



Scripts Execution

Screenshots of the execution of the scripts written

Creation of the database:

```
hive> create database ccfraud_capstone;

OK
Time taken: 0.481 seconds
hive> use ccfraud_capstone;

OK
Time taken: 0.083 seconds
hive>
```

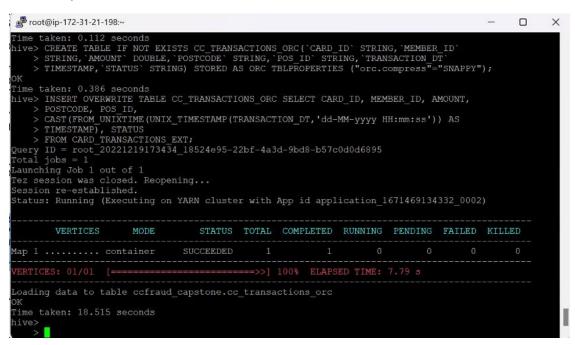
Creating an external table CARD_TRANSACTIONS_EXT

Creating table " CC_TRANSACTIONS_ORC" in ORC format for better performance.





Loading the data in "CC_TRANSACTIONS_ORC" table and type casting transaction_dt column to timestampformat



Verifying transaction_dt and year columns in " CC_TRANSACTIONS_ORC " table.





Creating an integrated hive - hbase table that will be visible in HBase as well." **CC_TRANSACTIONS_HBASE** " table

```
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
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
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
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
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
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
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
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
2018 2018-02-11 00:00:00
2018 2018-02-11 00:00
```

Loading data in "CC_TRANSACTIONS_HBASE" table that will be visible in HBase as well with table name as "cc_transactions_hive". Using randomUUID to populate TRANSACTION_ID field (row key).





Verify data in "cc_transactions_hbase" table.

```
₽ root@ip-172-31-21-198:~
                                                                                                                                                                            ×
ime taken: 15.654 seconds
live> select * from cc_transactions_hbase limit 10;
 001cef0-4d5b-4d88-b0ce-0da380409a66
                                                                                                                        413021843879928 1604268.0
             214696179763291 2016-03-27 05:29:37
 005ec01-c98b-4d41-9d17-b2cf0ff5e960
                                                                                                                        991784667284668 1590198.0
005ec01-c99b-4d41-9d17-b2c10115e960 5271

5 225865830430921 2017-05-30 20:07:20

10070787-5f76-4fb6-bb79-45aed2e3a5fd 6460

19 893307976868097 2017-01-12 15:47:50

1007767e-56f7-407f-b2fe-b50306d83dfb 5175

14 215945504043719 2018-02-11 00:00:00

1007c0d3-3b77-4304-b2b6-cdd6e2333937 6223
                                                                                                                        266629518896098 4366995.0
                                                                                                                        227686059795799 2082575.0
                                                                                                                        736137659378265 1714630.0
             748648404529559 2017-07-29 05:40:05
                                                                                         GENUINE
  00916dc-4a82-4068-9062-02bb410fa5c4
000916dc-4a82-4068-9062-02bb410fa5c4 5134:
00 569335687828707 2016-05-27 03:54:25
000a5c44-d3f7-4650-9e03-55d0747a856b 4389;
07 911478839996964 2017-04-17 05:47:08
000b85c8-16ff-4f9a-9d18-f13292fa4005 6011:
2 478785382784557 2017-12-14 22:16:57
000baa11-7d7b-4682-94dc-5663885bc269 3417:
082736002337 2017-02-06 05:40:42 GENU:
                                                                                                                        295554828848966 6305493.0
                                                                                                                        199243620982420 9732433.0
                                                                         341722035429601 979218131207765 8325634.0
                                                                                                                        577907767500023 1988126.0
             814499910586436 2016-02-23 00:27:07
```

