• Start hive and create new database named **ccfd_capstone_project** and switch to **ccfd_capstone_project** database.

```
create database ccfd_capstone_project;
use ccfd_capstone_project;
```

Hive Parameter

```
set hive.auto.convert.join=false;
set hive.stats.autogather=true;
set orc.compress=SNAPPY;
set hive.exec.compress.output=true;
set
mapred.output.compression.codec=org.apache.hadoop.io.compress.SnappyCod
ec; set mapred.output.compression.type=BLOCK;
set mapreduce.map.java.opts=-Xmx5G; set mapreduce.reduce.java.opts=-
Xmx5G;
set mapred.child.java.opts=-Xmx5G -XX:+UseConcMarkSweepGC -XX:-
UseGCOverheadLimit;
```

Create an external table "**card_transactions_ext**" which will point to HDFS location created in ccfd_capstone_project location.

```
CREATE EXTERNAL TABLE IF NOT EXISTS CARD_TRANSACTIONS_EXT(
`CARD_ID` STRING,`MEMBER_ID` STRING,`AMOUNT` DOUBLE,`POSTCODE`
STRING,`POS_ID` STRING,`TRANSACTION_DT` STRING,`STATUS` STRING)ROW
FORMAT DELIMITED FIELDS TERMINATED BY ',' LOCATION '/
ccfd_capstone_project/card_transactions'
TBLPROPERTIES("skip.header.line.count"="1");
```

Create table "**card_transactions_orc**" in ORC format for better performance.

```
CREATE TABLE IF NOT EXISTS CARD_TRANSACTIONS_ORC(`CARD_ID`
STRING,`MEMBER_ID` STRING,`AMOUNT` DOUBLE,`POSTCODE`
```

STRING,`POS_ID` STRING,`TRANSACTION_DT` TIMESTAMP,`STATUS` STRING) STORED AS ORC TBLPROPERTIES ("orc.compress"="SNAPPY");

Create hive-hbase integrated table which will be visible in HBase as well. "**card_transactions_hbase**" table

```
CREATE TABLE CARD_TRANSACTIONS_HBASE(`TRANSACTION_ID`
STRING,`CARD_ID` STRING,`MEMBER_ID` STRING,`AMOUNT`
DOUBLE,`POSTCODE` STRING,`POS_ID` STRING,`TRANSACTION_DT`
TIMESTAMP,
`STATUS` STRING) ROW FORMAT DELIMITED STORED BY
'org.apache.hadoop.hive.hbase.HBaseStorageHandler' WITH
SERDEPROPERTIES
("hbase.columns.mapping"=":key, card_transactions_family:card_id,
card_transactions_family:member_id, card_transactions_family:amount,
card_transactions_family:postcode, card_transactions_family:pos_id,
card_transactions_family:transaction_dt, card_transactions_family:status")
TBLPROPERTIES ("hbase.table.name"="card_transactions_hive");
```

Load data in "**card_transactions_orc**" table and type cast **transaction_dt** column in timestamp format

```
INSERT OVERWRITE TABLE CARD_TRANSACTIONS_ORC SELECT CARD_ID,
MEMBER_ID, AMOUNT, POSTCODE,
POS_ID,CAST(FROM_UNIXTIME(UNIX_TIMESTAMP(TRANSACTION_DT,'dd-MM-
yyyy HH:mm:ss')) AS TIMESTAMP), STATUS FROM CARD_TRANSACTIONS_EXT;
```

Verify **transaction_dt** and **year** columns in "**card_transactions_orc**" table.

```
select year(transaction_dt), transaction_dt from card_transactions_orc limit 10;
```

Load data in "**card_transactions_hbase**" table which will be visible in HBase as well with table name as "**card_transactions_hive**".Using randomUUID to populate TRANSACTION_ID field (row key).

INSERT OVERWRITE TABLE CARD_TRANSACTIONS_HBASE SELECT reflect('java.util.UUID', 'randomUUID') as TRANSACTION_ID, CARD_ID, MEMBER_ID, AMOUNT, POSTCODE, POS_ID, TRANSACTION_DT, STATUS FROM CARD_TRANSACTIONS_ORC;

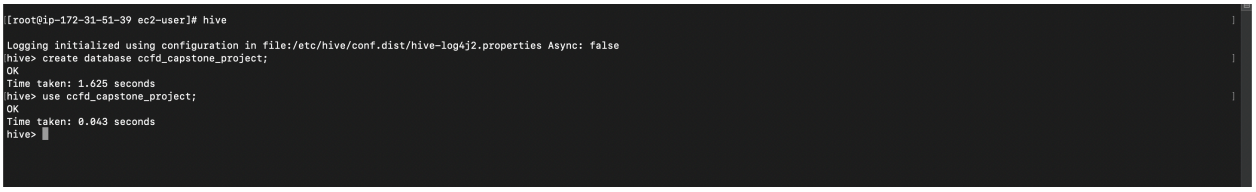
Verify data in "**card_transactions_hbase**" table.
select * from card_transactions_hbase limit 10;

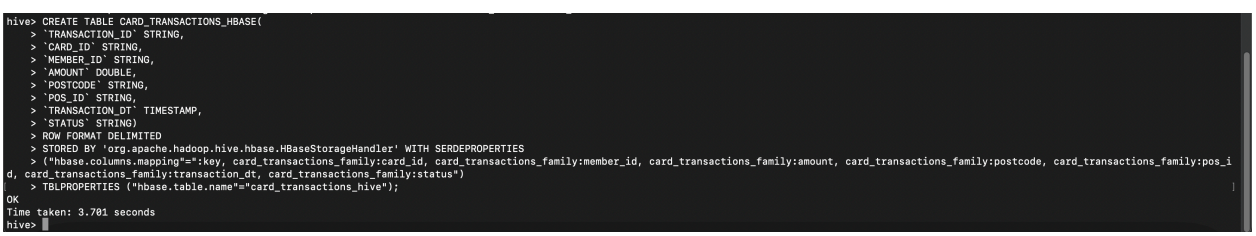
Verify details of "**card_transactions_hive**" table (hive-hbase integrated table).
describe 'card_transactions_hive'

