

Scripts Execution

Screenshots of the execution of the scripts written

Clone
Terminate
AWS CLI export

Cluster: CCFD_v10
Waiting
Cluster ready to run steps.

Summary
Application user interfaces
Monitoring
Hardware
Configurations
Events
Steps
Bootstrap actions

Summary

ID: j-33XMBN5WLAZ1A
Creation date: 2023-02-14 01:52 (UTC+5:30)
Elapsed time: 35 minutes
After last step completes: Cluster waits
Termination protection: Off [Change](#)
Tags: -- [View All](#) / [Edit](#)
Master public DNS: ec2-52-90-144-110.compute-1.amazonaws.com [Connect to the Master Node Using SSH](#)

Application user interfaces

Persistent user interfaces [🔗](#): Spark history server, YARN timeline server, Tez UI
On-cluster user interfaces [🔗](#): Not Enabled [Enable an SSH Connection](#)

Security and access

Key name: Abhi_win2
EC2 instance profile: EMR_EC2_DefaultRole
EMR role: EMR_DefaultRole
Auto Scaling role: EMR_AutoScaling_DefaultRole
Visible to all users: All [Change](#)
Security groups for Master: [sg-0ebec1af2688a0130](#) [🔗](#) (ElasticMapReduce-master)
Security groups for Core & Task: [sg-00b5e953ddfa55720](#) [🔗](#) (ElasticMapReduce-slave)

Configuration details

Release label: emr-5.36.0
Hadoop distribution: Amazon 2.10.1
Applications: Hive 2.3.9, Hue 4.10.0, Spark 2.4.8, Sqoop 1.4.7, ZooKeeper 3.4.14, HCatalog 2.3.9, HBase 1.4.13
Log URI: --
EMRFS consistent view: Disabled
Custom AMI ID: --
Amazon Linux Release: 2.0.20230119.1 [Learn more](#) [🔗](#)

Network and hardware

Availability zone: us-east-1c
Subnet ID: [subnet-009878956a2ba0e8e](#) [🔗](#)
Master: Running 1 m4.xlarge
Core: --
Task: --
Cluster scaling: Not enabled
Auto-termination: Not enabled

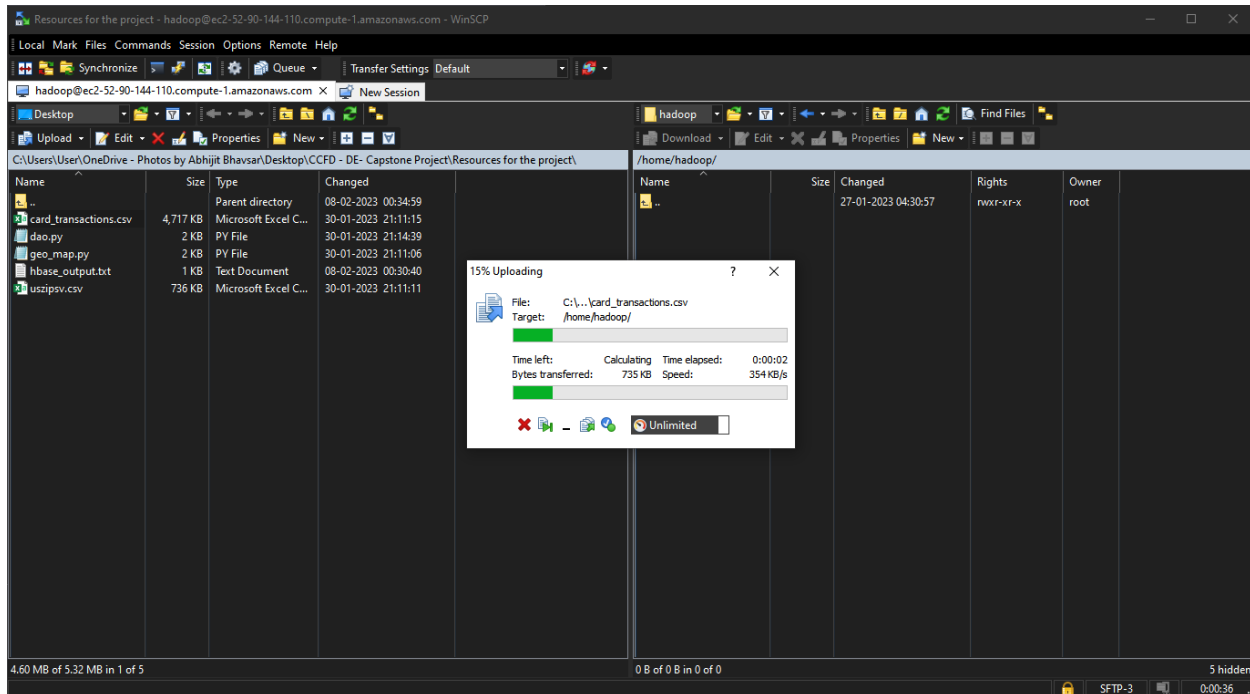
Console Screenshot 1: Created EMR with Hadoop, Hive, HBase, Hcatalog, Spark, Hue and Sqoop.