Verify count of "cc_transactions_hive" table





```
Current count: 28000, row: 87685a0a-05c3-4a42-a279-210baadf35bc
Current count: 29000, row: 8c4e959d-c385-453f-8f62-e0d01217173e
Current count: 30000, row: 9c8731b-3225-4f70-a5cf-3db1156bce4a
Current count: 31000, row: 95a87f67-0b18-4846-8d67-3366421bce46
Current count: 32000, row: 9f16276a-f6ad-4e22-9850-e71f1dc869de
Current count: 33000, row: 9f16276a-f6ad-4e22-9850-e71f1dc869de
Current count: 35000, row: a89aadle-d088-4909-a305-1dee9cf09ac6
Current count: 35000, row: a89aadle-d088-4909-a305-1dee9cf09ac6
Current count: 37000, row: a89aadle-d088-4909-a305-1dee9cf09ac6
Current count: 37000, row: b6956bff-e40d-4a07-a245-09887745840f
Current count: 39000, row: b6956bff-e40d-4a07-a245-09887745840f
Current count: 40000, row: c9565bff-e40d-4a07-a245-09887745840f
Current count: 40000, row: c9565bff-e40d-4a07-a245-09887745840f
Current count: 41000, row: c9565bff-e70d-8d-2ed-a62d-9df2865e9fd6
Current count: 41000, row: c9565bff-e70d-8d-2ed-a62d-9df2865e9fd6
Current count: 41000, row: c9565bff-e70d-8d-2ed-a62d-9df2865e9fd6
Current count: 40000, row: c9565bff-e70f3-a565bff-e80d-4d-2ed-a62d-9df2865e9fd6
Current count: 40000, row: c9565bff-e70f3-a565bff-e80d-4df15bf875b4
Current count: 40000, row: c61b5ff8-9951-dc27-9163-a401fb6875b4
Current count: 45000, row: c61b5ff8-9951-dc27-9163-a401fb6875b4
Current count: 45000, row: dc652057-071e-4ad7-bab3-5bef2f0d247f
Current count: 45000, row: dc66665e-715e-335-996-684021f89795
Current count: 47000, row: dc66665e-715e-335-996-684021f89795
Current count: 50000, row: c6ab665e-715e-335-996-684021f89795
Current count: 50000, row: c6ab665e-715e-335-996-684021f89795
Current count: 50000, row: dc66666-444-4214-9780-c24131b08f51
Current count: 50000, row: f50c5b66-444-4214-9780-c24131b08f51
Current count: 53000, row: c6ab665e-444-4214-9780-c24131b08f51
Current count: 53000, row: f9c8db4c-6bc0-4a24-8734-1a4edc19b3dd
Current count: 53000, row: f9c8db4c-6bc0-4a24-8734-1a4edc19b3dd
Current count: 53000, row: f9c8db4c-6bc0-6bc0-6a24-8734-1a4edc19b3dd
Current count: 53000, row: f9c8db4c-6bc0-6bc
```

Count of the "cc_transactions_hive" table is 53,292 which is matching with given requirement

Running Sqoop command to import "member_score" table from RDS to HDFS.

```
### C2/USF(0)p-172-31-21-198*

### C2/USF(0)p-172-31-21-198*
```

Running Sqoop command to import "card_member" table from RDS to HDFS.





```
ec2-user@ip-172-31-21-198:~
                                                                                                                                                                  - 🗆 X
22/12/19 18:13:22 INFO mapreduce.Job: Job job_1671469134332_0006 completed successfully 22/12/19 18:13:22 INFO mapreduce.Job: Counters: 30
              File System Counters
                               FILE: Number of bytes written=189979
                              FILE: Number of bytes written=169979
FILE: Number of read operations=0
FILE: Number of large read operations=0
HDFS: Number of bytes read=87
HDFS: Number of bytes written=85081
HDFS: Number of read operations=4
                              HDFS: Number of read operations=4
HDFS: Number of large read operations=0
                              HDFS: Number of write operations=2
                              Other local map tasks=1
Total time spent by all maps in occupied slots (ms)=216960
Total time spent by all reduces in occupied slots (ms)=0
Total time spent by all map tasks (ms)=4520
Total vcore-milliseconds taken by all map tasks=4520
Total megabyte-milliseconds taken by all map tasks=6942720
              Map-Reduce Framework
                              Map input records=999
                              Map output records=999
Input split bytes=87
Spilled Records=0
Failed Shuffles=0
                              Merged Map outputs=0
GC time elapsed (ms)=85
                               CPU time spent (ms)=3070
                              Physical memory (bytes) snapshot=284651520
Virtual memory (bytes) snapshot=3302727680
Total committed heap usage (bytes)=246415360
              Bytes Read=0
File Output Format Counters
Bytes Written=85081
 2/12/19 18:13:22 INFO mapreduce.ImportJobBase: Transferred 83.0869 KB in 19.3302 seconds (4.2
 B3 KB/sec)
 2/12/19 18:13:22 INFO mapreduce.ImportJobBase: Retrieved 999 records. ec2-user@ip-172-31-21-198 \sim]$ \cap
```

