Loading Historical Transactions Data into NoSQL Database

(Commands to load the past transactions data into NoSQL database**)**

Purpose of document: To provide commands to load the transactions history data (card_transactions.csv) in a NoSQL database.

Following steps are followed to populate card_transactions table in hive integrated Hbase table

---

STEP 1 – COPY THE CARD_TRANSACTIONS.CSV FILE (SOURCE)

---

hdfs dfs -mkdir hive
hdfs dfs -copyFromLocal /home/hadoop/card_transactions.csv /user/hadoop/hive/card_transactions.csv

---

Step 2 – Create and load data on to card_transactions staging table:

STAGING TABLE CREATION

```
CREATE EXTERNAL TABLE IF NOT EXISTS card_transactions_stg (
  card_id STRING,
  member_id STRING,
  amount DOUBLE,
  postcode STRING,
  pos_id STRING,
  transaction_dt STRING,
  status STRING
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
LOCATION '/user/hadoop/historical/card_transactions'
tblproperties("skip.header.line.count"="1");
```

LOADING DATA ON TO STAGING TABLE:

```
LOAD DATA INPATH 'hdfs:/user/hadoop/hive/card_transactions.csv' OVERWRITE INTO TABLE card_transactions_stg;
```

```
hive> LOAD DATA INPATH 'hdfs:/user/hadoop/hive/card_transactions.csv' OVERWRITE INTO TABLE card_transactions_stg;
Loading data to table default.card_transactions_stg
OK
Time taken: 1.259 seconds
hive> []
```

CREATE HBASE CARD_TRANSACITONS TABLE

echo "create 'card_transactions', 'TD'" | hbase shell -n

```
[hadoop@ip-172-31-64-25 ~]$ hdfs dfs -mkdir hive
[hadoop@ip-172-31-64-25 ~]$ hdfs dfs -copyFromLocal /home/hadoop/card_transactio
ns.csv /user/hadoop/hive/card_transactions.csv
[hadoop@ip-172-31-64-25 ~]$ echo "create 'card_transactions', 'TD'" | hbase shel
l -n
```

CREATING HIVE INTEGRATED HBASE TABLE FOR CARD_TRANSACTIONS:

CREATE EXTERNAL TABLE IF NOT EXISTS card_transactions (row_key struct<card_id:string, pos_id:string, transaction_dt:string, amount:double>, card_id STRING, pos_id STRING, transaction_dt STRING, member_id STRING,amount DOUBLE, postcode STRING,status STRING)
ROW FORMAT DELIMITED
COLLECTION ITEMS TERMINATED BY '~'
STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
WITH SERDEPROPERTIES ('hbase.columns.mapping'='TD:card_id, TD:pos_id,
TD:transaction_dt,TD:member_id,TD:amount, TD:postcode, TD:status')
TBLPROPERTIES ("hbase.table.name" = "card_transactions");

```
hive> CREATE EXTERNAL TABLE IF NOT EXISTS card_transactions (row_key struct<card_id:string, pos_id:string, transaction_dt:string,
amount:double>, card_id STRING, pos_id STRING, transaction_dt STRING, member_id STRING,amount DOUBLE, postcode STRING,status STRIN
G)
    > ROW FORMAT DELIMITED
    > COLLECTION ITEMS TERMINATED BY '~'
    > STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
    > WITH SERDEPROPERTIES ('hbase.columns.mapping'='TD:card_id, TD:pos_id, TD:transaction_dt,TD:member_id,TD:amount, TD:postcode,
 TD:status')
    > TBLPROPERTIES ("hbase.table.name" = "card_transactions");
OK
Time taken: 0.161 seconds
hive> []
```

LOAD DATA TO CARD_TRANSACTIONS TABLE FROM STAGING TABLE:

Inserting data from staging table to main card_transactions table:
Insert into card_transactions select NAMED_STRUCT('card_id',card_id,'pos_id',pos_id,'transaction_dt', transaction_dt, 'amount', amount) as row_key,card_id,pos_id,transaction_dt, member_id, amount, postcode, status from card_transactions_stg;

```
hive> Insert into card_transactions  select NAMED_STRUCT('card_id',card_id,'pos_id',pos_id,'transaction_dt', transaction_dt, 'amou
nt', amount) as row_key,card_id,pos_id,transaction_dt, member_id, amount, postcode, status from card_transactions_stg;
Query ID = hadoop_20230729110403_d8413718-673c-4c95-b14e-5b9dc0cb8644
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_0002)

--------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 .......... container      SUCCEEDED      1          1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 01/01  [==========================>>] 100%  ELAPSED TIME: 10.68 s
--------------------------------------------------------------------------------
OK
Time taken: 24.095 seconds
```

<Command to list the table in which the data is loaded and the command to get the count of the rows of the table>

COMMAND: 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>
```

COUNT ROWS - CARD_TRANSACTIONS:

```
hive> select count(*) from card_transactions;
Query ID = hadoop_20230729111028_28eaabf3-8efa-47f1-a96c-6fac02f6f614
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_0003)

----------------------------------------------------------------------------------------------
        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: 8.20 s
----------------------------------------------------------------------------------------------
OK
53292
Time taken: 17.63 seconds, Fetched: 1 row(s)
hive>
```

SAMPLE DATA - CARD_TRANSACITONS:

