

Loading the Lookup Table

Commands to load the relevant data in the Lookup Table

```
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;
```

Command to see the table created and it's content

```
SELECT COUNT(*) FROM LOOKUP_DATA_HBASE;
```

```
COUNT 'LOOKUP_DATA_HIVE'
```

```
SCAN 'LOOKUP_DATA_HIVE'
```

Screenshot of the created 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 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_20250920103454_525a5786-991c-4d6a-9d20-4d2402257ca9
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1758361848317_0004)

-----
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: 8.57 s
-----
```

Console Screenshot 1: Loading data into Lookup Table

```
hive> SELECT COUNT(*) FROM LOOKUP_DATA_HBASE;SELECT COUNT(*) FROM LOOKUP_DATA_HBASE;
Query ID = hadoop_20250920103626_9998c63c-fbc6-4dd8-9734-7abd83bea117
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1758361848317_0004)
```

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: 6.14 s

```
OK
999
Time taken: 15.01 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: 0.13 seconds, Fetched: 10 row(s)
hive>
```

Console Screenshot 2: Verifying the data

```
hbase(main):004:0> count 'lookup_data_hive'
999 row(s) in 0.1240 seconds

=> 999
hbase(main):005:0>
```

Console Screenshot 3: Verifying the data

```
6595638658736751 column=lookup_card_family:score, timestamp=1758364511884, value=310
6595638658736751 column=lookup_card_family:ucl, timestamp=1758364511884, value=1.356629177577566E7
6595638658736751 column=lookup_transaction_family:postcode, timestamp=1758364511884, value=68328
6595638658736751 column=lookup_transaction_family:transaction_dt, timestamp=1758364511884, value=2018-01-30 10:50:34

6595814135833988 column=lookup_card_family:score, timestamp=1758364511884, value=210
6595814135833988 column=lookup_card_family:ucl, timestamp=1758364511884, value=1.3926273240525039E7
6595814135833988 column=lookup_transaction_family:postcode, timestamp=1758364511884, value=22508
6595814135833988 column=lookup_transaction_family:transaction_dt, timestamp=1758364511884, value=2018-01-30 02:03:54

6595928469079750 column=lookup_card_family:score, timestamp=1758364511884, value=412
6595928469079750 column=lookup_card_family:ucl, timestamp=1758364511884, value=1.142797041440079E7
6595928469079750 column=lookup_transaction_family:postcode, timestamp=1758364511884, value=98349
6595928469079750 column=lookup_transaction_family:transaction_dt, timestamp=1758364511884, value=2018-01-24 12:38:22

6597703848279563 column=lookup_card_family:score, timestamp=1758364511884, value=218
6597703848279563 column=lookup_card_family:ucl, timestamp=1758364511884, value=1.4718634149498457E7
6597703848279563 column=lookup_transaction_family:postcode, timestamp=1758364511884, value=95699
6597703848279563 column=lookup_transaction_family:transaction_dt, timestamp=1758364511884, value=2018-01-27 10:51:49

6598830758632447 column=lookup_card_family:score, timestamp=1758364511884, value=293
6598830758632447 column=lookup_card_family:ucl, timestamp=1758364511884, value=1.2227949982601807E7
6598830758632447 column=lookup_transaction_family:postcode, timestamp=1758364511884, value=19421
6598830758632447 column=lookup_transaction_family:transaction_dt, timestamp=1758364511884, value=2018-01-30 00:18:34

6599900931314251 column=lookup_card_family:score, timestamp=1758364511884, value=297
6599900931314251 column=lookup_card_family:ucl, timestamp=1758364511884, value=1.2121408572464656E7
6599900931314251 column=lookup_transaction_family:postcode, timestamp=1758364511884, value=97423
6599900931314251 column=lookup_transaction_family:transaction_dt, timestamp=1758364511884, value=2018-01-31 11:25:16

999 row(s) in 0.6640 seconds
```

Console Screenshot 4: Results of Scan 'lookup_data_hive'