Loading the Lookup Table

Commands to load the relevant data in the Lookup Table

Here in below sql we load the data into card_member_lookup table.
We are populating following 4 columns:

1) Card_id
2) UCL
3) POSTCODE
4) Last transaction Date
5) Member score

For calculating these values we need data loaded into tables 1. Transaction, 2. Card_member 3. Member_score.

- Before running above query we would need to create member_score and card_member table to be able to inner join. These tables are created with below commands:

- *CREATE EXTERNAL TABLE IF NOT EXISTS member_score(*
  MEMBER_ID String,
  score String)
  ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
  LINES TERMINATED BY '\n'
  LOCATION '/user/capstone/member_score';
- CREATE EXTERNAL TABLE IF NOT EXISTS card_member(
  card_id string,
  MEMBER_ID string,
  score string,
  tr_date string,
  exp_date string,
  country string,
  area string)
  ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
  LINES TERMINATED BY '\n'
  LOCATION '/user/capstone/card_member';

Command to check data from lookup table:

*SELECT * FROM card_member_lookup;*

Screenshot of the created table

- Screenshot of member_score table creation

```
Time taken: 0.056 seconds
hive> CREATE EXTERNAL TABLE IF NOT EXISTS member_score(
    > MEMBER_ID String,
    > score String)
    > ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
    > LINES TERMINATED BY '\n'
    > LOCATION '/user/capstone/member_score';
OK
Time taken: 0.034 seconds
hive> select * from member_score 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: 0.087 seconds, Fetched: 10 row(s)
hive>
```

- Screenshot of card_member table creation

```
hive> CREATE EXTERNAL TABLE IF NOT EXISTS card_member(
    > card_id string,
    > MEMBER_ID string,
    > score string,
    > tr_date string,
    > exp_date string,
    > country string,
    > area string)
    > ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
    > LINES TERMINATED BY '\n'
    > LOCATION '/user/capstone/card_member';
OK
Time taken: 0.045 seconds
```

- Query:

UCL Is calculated from rolling last 10 transactions as (AVG amount) + 3 x Standard deviation

```
INSERT OVERWRITE TABLE card_member_lookup
SELECT trans.card_id,
    trans.moving_average+3*standard_deviation as UCL,
    POSTCODE,
    transaction_dt,
```

```
    member_score.score
FROM
(
SELECT
  card_id,
  AVG(amount)
      OVER(PARTITION BY card_id ORDER BY transaction_dt ROWS BETWEEN 9 PRECEDING AND CURRENT
ROW)
      AS moving_average,
  STDDEV(amount)
      OVER(PARTITION BY card_id ORDER BY transaction_dt ROWS BETWEEN 9 PRECEDING AND CURRENT
ROW)
      AS standard_deviation,
   transaction_dt,
   POSTCODE,
   ROW_NUMBER() OVER(PARTITION BY card_id ORDER BY transaction_dt DESC ) RN
FROM transactions_formatted
WHERE STATUS = 'GENUINE'
)trans
inner JOIN card_member on (trans.card_id=card_member.card_id)
inner JOIN member_score on (member_score.MEMBER_ID=card_member.MEMBER_ID)
WHERE RN=1;
```

- Screenshot of logic and insertion query into lookup table.

```
hive> INSERT OVERWRITE TABLE LOOKUP_DATA_HBASE
    > SELECT trans.card_id,
    >        trans.moving_average+3*standard_deviation as UCL,
    >        POSTCODE,
    >        transaction_dt,
    >        member_score.score
    > FROM
    > (
    > SELECT
    >    card_id,
    >    AVG(amount)
    >          OVER(PARTITION BY card_id ORDER BY transaction_dt ROWS BETWEEN 9 PRECEDING AND CURRENT ROW)
    >          AS moving_average,
    >    STDDEV(amount)
    >          OVER(PARTITION BY card_id ORDER BY transaction_dt ROWS BETWEEN 9 PRECEDING AND CURRENT ROW)
    >          AS standard_deviation,
    >     transaction_dt,
    >     POSTCODE,
    >     ROW_NUMBER() OVER(PARTITION BY card_id ORDER BY transaction_dt DESC ) RN
    > FROM transactions_formatted
    > WHERE STATUS = 'GENUINE'
    > )trans
    > inner JOIN card_member on (trans.card_id=card_member.card_id)
    > inner JOIN member_score on (member_score.MEMBER_ID=card_member.MEMBER_ID)
    > WHERE RN=1;
No Stats for capstone@transactions_formatted, Columns: amount, postcode, transaction_dt, card_id, status
No Stats for capstone@card_member, Columns: member_id, card_id
No Stats for capstone@member_score, Columns: member_id, score
Query ID = hadoop_20231209111801_337a14d4-a8b5-421e-a671-09bd89e6e9e2
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_1702114013497_0011)

--------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      1          1        0        0       0       0
Map 4 .......... container     SUCCEEDED      1          1        0        0       0       0
Map 5 .......... container     SUCCEEDED      1          1        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      2          2        0        0       0       0
Reducer 3 ...... container     SUCCEEDED      2          2        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 05/05  [==========================>>] 100%  ELAPSED TIME: 19.13 s
--------------------------------------------------------------------------------
OK
Time taken: 30.255 seconds
```

- Screenshot of data loaded into lookup table

```
Time taken: 21.455 seconds
hive> select * from card_member_lookup;
OK
340028465709212 1.6331555548882348E7    24658   2018-01-02 03:25:35     233
340054675199675 1.4156079786189131E7    50140   2018-01-15 19:43:23     631
340082915339645 1.5285685330791473E7    17844   2018-01-26 19:03:47     407
340134186926007 1.5239767522438556E7    67576   2018-01-18 23:12:50     614
340265728490548 1.608491671255562E7     72435   2018-01-21 02:07:35     202
340268219434811 1.2507323937605347E7    62513   2018-01-16 04:30:05     415
340379737226464 1.4198310998368107E7    26656   2018-01-27 00:19:47     229
```

Result -> Above query shows calculated data is loaded successfully into card_member_lookup table.