© Copyright 2020. upGrad Education Pvt. Ltd. All rights reserved

```
hadoop@ip-172-31-26-104:~
[hadoop@ip-172-31-26-104 ~]$ hive

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j2.
properties Async: false
hive> create database ccfd;
OK
Time taken: 1.741 seconds
hive> use ccfd;
OK
Time taken: 0.05 seconds
hive> █
```

Console Screenshot 2: Created the Database in Hive



Console Screenshot 3: Transferring Historical data to Hadoop local ec2 machine.

```
hadoop@ip-172-31-26-104:~  
[hadoop@ip-172-31-26-104 ~]$ hadoop fs -mkdir /user/CCFD_project  
[hadoop@ip-172-31-26-104 ~]$  
[hadoop@ip-172-31-26-104 ~]$ hadoop fs -put /home/hadoop/card_transactions.csv /user/CCFD_project/card_transactions.csv  
[hadoop@ip-172-31-26-104 ~]$
```

Console Screenshot 4: Moving data from EC2 machine to HDFS.

```
hadoop@ip-172-31-26-104:~  
[hadoop@ip-172-31-26-104 ~]$ hive  
Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j2.properties Async: false  
hive> use ccfcd;  
OK  
Time taken: 1.018 seconds  
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 '/user/CCFD_Project/card_transactions.csv'  
  > TBLPROPERTIES ("skip.header.line.count"="1");  
OK  
Time taken: 0.567 seconds  
hive>
```

Console Screenshot 5: Creating historical data External table.

```
hadoop@ip-172-31-26-104:~
> `POSTCODE` STRING,
> `POS_ID` STRING,
> `TRANSACTION_DT` STRING,
> `STATUS` STRING)
> ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
> LOCATION '/user/CCFD_Project/card_transactions.csv'
> TBLPROPERTIES ("skip.header.line.count"="1");
OK
Time taken: 0.567 seconds
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;
OK
Time taken: 0.875 seconds
hive>
```

Console Screenshot 6: Creating ORC table from external table.

```
Time taken: 0.833 seconds
hive> LOAD DATA INPATH '/user/CCFD_project/card_transactions.csv' INTO TABLE card_transactions_ext;
Loading data to table ccfd.card_transactions_ext
OK
Time taken: 1.066 seconds
hive>
```

Console Screenshot 7: loading data into external table.

```
hadoop@ip-172-31-26-104:~
OK
Time taken: 1.066 seconds
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 = hadoop_20230213212431_cfce1ea7-f6ce-49b7-85e4-a06e291d55a6
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1676320060972_0003)

-----
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: 7.49 s
-----
Loading data to table ccfd.card_transactions_orc
OK
Time taken: 12.597 seconds
hive>
```

Console Screenshot 8: Inserting data into orc table.

```
hadoop@ip-172-31-26-104:~
-----
Map 1 ..... container      SUCCEEDED      1      1      0      0      0      0
-----
VERTICES: 01/01  [=====>>] 100%  ELAPSED TIME: 7.49 s
-----
Loading data to table ccfd.card_transactions_orc
OK
Time taken: 12.597 seconds
hive> select count(*) from card_transactions_ext;
Query ID = hadoop_20230213212529_02b24911-f5b5-4152-a3e9-d0b56902b5cc
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1676320060972_0003)

Map 1: 0/1      Reducer 2: 0/1
Map 1: 0(+1)/1  Reducer 2: 0/1
Map 1: 1/1      Reducer 2: 0(+1)/1
Map 1: 1/1      Reducer 2: 1/1
OK
53292
Time taken: 6.029 seconds, Fetched: 1 row(s)
hive>
```

Console Screenshot 9: Validating external table records.

```
Time taken: 0.371 seconds, Fetched: 5 row(s)
hive> SELECT * FROM CARD_TRANSACTIONS_ORC LIMIT 5;
OK
348702330256514 000037495066290 9084849.0      33946      614677375609919 2018-02-11 00:00:00      GENUINE
348702330256514 000037495066290 330148.0      33946      614677375609919 2018-02-11 00:00:00      GENUINE
348702330256514 000037495066290 136052.0      33946      614677375609919 2018-02-11 00:00:00      GENUINE
348702330256514 000037495066290 4310362.0     33946      614677375609919 2018-02-11 00:00:00      GENUINE
348702330256514 000037495066290 9097094.0     33946      614677375609919 2018-02-11 00:00:00      GENUINE
Time taken: 2.017 seconds, Fetched: 5 row(s)
hive>
```

Console Screenshot 10: Validating external table records.

```
hadoop@ip-172-31-26-104:~
Map 1: 0/1      Reducer 2: 0/1
Map 1: 0(+1)/1 Reducer 2: 0/1
Map 1: 1/1      Reducer 2: 0(+1)/1
Map 1: 1/1      Reducer 2: 1/1
OK
53292
Time taken: 6.029 seconds, Fetched: 1 row(s)
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: 0.176 seconds, Fetched: 10 row(s)
hive>
```