```
hive> select * from card_transactions limit 20;
OK
card_transactions.row_key        card_transactions.card_id        card_transactions.pos_id        card_transactions.transaction_dt        card_transactions.member_id c
ard_transactions.amount card_transactions.postcode      card_transactions.status
{"card_id":"340028465709212","pos_id":"033981484018726","transaction_dt":"27-05-2016 17:57:05","amount":5551324.0}        340028465709212 033981484018726 27-05-2016 17
:57:05  0092506698176266 5551324.0        25915   GENUINE
{"card_id":"340028465709212","pos_id":"044677148710313","transaction_dt":"08-05-2016 03:27:04","amount":4748901.0}        340028465709212 044677148710313 08-05-2016 03
:27:04  0092506698176266 4748901.0        58641   GENUINE
{"card_id":"340028465709212","pos_id":"057678222018909","transaction_dt":"12-08-2017 08:29:54","amount":9291309.0}        340028465709212 057678222018909 12-08-2017 08
:29:54  0092506698176266 9291309.0        31322   GENUINE
{"card_id":"340028465709212","pos_id":"072517912051441","transaction_dt":"09-11-2017 07:18:21","amount":6503191.0}        340028465709212 072517912051441 09-11-2017 07
:18:21  0092506698176266 6503191.0        84776   GENUINE
{"card_id":"340028465709212","pos_id":"084349436605018","transaction_dt":"09-03-2017 07:48:26","amount":1325446.0}        340028465709212 084349436605018 09-03-2017 07
:48:26  0092506698176266 1325446.0        82630   GENUINE
{"card_id":"340028465709212","pos_id":"159876448781622","transaction_dt":"27-01-2017 08:40:20","amount":5093181.0}        340028465709212 159876448781622 27-01-2017 08
:40:20  0092506698176266 5093181.0        56459   GENUINE
{"card_id":"340028465709212","pos_id":"219307586884330","transaction_dt":"14-01-2017 16:57:13","amount":8737772.0}        340028465709212 219307586884330 14-01-2017 16
:57:13  0092506698176266 8737772.0        82933   GENUINE
{"card_id":"340028465709212","pos_id":"236114661616473","transaction_dt":"15-03-2017 02:29:58","amount":1730320.0}        340028465709212 236114661616473 15-03-2017 02
:29:58  0092506698176266 1730320.0        54545   GENUINE
{"card_id":"340028465709212","pos_id":"246987608008994","transaction_dt":"02-01-2018 03:25:35","amount":8696557.0}        340028465709212 246987608008994 02-01-2018 03
:25:35  0092506698176266 8696557.0        24658   GENUINE
{"card_id":"340028465709212","pos_id":"305578834263571","transaction_dt":"01-06-2017 19:31:24","amount":7381252.0}        340028465709212 305578834263571 01-06-2017 19
:31:24  0092506698176266 7381252.0        63951   GENUINE
{"card_id":"340028465709212","pos_id":"324423694853335","transaction_dt":"22-05-2017 09:47:34","amount":5682251.0}        340028465709212 324423694853335 22-05-2017 09
:47:34  0092506698176266 5682251.0        20609   GENUINE
{"card_id":"340028465709212","pos_id":"325960823570918","transaction_dt":"01-03-2016 15:02:02","amount":983262.0}        340028465709212 325960823570918 01-03-2016 15
:02:02  0092506698176266 983262.0        13485   GENUINE
{"card_id":"340028465709212","pos_id":"349764614809247","transaction_dt":"08-04-2016 12:26:02","amount":3433123.0}        340028465709212 349764614809247 08-04-2016 12
:26:02  0092506698176266 3433123.0        11952   GENUINE
{"card_id":"340028465709212","pos_id":"377613506830624","transaction_dt":"29-02-2016 19:43:20","amount":107640.0}        340028465709212 377613506830624 29-02-2016 19
:43:20  0092506698176266 107640.0        43812   GENUINE
{"card_id":"340028465709212","pos_id":"418287518557763","transaction_dt":"28-05-2016 13:15:47","amount":1419025.0}        340028465709212 418287518557763 28-05-2016 13
:15:47  0092506698176266 1419025.0        43045   GENUINE
{"card_id":"340028465709212","pos_id":"453792673786320","transaction_dt":"30-03-2017 04:09:10","amount":5118701.0}        340028465709212 453792673786320 30-03-2017 04
:09:10  0092506698176266 5118701.0        12045   GENUINE
{"card_id":"340028465709212","pos_id":"485353426643646","transaction_dt":"13-02-2017 16:44:56","amount":5829887.0}        340028465709212 485353426643646 13-02-2017 16
:44:56  0092506698176266 5829887.0        24078   GENUINE
{"card_id":"340028465709212","pos_id":"489939408321423","transaction_dt":"19-10-2016 01:44:49","amount":9230684.0}        340028465709212 489939408321423 19-10-2016 01
:44:49  0092506698176266 9230684.0        98588   GENUINE
{"card_id":"340028465709212","pos_id":"498255481028848","transaction_dt":"05-07-2017 11:05:55","amount":9687739.0}        340028465709212 498255481028848 05-07-2017 11
:05:55  0092506698176266 9687739.0        51542   GENUINE
{"card_id":"340028465709212","pos_id":"509545401949131","transaction_dt":"17-05-2017 12:36:12","amount":581323.0}        340028465709212 509545401949131 17-05-2017 12
:36:12  0092506698176266 581323.0        46182   GENUINE
Time taken: 0.201 seconds, Fetched: 20 row(s)
hive> []
```

Capstone project – Credit Card Fraud Detection (mid submission)
by
Susil Patro, Krishna Mohan & Ashmeet Singh Deol