

## Loading Historical Transactions Data into NoSQL Database

### Commands to load the past transactions data into NoSQL database

#### Note:

I am following an approach to read all card history transaction data by using Spark. Hence as confirmed by instructor in Mid Submission Session I am proceeding with pyspark approach.

Still, for submission, I coded a program in .py to load CSV to hbase and create a table card\_transaction.

This below code is going to load all CSV into NoSQL in batch mode.

```
import happybase
#create connection
connection = happybase.Connection('localhost', port=9090 ,autoconnect=False)
#open connection to perform operations
def open_connection():
    connection.open()
#close the opened connection
def close_connection():
    connection.close()
#list all tables in Hbase
def list_tables():
    print "fetching all table"
    open_connection()
    tables = connection.tables()
    close_connection()
    print "all tables fetched"
    return tables

#create the required table
def create_table(name,cf):
    print "creating table " + name
    tables = list_tables()
    if name not in tables:
        open_connection()
        connection.create_table(name, cf)
        close_connection()
        print "table created"
```

```

else:
    print "table already present"
#get the pointer to a table
def get_table(name):
    open_connection()
    table = connection.table(name)
    close_connection()
    return table
#batch insert data in events table
def batch_insert_data(filename,tableName):
    print "starting batch insert of events"
    file = open(filename, "r")
    table = get_table(tableName)
    open_connection()
    i=0
    for line in file:

        temp = line.strip().split(",")
        #Skip the first row
        if temp[0]!='card_id':
            table.put(bytes(i) , { 'info:card_id':bytes(temp[0]),
                                   'info:member_id':bytes(temp[1]),
                                   'info:amount':bytes(temp[2]),
                                   'info:postcode':bytes(temp[3]),
                                   'info:pos_id':bytes(temp[4]),
                                   'info:transaction_dt':bytes(temp[5]),
                                   'info:status':bytes(temp[6])})

            i=i+1

    file.close()

    print "batch insert done"
    close_connection()
# Batch insert data of card_transactions.csv file.
create_table('card_transactions', {'info' : dict(max_versions=5) })
batch_insert_data('card_transactions.csv','card_transactions')

```

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

- 1) Login to Putty and enter as root user.
- 2) Run thrift server
- 3) Give Hbase Shell and press enter.
- 4) Give "list" to see all tables in hbase

```
hbase(main):001:0> list
TABLE
card_transactions
employee
look_up_table
3 row(s) in 0.1970 seconds

=> ["card_transactions", "employee", "look_up_table"]
hbase(main):002:0> 
```

- 5) Give “ count ‘card\_transactions’ “ in command line.

```
=> ["card_transactions", "employee", "look_up_table"]
hbase(main):002:0> count 'card_transactions'
Current count: 1000, row: 10898
```

```
Current count: 53000, row: 9735
53292 row(s) in 4.1730 seconds
```

## Screenshot of the table created

- 1) Scan ‘card\_transactions’

```
root@vip-10-u-u-02/~/ec2-user:~$ hbase shell
hbase(main):001:0> scan 'card_transactions'
9736 column=info:pos_id, timestamp=1607957944969, value=546062979209182
9736 column=info:postcode, timestamp=1607957944969, value=95991
9736 column=info:status, timestamp=1607957944969, value=GENUINE
9736 column=info:transaction_dt, timestamp=1607957944969, value=03-01-2018 12:46:58
9737 column=info:amount, timestamp=1607957944972, value=816027
9737 column=info:card_id, timestamp=1607957944972, value=6011706374151856
9737 column=info:member_id, timestamp=1607957944972, value=199243620982420
9737 column=info:pos_id, timestamp=1607957944972, value=546062979209182
9737 column=info:postcode, timestamp=1607957944972, value=95991
9737 column=info:status, timestamp=1607957944972, value=GENUINE
9737 column=info:transaction_dt, timestamp=1607957944972, value=03-01-2018 23:18:59
9738 column=info:amount, timestamp=1607957944973, value=4042654
9738 column=info:card_id, timestamp=1607957944973, value=6011706374151856
9738 column=info:member_id, timestamp=1607957944973, value=199243620982420
9738 column=info:pos_id, timestamp=1607957944973, value=599304777798828
9738 column=info:postcode, timestamp=1607957944973, value=96727
9738 column=info:status, timestamp=1607957944973, value=GENUINE
9738 column=info:transaction_dt, timestamp=1607957944973, value=03-12-2017 08:53:52
9739 column=info:amount, timestamp=1607957944974, value=4034933
9739 column=info:card_id, timestamp=1607957944974, value=6011706374151856
9739 column=info:member_id, timestamp=1607957944974, value=199243620982420
9739 column=info:pos_id, timestamp=1607957944974, value=068028537978614
9739 column=info:postcode, timestamp=1607957944974, value=14511
9739 column=info:status, timestamp=1607957944974, value=GENUINE
9739 column=info:transaction_dt, timestamp=1607957944974, value=05-11-2017 10:15:33
974 column=info:amount, timestamp=1607957932834, value=5093161
974 column=info:card_id, timestamp=1607957932834, value=4380357644954006
974 column=info:member_id, timestamp=1607957932834, value=012871894714576
974 column=info:pos_id, timestamp=1607957932834, value=11443968682626
974 column=info:postcode, timestamp=1607957932834, value=10919
974 column=info:status, timestamp=1607957932834, value=GENUINE
974 column=info:transaction_dt, timestamp=1607957932834, value=05-10-2017 01:06:28
9740 column=info:amount, timestamp=1607957944975, value=3173523
9740 column=info:card_id, timestamp=1607957944975, value=6011706374151856
9740 column=info:member_id, timestamp=1607957944975, value=199243620982420
9740 column=info:pos_id, timestamp=1607957944975, value=771819356785069
9740 column=info:postcode, timestamp=1607957944975, value=72143
9740 column=info:status, timestamp=1607957944975, value=GENUINE
9740 column=info:transaction_dt, timestamp=1607957944975, value=05-12-2017 15:19:17
9741 column=info:amount, timestamp=1607957944976, value=2070998
9741 column=info:card_id, timestamp=1607957944976, value=6011706374151856
9741 column=info:member_id, timestamp=1607957944976, value=199243620982420
9741 column=info:pos_id, timestamp=1607957944976, value=771819356785069
9741 column=info:postcode, timestamp=1607957944976, value=72143
```