Console Screenshot 11: Validating timestamp column in orc table.

```
hadoop@ip-172-31-26-104:~
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: 0.176 seconds, Fetched: 10 row(s)
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_fam
ily:member_id, card_transactions_family:amount, card_transactions_family:postcode, card_trans
actions_family:pos_id, card_transactions_family:transaction_dt, card_transactions_family:stat
us")
> TBLPROPERTIES ("hbase.table.name"="card_transactions_hive");
OK
Time taken: 2.314 seconds
hive>
```

Console Screenshot 12: Creating card_transactions_hbase hive-hbase integrated table.

```
hadoop@ip-172-31-26-104:~
OK
Time taken: 2.314 seconds
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 = hadoop_20230213212714_15552874-efa7-4b8d-93bf-58ba88359f34
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1676320060972_0003)

Map 1: 0/1
Map 1: 0/1
Map 1: 0(+1)/1
Map 1: 0(+1)/1
Map 1: 0(+1)/1
Map 1: 1/1
OK
Time taken: 13.476 seconds
hive>
```

Console Screenshot 13: Inserting data into from orc table.

```
hadoop@ip-172-31-26-104:~
Time taken: 13.476 seconds
hive> select * from card_transactions_hbase limit 10;
OK
00010243-4b22-4584-98ff-1c7ea98c79a8      6227071613331391      774599835915613 6377902.0    9
6768      877803162288642 2017-06-19 08:52:02      GENUINE
00021a29-1a78-4592-b617-3987ba71a2e1      370628435959605 100655658028368 7994608.0    720380
59157340802199 2018-01-11 08:37:55      GENUINE
00031e06-89c3-43b3-9775-89be9d44016b      6011989509446330      015509173543086 8341899.0    4
9948      702002414547753 2016-06-28 13:38:13      GENUINE
00034899-22bf-4b71-99b0-f9dc6a306f4f      5596710122356317      735012662601476 6027804.0    9
9634      903017004769385 2018-01-11 00:00:00      GENUINE
0006a019-379b-440a-b1d3-7cf2c6bec486      4586353565904288      947445762190536 2461658.0    4
3008      304144829675370 2018-01-11 08:37:55      GENUINE
00073ce6-d03b-4287-bb09-1d2d553c8980      379321864695232 082567374418739 904794.0     157675
58931881914719 2017-07-30 09:45:27      GENUINE
0007d0db-aeac-44da-992e-8bb813425638      342136378319429 756971727303005 332062.0     546572
28416461928688 2017-12-10 00:39:27      GENUINE
00083fdd-1311-45f2-9f28-70951b67c6b0      348684315090900 293516519272933 4431821.0     583218
62123266998348 2017-09-26 06:26:09      GENUINE
0009ab94-30ca-4abb-bb00-dbe9500092a3      5347795209930318      759563212896397 3912589.0    1
2051      450743412591954 2017-12-26 20:26:42      GENUINE
0009b3c6-4944-4480-a3e0-9e0781816a05      4173451524840806      919377025657682 2426689.0    5
7001      627975053852882 2017-09-23 18:44:44      GENUINE
Time taken: 0.192 seconds, Fetched: 10 row(s)
hive>
```

Console Screenshot 14: Validating HBase table data.

```
Time taken: 0.192 seconds, Fetched: 10 row(s)
hive> CREATE TABLE LOOKUP_DATA_HBASE(`CARD_ID` STRING,`UCL` DOUBLE, `SCORE` INT, `POSTCODE` STRI
NG, `TRANSACTION_DT` TIMESTAMP)
  > STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
  > WITH SERDEPROPERTIES ("hbase.columns.mapping"=":key, lookup_card_family:ucl, lookup_card_f
amily:score, lookup_transaction_family:postcode, lookup_transaction_family:transaction_dt")
  > TBLPROPERTIES ("hbase.table.name" = "lookup_data_hive");
OK
Time taken: 2.363 seconds
hive>
```

Console Screenshot 15: Creating Lookup Table

```
hive> DESCRIBE lookup_data_hbase;
OK
card_id          string
ucl              double
score            int
postcode         string
transaction_dt   timestamp
Time taken: 0.371 seconds, Fetched: 5 row(s)
hive>
```

Console Screenshot 16: Validating lookup 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 => 'F
ALSE', DATA_BLOCK_ENCODING => 'NONE', TTL => 'FOREVER', COMPRESSION => 'NONE', MIN_VERSIONS => '0', BLOCKCACHE => 'true',
BLOCKSIZE => '65536', REPLICATION_SCOPE => '0'}
1 row(s) in 0.3570 seconds
hbase(main):002:0>
```

Console Screenshot 17: Checking details of card_transactions_hive hive-HBase integrated table


