table orders_no_partition;
Loading data to table demo_db.orders_no_partition
Table demo_db.orders_no_partition stats: [numFiles=1, totalSize=100]
OK
Time taken: 0.885 seconds
hive> load data local inpath "/home/hduser/order_ct_w_state.csv" into table orders_no_partition;
Loading data to table demo_db.orders_no_partition
Table demo_db.orders_no_partition stats: [numFiles=2, totalSize=204]
OK
Time taken: 0.903 seconds
```

```
hive> select * from orders_no_partition;
OK
o1      c1      p1      1       1.11    50001   CA
o2      c2      p2      1       2.12    50002   CA
o3      c3      p3      1       3.31    50003   CA
o4      c4      p4      1       4.44    50004   CA
o5      c5      p5      1       10.11   50011   CT
o6      c6      p6      3       22.12   50022   CT
o7      c7      p7      2       33.31   50033   CT
o8      c8      p8      1       48.44   50034   CT
Time taken: 0.225 seconds, Fetched: 8 row(s)
```

```
[hduser@localhost ~]$ hdfs dfs -ls "/user/hive/warehouse/demo_db.db/orders_no_partition"
22/07/04 16:40:24 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r--   1 hduser supergroup        100 2022-07-04 16:35 /user/hive/warehouse/demo_db.db/orders_no_partition/order_ca_w_state.csv
-rw-r--r--   1 hduser supergroup        104 2022-07-04 16:35 /user/hive/warehouse/demo_db.db/orders_no_partition/order_ct_w_state.csv
```

```
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: 0.286 seconds
```

```
hive> insert into table orders_new
    > partition (state)
    > select * from orders_no_partition;
Query ID = hduser_20220704164354_b41c3279-81ad-48e9-8a81-5e557107d478
Total jobs = 3
```

```
hive> show partitions orders_new;
OK
state=CA
state=CT
Time taken: 0.288 seconds, Fetched: 2 row(s)
```

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: 0.269 seconds
hive> load data local inpath "/home/hduser/products_new" into table products_no_bucket;
Loading data to table demo_db.products_no_bucket
Table demo_db.products_no_bucket stats: [numFiles=1, totalSize=90]
OK
Time taken: 0.424 seconds
```

```
hive> create table products_w_bucket(
    > id int,
    > name string,
    > cost double,
    > category string)
    > clustered by (id) into 4 buckets;
OK
Time taken: 0.285 seconds
hive> insert into table products_w_bucket
    > select * from products_no_bucket;
Query ID = hduser_20220704212238_ae81b7d7-2d7f-40a6-a03d-ddac3baac72d
Total jobs = 1
[hduser@localhost ~]$ hdfs dfs -ls "/user/hive/warehouse/demo_db.db/products_w_bucket"
22/07/04 21:26:16 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 4 items
-rw-r--r--   1 hduser supergroup         24 2022-07-04 21:25 /user/hive/warehouse/demo_db.db/products_w_bucket/000000_0
-rw-r--r--   1 hduser supergroup         25 2022-07-04 21:25 /user/hive/warehouse/demo_db.db/products_w_bucket/000001_0
-rw-r--r--   1 hduser supergroup         25 2022-07-04 21:25 /user/hive/warehouse/demo_db.db/products_w_bucket/000002_0
-rw-r--r--   1 hduser supergroup         24 2022-07-04 21:25 /user/hive/warehouse/demo_db.db/products_w_bucket/000003_0
```

```
hive> select * from products_w_bucket tablesample(bucket 1 out of 4);
OK
4       dell    1000012.0       laptop
Time taken: 0.435 seconds, Fetched: 1 row(s)
hive> select * from products_w_bucket tablesample(bucket 2 out of 4);
OK
1       iphone  100000.0        mobile
Time taken: 0.176 seconds, Fetched: 1 row(s)
hive> select * from products_w_bucket tablesample(bucket 3 out of 4);
OK
2       samsung 20000.0 mobile
Time taken: 0.17 seconds, Fetched: 1 row(s)
hive> select * from products_w_bucket tablesample(bucket 4 out of 4);
OK
3       boat    1000.0  earphones
Time taken: 0.148 seconds, Fetched: 1 row(s)
hive>
```

