

Scripts Execution

Screenshots of the execution of the scripts written

- a) **Task 1 : Load the transactions history data (card_transactions.csv) in a NoSQL database.**

----- Hive Operations: Starts Here -----

1. Start hive and create new database named capstone_project -> switch to

```
[root@ip-172-31-49-185 ec2-user]# hive
Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j2.properties Async: false
hive> create database capstone_project;
OK
Time taken: 1.09 seconds
hive> use capstone_project;
OK
Time taken: 0.065 seconds
hive> █
```

2. Create an external table “card_transactions_ext”

```
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> █
```

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

```
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>
```

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

```
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>
```

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

```
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>
```

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

```
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>
```

7. 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).

```
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>
```

8. Verify data in "card_transactions_hbase" table.

```
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-b4a8-35da1a95c3bd 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 -----

1. Start HBase and verify details of "card_transactions_hive" table (hive-hbase integrated table).

```
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>
```

2. Verify count of "card_transactions_hive" table

```
Current count: 30000, row: 950be881-d851-4c85-bd67-a3ca25337e6d
Current count: 31000, row: 99a85383-880b-4daa-95d7-6b51810c6582
Current count: 32000, row: 9ead90f1-8577-4d7f-8516-c6989d5db08a
Current count: 33000, row: a3800bff-c8bb-42cd-8a67-f1eef171e6fb
Current count: 34000, row: a8577412-e5ce-468b-b9cd-d2d1f911f041
Current count: 35000, row: ad3e64bd-aaaa-473d-acd8-b92e30de83f5
Current count: 36000, row: b2347a6e-b0b9-476b-9475-e2a01858b077
Current count: 37000, row: b6980f81-ba32-4577-b5e7-e03d600eb0fb
Current count: 38000, row: bb74b15d-e75b-425d-8fda-9db0a306586e
Current count: 39000, row: c067b635-4581-4879-9228-a49d54a32a8e
Current count: 40000, row: c5131c6a-8799-486c-8b3f-8df5a54da1ca
Current count: 41000, row: c9db5496-87dd-41ff-a15c-e5e552c56645
Current count: 42000, row: ceb9e50e-0e66-4490-a1d0-5f9925bbb84c
Current count: 43000, row: d39c50af-7ece-44aa-a741-5bcb347a6a3d
Current count: 44000, row: d8874b91-bf15-4701-a8b5-7fef8e476d82
Current count: 45000, row: dd21397d-5fe5-4762-9b66-305703c8e529
Current count: 46000, row: elb3dfbd-05e8-43a4-93e7-974d591f4dd9
Current count: 47000, row: e6a70968-2c66-4d7c-82ba-ba0f9eb9a362
Current count: 48000, row: eb7d3027-cdc2-43a9-936e-7b061a13bcba
Current count: 49000, row: f03e7c81-d30e-464e-a5eb-6e6bab229fc5
Current count: 50000, row: f4bf944b-e967-44ef-a5f6-ea42aeafbae3
Current count: 51000, row: f98df591-56c0-4079-806f-0955d76a6e92
Current count: 52000, row: fe77da53-9e59-4f7a-bd97-cad375b52d7f
53292 row(s) in 3.6110 seconds

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

Total number for record is **53292** which is matching with given requirement.

b) Task 2: Ingest the relevant data from AWS RDS to Hadoop.