```
hadoop@ip-172-31-26-104:~
hbase(main):002:0> count 'card_transactions_hive'
Current count: 1000, row: 04d77926-52a5-4780-9626-975037fc675f
Current count: 2000, row: 09a98eb7-6765-4388-84c2-188c432d00cd
Current count: 3000, row: 0e9b2a73-c536-4636-a26f-e041eb0a750a
Current count: 4000, row: 1384037c-7476-41a4-8add-97bf6f0f12d3
Current count: 5000, row: 189f4181-834a-41e0-b0be-5a24c2ecf4d4
Current count: 6000, row: 1d7050b2-d6af-423b-961e-b7b9db9973c3
Current count: 7000, row: 2259a721-0b59-412f-b4a6-37b37e334d70
Current count: 8000, row: 272fbc19-125e-494c-85e8-cc7cca5b8942
Current count: 9000, row: 2c5b0d52-c6ac-4eac-a9a0-1d8db30f44cb
Current count: 10000, row: 3158eabb-170b-4fef-b108-35189e3db99a
Current count: 11000, row: 3651744b-90e4-4c3d-8c91-fb536141df03
Current count: 12000, row: 3b07e99c-ee57-4837-971c-745f4082143a
Current count: 13000, row: 3fcb522a-8377-4a66-8dac-10c911c45cc8
Current count: 14000, row: 44b7d9e4-c018-499b-8213-eae818ec072c
Current count: 15000, row: 498d63d8-1c35-486c-851e-d6b031b2606a
Current count: 16000, row: 4e8cee31-ec81-475b-bd0a-36c601597127
Current count: 17000, row: 534486da-adaf-4841-8ce9-0d2151024ecb
Current count: 18000, row: 57d3bf2d-971d-48b6-8c97-4ed51da9fb46
Current count: 19000, row: 5c54041d-089c-4183-9ed5-879e3a09776d
Current count: 20000, row: 60e828db-1c03-4ade-a498-489e2dc66b56
Current count: 21000, row: 659cc46d-0009-4fc9-9891-450a9600ca7b
Current count: 22000, row: 6a565d1b-e0dd-4f21-bb52-73415ee2a612
Current count: 23000, row: 6f15f6dc-49be-4863-9f96-9b44c2d9dbe7
Current count: 24000, row: 7409e355-49db-41eb-9f41-e42a394612f0
Current count: 25000, row: 78907982-d843-4b97-b183-9d2f8afa4053
Current count: 26000, row: 7d8efeb1-d722-4ba9-88e3-88c8673d6571
Current count: 27000, row: 8241c73f-daa2-428b-925d-83c8a47a5dab
Current count: 28000, row: 870d563a-bfd1-42d7-830f-1a50a151a262
Current count: 29000, row: 8bf916fa-652e-4f2d-8b78-c49d22ed847c
Current count: 30000, row: 90c39cca-7ba3-4e28-a1bf-8f710270f385
Current count: 31000, row: 95841e31-e169-4335-bff7-50a474dfdec9
Current count: 32000, row: 9a2f469a-1b08-4f7c-a80e-45c398983344
Current count: 33000, row: 9ee72238-0885-4a1b-868f-8a32a8da0f08
Current count: 34000, row: a36b6d90-0c00-4d38-b8c0-ed68b6f0415d
Current count: 35000, row: a8850ed0-ed9a-4e04-b907-48e03be36677
Current count: 36000, row: ad239cc8-f865-4c8c-b9ea-cf596dbe80f3
Current count: 37000, row: b1f74bea-5ac2-44ba-a821-e6752d25fa1d
Current count: 38000, row: b6a8827d-8151-43d6-86b9-a8bd8412b00f
Current count: 39000, row: bb57a00f-a5a0-4c25-ab05-d40c27dc69cc
Current count: 40000, row: c06b77f9-7b34-40ef-ac76-f558c98a2647
Current count: 41000, row: c5439389-d7da-4f93-b8d9-52893d6f48c4
Current count: 42000, row: c9f3067f-45c3-4e67-8e80-9c8610c823e5
Current count: 43000, row: ceac24b6-f9cb-4795-b338-3328a13a8852
Current count: 44000, row: d3bcb785-f8d6-4404-b68d-8ebc01ddd8a5
Current count: 45000, row: d86f048d-fa9c-4b3e-9470-83571bd21b10
Current count: 46000, row: dd2f3709-79fb-4668-9f8f-e22ebf973499
Current count: 47000, row: e1c87c64-3425-492b-8810-3a394337f6d8
Current count: 48000, row: e6a90bf2-2310-4848-9540-884070b64290
Current count: 49000, row: eb64b7c8-7ac6-4300-88e8-684cafbd48e0
Current count: 50000, row: f0158a52-defa-40a0-af6e-1a2588e8aee8
Current count: 51000, row: f4e850a9-9ccd-4280-90ad-72662756c522
Current count: 52000, row: f9824d77-8879-4af3-b18b-bb89f3859dbd
Current count: 53000, row: fe67799b-54f3-4413-91c7-68c4f14d5c60
53292 row(s) in 4.0110 seconds
```

Console Screenshot 18: Checking the count of the table.

