

## Loading Historical Transactions Data into NoSQL Database

### Commands to load the past transactions data into NoSQL database

#### 1. \_\_\_\_\_Hive operation starts here\_\_\_\_\_

- Start hive and create new database named **capstone\_project** and switch to **capstone\_project** database .

```
create database capstone_project;  
use capstone_project;
```

#### Set below parameters for the hive session

```
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.SnappyCodec;  
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 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 '/capstone_project/card_transactions'  
TBLPROPERTIES ("skip.header.line.count"="1"); >
```

#### Screenshot of the table created

```
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 '/capstone_project/card_transactions' TBLPROPERTIES  
  > ("skip.header.line.count"="1");  
OK  
Time taken: 0.19 seconds  
hive> █
```

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");
```

```
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: 0.069 seconds  
hive> █
```

- **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");
```

```
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_id,
> card_transactions_family:transaction_dt, card_transactions_family:status")
> TBLPROPERTIES ("hbase.table.name"="card_transactions_hive");
OK
Time taken: 3.702 seconds
hive>
```

\_\_\_\_\_ Hive operations : ends here \_\_\_\_\_

2. < command to list the table in which the data is loaded and the command to get the count of the rows of the tables >

\_\_\_\_\_ Hive operations : starts here \_\_\_\_\_

- 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;

```
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_20240715205447_b44358c6-0584-4ebe-8f97-6ee489c51929
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1721070765306_0006)

-----
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: 7.84 s
-----
Loading data to table capstone_project.card_transactions_orc
OK
Time taken: 18.641 seconds
hive>
```

- Verify **transaction\_dt** and year columns in "**card\_transactions\_orc**" table.  
select year(transaction\_dt), transaction\_dt from card\_transactions\_orc limit 10;

```
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: 0.189 seconds, Fetched: 10 row(s)
hive>
```

- 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;

```
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_20240715205834_4b11acc7-44e2-4546-8136-8297e57ed601
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1721070765306_0006)

-----
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: 10.81 s
-----
OK
Time taken: 15.476 seconds
hive>
```

- Verify data in "**card\_transactions\_hbase**" table.  
select \* from card\_transactions\_hbase limit 10;

```
hive> select * from card_transactions_hbase limit 10;
OK
0003be52-345a-4380-b523-81f13b97d297 344402144417599 920759851362201 5113055.0 10519 614421366821657 2017-08-24 13:41:01 GENUINE
0003bfc2-1204-4511-8423-37744880104c 4877961603808028 015235737327750 9822669.0 15681 357202104572567 2017-08-15 04:58:00 GENUINE
00046d01-f824-4e21-b2e6-d3fa59196071 5309060557968733 867154068331056 1435512.0 18622 623663762504748 2017-06-20 14:42:44 GENUINE
00049eb7-d40c-4cf7-adc2-096c45c285b1 343252340741735 055787011798644 9427157.0 25611 306400889270750 2016-04-24 23:26:37 GENUINE
000604ce-bd96-4107-a814-2190acbefa9f 4174085648252388 169940172498283 651566.0 10536 930025146344734 2017-10-08 02:50:34 GENUINE
00063d0f-4d19-4ca7-b4a8-35dala95c3bd 5400251558458125 006836124210484 3540256.0 28039 201514353456964 2017-09-12 14:08:21 GENUINE
0007e7d4-a309-4c9d-867d-a5beb80f6050 345864096406128 246259923541675 8051852.0 35060 612950113705968 2017-11-22 20:05:31 GENUINE
00085e9e-3303-4e71-80c5-6df7bb82b0d4 379415836364353 983390343419665 621976.0 26101 693046989957830 2017-07-08 11:04:42 GENUINE
0008b03c-f63c-4144-bac9-7d0b2a3f5218 6223681626729292 601816584611264 3635507.0 16853 121320191081888 2017-10-10 14:32:00 GENUINE
0009850e-5542-4e49-84ba-fc2c6d9d0d5a 4599707948756480 590646959017984 7899529.0 88063 456628347220724 2018-01-21 23:11:37 GENUINE
Time taken: 0.281 seconds, Fetched: 10 row(s)
hive>
```

----- Hive Operations: Ends Here-----

## -----Hbase Operations: Starts Here-----

- Verify details of "card\_transactions\_hive" table (hive-hbase integrated table).  
describe 'card\_transactions\_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_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0')
1 row(s) in 0.3430 seconds
hbase(main):002:0>
```

Check count of "card\_transactions\_hive" table.

count 'card\_transactions\_hive'

