

Loading the Lookup Table

Commands to load the relevant data in the Lookup Table:

Purpose of document: This document is intent to contain the script to calculate the moving average and standard deviation of the last 10 transactions for each card_id for the data present in Hadoop and NoSQL database.

It also should have logic If the total number of transactions for a particular card_id is less than 10, then calculate the parameters based on the total number of records available for that card_id.

The script should be able to extract and feed the other relevant data ('postcode', 'transaction_dt', 'score', etc.) for the look-up table along with card_id and UCL.

View referred for genuine transactions (rank wise).

```
CREATE VIEW IF NOT EXISTS last_ten_transactions
AS select card_id, member_id, amount, transaction_dt, postcode, rank() over (PARTITION BY card_id
ORDER BY unix_timestamp(transaction_dt , 'dd-MM-yyyy hh:mm:ss') desc, amount desc) as ranking
from card_transactions where status='GENUINE';
```

Logic explanation:

- Purpose to get last 10 latest genuine transactions for each card.
 - Rank applied for each cards with descending order of transaction time-stamp and amount
 - Filtered data only for genuine transactions
 - If last 10 transactions not available picked-up only available transactions for the card.
- View provides the attributes like card_d, member_id, amount, transaction_dt, postcode for each card id.

Inserting data into look up table:

```
Insert into look_up_table
select ltt.card_id, cm.card_purchase_dt, ltt.transaction_dt, ltt.member_id, member_joining_dt, country,
city, UCL, ltt.postcode, score
from last_ten_transactions ltt
inner join member_score ms on ltt.member_id=ms.member_id and ltt.ranking=1
inner join card_member cm on cm.member_id=ltt.member_id and ltt.ranking=1
inner join (select card_id, avg(amount) + (3* stddev(amount)) as UCL
from last_ten_transactions
where ranking<=10 group by card_id) as ucl
on ltt.card_id=ucl.card_id ;
```

Logic explanation:

- UCL: UCL value is calculated as “average amount + 3 * standard deviation for last 10 genuine transactions” for the card. In case last 10 transactions not available considered what is available for the card id (i.e. <10)
- Postcode: postcode derived from the latest transaction.
- Score: score derived from member_score table

```
hive> Insert into look_up_table select ltt.card_id, cm.card_purchase_dt, ltt.transaction_dt, ltt.member_id, member_joining_dt, country, city, UCL, ltt.postcode, score from last_ten_transactions ltt inner join member_score ms on ltt.member_id=ms.member_id and ltt.ranking=1 inner join card_member cm on cm.member_id=ltt.member_id and ltt.ranking=1 inner join (select card_id, avg(amount)+ (3* stddev(amount)) as UCL from last_ten_transactions where ranking<=10 group by card_id) as ucl on ltt.card_id=ucl.card_id ;
No Stats for default@card_transactions, Columns: member_id, amount, postcode, transaction_dt, card_id, status
No Stats for default@member_score, Columns: member_id, score
No Stats for default@card_member, Columns: member_id, country, member_joining_dt, city, card_purchase_dt
No Stats for default@card_transactions, Columns: amount, transaction_dt, card_id, status
Query ID = hadoop_20230729113122_ca40f2a6-11be-46c0-a2a1-682ad54742fc
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_1690626913936_0007)

-----
VERTICES      MODE      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container  SUCCEEDED    1         1         0         0         0         0
Map 3 ..... container  SUCCEEDED    1         1         0         0         0         1
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 6 ..... container  SUCCEEDED    2         2         0         0         0         0
Reducer 7 ..... container  SUCCEEDED    2         2         0         0         0         0
-----
VERTICES: 07/07 [=====>>>] 100% ELAPSED TIME: 25.30 s
-----
OK
Time taken: 41.888 seconds
hive>
```

Command to see the table created and it's content:

Show tables;

```
hive> show tables;
OK
card_member
card_transactions
card_transactions_stg
last_ten_transactions
look_up_table
member_score
Time taken: 0.058 seconds, Fetched: 6 row(s)
hive>
```

Check tables are populated with content:

Look_up_table:

Select * from look_up_table limit 20;

```
hive> select * from look_up_table limit 20;
OR
look_up_table.card_id    look_up_table.card_purchase_dt look_up_table.transaction_dt    look_up_table.member_id look_up_table.member_joining_dt look_up_table.country
look_up_table.city      look_up_table.ucl            look_up_table.postcode look_up_table.score
3400528465709212 05/13 02-01-2018 03:25:35 009250698176266 2012-02-08 06:04:13.0 United States Barberton 1.6331555548882348E7 24658 233
340054675199675 03/17 15-01-2018 19:43:23 835873341185231 2017-03-10 09:24:44.0 United States Fort Dodge 1.4156079786189131E7 50140 631
340082915339645 07/14 26-01-2018 19:03:47 512969555857346 2014-02-15 06:30:30.0 United States Graham 1.5285685330791473E7 17844 407
340134186926007 02/13 18-01-2018 23:12:50 887711945571282 2012-02-05 01:21:58.0 United States Dix Hills 1.5239767522438556E7 67576 614
340265728490549 11/14 21-01-2018 02:07:35 680324265406190 2014-03-29 07:49:14.0 United States Rancho Cucamonga 1.608491671255562E7 72435 202
340268219434811 08/12 16-01-2018 04:30:05 929799084911715 2012-07-08 02:46:08.0 United States San Francisco 1.2507323937605347E7 62513 415
340379737226464 09/10 27-01-2018 00:19:47 089615510858348 2010-03-10 00:06:42.0 United States Clinton 1.4198310998368107E7 26656 229
340383645652108 10/16 29-01-2018 01:29:12 181180599313885 2012-02-24 05:32:44.0 United States West New York 1.4091750460468251E7 34734 645
340803866934451 08/17 31-01-2018 04:23:57 417664728506297 2015-05-21 04:30:45.0 United States Beaverton 1.0843341196189412E7 87525 502
340889618969736 11/15 31-01-2018 21:57:18 459292914761635 2013-04-23 08:40:11.0 United States West Palm Beach 1.3217942365515321E7 61341 330
340924125838453 12/13 12-10-2017 15:41:23 188119365574843 2011-04-12 04:28:47.0 United States Scottsbluff 1.1712496575647112E7 46392 644
341005627432127 02/17 22-01-2018 14:59:38 872138964937565 2013-09-08 03:16:50.0 United States Chillum 1.2843150457220368E7 23306 629
341029651579925 08/12 17-01-2018 10:20:48 974087224071871 2011-01-14 00:20:25.0 United States Valley Station 1.384435212566918E7 18068 682
341311317050937 02/15 29-12-2017 23:56:09 561687420200207 2014-03-18 06:23:23.0 United States Vincennes 1.2104977216410311E7 29067 440
341344252914274 03/13 30-01-2018 21:22:35 685906467918552 2012-03-02 03:21:01.0 United States Columbine 1.0770448614621948E7 35491 643
341363850179050 04/14 21-01-2018 23:11:37 009190444424572 2012-02-19 05:16:44.0 United States Cheakowaga 1.2514602030066479E7 43967 559
341519629171379 01/15 21-01-2018 09:08:55 533670008048847 2013-05-13 07:59:32.0 United States Centennial 1.529211229141618E7 87513 408
341641153427489 11/15 17-12-2017 06:27:42 230523184584316 2013-03-25 08:51:19.0 United States Colchester 1.3421215715089727E7 38023 475
341719092861087 11/17 30-01-2018 09:05:14 304847505155781 2015-12-06 08:06:35.0 United States Vernon Hills 1.1521372646889605E7 95372 442
341722035429601 01/17 01-02-2018 01:27:59 979218131207765 2015-12-22 10:46:23.0 United States Elk Grove Village 1.2890883282580867E7 73554 552
Time taken: 0.229 seconds, Fetched: 20 row(s)
hive>
```

Select count(*) from look_up_table;

```
hive> select count(*) from look_up_table;
Query ID = hadoop_20230729125840_ccabfef2-f9bc-43de-b534-35b085dac966
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1690626913936_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: 8.11 s
-----
OK
999
Time taken: 10.534 seconds, Fetched: 1 row(s)
hive> █
```

Screenshot of the created table

Screenshot of the table created:

```
hbase(main):001:0> describe 'look_up_table'
Table look_up_table is ENABLED
look_up_table
COLUMN FAMILIES DESCRIPTION
(NAME => 'Location', 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 => 'Member_details', 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 => 'Rule_params', 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 => 'card_details', 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')
4 row(s) in 0.3170 seconds
hbase(main):002:0> █
```

Capstone project – Credit Card Fraud Detection (mid submission)

by

Susil Patro, Krishna Mohan & Ashmeet Singh Deol