```
hadoop@ip-172-31-26-104:~
Current count: 37000, row: b1f74bea-5ac2-44ba-a821-e6752d25fa1d
Current count: 38000, row: b6a8827d-8151-43d6-86b9-a8bd8412b00f
Current count: 39000, row: bb57a00f-a5a0-4c25-ab05-d40c27dc69cc
Current count: 40000, row: c06b77f9-7b34-40ef-ac76-f558c98a2647
Current count: 41000, row: c5439389-d7da-4f93-b8d9-52893d6f48c4
Current count: 42000, row: c9f3067f-45c3-4e67-8e80-9c8610c823e5
Current count: 43000, row: ceac24b6-f9cb-4795-b338-3328a13a8852
Current count: 44000, row: d3bcb785-f8d6-4404-b68d-8ebc01ddd8a5
Current count: 45000, row: d86f048d-fa9c-4b3e-9470-83571bd21b10
Current count: 46000, row: dd2f3709-79fb-4668-9f8f-e22ebf973499
Current count: 47000, row: e1c87c64-3425-492b-8810-3a394337f6d8
Current count: 48000, row: e6a90bf2-2310-4848-9540-884070b64290
Current count: 49000, row: eb64b7c8-7ac6-4300-88e8-684c4bd48e0
Current count: 50000, row: f0158a52-defa-40a0-af6e-1a2588e8aee8
Current count: 51000, row: f4e850a9-9ccd-4280-90ad-72662756c522
Current count: 52000, row: f9824d77-8879-4af3-b18b-bb89f3859dbd
Current count: 53000, row: fe67799b-54f3-4413-91c7-68c4f14d5c60
53292 row(s) in 4.0110 seconds

=> 53292
hbase(main):003:0> 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 0.0330 seconds

hbase(main):004:0>
```

Console Screenshot 19: Checking the details of lookup_data_hive hive-HBase integrated table.

```
hadoop@ip-172-31-26-104:~
=> 53292
hbase(main):003:0> 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 0.0330 seconds

hbase(main):004:0> alter 'lookup_data_hive', {NAME => 'lookup_transaction_family', VERSIONS => 10}
Updating all regions with the new schema...
1/1 regions updated.
Done.
0 row(s) in 1.9720 seconds

hbase(main):005:0> 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 => '10', 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 0.0200 seconds

hbase(main):006:0>
```

Console Screenshot 20: Altering the lookup_data_hive table and set VERSIONS to 10 for lookup_transaction_family & confirming with describe command.

```
hadoop@ip-172-31-26-104:~/mysql-connector-java-8.0.25
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/CollectionTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/CompressionTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/DevApiBaseTestCase.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/Ipv6SupportTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/MetadadataTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/ResultTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/SchemaTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/SecureSessionTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/SessionFailoverTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/SessionTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/TableDeleteTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/TableInsertTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/TableSelectTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/TableTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/TableUpdateTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/TransactionTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/package-info.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/internal/InternalXBaseTestCase.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/internal/MysqlxSessionTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/internal/XProtocolAsyncTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/internal/XProtocolAuthTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/internal/XProtocolTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/internal/package-info.java
[hadoop@ip-172-31-26-104 ~]$
[hadoop@ip-172-31-26-104 ~]$ cd mysql-connector-java-8.0.25/
[hadoop@ip-172-31-26-104 mysql-connector-java-8.0.25]$
[hadoop@ip-172-31-26-104 mysql-connector-java-8.0.25]$ sudo cp mysql-connector-java-8.0.25.jar /usr/lib/sqoop/lib/
[hadoop@ip-172-31-26-104 mysql-connector-java-8.0.25]$
```

Console Screenshot 21: For Scoop Import, we will install MySQL connector before starting with Apache Sqoop.

```
hadoop@ip-172-31-26-104:~/mysql-connector-java-8.0.25
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/TransactionTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/devapi/package-info.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/internal/InternalXBaseTestCase.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/internal/MysqlxSessionTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/internal/XProtocolAsyncTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/internal/XProtocolAuthTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/internal/XProtocolTest.java
mysql-connector-java-8.0.25/src/test/java/testsuite/x/internal/package-info.java
[hadoop@ip-172-31-26-104 ~]$
[hadoop@ip-172-31-26-104 ~]$ cd mysql-connector-java-8.0.25/
[hadoop@ip-172-31-26-104 mysql-connector-java-8.0.25]$
[hadoop@ip-172-31-26-104 mysql-connector-java-8.0.25]$ sudo cp mysql-connector-java-8.0.25.jar /usr/lib/sqoop/lib/
[hadoop@ip-172-31-26-104 mysql-connector-java-8.0.25]$ sqoop import \
> --connect jdbc:mysql://upgradawsrds1.cyaie1c9bmnf.us-east-1.rds.amazonaws.com/cred_financials_data \
> --username upgraduser \
> --password upgraduser \
> --table card_member \
> --target-dir /user/CCFD_project/card_member \
> -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.
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/lib/hadoop/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/lib/hive/lib/log4j-slf4j-impl-2.17.1.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]
23/02/13 21:42:00 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
23/02/13 21:42:00 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
```

Console Screenshot 22: Scoop import for card member table.