Verify data in "CC_MEMBER_ORC" table.

```
====>>] 100% ELAPSED TIME: 4.82 s
  oading data to table default.cc_member_score_orc
Time taken: 5.872 seconds
hive> SELECT * FROM CC_MEMBER_ORC LIMIT 10;
 340028465709212 009250698176266 2012-02-08 06:04:13
                                                                                                                                               United States
340028465709212 009250698176266 2012-02-08 06:04:13 340054675199675 835873341185231 2017-03-10 09:24:44 340082915339645 512969555857346 2014-02-15 06:30:30 340134186926007 887711945571282 2012-02-05 01:21:58 340265728490548 680324265406190 2014-03-29 07:49:14 340268219434811 929799084911715 2012-07-08 02:46:08 340379737226464 089615510858348 2010-03-10 00:06:42 340383645652108 181180599313885 2012-02-24 05:32:44 340803866934451 417664728506297 2015-05-21 04:30:45 340889618969736 459292914761635 2013-04-23 08:40:11 Time taken: 0.188 seconds.
                                                                                                                                                                                   Barberton
                                                                                                                                               United States
                                                                                                                                                                                   Fort Dodge
                                                                                                                                               United States
                                                                                                                                               United States
                                                                                                                                                                                  San Francisco
Clinton
                                                                                                                            09/10
10/16
08/17
                                                                                                                                               United States
                                                                                                                                               United States
                                                                                                                                                                                   Beaverton
                                                                                                                                               United States
                                                                                                                                                                                   West Palm Beach
  ime taken: 0.198 seconds, Fetched: 10 row(s)
```

Verify data in "CC_MEMBER_SCORE_ORC" table.





```
Proot@ip-172-31-21-198:/home/ec2-user
 otal jobs
 aunching Job 1 out of 1
 Status: Running (Executing on YARN cluster with App id application_1671469134332_0008)
                                     MODE
                                                          STATUS TOTAL COMPLETED RUNNING PENDING FAILED KILLED
 oading data to table default.cc member score orc
Time taken: 5.872 seconds
hive> SELECT * FROM CC_MEMBER_ORC LIMIT 10;
 40028465709212 009250698176266 2012-02-08 06:04:13
                                                                                                               United States
                                                                                                                                           Barberton
340028465709212 009250698176266 2012-02-08 06:04:13 340054675199675 835873341185231 2017-03-10 09:24:44 340082915339645 512969555857346 2014-02-15 06:30:30 340134186926007 887711945571282 2012-02-05 01:21:58 340265728490548 680324265406190 2014-03-29 07:49:14 340268219434811 929799084911715 2012-07-08 02:46:08 340379737226464 089615510858348 2010-03-10 00:06:42 340383645652108 181180599313885 2012-02-24 05:32:44 3408803866934451 417664728506297 2015-05-21 04:30:45 340889618969736 45992914761635 2013-04-23 08:40:11
                                                                                                 03/17
07/14
02/13
11/14
08/12
09/10
10/16
                                                                                                                United States
                                                                                                                                           Fort Dodge
                                                                                                                                           Dix Hills
                                                                                                               United States
United States
                                                                                                                                           San Francisco
                                                                                                               United States
                                                                                                               United States
                                                                                                                                           West New York
                                                                                                               United States
                                                                                                                                           Beaverton
 40889618969736 459292914761635 2013-04-23 08:40:11
                                                                                                               United States
                                                                                                                                           West Palm Beach
nive> SELECT * FROM CC_MEMBER_SCORE_ORC LIMIT 10;
000037495066290 339
000117826301530 289
 01147922084344 393
 01314074991813 225
 01739553947511 642
 03761426295463 413
006836124210484 504
006836124210484 504
006991872634058 697
007955566230397 372
Fime taken: 0.117 seconds, Fetched: 10 row(s)
```