Check count of "**card_transactions_hive**" table.

count 'card_transactions_hive''

Expected output : 53292 rows





```
hive> CREATE TABLE IF NOT EXISTS CARD_TRANSACTIONS_ORC(`CARD_ID` STRING,`MEMBER_ID` STRING,`AMOUNT` DOUBLE,`POSTCODE` STRING,`POS_ID` STRING,`TRANSACTION_DT` TIMESTAMP,`STATUS` STRING) STORED AS ORC TBLPROP
ERTIES ("orc.compress"="SNAPPY");
OK
Time taken: 0.445 seconds
hive>
```

```
hive> INSERT OVERWRITE TABLE CARD_TRANSACTIONS_HBASE SELECT
    > reflect('java.util.UUID', 'randomUUID') as TRANSACTION_ID, CARD_ID, MEMBER_ID, AMOUNT, POSTCODE, POS_ID, TRANSACTION_DT, STATUS
    > FROM CARD_TRANSACTIONS_ORC;
Query ID = root_20220919075508_31c646d0-3e81-4eb0-b571-e2ae9b520902
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1663572699218_0001)

--------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 ......... container     SUCCEEDED      1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 01/01  [==========================>>] 100%  ELAPSED TIME: 10.15 s
--------------------------------------------------------------------------------
OK
Time taken: 13.008 seconds
hive>
```

```
hive> INSERT OVERWRITE TABLE CARD_TRANSACTIONS_ORC SELECT CARD_ID, MEMBER_ID, AMOUNT, POSTCODE, POS_ID,
    > CAST(FROM_UNIXTIME(UNIX_TIMESTAMP(TRANSACTION_DT,'dd-MM-yyyy HH:mm:ss')) AS TIMESTAMP), STATUS
    > FROM CARD_TRANSACTIONS_EXT;
Query ID = root_20220919075221_0610f937-fbab-418f-ae2b-c81d642e110b
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1663572699218_0001)

--------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 ......... container     SUCCEEDED      1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 01/01  [==========================>>] 100%  ELAPSED TIME: 8.63 s
--------------------------------------------------------------------------------
Loading data to table ccfd_capstone_project.card_transactions_orc
OK
Time taken: 13.919 seconds
hive>
```

```
hive> select * from card_transactions_hbase limit 10;
OK
00007260-aaf8-4e9a-981e-d77e04b3df78    4492996315795951        626213673450614 9173027.0       20862   588879563387927 2017-06-20 05:56:03     GENUINE
00009b0c-2a5e-4115-b297-212a2426adbd    6553602843717324        489363495489840 6580216.0       47578   552792990977827 2017-04-05 14:38:43     GENUINE
00027002-4437-4f9e-a615-9246a0931f85    4045133934926419        195470545210125 9433555.0       36584   074727175934727 2018-01-09 19:45:52     GENUINE
0003eff1-795c-4d31-9435-70ec4eeab61f    4639774182867649        121565329352738 5873696.0       28552   766017189003893 2017-04-09 08:10:51     GENUINE
000477f1-3ec4-48eb-91a0-5b485aadd276    5447253073779618        044484418172299 4332923.0       54202   891043855542738 2017-05-12 16:44:02     GENUINE
00083afa-3aa2-4e7e-b711-3010b7852e9f    4039137046637960        968236238251435 2458569.0       31013   069820400514958 2018-01-15 10:56:44     GENUINE
0008db58-531d-4a4f-a68f-ab0c1bc793b4    3710843256845452 326158703866139 9451553.0       40060   626564244657140 2017-08-01 00:22:51      GENUINE
0009424f-8b6a-4ca0-b497-05aceb4e5d90    4593066224121489        110100522784085 2276006.0       44682   677950352387976 2017-11-07 02:38:18      GENUINE
000997f8-8e93-42e3-9aeb-78c7f7d5097d    4051600947391295        378428796688739 3192962.0       58647   875343854926444 2017-06-06 17:08:54      GENUINE
000ad4b6-55af-4270-9bc2-2b5c9b84950b    3473611606702352 270322592246125 8253483.0       81144   986503382359430 2016-06-09 06:09:32      GENUINE
Time taken: 0.279 seconds, Fetched: 10 row(s)
hive>
```

```
hive> select year(transaction_dt), transaction_dt from card_transactions_orc limit 10;
OK
2018    2018-02-11 00:00:00
2018    2018-02-11 00:00:00
2018    2018-02-11 00:00:00
2018    2018-02-11 00:00:00
2018    2018-02-11 00:00:00
2018    2018-02-11 00:00:00
2018    2018-02-11 00:00:00
2018    2018-02-11 00:00:00
2018    2018-02-11 00:00:00
2018    2018-02-11 00:00:00
Time taken: 0.195 seconds, Fetched: 10 row(s)
hive>
```

```
hbase(main):001:0> describe 'card_transactions_hive'
Table card_transactions_hive is ENABLED
card_transactions_hive
COLUMN FAMILIES DESCRIPTION
{NAME => 'card_transactions_family', BLOOMFILTER => 'ROW', VERSIONS => '1', IN_MEMORY => 'false', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', TTL => 'FOREVER', COMPRESSION => 'NONE', MIN_V
ERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0'}
1 row(s) in 0.4000 seconds

hbase(main):002:0>
```