```
hadoop@ip-172-31-26-104:~/mysql-connector-java-8.0.25
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)=198240
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=4130
  Total vcore-milliseconds taken by all map tasks=4130
  Total megabyte-milliseconds taken by all map tasks=6343680
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)=91
  CPU time spent (ms)=2190
  Physical memory (bytes) snapshot=326868992
  Virtual memory (bytes) snapshot=3313852416
  Total committed heap usage (bytes)=233832448
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=85081
23/02/13 21:42:23 INFO mapreduce.ImportJobBase: Transferred 83.0869 KB in 18.2504 seconds (4.5526 KB/sec)
23/02/13 21:42:23 INFO mapreduce.ImportJobBase: Retrieved 999 records.
[hadoop@ip-172-31-26-104 mysql-connector-java-8.0.25]$
```

Console Screenshot 23: 999 records imported.

```
hadoop@ip-172-31-26-104:~/mysql-connector-java-8.0.25
Input split bytes=87
Spilled Records=0
Failed Shuffles=0
Merged Map outputs=0
GC time elapsed (ms)=91
CPU time spent (ms)=2190
Physical memory (bytes) snapshot=326868992
Virtual memory (bytes) snapshot=3313852416
Total committed heap usage (bytes)=233832448
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=85081
23/02/13 21:42:23 INFO mapreduce.ImportJobBase: Transferred 83.0869 KB in 18.2504 seconds (4.5526 KB/sec)
23/02/13 21:42:23 INFO mapreduce.ImportJobBase: Retrieved 999 records.
[hadoop@ip-172-31-26-104 mysql-connector-java-8.0.25]$ sqoop import \
> --connect jdbc:mysql://upgradawsrds1.cyaie1c9bmfn.us-east-1.rds.amazonaws.com/cred_financials_data \
> --username upgraduser \
> --password upgraduser \
> --table member_score \
> --target-dir /user/CCFD_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.
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/lib/hadoop/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/lib/hive/lib/log4j-slf4j-impl-2.17.1.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
```

Console Screenshot 24: Scoop import for member score table.

```
hadoop@ip-172-31-26-104:~/mysql-connector-java-8.0.25
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)=197328
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=4111
  Total vcore-milliseconds taken by all map tasks=4111
  Total megabyte-milliseconds taken by all map tasks=6314496
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)=91
  CPU time spent (ms)=1590
  Physical memory (bytes) snapshot=335671296
  Virtual memory (bytes) snapshot=3305467904
  Total committed heap usage (bytes)=326631424
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=19980
23/02/13 21:44:06 INFO mapreduce.ImportJobBase: Transferred 19.5117 KB in 20.2668 seconds (985.8498 bytes/sec)
23/02/13 21:44:06 INFO mapreduce.ImportJobBase: Retrieved 999 records.
[hadoop@ip-172-31-26-104 mysql-connector-java-8.0.25]$
```

Console Screenshot 25: 999 records imported.

```
hadoop@ip-172-31-26-104:~/mysql-connector-java-8.0.25
[hadoop@ip-172-31-26-104 mysql-connector-java-8.0.25]$ hadoop fs -ls /user/CCFD_project/card_member
Found 2 items
-rw-r--r-- 1 hadoop hdfsadmingroup 0 2023-02-13 21:42 /user/CCFD_project/card_member/_SUCCESS
-rw-r--r-- 1 hadoop hdfsadmingroup 85081 2023-02-13 21:42 /user/CCFD_project/card_member/part-m-00000
[hadoop@ip-172-31-26-104 mysql-connector-java-8.0.25]$ hadoop fs -ls /user/CCFD_project/member_score
Found 2 items
-rw-r--r-- 1 hadoop hdfsadmingroup 0 2023-02-13 21:44 /user/CCFD_project/member_score/_SUCCESS
-rw-r--r-- 1 hadoop hdfsadmingroup 19980 2023-02-13 21:44 /user/CCFD_project/member_score/part-m-00000
[hadoop@ip-172-31-26-104 mysql-connector-java-8.0.25]$
```

Console Screenshot 26: Checking the scoop imported files.

```
hadoop@ip-172-31-26-104:~/mysql-connector-java-8.0.25
-rw-r--r-- 1 hadoop hdfsadmingroup 0 2023-02-13 21:42 /user/CCFD_project/card_member/_SUCCESS
-rw-r--r-- 1 hadoop hdfsadmingroup 85081 2023-02-13 21:42 /user/CCFD_project/card_member/part-m-00000
[hadoop@ip-172-31-26-104 mysql-connector-java-8.0.25]$ hadoop fs -ls /user/CCFD_project/member_score
Found 2 items
-rw-r--r-- 1 hadoop hdfsadmingroup 0 2023-02-13 21:44 /user/CCFD_project/member_score/_SUCCESS
-rw-r--r-- 1 hadoop hdfsadmingroup 19980 2023-02-13 21:44 /user/CCFD_project/member_score/part-m-00000
[hadoop@ip-172-31-26-104 mysql-connector-java-8.0.25]$ hive