- Run Sqoop command to import "member_score" table from RDS to HDFS.

```
[hadoop@ip-172-31-58-228 ~]$ sqoop import --connect jdbc:mysql://upgradawsrds1.cyaie1c9bmnf.us-east-1.rds.amazonaws.com/cred_financials_data \
> --username upgraduser \
> --password upgraduser \
> --table member_score \
> --null-string 'NA' \
> --null-non-string '\\N' \
> --delete-target-dir \
> --target-dir '/capstone_project/member_score' \
> -m 1
Warning: /usr/lib/sqoop/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
24/07/16 05:51:49 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/lib/hadoop/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/share/aws/redshift/jdbc/redshift-jdbc42-1.2.37.1061.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/lib/hive/lib/log4j-slf4j-impl-2.6.2.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
24/07/16 05:51:49 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
24/07/16 05:51:49 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
24/07/16 05:51:49 INFO tool.CodeGenTool: Beginning code generation
24/07/16 05:51:50 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'member_score' AS t LIMIT 1
24/07/16 05:51:50 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'member_score' AS t LIMIT 1
24/07/16 05:51:50 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-hadoop/compile/3a4cd039d9e7e3793224bf319fbb729/member_score.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
24/07/16 05:51:53 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-hadoop/compile/3a4cd039d9e7e3793224bf319fbb729/member_score.jar
24/07/16 05:51:54 INFO tool.ImportTool: Destination directory /capstone_project/member_score is not present, hence not deleting.
24/07/16 05:51:54 WARN manager.MySQLManager: It looks like you are importing from mysql.
24/07/16 05:51:54 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
24/07/16 05:51:54 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
24/07/16 05:51:54 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
```

```
24/07/16 05:52:07 INFO mapreduce.Job: Job job_1721107016086_0003 running in uber mode : false
24/07/16 05:52:07 INFO mapreduce.Job: map 0% reduce 0%
24/07/16 05:52:13 INFO mapreduce.Job: map 100% reduce 0%
24/07/16 05:52:13 INFO mapreduce.Job: Job job_1721107016086_0003 completed successfully
24/07/16 05:52:14 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=0
    FILE: Number of bytes written=189949
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=87
    HDFS: Number of bytes written=19980
    HDFS: Number of read operations=4
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
  Job Counters
    Launched map tasks=1
    Other local map tasks=1
    Total time spent by all maps in occupied slots (ms)=170736
    Total time spent by all reduces in occupied slots (ms)=0
    Total time spent by all map tasks (ms)=3557
    Total vcore-milliseconds taken by all map tasks=3557
    Total megabyte-milliseconds taken by all map tasks=5463552
  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)=71
    CPU time spent (ms)=2220
    Physical memory (bytes) snapshot=263106560
    Virtual memory (bytes) snapshot=3285831680
    Total committed heap usage (bytes)=245891072
  File Input Format Counters
    Bytes Read=0
  File Output Format Counters
    Bytes Written=19980
24/07/16 05:52:14 INFO mapreduce.ImportJobBase: Transferred 19.5117 KB in 19.6411 seconds (1,017.2571 bytes/sec)
24/07/16 05:52:14 INFO mapreduce.ImportJobBase: Retrieved 999 records.
[hadoop@ip-172-31-58-228 ~]$
```

- Run Sqoop command to import “**card_member**” table from RDS to HDFS.

```
[hadoop@ip-172-31-58-228 ~]$ sqoop import --connect jdbc:mysql://upgradawsrds1.cyaielc9bmnf.us-east-1.rds.amazonaws.com/cred_financials_data \
--delete-target-dir \
--target-dir '/capstone_project/card_member' \
-m l> --username upgraduser \
> --password upgraduser \
> --table card_member \
> --null-string 'NA' \
> --null-non-string '\\N' \
> --delete-target-dir \
> --target-dir '/capstone_project/card_member' \
> -m l
Warning: /usr/lib/sqoop/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
24/07/16 05:54:00 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/lib/hadoop/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/share/aws/redshift/jdbc/redshift-jdbc42-1.2.37.1061.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/lib/hive/lib/log4j-slf4j-impl-2.6.2.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
24/07/16 05:54:00 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
24/07/16 05:54:00 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
```

```
24/07/16 05:54:29 INFO mapreduce.Job: map 100% reduce 0%
24/07/16 05:54:30 INFO mapreduce.Job: Job job_1721107016086_0004 completed successfully
24/07/16 05:54:31 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=0
    FILE: Number of bytes written=190038
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=87
    HDFS: Number of bytes written=85081
    HDFS: Number of read operations=4
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
  Job Counters
    Launched map tasks=1
    Other local map tasks=1
    Total time spent by all maps in occupied slots (ms)=275472
    Total time spent by all reduces in occupied slots (ms)=0
    Total time spent by all map tasks (ms)=5739
    Total vcore-milliseconds taken by all map tasks=5739
    Total megabyte-milliseconds taken by all map tasks=8815104
  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)=100
    CPU time spent (ms)=2600
    Physical memory (bytes) snapshot=272175104
    Virtual memory (bytes) snapshot=3281952768
    Total committed heap usage (bytes)=238026752
  File Input Format Counters
    Bytes Read=0
  File Output Format Counters
    Bytes Written=85081
24/07/16 05:54:31 INFO mapreduce.ImportJobBase: Transferred 83.0869 KB in 24.246 seconds (3.4268 KB/sec)
24/07/16 05:54:31 INFO mapreduce.ImportJobBase: Retrieved 999 records.
[hadoop@ip-172-31-58-228 ~]$
```

- Verify data in “card_member_orc” table.

```
hive> SELECT * FROM CARD_MEMBER_ORC LIMIT 10;
OK
340028465709212 009250698176266 2012-02-08 06:04:13 05/13 United States Barberton
340054675199675 835873341185231 2017-03-10 09:24:44 03/17 United States Fort Dodge
340082915339645 512969555857346 2014-02-15 06:30:30 07/14 United States Graham
340134186926007 887711945571282 2012-02-05 01:21:58 02/13 United States Dix Hills
340265728490548 680324265406190 2014-03-29 07:49:14 11/14 United States Rancho Cucamonga
340268219434811 929799084911715 2012-07-08 02:46:08 08/12 United States San Francisco
340379737226464 089615510858348 2010-03-10 00:06:42 09/10 United States Clinton
340383645652108 181180599313885 2012-02-24 05:32:44 10/16 United States West New York
340803866934451 417664728506297 2015-05-21 04:30:45 08/17 United States Beaverton
340889618969736 459292914761635 2013-04-23 08:40:11 11/15 United States West Palm Beach
Time taken: 0.172 seconds, Fetched: 10 row(s)
hive>
```