Partition with 2 columns :

```
hive> create table orders_two_no_partition(
    > id string,
    > customer_id string,
    > product_id string,
    > quantity int,
    > amount double,
    > zipcode char(5),
    > state char(2),
    > country char(2))
    > row format delimited
    > fields terminated by ",";
hive> load data local inpath '/home/hduser/order_new.csv' into table orders_two_no_partition;
Loading data to table demo_db.orders_two_no_partition
Table demo_db.orders_two_no_partition stats: [numFiles=1, totalSize=232]
OK
Time taken: 0.748 seconds
```

```
hive> create table orders_two_partition(
    > id string,
    > customer_id string,
    > product_id string,
    > quantity int,
    > amount double,
    > zipcode string)
    > partitioned by (state string, country string)
    > row format delimited
    > fields terminated by ",";
OK
Time taken: 0.316 seconds
hive> insert into orders_two_partition
    > partition(state, country)
    > select * from orders_two_no_partition;
Query ID = hduser_20220704220654_d2a2017b-08a3-4ddf-9ae3-66b05dd5f137
Total jobs = 3
```

```
[hduser@localhost ~]$ hdfs dfs -ls -R '/user/hive/warehouse/demo_db.db/orders_two_partition'
22/07/04 22:12:44 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 22:07 /user/hive/warehouse/demo_db.db/orders_two_partition/state=CA
drwxr-xr-x   - hduser supergroup          0 2022-07-04 22:07 /user/hive/warehouse/demo_db.db/orders_two_partition/state=CA/country=US
-rw-r--r--   1 hduser supergroup         88 2022-07-04 22:07 /user/hive/warehouse/demo_db.db/orders_two_partition/state=CA/country=US/000000_0
drwxr-xr-x   - hduser supergroup          0 2022-07-04 22:07 /user/hive/warehouse/demo_db.db/orders_two_partition/state=CT
drwxr-xr-x   - hduser supergroup          0 2022-07-04 22:07 /user/hive/warehouse/demo_db.db/orders_two_partition/state=CT/country=US
-rw-r--r--   1 hduser supergroup         92 2022-07-04 22:07 /user/hive/warehouse/demo_db.db/orders_two_partition/state=CT/country=US/000000_0
[hduser@localhost ~]$
```

Partition with bucketing :

```
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: 0.648 seconds
hive> insert into table products_partitioned_buckets
    > partition (category)
    > select * from products_no_bucket;
Query ID = hduser_20220704223806_137b2ead-19ab-43ec-a606-7a5d0e27796c
Total jobs = 1
```

```
[hduser@localhost ~]$ hdfs dfs -ls -R '/user/hive/warehouse/demo_db.db/products_partitioned_buckets'
22/07/04 22:43:44 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 22:40 /user/hive/warehouse/demo_db.db/products_partitioned_buckets/category=earphones
-rw-r--r--   1 hduser supergroup          0 2022-07-04 22:40 /user/hive/warehouse/demo_db.db/products_partitioned_buckets/category=earphones/000000_0
-rw-r--r--   1 hduser supergroup          0 2022-07-04 22:40 /user/hive/warehouse/demo_db.db/products_partitioned_buckets/category=earphones/000001_0
-rw-r--r--   1 hduser supergroup          0 2022-07-04 22:40 /user/hive/warehouse/demo_db.db/products_partitioned_buckets/category=earphones/000002_0
-rw-r--r--   1 hduser supergroup         14 2022-07-04 22:40 /user/hive/warehouse/demo_db.db/products_partitioned_buckets/category=earphones/000003_0
drwxr-xr-x   - hduser supergroup          0 2022-07-04 22:40 /user/hive/warehouse/demo_db.db/products_partitioned_buckets/category=laptop
-rw-r--r--   1 hduser supergroup         17 2022-07-04 22:40 /user/hive/warehouse/demo_db.db/products_partitioned_buckets/category=laptop/000000_0
-rw-r--r--   1 hduser supergroup          0 2022-07-04 22:40 /user/hive/warehouse/demo_db.db/products_partitioned_buckets/category=laptop/000001_0
-rw-r--r--   1 hduser supergroup          0 2022-07-04 22:40 /user/hive/warehouse/demo_db.db/products_partitioned_buckets/category=laptop/000002_0
-rw-r--r--   1 hduser supergroup          0 2022-07-04 22:40 /user/hive/warehouse/demo_db.db/products_partitioned_buckets/category=laptop/000003_0
drwxr-xr-x   - hduser supergroup          0 2022-07-04 22:40 /user/hive/warehouse/demo_db.db/products_partitioned_buckets/category=mobile
-rw-r--r--   1 hduser supergroup          0 2022-07-04 22:40 /user/hive/warehouse/demo_db.db/products_partitioned_buckets/category=mobile/000000_0
-rw-r--r--   1 hduser supergroup         18 2022-07-04 22:40 /user/hive/warehouse/demo_db.db/products_partitioned_buckets/category=mobile/000001_0
-rw-r--r--   1 hduser supergroup         18 2022-07-04 22:40 /user/hive/warehouse/demo_db.db/products_partitioned_buckets/category=mobile/000002_0
-rw-r--r--   1 hduser supergroup          0 2022-07-04 22:40 /user/hive/warehouse/demo_db.db/products_partitioned_buckets/category=mobile/000003_0
[hduser@localhost ~]$
```

```
hive> show partitions products_partitioned_buckets;
OK
category=earphones
category=laptop
category=mobile
Time taken: 0.282 seconds, Fetched: 3 row(s)
hive> select * from products_partitioned_buckets tablesample (bucket 1 out of 4);
OK
4       dell    1000012.0       laptop
Time taken: 0.662 seconds, Fetched: 1 row(s)
hive> select * from products_partitioned_buckets tablesample (bucket 2 out of 4);
OK
1       iphone  100000.0        mobile
Time taken: 0.212 seconds, Fetched: 1 row(s)
hive> select * from products_partitioned_buckets tablesample (bucket 3 out of 4);
OK
2       samsung 20000.0 mobile
Time taken: 0.186 seconds, Fetched: 1 row(s)
hive> select * from products_partitioned_buckets tablesample (bucket 4 out of 4);
OK
3       boat    1000.0  earphones
Time taken: 0.192 seconds, Fetched: 1 row(s)
hive>
```