# Loading the Lookup Table

## Commands to load the relevant data in the Lookup Table ——————————HIVE

- **Creating ranked card transaction table to store last 10 transaction** 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");

```
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: 0.807 seconds
```

- **Creating table for UCL**

CREATE TABLE IF NOT EXISTS
CARD_UCL_ORC( `CARD_ID` STRING,
`UCL` DOUBLE)
STORED AS ORC

TBLPROPERTIES ("orc.compress"="SNAPPY");

```
hive> CREATE TABLE IF NOT EXISTS CARD_UCL_ORC(
    > `CARD_ID` STRING,
    > `UCL` DOUBLE)
    > STORED AS ORC
    > TBLPROPERTIES ("orc.compress"="SNAPPY");
OK
Time taken: 0.068 seconds
```

**- Inserting last 10 data into ranked card transaction**

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_TRANSACTION_HBASE
WHERE
STATUS = 'GENUINE') A ) B WHERE B.RANK <= 10;

```
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_TRANSACTION_HBASE WHERE
    > STATUS = 'GENUINE') A ) B WHERE B.RANK <= 10;
Query ID = root_20220513193100_0b90689b-bff7-4203-81e7-14dc9170b79c
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1652465715257_0005)

--------------------------------------------------------------------------------
        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: 18.75 s
--------------------------------------------------------------------------------
Loading data to table capstone_project.ranked_card_transactions_orc
OK
Time taken: 25.425 seconds
```

**- Inserting data into UCL**
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;

```
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_20220513193127_e9034fa4-4877-4c7d-99a9-90acccd2b472
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1652465715257_0005)

--------------------------------------------------------------------------------
        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: 1.87 s
--------------------------------------------------------------------------------
Loading data to table capstone_project.card_ucl_orc
OK
Time taken: 3.204 seconds
```

## - Inserting data into 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;

```
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 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_20220513193236_913ab00a-12d6-4a3a-aa80-4d1a4264998e
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1652465715257_0005)

----------------------------------------------------------------------------------------
        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: 14.40 s
----------------------------------------------------------------------------------------
OK
Time taken: 21.529 seconds
```

# Command to see the table created and it's content
## - Counting rows in lookup table

select count(*) from lookup_data_hbase;

```
hive> select count(*) from lookup_data_hbase;
Query ID = root_20220513193343_a2db2ecd-3433-4898-adb5-291a0f7eccdc
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1652465715257_0005)

----------------------------------------------------------------------------------------
        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: 7.20 s
----------------------------------------------------------------------------------------
OK
999
```

- **Printing top 10 rows of lookup table** select * from lookup_data_hbase limit 10;

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

# ─────────────────────────────HBASE

## ───────────────────────── - Counting rows in lookup table in HBASE

count 'lookup_data_hive'

```
hbase(main):001:0> count 'lookup_data_hive'
999 row(s) in 0.6450 seconds

=> 999
hbase(main):002:0> ▐
```