Creating "CC_lookup_data_hbase" table (hive-hbase integrated table) which will be visible in HBase(lookup_data_hive).

```
hive> CREATE TABLE CC_LOOKUP_DATA_HBASE(`CARD_ID` STRING,`UCL` DOUBLE, `SCORE` INT, `POSTCODE`

> STRING, `TRANSACTION_DT` TIMESTAMP) STORED BY

> 'org.apache.hadoop.hive.hbase.HBaseStorageHandler' WITH SERDEPROPERTIES

> ("hbase.columns.mapping"=":key, lookup_card_family:ucl, lookup_card_family:score,

> lookup_transactions_family:postcode, lookup_transactions_family:transactions_dt") TBLPROPERTIES

> ("hbase.table.name" = "cc_lookup_data_hive");

OK

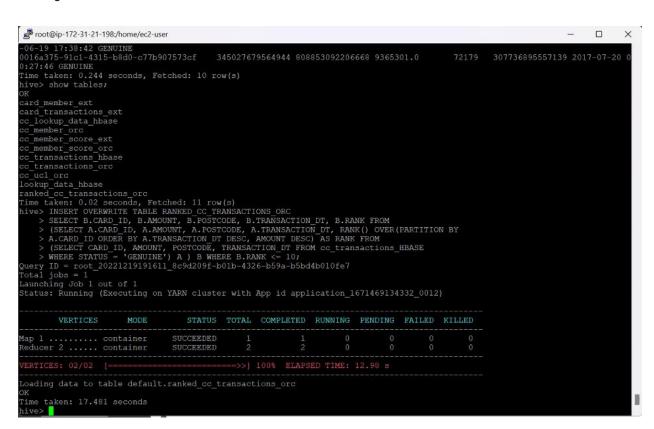
Time taken: 1.65 seconds
hive>
```

Verifying details of **cc_lookup_data_hive** (hive-hbase integrated) table:





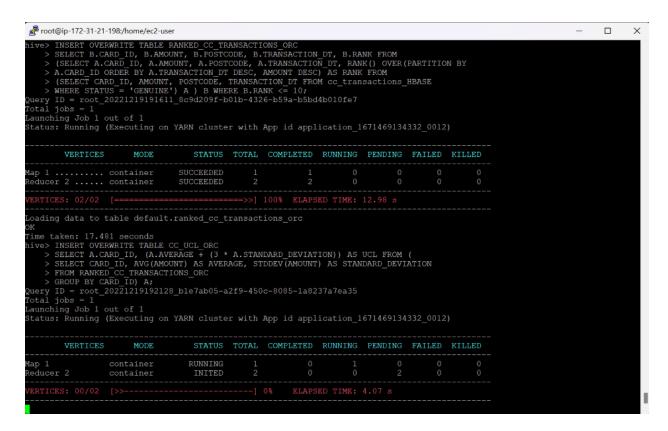
Loading data in "ranked_cc_transactions_orc" table



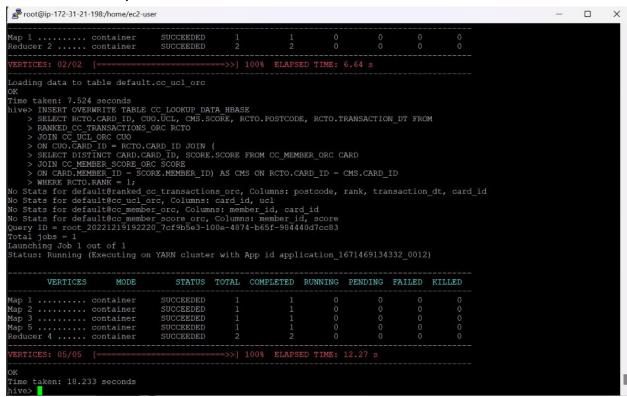
Loading data in "cc_ucl_orc" table







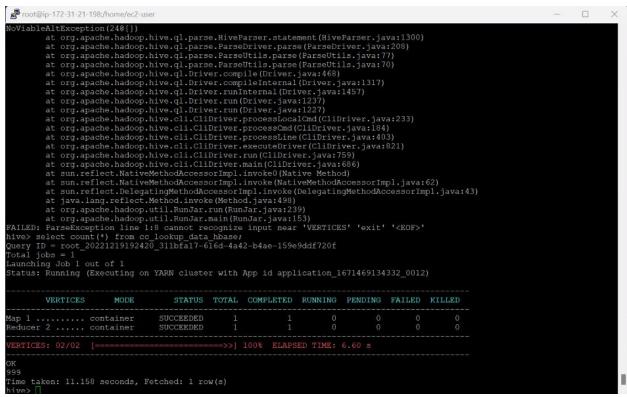
Load data in cc_lookup_data_hbase table.







