## Scripts Execution

**Screenshots of the execution of the scripts written**

Once data ingestion is done from RDS and card\_transaction.csv is loaded into hbase

**Step 1: Load card\_transaction into hive**

CREATE SCHEMA transactions;

CREATE TABLE IF NOT EXISTS transactions.card\_transactions ( card\_id string, member\_id string, amount int, postcode string, pos\_id string, transaction\_dt string, status string) COMMENT 'Transaction Table' ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' tblproperties("skip.header.line.count"="1");

LOAD DATA INPATH '/user/hbase/data/card\_transactions.csv' INTO TABLE transactions.card\_transactions ;

**Step 2: Load Member score into Hive**

hive> CREATE TABLE IF NOT EXISTS transactions.member\_score ( member\_id string, score int) COMMENT 'Score Table' ROW FORMAT DELIMITED FIELDS TERMINATED BY ',';

OK

Time taken: 0.089 seconds

hive> LOAD DATA INPATH '/user/hadoop/member\_score/part-m-00000' INTO TABLE transactions.member\_score ;

Loading data to table transactions.member\_score

OK

**Step 3: Create lookup table**

CREATE TABLE transactions.look\_up\_table\_hbase ( card\_id string, ucl int, postcode string, transaction\_dt string, score int ) STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler' WITH SERDEPROPERTIES ("hbase.columns.mapping" = ":key,cf:ucl,cf:postcode,cf:transaction\_dt,cf:score") TBLPROPERTIES ("hbase.table.name" = "look\_up\_table\_hbase");

**Step 4: Load processed data into lookup table**

insert into transactions.look\_up\_table\_hbase ( card\_id, ucl, postcode, transaction\_dt, score)

select t1.card\_id, us.ucl, t1.postcode, t1.transaction\_dt, t1.score

from

(SELECT t.card\_id

,t.postcode

,t.transaction\_dt

,t.score

FROM (

SELECT s.member\_id, c.card\_id

,c.postcode

,c.transaction\_dt

,s.score

,ROW\_NUMBER() OVER (

PARTITION BY card\_id ORDER BY unix\_timestamp(transaction\_dt,'dd-MM-yyyy hh:mm:ss') DESC

) AS ROW\_NUMBER

FROM transactions.card\_transactions c

join transactions.member\_score s on c.member\_id=s.member\_id

order by c.card\_id

) t

WHERE t.ROW\_NUMBER <= 1) t1

join (SELECT t.card\_id, cast (avg(t.amount)+3\*STDDEV(t.amount) as int) as ucl

FROM (

SELECT s.member\_id, c.card\_id

,c.postcode

,c.amount

,c.transaction\_dt

,s.score

,ROW\_NUMBER() OVER (

PARTITION BY card\_id ORDER BY unix\_timestamp(transaction\_dt,'dd-MM-yyyy hh:mm:ss') DESC

) AS ROW\_NUMBER

FROM transactions.card\_transactions c

join transactions.member\_score s on c.member\_id=s.member\_id

order by c.card\_id

) t

WHERE t.ROW\_NUMBER <= 10

group by t.card\_id) us on us.card\_id=t1.card\_id;

**Result:**

See the table and entries created in hbase

hbase(main):001:0> list

TABLE

look\_up\_table\_hbase

1 row(s) in 0.4160 seconds

=> ["look\_up\_table\_hbase"]

hbase(main):002:0> count look\_up\_table\_hbase

NameError: undefined local variable or method `look\_up\_table\_hbase' for #<Object:0x787d1f9c>

hbase(main):003:0> count 'look\_up\_table\_hbase'

999 row(s) in 0.4490 seconds

=> 999

hbase(main):004:0> list

TABLE

look\_up\_table\_hbase

1 row(s) in 0.0150 seconds

=> ["look\_up\_table\_hbase"]

hbase(main):005:0> scan 'look\_up\_table\_hbase'

ROW COLUMN+CELL

340028465709212 column=cf:postcode, timestamp=1666615518424, value=24658

340028465709212 column=cf:score, timestamp=1666615518424, value=233

340028465709212 column=cf:transaction\_dt, timestamp=1666615518424, value=02-01-2018 03:25:35

340028465709212 column=cf:ucl, timestamp=1666615518424, value=319613580

340054675199675 column=cf:postcode, timestamp=1666615518424, value=50140

340054675199675 column=cf:score, timestamp=1666615518424, value=631

340054675199675 column=cf:transaction\_dt, timestamp=1666615518424, value=15-01-2018 19:43:23

340054675199675 column=cf:ucl, timestamp=1666615518424, value=14156079

340082915339645 column=cf:postcode, timestamp=1666615518424, value=17844

340082915339645 column=cf:score, timestamp=1666615518424, value=407

340082915339645 column=cf:transaction\_dt, timestamp=1666615518424, value=26-01-2018 19:03:47

340082915339645 column=cf:ucl, timestamp=1666615518424, value=15285685

340134186926007 column=cf:postcode, timestamp=1666615518424, value=67576

340134186926007 column=cf:score, timestamp=1666615518424, value=614

340134186926007 column=cf:transaction\_dt, timestamp=1666615518424, value=18-01-2018 23:12:50

340134186926007 column=cf:ucl, timestamp=1666615518424, value=909246042

340265728490548 column=cf:postcode, timestamp=1666615518424, value=72435

340265728490548 column=cf:score, timestamp=1666615518424, value=202

340265728490548 column=cf:transaction\_dt, timestamp=1666615518424, value=21-01-2018 02:07:35

340265728490548 column=cf:ucl, timestamp=1666615518424, value=16084916

340268219434811 column=cf:postcode, timestamp=1666615518424, value=62513

340268219434811 column=cf:score, timestamp=1666615518424, value=415

340268219434811 column=cf:transaction\_dt, timestamp=1666615518424, value=16-01-2018 04:30:05

340268219434811 column=cf:ucl, timestamp=1666615518424, value=12507323