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j2.properties Async: false
hive> CREATE EXTERNAL TABLE IF NOT EXISTS CARD_MEMBER_EXT(
  > `CARD_ID` STRING,
  > `MEMBER_ID` STRING,
  > `MEMBER_JOINING_DT` TIMESTAMP,
  > `CARD_PURCHASE_DT` STRING,
  > `COUNTRY` STRING,
  > `CITY` STRING)
  > ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
  > LOCATION '/user/CCFD_project/card_member';
OK
Time taken: 1.281 seconds
hive> CREATE EXTERNAL TABLE IF NOT EXISTS MEMBER_SCORE_EXT(
  > `MEMBER_ID` STRING,
  > `SCORE` INT)
  > ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
  > LOCATION '/user/CCFD_project/member_score';
OK
Time taken: 0.21 seconds
hive>
```

Console Screenshot 27: Creating external tables for RDS data.

```
OK
Time taken: 0.21 seconds
hive> CREATE TABLE IF NOT EXISTS CARD_MEMBER_ORC(
  > `CARD_ID` STRING,
  > `MEMBER_ID` STRING,
  > `MEMBER_JOINING_DT` TIMESTAMP,
  > `CARD_PURCHASE_DT` STRING,
  > `COUNTRY` STRING,
  > `CITY` STRING)
  > STORED AS ORC
  > TBLPROPERTIES ("orc.compress"="SNAPPY");
OK
Time taken: 0.552 seconds
hive> CREATE TABLE IF NOT EXISTS MEMBER_SCORE_ORC(
  > `MEMBER_ID` STRING,
  > `SCORE` INT)
  > STORED AS ORC
  > TBLPROPERTIES ("orc.compress"="SNAPPY");
OK
Time taken: 0.112 seconds
```

Console Screenshot 28: Creating ORC tables from the external tables.

```
hadoop@ip-172-31-26-104:~/mysql-connector-java-8.0.25
OK
Time taken: 0.112 seconds
hive> INSERT OVERWRITE TABLE CARD_MEMBER_ORC
> SELECT CARD_ID, MEMBER_ID, MEMBER_JOINING_DT, CARD_PURCHASE_DT, COUNTRY, CITY FROM CARD_MEMBER_EXT;
Query ID = hadoop_20230213215911_7bae224e-156b-4693-bada-00ebb325f74a
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1676320060972_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: 4.87 s
-----
Loading data to table default.card_member_orc
OK
Time taken: 8.738 seconds
hive> INSERT OVERWRITE TABLE MEMBER_SCORE_ORC
> SELECT MEMBER_ID, SCORE FROM MEMBER_SCORE_EXT;
Query ID = hadoop_20230213220007_d322f16e-563d-42f7-a2c6-a51527f19af6
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1676320060972_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: 4.26 s
-----
Loading data to table default.member_score_orc
OK
Time taken: 5.266 seconds
hive> █
```

Console Screenshot 29: Inserting data into ORC tables from external tables.

```
hadoop@ip-172-31-26-104:~  
Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j2.  
properties Async: false  
hive> use ccfd;  
OK  
Time taken: 0.745 seconds  
hive> CREATE TABLE IF NOT EXISTS RANKED_CARD_TRANSACTIONS_ORC(  
  > `CARD_ID` STRING,  
  > `AMOUNT` DOUBLE,  
  > `POSTCODE` STRING,  
  > `TRANSACTION_DT` TIMESTAMP,  
  > `RANK` INT)  
  > STORED AS ORC  
  > TBLPROPERTIES ("orc.compress"="SNAPPY");  
OK  
Time taken: 0.855 seconds  
hive> CREATE TABLE IF NOT EXISTS CARD_UCL_ORC(  
  > `CARD_ID` STRING,  
  > `UCL` DOUBLE)  
  > STORED AS ORC  
  > TBLPROPERTIES ("orc.compress"="SNAPPY");  
OK  
Time taken: 0.125 seconds  
hive> █
```

Console Screenshot 30: Creating Ranking ORC table of last 10 customer transactions & Creating Card UCL ORC table for each card.


```
hadoop@ip-172-31-26-104:~
Time taken: 0.125 seconds
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(PAR
TITION 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_H
BASE WHERE
> STATUS = 'GENUINE') A ) B WHERE B.RANK <= 10;
Query ID = hadoop_20230213222121_6f46458d-98a8-4113-a9dd-ca7085935cfc
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_1676320060972_0008)

Map 1: -/-      Reducer 2: 0/2
Map 1: 0/1      Reducer 2: 0/2
Map 1: 0/1      Reducer 2: 0/2
Map 1: 0(+1)/1  Reducer 2: 0/2
Map 1: 0(+1)/1  Reducer 2: 0/2
Map 1: 1/1      Reducer 2: 0/2
Map 1: 1/1      Reducer 2: 0(+1)/2
Map 1: 1/1      Reducer 2: 1(+1)/2
Map 1: 1/1      Reducer 2: 2/2
Loading data to table ccfd.ranked_card_transactions_orc
OK
Time taken: 27.113 seconds
hive>
```

Console Screenshot 31: Inserting data into ranked_card_transactions_orc table with conditions.