Activate Windows  
Go to Settings to activate Windows.

```

9994 column=info:card_id, timestamp=1607957945222, value=6515567258324915
9994 column=info:member_id, timestamp=1607957945222, value=203259382349255
9994 column=info:pos_id, timestamp=1607957945222, value=752540960094399
9994 column=info:postcode, timestamp=1607957945222, value=29375
9994 column=info:status, timestamp=1607957945222, value=GENUINE
9994 column=info:transaction_dt, timestamp=1607957945222, value=19-12-2017 07:37:06
9995 column=info:amount, timestamp=1607957945223, value=3199042
9995 column=info:card_id, timestamp=1607957945223, value=6515567258324915
9995 column=info:member_id, timestamp=1607957945223, value=203259382349255
9995 column=info:pos_id, timestamp=1607957945223, value=716478169896133
9995 column=info:postcode, timestamp=1607957945223, value=65766
9995 column=info:status, timestamp=1607957945223, value=GENUINE
9995 column=info:transaction_dt, timestamp=1607957945223, value=20-05-2016 06:52:49
9996 column=info:amount, timestamp=1607957945223, value=2720038
9996 column=info:card_id, timestamp=1607957945223, value=6515567258324915
9996 column=info:member_id, timestamp=1607957945223, value=203259382349255
9996 column=info:pos_id, timestamp=1607957945223, value=279737182885278
9996 column=info:postcode, timestamp=1607957945223, value=12063
9996 column=info:status, timestamp=1607957945223, value=GENUINE
9996 column=info:transaction_dt, timestamp=1607957945223, value=20-10-2016 12:51:32
9997 column=info:amount, timestamp=1607957945224, value=1712819
9997 column=info:card_id, timestamp=1607957945224, value=6515567258324915
9997 column=info:member_id, timestamp=1607957945224, value=203259382349255
9997 column=info:pos_id, timestamp=1607957945224, value=504297931562759
9997 column=info:postcode, timestamp=1607957945224, value=12486
9997 column=info:status, timestamp=1607957945224, value=GENUINE
9997 column=info:transaction_dt, timestamp=1607957945224, value=20-11-2016 09:43:17
9998 column=info:amount, timestamp=1607957945225, value=6443557
9998 column=info:card_id, timestamp=1607957945225, value=6515567258324915
9998 column=info:member_id, timestamp=1607957945225, value=203259382349255
9998 column=info:pos_id, timestamp=1607957945225, value=793759051265381
9998 column=info:postcode, timestamp=1607957945225, value=17250
9998 column=info:status, timestamp=1607957945225, value=GENUINE
9998 column=info:transaction_dt, timestamp=1607957945225, value=21-03-2017 02:27:31
9999 column=info:amount, timestamp=1607957945226, value=9656436
9999 column=info:card_id, timestamp=1607957945226, value=6515567258324915
9999 column=info:member_id, timestamp=1607957945226, value=203259382349255
9999 column=info:pos_id, timestamp=1607957945226, value=736124761893873
9999 column=info:postcode, timestamp=1607957945226, value=14453
9999 column=info:status, timestamp=1607957945226, value=GENUINE
9999 column=info:transaction_dt, timestamp=1607957945226, value=21-04-2016 05:54:28

```

53292 row(s) in 53.8490 seconds

Activate Windows  
Go to Settings to activate Windows.