Verifying count in "cc_lookup_data_hbase" table.



Total number of records is 999 which is matching with given requirement.

Verify some data in "cc_lookup_data_hbase" table.





Start HBase shell session and verifying count in "cc_lookup_data_hive" table.

```
hbase(main):002:0> count 'cc_lookup_data_hive'
999 row(s) in 0.3560 seconds
-> 999
hbase(main):003:0>
```

Total number of records is 999 which is matching with given requirement.

Verifying data in "cc_lookup_data_hive" table.





Proot@ip-172-31-21-198:/home/ec2-user	- D X
6592184145413632 6592184145413632	<pre>column=lookup_transactions_family:postcode, timestamp=1671477758927, value=53186 column=lookup_transactions_family:transactions_dt, timestamp=1671477758927, value=2018-01-2 8 00:54:30</pre>
6594248319343442	column=lookup card family:score, timestamp=1671477758927, value=350
6594248319343442	column=lookup card family:ucl, timestamp=1671477758927, value=1.4567957140418548E7
6594248319343442	column=lookup transactions family:postcode, timestamp=1671477758927, value=24927
6594248319343442	<pre>column=lookup_transactions_family:transactions_dt, timestamp=1671477758927, value=2018-01-3 1 23:42:38</pre>
6595638658736751	column=lookup card family:score, timestamp=1671477758927, value=310
6595638658736751	column=lookup card family:ucl, timestamp=1671477758927, value=1.356629177577566E7
6595638658736751	column=lookup transactions family:postcode, timestamp=1671477758927, value=68328
6595638658736751	<pre>column=lookup_transactions_family:transactions_dt, timestamp=1671477758927, value=2018-01-3 0 10:50:34</pre>
6595814135833988	column=lookup card family:score, timestamp=1671477758927, value=210
6595814135833988	column=lookup card family:ucl, timestamp=1671477758927, value=1.3926273240525039E7
6595814135833988	column=lookup transactions family:postcode, timestamp=1671477758927, value=22508
6595814135833988	<pre>column=lookup_transactions_family:transactions_dt, timestamp=1671477758927, value=2018-01-3 0 02:03:54</pre>
6595928469079750	column=lookup card family:score, timestamp=1671477758927, value=412
6595928469079750	column=lookup card family:ucl, timestamp=1671477758927, value=1.142797041440079E7
6595928469079750	column=lookup transactions family:postcode, timestamp=1671477758927, value=98349
6595928469079750	column=lookup_transactions_family:transactions_dt, timestamp=1671477758927, value=2018-01-2 4 12:38:22
6597703848279563	column=lookup card family:score, timestamp=1671477758927, value=218
6597703848279563	column=lookup card family:ucl, timestamp=1671477758927, value=1.4718634149498457E7
6597703848279563	column=lookup transactions family:postcode, timestamp=1671477758927, value=95699
6597703848279563	<pre>column=lookup_transactions_family:transactions_dt, timestamp=1671477758927, value=2018-01-2 7 10:51:49</pre>
6598830758632447	column=lookup card family:score, timestamp=1671477758927, value=293
6598830758632447	column=lookup card family:ucl, timestamp=1671477758927, value=1.2227949982601807E7
6598830758632447	column=lookup transactions family:postcode, timestamp=1671477758927, value=19421
6598830758632447	<pre>column=lookup_transactions_family:transactions_dt, timestamp=1671477758927, value=2018-01-3 0 00:18:34</pre>
6599900931314251	column=lookup card family:score, timestamp=1671477758927, value=297
6599900931314251	column=lookup card family:ucl, timestamp=1671477758927, value=1.2121408572464656E7
6599900931314251	column=lookup transactions family:postcode, timestamp=1671477758927, value=97423
6599900931314251	column=lookup_transactions_family:transactions_dt, timestamp=1671477758927, value=2018-01-3 11:25:16
999 row(s) in 5.6360 seconds	
hbase(main):006:0> 🗌	