```
hadoop@ip-172-31-26-104:~
Map 1: 1/1      Reducer 2: 0/2
Map 1: 1/1      Reducer 2: 0(+1)/2
Map 1: 1/1      Reducer 2: 1(+1)/2
Map 1: 1/1      Reducer 2: 2/2
Loading data to table ccfd.ranked_card_transactions_orc
OK
Time taken: 27.113 seconds
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 = hadoop_20230213222251_3d3e74e2-9d4a-4f02-acde-2f5c8783ee69
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1676320060972_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: 10.67 s
-----
Loading data to table ccfd.card_ucl_orc
OK
Time taken: 11.659 seconds
hive>
```

Console Screenshot 32: Inserting data into card_ucl_orc table with conditions.

```

hadoop@ip-172-31-26-104:~
Time taken: 8.278 seconds
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 SCORE
> ON CARD.MEMBER_ID = SCORE.MEMBER_ID) AS CMS
> ON RCTO.CARD_ID = CMS.CARD_ID
> WHERE RCTO.RANK = 1;
No Stats for ccfd@ranked_card_transactions_orc, Columns: postcode, rank, transaction_dt, card_id
No Stats for ccfd@card_ucl_orc, Columns: card_id, ucl
No Stats for ccfd@card_member_orc, Columns: member_id, card_id
No Stats for ccfd@member_score_orc, Columns: member_id, score
Query ID = hadoop_20230213223039_af64a4db-ea8c-4511-ab38-4044e72b90df
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1676320060972_0009)

-----
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: 14.04 s
-----
OK
Time taken: 21.429 seconds
hive> █

```

Console Screenshot 33: Loading data in lookup_data_hbase table.

```
hadoop@ip-172-31-26-104:~
-----
VERTICES: 05/05 [=====>>] 100% ELAPSED TIME: 14.04 s
-----
OK
Time taken: 21.429 seconds
hive> SELECT COUNT(*) FROM LOOKUP_DATA_HBASE;
Query ID = hadoop_20230213223337_191850c0-0c80-4223-9323-2bc4aa6d5e46
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1676320060972_0009)

-----
VERTICES      MODE           STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container    SUCCEEDED      1          1          0          0          0          0
Reducer 2 ..... container  SUCCEEDED      1          1          0          0          0          0
-----
VERTICES: 02/02 [=====>>] 100% ELAPSED TIME: 6.39 s
-----
OK
999
Time taken: 9.518 seconds, Fetched: 1 row(s)
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: 0.305 seconds, Fetched: 10 row(s)
hive>
```

Console Screenshot 34: Verifying lookup table data.

```
at org.jruby.Main.main(Main.java:188)
log4j:ERROR Either File or DatePattern options are not set f
HBase Shell
Use "help" to get list of supported commands.
Use "exit" to quit this interactive shell.
Version 1.4.13, rUnknown, Wed Jun  8 00:30:30 UTC 2022

hbase(main):001:0> count 'lookup_data_hive'
999 row(s) in 0.4720 seconds

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

Console Screenshot 35: Checking count in lookup_data_hive table.

```
hadoop@ip-172-31-26-104:~
079E7
6595928469079750 column=lookup_transaction_family:postcode, timestamp=1676327460635, value=98349
6595928469079750 column=lookup_transaction_family:transaction_dt, timestamp=1676327460635, value=2018-01-24 12:38:22
6597703848279563 column=lookup_card_family:score, timestamp=1676327460635, value=218
6597703848279563 column=lookup_card_family:ucl, timestamp=1676327460635, value=1.4718634149498457E7
6597703848279563 column=lookup_transaction_family:postcode, timestamp=1676327460635, value=95699
6597703848279563 column=lookup_transaction_family:transaction_dt, timestamp=1676327460635, value=2018-01-27 10:51:49
6598830758632447 column=lookup_card_family:score, timestamp=1676327460635, value=293
6598830758632447 column=lookup_card_family:ucl, timestamp=1676327460635, value=1.2227949982601807E7
6598830758632447 column=lookup_transaction_family:postcode, timestamp=1676327460635, value=19421
6598830758632447 column=lookup_transaction_family:transaction_dt, timestamp=1676327460635, value=2018-01-30 00:18:34
6599900931314251 column=lookup_card_family:score, timestamp=1676327460635, value=297
6599900931314251 column=lookup_card_family:ucl, timestamp=1676327460635, value=1.2121408572464656E7
6599900931314251 column=lookup_transaction_family:postcode, timestamp=1676327460635, value=97423
6599900931314251 column=lookup_transaction_family:transaction_dt, timestamp=1676327460635, value=2018-01-31 11:25:16
999 row(s) in 2.7220 seconds
hbase(main):003:0>
```

Console Screenshot 36: Checking data in lookup_data_hive table.