

## TASK 1

A PDF document (**LoadNoSQL.pdf**) containing the commands to load the transactions history data (card\_transactions.csv) in a NoSQL database.

### **Loading Historical Transactions Data into NoSQL Database**

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

---

```
import happybase

# To create connection
connection = happybase.Connection('localhost', port=9090, autoconnect=False)

# To open connection and perform operations
def open_connection():
    connection.open()

# To close the opened connection
def close_connection():
    connection.close()

# To list all tables in HBase
def list_tables():
    print("fetching all tables")
    open_connection()
    tables = connection.tables()
    close_connection()
```

```
print("all tables fetched")  
return tables
```

#### **# To create the required table**

```
def create_table(name, cf):  
    print("creating table " + name)  
    tables = list_tables()  
    if name in tables:  
        print("table already exists, deleting table " + name)  
        open_connection()  
        connection.disable_table(name)  
        connection.delete_table(name)  
        close_connection()  
        print("table deleted")  
    open_connection()  
    connection.create_table(name, cf)  
    close_connection()  
    print("table created")
```

#### **# To get the created table**

```
def get_table(name):  
    open_connection()  
    table = connection.table(name)  
    close_connection()  
    return table
```

#### **# To batch insert data**

```
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(",")  
        # To skip first row  
        if temp[0] != 'card_id':  
            table.put(bytes(str(i), 'utf-8'), {  
                'info:card_id': bytes(temp[0], 'utf-8'),  
                'info:member_id': bytes(temp[1], 'utf-8'),  
                'info:amount': bytes(temp[2], 'utf-8'),  
                'info:postcode': bytes(temp[3], 'utf-8'),  
                'info:pos_id': bytes(temp[4], 'utf-8'),  
                'info:transaction_dt': bytes(temp[5], 'utf-8'),
```

```
'info:status': bytes(temp[6], 'utf-8')
    })
    i += 1
    file.close()
    print("batch insert done")
    close_connection()
```

#### # To batch insert data of card\_transactions.csv

```
create_table('card_transactions', {'info': dict()})
```

```
batch_insert_data('/home/hadoop/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>**

### Command to list the table : list

```
[root@ip-172-31-5-189 ~]# hbase shell
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/lib/hadoop/lib/slf4j-reload4j-1.7.36.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/lib/hbase/lib/client-facing-thirdparty/slf4j-reload4j-1.7.33.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Reload4jLoggerFactory]
HBase Shell
Use "help" to get list of supported commands.
Use "exit" to quit this interactive shell.
For Reference, please visit: http://hbase.apache.org/2.0/book.html#shell
Version 2.4.17-amzn-6, rUnknown, Sat Jun 1 04:35:21 AM UTC 2024
Took 0.0020 seconds
hbase:001:0> list
TABLE
card_transactions
1 row(s)
Took 0.4245 seconds
=> ["card_transactions"]
hbase:002:0>
```

### Count 'card\_transactions'

```
hbase:002:0> count 'card_transactions'
Current count: 1000, row: 10897
Current count: 2000, row: 11797
Current count: 3000, row: 12697
Current count: 4000, row: 13597
Current count: 5000, row: 14497
Current count: 6000, row: 15397
Current count: 7000, row: 16297
Current count: 8000, row: 17197
Current count: 9000, row: 18097
Current count: 10000, row: 18998
Current count: 11000, row: 19898
Current count: 12000, row: 20797
Current count: 13000, row: 21697
Current count: 14000, row: 22597
Current count: 15000, row: 23497
Current count: 16000, row: 24397
Current count: 17000, row: 25297
Current count: 18000, row: 26197
Current count: 19000, row: 27097
Current count: 20000, row: 27998
Current count: 21000, row: 28898
Current count: 22000, row: 29798
Current count: 23000, row: 30697
Current count: 24000, row: 31597
Current count: 25000, row: 32497
Current count: 26000, row: 33397
```