- **Printing top 10 rows of lookup table in HBASE** scan 'lookup_data_hive'

```
hbase(main):002:0> scan 'lookup_data_hive'
ROW                              COLUMN+CELL
 340028465709212                 column=lookup_card_family:score, timestamp=1652470377779, value=233
 340028465709212                 column=lookup_card_family:ucl, timestamp=1652470377779, value=1.6331555548882348E7
 340028465709212                 column=lookup_transaction_family:postcode, timestamp=1652470377779, value=24658
 340028465709212                 column=lookup_transaction_family:transaction_dt, timestamp=1652470377779, value=2018-01-02 03:25:35
 340054675199675                 column=lookup_card_family:score, timestamp=1652470377779, value=631
 340054675199675                 column=lookup_card_family:ucl, timestamp=1652470377779, value=1.4156079786189131E7
 340054675199675                 column=lookup_transaction_family:postcode, timestamp=1652470377779, value=50140
 340054675199675                 column=lookup_transaction_family:transaction_dt, timestamp=1652470377779, value=2018-01-15 19:43:23
 340082915339645                 column=lookup_card_family:score, timestamp=1652470377779, value=407
 340082915339645                 column=lookup_card_family:ucl, timestamp=1652470377779, value=1.5285685330791473E7
 340082915339645                 column=lookup_transaction_family:postcode, timestamp=1652470377779, value=17844
 340082915339645                 column=lookup_transaction_family:transaction_dt, timestamp=1652470377779, value=2018-01-26 19:03:47
 340134186926007                 column=lookup_card_family:score, timestamp=1652470377779, value=614
 340134186926007                 column=lookup_card_family:ucl, timestamp=1652470377779, value=1.5239767522438556E7
 340134186926007                 column=lookup_transaction_family:postcode, timestamp=1652470377779, value=67576
 340134186926007                 column=lookup_transaction_family:transaction_dt, timestamp=1652470377779, value=2018-01-18 23:12:50
 340265728490548                 column=lookup_card_family:score, timestamp=1652470377779, value=202
 340265728490548                 column=lookup_card_family:ucl, timestamp=1652470377779, value=1.608491671255562E7
 340265728490548                 column=lookup_transaction_family:postcode, timestamp=1652470377779, value=72435
 340265728490548                 column=lookup_transaction_family:transaction_dt, timestamp=1652470377779, value=2018-01-21 02:07:35
 340268219434811                 column=lookup_card_family:score, timestamp=1652470377779, value=415
 340268219434811                 column=lookup_card_family:ucl, timestamp=1652470377779, value=1.2507323937605347E7
 340268219434811                 column=lookup_transaction_family:postcode, timestamp=1652470377779, value=62513
 340268219434811                 column=lookup_transaction_family:transaction_dt, timestamp=1652470377779, value=2018-01-16 04:30:05
 340379737226464                 column=lookup_card_family:score, timestamp=1652470377779, value=229
 340379737226464                 column=lookup_card_family:ucl, timestamp=1652470377779, value=1.4198310998368107E7
 340379737226464                 column=lookup_transaction_family:postcode, timestamp=1652470377779, value=26656
 340379737226464                 column=lookup_transaction_family:transaction_dt, timestamp=1652470377779, value=2018-01-27 00:19:47
 340383645652108                 column=lookup_card_family:score, timestamp=1652470377779, value=645
 340383645652108                 column=lookup_card_family:ucl, timestamp=1652470377779, value=1.4091750460468251E7
 340383645652108                 column=lookup_transaction_family:postcode, timestamp=1652470377779, value=34734
 340383645652108                 column=lookup_transaction_family:transaction_dt, timestamp=1652470377779, value=2018-01-29 01:29:12
 340803866934451                 column=lookup_card_family:score, timestamp=1652470377779, value=502
 340803866934451                 column=lookup_card_family:ucl, timestamp=1652470377779, value=1.0843341196185412E7
 340803866934451                 column=lookup_transaction_family:postcode, timestamp=1652470377779, value=87525
 340803866934451                 column=lookup_transaction_family:transaction_dt, timestamp=1652470377779, value=2018-01-31 04:23:57
 340889618969736                 column=lookup_card_family:score, timestamp=1652470377779, value=330
 340889618969736                 column=lookup_card_family:ucl, timestamp=1652470377779, value=1.3217942365515321E7
 340889618969736                 column=lookup_transaction_family:postcode, timestamp=1652470377779, value=61341
 340889618969736                 column=lookup_transaction_family:transaction_dt, timestamp=1652470377779, value=2018-01-31 21:57:18
 340924125838453                 column=lookup_card_family:score, timestamp=1652470377779, value=644
 340924125838453                 column=lookup_card_family:ucl, timestamp=1652470377779, value=1.1712496575647112E7
 340924125838453                 column=lookup_transaction_family:postcode, timestamp=1652470377779, value=46392
 340924125838453                 column=lookup_transaction_family:transaction_dt, timestamp=1652470377779, value=2017-10-12 15:41:23
 341005627432127                 column=lookup_card_family:score, timestamp=1652470377779, value=629
 341005627432127                 column=lookup_card_family:ucl, timestamp=1652470377779, value=1.2843150457220368E7
 341005627432127                 column=lookup_transaction_family:postcode, timestamp=1652470377779, value=23306
 341005627432127                 column=lookup_transaction_family:transaction_dt, timestamp=1652470377779, value=2018-01-22 14:59:38
```