

- 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

```
[root@ip-172-31-51-39 ec2-user]# hive
Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j2.properties Async: false
hive> create database ccfd_capstone_project;
OK
Time taken: 1.625 seconds
hive> use ccfd_capstone_project;
OK
Time taken: 0.843 seconds
hive>
```

```
hive> 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");
OK
Time taken: 0.349 seconds
hive>
```

```
hive> 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_i
d, card_transactions_family:transaction_dt, card_transactions_family:status")
> TBLPROPERTIES ("hbase.table.name"="card_transactions_hive");
OK
Time taken: 3.701 seconds
hive>
```

```
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 TBLPROPERTIES ('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_31c646d8-3a81-4eb0-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_202209190755221_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: 0.63 s

Loading data to table ccf4_capstone_project.card_transactions_orc
OK
Time taken: 13.919 seconds
hive>

```
hive> select * from card_transactions_hbase limit 10;
OK
00007260-aaf0-4e9a-981e-d77e04b3df78 4492996315795951 626213673450614 9173027.0 20862 588879563387927 2017-06-20 05:56:03 GENUINE
0000700c-2a5e-411b-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-4d31-9435-70ecdeeb051f 4639774102867649 121565329352738 5873696.0 28552 766017189083093 2017-04-09 08:18:51 GENUINE
000477f1-3e0c-40eb-91a0-5b405a0d276 5447253072779610 044404431972299 4392923.0 54202 071043085562738 2017-05-12 16:44:02 GENUINE
00083afa-3aa7-4a7e-b711-3010b785269f 4039137046637960 968236238251435 2458569.0 31013 069820400514958 2018-01-15 10:55:44 GENUINE
0008db50-531d-4a4f-a68f-ab0c0bc793b4 37108432504552 326158703846139 9451553.0 40060 626564244657140 2017-08-01 00:22:51 GENUINE
0009424f-bb6a-4ca0-b497-05aceb4e5d90 4593066224121489 110100522784085 2276006.0 44682 677950352387976 2017-11-07 02:38:18 GENUINE
000977f8-0a93-42e3-9aeb-78c77d0097d 4051600947391295 370428796688739 3192962.0 58647 875343854926444 2017-06-06 17:08:54 GENUINE
000ad40c-b5af-427e-00c2-28c0b04900b 347361106702352 270322592246125 0253403.0 01144 906508382359430 2016-06-09 00:09:32 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>
```

```
Current count: 14800, row: 43dcAc89-a1e1-42c7-996f-a8ab6b3b5fc8
Current count: 15000, row: 488712b9-7d7a-42e2-ad1a-b1d7e0b8bb66
Current count: 16000, row: 4d185816-1d35-4c57-a197-e6332865d987
Current count: 17000, row: 521af65a-179c-46b6-a60a-e6072dc9bba8
Current count: 18000, row: 57086c20-7dad-4177-b591-bd22941471b0
Current count: 19000, row: 5bc7ca8c-9ba6-4e32-a9e1-6cb9a54ed854
Current count: 20000, row: 69a2a3bd-dc24-4f8c-abb6-380d08e62528
Current count: 21000, row: 652c623d-4860-4d7c-930a-3ae43c73ad9e
Current count: 22000, row: 6a196d8e-sc2a-4440-9757-987b03147d68
Current count: 23000, row: 6f2f74d4-162a-4bfe-99f1-a8e006ae181c
Current count: 24000, row: 73e71d49-a17b-4cbf-a878-f3e5c471011d
Current count: 25000, row: 78d7d327-d844-4da4-84be-db2ed6237c5c
Current count: 26000, row: 7d8d3642-9b8f-4ccf-9e1a-a69525fb22c
Current count: 27000, row: 82870543-a0bd-4ffa-86f7-4cb9b9cac907
Current count: 28000, row: 87628e4f-37ce-43de-aaff-f2b31a836faf
Current count: 29000, row: 8c454538-2ba6-47d5-a0f3-5a40993ef8b6
Current count: 30000, row: 912ae272-5f01-4cd8-b4b7-68f815f4ae02
Current count: 31000, row: 9608a485-d04e-4118-8836-45afb61818e6
Current count: 32000, row: 9ab6e9be-9209-479b-b267-ff76a60b4975
Current count: 33000, row: 9fa69dbd-b1c2-4415-93aa-8be804480d14
Current count: 34000, row: a43d0dac-3db3-4c34-b292-64ee2fe39c7d
Current count: 35000, row: a9189a5f-a336-483d-8f9c-9b157d635490
Current count: 36000, row: ae02cb67-d02b-4338-b3d0-1a3e04f79550
Current count: 37000, row: b2d482ff-cbcb-48dd-a833-3e2a4bf061b9
Current count: 38000, row: b79a8ab1-aed7-4b65-a01f-fa9467b3c2d
Current count: 39000, row: bc5f47a9-0ada-4446-96ee-bd60ae9f4f59
Current count: 40000, row: c0e07f5c-d1e9-4f7f-b95d-64f9ebeb28
Current count: 41000, row: c5960394-115e-4995-b42f-191c489b1184
Current count: 42000, row: ca31c7e8-d826-4033-849b-55a0029e0e51
Current count: 43000, row: ce087bf4-8b5a-44ca-96f7-f6585bfa73a7
Current count: 44000, row: d3cdc79-dab1-402c-8256-a69b08063afe
Current count: 45000, row: d8789529-7e7e-4f3c-b66f-4ffb7f2228e2
Current count: 46000, row: d05d8f8-70fc-452c-902f-5773e8ce4e87
Current count: 47000, row: e215794a-d0dc-4532-92d6-797cfd4a4f4c
Current count: 48000, row: e6fae5e5-9d9b-aecc-851a-ed890332fed2
Current count: 49000, row: eb9cc56e-4a45-48e8-89da-cae6227739cb
Current count: 50000, row: f02f2877-539e-46b6-8ee0-c918fd6db2ff
Current count: 51000, row: faf6a181-74a8-4806-9a63-50ff44ddc520
Current count: 52000, row: f9d77042-1dfa-413c-a7ad-c09b4ddd8c88
Current count: 53000, row: fe93fbcd-45a1-44cf-be06-084bd7a49c41
53292 row(s) in 4.7540 seconds

-> 53292
hbase(main):003:0>
```