```
hbase(main):002:0> count 'card_transactions_hive'
Current count: 1000, row: 04d6b6d3-2c2d-4f05-9846-b2ad163de3ed
Current count: 2000, row: 09472dcf-74f7-46f8-bde1-bcd88572cc06
Current count: 3000, row: 0ddcf511-d9e9-43c0-9c28-e7cfc2b95e30
Current count: 4000, row: 12ebfd47-9439-41b9-a8e6-919348c1edd9
Current count: 5000, row: 17ba400d-3ec5-42af-8e51-f46a9eba1f27
Current count: 6000, row: 1c874533-7a76-45d1-a76e-6361f9d60060
Current count: 7000, row: 214b939f-6af5-4ab7-bf4a-79aeae16b329
Current count: 8000, row: 26209827-d769-4f37-b8c0-f165fe60c4c9
Current count: 9000, row: 2af185f2-6784-4108-a382-e6adbd5fffc8d
Current count: 10000, row: 2f954aba-0fe4-47a7-b66e-7d7849cab8f7
Current count: 11000, row: 34334440-3af6-42c3-9ca0-15d537f7bfcf
Current count: 12000, row: 392025df-0353-4e8f-b131-10ba30c6e7ca
Current count: 13000, row: 3ddeb3d5-a7f3-4541-b965-a79dda0deca3
Current count: 14000, row: 42a387a1-b165-4f0f-90dc-53e187a12740
Current count: 15000, row: 47c04508-ef5f-4db7-be7f-81653f5f2765
Current count: 16000, row: 4cac802b-72fc-4747-bbe6-08cd0368b3cb
Current count: 17000, row: 518e1590-663f-44c2-bf6f-33755f272792
Current count: 18000, row: 56650426-f85f-4d30-bdb2-55f3d03cd0ea
Current count: 19000, row: 5b4457b3-81b8-4376-ba1f-76fb70fe2ab0
Current count: 20000, row: 5fc93310-71fa-40b5-8fef-e7d595addb86
Current count: 21000, row: 6497cfe2-a462-47d1-886d-c0b855eb103b
Current count: 22000, row: 691824b4-31c3-4a45-8817-ceb7550589e3
Current count: 23000, row: 6e01aee8-fbcc-4ce6-b14a-e2caf8a51c56
Current count: 24000, row: 72ca7531-e2a8-4300-98ec-663468e567bd
Current count: 25000, row: 77cfe998-30ff-4f68-be82-0e7ab515c78a
Current count: 26000, row: 7cbea54e-3f09-4ba6-9d2b-7836a5ca60b5
Current count: 27000, row: 819ed109-60da-4a21-9e6f-0311107605fb
Current count: 28000, row: 86808a91-768e-48dd-b822-18498ea2ef3a
Current count: 29000, row: 8b8befed-aa6d-4c58-b3f8-a0918657890a
Current count: 30000, row: 903c62de-6947-4d1d-90cf-e01a803b9c9e
Current count: 31000, row: 950be881-d851-4c85-bd67-a3ca25337e6d
Current count: 32000, row: 99a85383-880b-4daa-95d7-6b51810c6582
Current count: 33000, row: 9ead90f1-8577-4d7f-8516-c6989d5db08a
```

```
Current count: 30000, row: 950be881-d851-4c85-bd67-a3ca25337e6d
Current count: 31000, row: 99a85383-880b-4daa-95d7-6b51810c6582
Current count: 32000, row: 99a85383-880b-4daa-95d7-6b51810c6582
Current count: 33000, row: 9ead90f1-8577-4d7f-8516-c6989d5db08a
Current count: 34000, row: a3800bff-c8bb-42cd-8a67-f1eef171e6fb
Current count: 35000, row: a8577412-e5ce-468b-b9cd-d2d1f911f041
Current count: 36000, row: ad3e64bd-aeee-473d-acd8-b92e30de83f5
Current count: 37000, row: b2347a6e-b0b9-476b-9475-e2a01858b077
Current count: 38000, row: b6980f81-ba32-4577-b5e7-e03d600eb0fb
Current count: 39000, row: bb74b15d-e75b-425d-8fda-9db0a306586e
Current count: 40000, row: c067b635-4581-4879-9228-a49d54a32a8e
Current count: 41000, row: c5131c6a-8799-486c-8b3f-8df5a54dalca
Current count: 42000, row: c9db5496-87dd-41ff-a15c-e5e552c56645
Current count: 43000, row: ceb9e50e-0e66-4490-a1d0-5f9925bbb84c
Current count: 44000, row: d39c50af-7ece-44aa-a741-5bcb347a6a3d
Current count: 45000, row: d8874b91-bf15-4701-a8b5-7fef8e476d82
Current count: 46000, row: dd21397d-5fe5-4762-9b66-305703c8e529
Current count: 47000, row: elb3dfbd-05e8-43a4-93e7-974d591f4dd9
Current count: 48000, row: e6a70968-2c66-4d7c-82ba-ba0f9eb9a362
Current count: 49000, row: eb7d3027-cdc2-43a9-936e-7b061a13bcba
Current count: 50000, row: f03e7c81-d30e-464e-a5eb-6e6bab229fc5
Current count: 51000, row: f4bf944b-e967-44ef-a5f6-ea42aeafbae3
Current count: 52000, row: f98df591-56c0-4079-806f-0955d76a6e92
Current count: 53000, row: fe77da53-9e59-4f7a-bd97-cad375b52d7f
53292 row(s) in 3.6110 seconds

=> 53292
hbase(main):003:0> █
```

----- Hbase Operations: Ends Here-----

Total number for record is 53292