- Verify data in "member_score_orc" table.


```
hive> SELECT * FROM MEMBER_SCORE_ORC LIMIT 10;
OK
000037495066290 339
000117826301530 289
001147922084344 393
001314074991813 225
001739553947511 642
003761426295463 413
004494068832701 217
006836124210484 504
006991872634058 697
007955566230397 372
Time taken: 0.131 seconds, Fetched: 10 row(s)
hive>
```

c) **Task 3: Create a look-up table with columns specified earlier in the problem statement.**

- Create “lookup_data_hbase” table (hive-hbase integrated table) which will be visible in HBase (lookup_data_hive).

```
hive> CREATE TABLE 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_transaction_family:postcode, lookup_transaction_family:transaction_dt") TBLPROPERTIES
> ("hbase.table.name" = "lookup_data_hive");
OK
Time taken: 3.717 seconds
hive>
```

- Verify details of **lookup_data_hive** (hive-hbase integrated) table :

```
[root@ip-172-31-58-228 ec2-user]# hbase shell
HBase Shell
Use "help" to get list of supported commands.
Use "exit" to quit this interactive shell.
Version 1.4.13, rUnknown, Fri Apr 17 15:18:24 UTC 2020

hbase(main):001:0> describe 'lookup_data_hive'
Table lookup_data_hive is ENABLED
lookup_data_hive
COLUMN FAMILIES DESCRIPTION
{NAME => 'lookup_card_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'}
{NAME => 'lookup_transaction_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'}
2 row(s) in 0.3360 seconds

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


d) **Task 4: After creating the table, you need to load the relevant data in the lookup table.**

- Load data in “ranked_card_transactions_orc” table

```
hive> INSERT OVERWRITE TABLE RANKED_CARD_TRANSACTIONS_ORC SELECT B.CARD_ID, B.AMOUNT, B.POSTCODE, B.TRANSACTION_DT, B.RANK FROM (SELECT A.CARD_ID, A.AMOUNT, A.POSTCODE, A.TRANSACTION_DT, RANK() OVER (PARTITION BY A.CARD_ID ORDER BY A.TRANSACTION_DT DESC, AMOUNT DESC) AS RANK FROM (SELECT CARD_ID, AMOUNT, POSTCODE, TRANSACTION_DT FROM CARD_TRANSACTIONS_HBASE WHERE STATUS = 'GENUINE')) A ) B WHERE B.RANK <= 10;
Query ID = root_20240716061632_d02d31ac-9339-428d-93cb-ba93e668cfc7
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_1721107016086_0008)
```

VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1	container	SUCCEEDED	1	1	0	0	0	0
Reducer 2	container	SUCCEEDED	2	2	0	0	0	0

```
VERTICES: 02/02 [=====]>>>] 100% ELAPSED TIME: 11.36 s
Loading data to table capstone_project.ranked_card_transactions_orc
OK
Time taken: 25.016 seconds
hive>
```

- Load data in "card_ucl_orc" table

```
hive> INSERT OVERWRITE TABLE CARD_UCL_ORC
> SELECT A.CARD_ID, (A.AVERAGE + (3 * A.STANDARD_DEVIATION)) AS UCL FROM (
> SELECT CARD_ID, AVG(AMOUNT) AS AVERAGE, STDDEV(AMOUNT) AS STANDARD_DEVIATION
> FROM RANKED_CARD_TRANSACTIONS_ORC
> GROUP BY CARD_ID) A;
Query ID = root_20240716061734_bc794501-7d10-4d61-8fb7-30bf9af0af9a
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1721107016086_0008)
```

VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1	container	SUCCEEDED	1	1	0	0	0	0
Reducer 2	container	SUCCEEDED	2	2	0	0	0	0

```
VERTICES: 02/02 [=====]>>>] 100% ELAPSED TIME: 6.55 s
Loading data to table capstone_project.card_ucl_orc
OK
Time taken: 8.562 seconds
hive>
```

- Load data in lookup_data_hbase table.

