

Loading Historical Transactions Data into NoSQL Database

1. Transfer the CSV from s3 bucket to the HDFS

```
aws s3 cp s3://capstonecreditcard/card_transactions.csv .
```

2. Commands to load the past transactions data into NoSQL database

```
import happybase
import uuid

connection=happybase.Connection('localhost')

connection.open()
print('Connected to ', connection.host)

try:
    connection.create_table('card_transactions',
    {
        'card_info': dict(),
        'member_info':dict(),
        'transaction_info':dict()
    })

    print('table created : card_transactions')

except Exception as e :
    print(e)
    connection.close()
    print('connection closed')

card_transactions=connection.table('card_transactions')

# put data to the table
with open('card_transactions.csv', 'r') as data:
```

```
header= data.readline()
lines= data.readlines()
for line in lines:
    line=line.strip()
    line= line.split(",")
    card_transactions.put(uuid.uuid1().bytes, {'card_info:card_id': line[0],
'member_info:member_id': line[1],
'transaction_info:amount': line[2],
'transaction_info:postcode': line[3],
'transaction_info:pos_id': line[4],
'transaction_info:transaction_dt': line[5],
'transaction_info:status': line[6]
    })

print('Data inserted: card_transactions')

# close connection to the HBase
connection.close()
print('connection closed')
```

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

List : to list the table in which the data is loaded

count 'card_transactions': get the count of the rows of the table

Screenshot of the table created

```
[hbase(main):001:0> list
TABLE
card_transactions
lookup
2 row(s) in 0.2210 seconds

=> ["card_transactions", "lookup"]
```

```
[hbase(main):002:0> count 'card_transactions'
Current count: 1000, row: \xA5\x10G(\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 2000, row: \xA6W\xE3\x98\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 3000, row: \xA7\xC9K\xAE\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 4000, row: \xA8\x98\x84\xB4\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 5000, row: \xA9^\xFF\xF4\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 6000, row: \xAAG\xD1\xD2\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 7000, row: \xAB\xB7\xC4\xFC\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 8000, row: \xAC\x94a\x00\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 9000, row: \xAD\x9BL\xDA\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 10000, row: \xAE\xA01J\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 11000, row: \xB05\xCB\xF0\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 12000, row: \xB0\xEE\xFB*\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 13000, row: \xB1\xA1\xB9@\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 14000, row: \xB2z\xF3\xEA\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 15000, row: \xB3)I\x90\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 16000, row: \xB3\xD7\xAC~\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 17000, row: \xB4\x96/\xE6\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 18000, row: \xB5=\x03\xC0\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 19000, row: \xB5\xEE\x0AN\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 20000, row: \xB6\x9Cv\xBA\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 21000, row: \xB7\x1Bz\xA0\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 22000, row: \xB7\x96\x9C\x1C\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 23000, row: \xB8.r\x12\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 24000, row: \xB8\xAC46\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 25000, row: \xB9'J\x1E\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 26000, row: \xB9\xB6)\xA4\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 27000, row: \xBA68\x10\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 28000, row: \xBA\xB2T|\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 29000, row: \xBBO\x00\xB0\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 30000, row: \xBB\xDD\xE6\x86\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 31000, row: \xBCZ_\xCC\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 32000, row: \xBC\xE0&\xE\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 33000, row: \xBDD\xD5P\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 34000, row: \xBD\xE0\x92\x12\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 35000, row: \xBE~\xD5\x12\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 36000, row: \xBF9\x06\xBC\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 37000, row: \xBF\xCD\x8A\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 38000, row: \xC0Lw\x00\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 39000, row: \xC0\xDA\xDB\x08\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 40000, row: \xC1S\xF9f\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 41000, row: \xC1\xCF\x82H\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 42000, row: \xC2bJ\xF2\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 43000, row: \xC2\xDC^\xF4\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 44000, row: \xC3Y\xD4.\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 45000, row: \xC3\xEAX\x82\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 46000, row: \xC4c{\xC2\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 47000, row: \xC4\xDEd|\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 48000, row: \xC5i\x8AH\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 49000, row: \xC5\xE8KD\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 50000, row: \xC6aE\x80\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 51000, row: \xC6\xE5\xD2\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 52000, row: \xC7l5\xCA\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
Current count: 53000, row: \xC7\xE4\x19\x14\x95\xA6\x11\xED\xB1\xC3\x0A\x99i\x11\xD0\xB9
53292 row(s) in 3.8410 seconds

=> 53292
```