```
hive> INSERT OVERWRITE TABLE LOOKUP_DATA_HBASE SELECT RCTO.CARD_ID, CUO.UCL, CMS.SCORE, RCTO.POSTCODE, RCTO.TRANSACTION_DT FROM RANKED_CARD_TRANSACTIONS_ORC
RCTO JOIN CARD_UCL_ORC CUO ON CUO.CARD_ID = RCTO.CARD_ID JOIN ( SELECT DISTINCT CARD.CARD_ID, SCORE.SCORE FROM CARD_MEMBER_ORC CARD JOIN MEMBER_SCORE_ORC SCO
RE ON CARD.MEMBER_ID = SCORE.MEMBER_ID) AS CMS ON RCTO.CARD_ID = CMS.CARD_ID WHERE RCTO.RANK = 1;
No Stats for capstone_project@ranked_card_transactions_orc, Columns: postcode, rank, transaction_dt, card_id
No Stats for capstone_project@card_ucl_orc, Columns: card_id, ucl
No Stats for capstone_project@card_member_orc, Columns: member_id, card_id
No Stats for capstone_project@member_score_orc, Columns: member_id, score
Query ID = root_20240716061946_b94fa709-e503-43c5-a074-bf1232413eca
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1721107016086_0008)
```

VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1	container	SUCCEEDED	1	1	0	0	0	0
Map 2	container	SUCCEEDED	1	1	0	0	0	0
Map 3	container	SUCCEEDED	1	1	0	0	0	0
Map 5	container	SUCCEEDED	1	1	0	0	0	0
Reducer 4	container	SUCCEEDED	2	2	0	0	0	0

```
VERTICES: 05/05 [=====]>>>] 100% ELAPSED TIME: 18.01 s
OK
Time taken: 22.188 seconds
hive>
```

- Verify some data in “lookup_data_hbase” table.

```
hive> select * from lookup_data_hbase limit 10;
OK
340028465709212 1.6331555548882348E7 233 24658 2018-01-02 03:25:35
340054675199675 1.4156079786189131E7 631 50140 2018-01-15 19:43:23
340082915339645 1.5285685330791473E7 407 17844 2018-01-26 19:03:47
340134186926007 1.5239767522438556E7 614 67576 2018-01-18 23:12:50
340265728490548 1.608491671255562E7 202 72435 2018-01-21 02:07:35
340268219434811 1.2507323937605347E7 415 62513 2018-01-16 04:30:05
340379737226464 1.4198310998368107E7 229 26656 2018-01-27 00:19:47
340383645652108 1.4091750460468251E7 645 34734 2018-01-29 01:29:12
340803866934451 1.0843341196185412E7 502 87525 2018-01-31 04:23:57
340889618969736 1.3217942365515321E7 330 61341 2018-01-31 21:57:18
Time taken: 0.304 seconds, Fetched: 10 row(s)
hive>
```

- Verify count in “lookup_data_hive” table. Record count is 999 which matches with the given requirement

```
hbase(main):001:0> count 'lookup_data_hive'
999 row(s) in 0.4200 seconds

=> 999
hbase(main):002:0>
```

Total number for record is **999** which is matching with given requirement.

- Verify data in “lookup_data_hive” table. Record count is 999 which matches with the given requirement

```
6595928469079750 column=lookup_card_family:score, timestamp=1721110808662, value=412
6595928469079750 column=lookup_card_family:ucl, timestamp=1721110808662, value=1.142797041440079E7
6595928469079750 column=lookup_transaction_family:postcode, timestamp=1721110808662, value=98349
6595928469079750 column=lookup_transaction_family:transaction_dt, timestamp=1721110808662, value=2018-01-24 12:38:22
6597703848279563 column=lookup_card_family:score, timestamp=1721110808662, value=218
6597703848279563 column=lookup_card_family:ucl, timestamp=1721110808662, value=1.4718634149498457E7
6597703848279563 column=lookup_transaction_family:postcode, timestamp=1721110808662, value=95699
6597703848279563 column=lookup_transaction_family:transaction_dt, timestamp=1721110808662, value=2018-01-27 10:51:49
6598830758632447 column=lookup_card_family:score, timestamp=1721110808662, value=293
6598830758632447 column=lookup_card_family:ucl, timestamp=1721110808662, value=1.2227949982601807E7
6598830758632447 column=lookup_transaction_family:postcode, timestamp=1721110808662, value=19421
6598830758632447 column=lookup_transaction_family:transaction_dt, timestamp=1721110808662, value=2018-01-30 00:18:34
6599900931314251 column=lookup_card_family:score, timestamp=1721110808662, value=297
6599900931314251 column=lookup_card_family:ucl, timestamp=1721110808662, value=1.2121408572464656E7
6599900931314251 column=lookup_transaction_family:postcode, timestamp=1721110808662, value=97423
6599900931314251 column=lookup_transaction_family:transaction_dt, timestamp=1721110808662, value=2018-01-31 11:25:16
999 row(s) in 6.2530 seconds
hbase(main):003:0>
```