```
Current count: 24000, row: 31597
Current count: 25000, row: 32497
Current count: 26000, row: 33397
Current count: 27000, row: 34297
Current count: 28000, row: 35197
Current count: 29000, row: 36097
Current count: 30000, row: 36998
Current count: 31000, row: 37898
Current count: 32000, row: 38798
Current count: 33000, row: 39698
Current count: 34000, row: 40597
Current count: 35000, row: 41497
Current count: 36000, row: 42397
Current count: 37000, row: 43297
Current count: 38000, row: 44197
Current count: 39000, row: 45097
Current count: 40000, row: 45998
Current count: 41000, row: 46898
Current count: 42000, row: 47798
Current count: 43000, row: 48698
Current count: 44000, row: 49598
Current count: 45000, row: 50497
Current count: 46000, row: 51397
Current count: 47000, row: 52297
Current count: 48000, row: 53197
Current count: 49000, row: 6134
Current count: 50000, row: 7034
Current count: 51000, row: 7935
Current count: 52000, row: 8835
Current count: 53000, row: 9735
53292 row(s)
Took 4.1537 seconds
=> 53292
hbase:003:0> █
```

**As seen 53292 rows were retrieved**

**Screenshot of the table created**

```

26457 column=info:transaction_dt, timestamp=2024-07-28T16:58:37.960, value=24-11-2016 05:00:46
26458 column=info:amount, timestamp=2024-07-28T16:58:37.961, value=2294690
26458 column=info:card_id, timestamp=2024-07-28T16:58:37.961, value=373392918838641
26458 column=info:member_id, timestamp=2024-07-28T16:58:37.961, value=491669340987836
26458 column=info:pos_id, timestamp=2024-07-28T16:58:37.961, value=631843556129810
26458 column=info:postcode, timestamp=2024-07-28T16:58:37.961, value=12060
26458 column=info:status, timestamp=2024-07-28T16:58:37.961, value=GENUINE
26458 column=info:transaction_dt, timestamp=2024-07-28T16:58:37.961, value=24-11-2016 15:32:46
26459 column=info:amount, timestamp=2024-07-28T16:58:37.961, value=6105548
26459 column=info:card_id, timestamp=2024-07-28T16:58:37.961, value=373392918838641
26459 column=info:member_id, timestamp=2024-07-28T16:58:37.961, value=491669340987836
26459 column=info:pos_id, timestamp=2024-07-28T16:58:37.961, value=074248319790872
26459 column=info:postcode, timestamp=2024-07-28T16:58:37.961, value=30285
26459 column=info:status, timestamp=2024-07-28T16:58:37.961, value=GENUINE
26459 column=info:transaction_dt, timestamp=2024-07-28T16:58:37.961, value=24-12-2017 03:29:16
2646 column=info:amount, timestamp=2024-07-28T16:58:17.535, value=7108213
2646 column=info:card_id, timestamp=2024-07-28T16:58:17.535, value=6445500224944594
2646 column=info:member_id, timestamp=2024-07-28T16:58:17.535, value=052090505482683
2646 column=info:pos_id, timestamp=2024-07-28T16:58:17.535, value=322354280827151
2646 column=info:postcode, timestamp=2024-07-28T16:58:17.535, value=13754
2646 column=info:status, timestamp=2024-07-28T16:58:17.535, value=GENUINE
2646 column=info:transaction_dt, timestamp=2024-07-28T16:58:17.535, value=28-07-2017 10:21:24
26460 column=info:amount, timestamp=2024-07-28T16:58:37.962, value=9538023
26460 column=info:card_id, timestamp=2024-07-28T16:58:37.962, value=373392918838641
26460 column=info:member_id, timestamp=2024-07-28T16:58:37.962, value=491669340987836
26460 column=info:pos_id, timestamp=2024-07-28T16:58:37.962, value=074248319790872
26460 column=info:postcode, timestamp=2024-07-28T16:58:37.962, value=30285
26460 column=info:status, timestamp=2024-07-28T16:58:37.962, value=GENUINE
26460 column=info:transaction_dt, timestamp=2024-07-28T16:58:37.962, value=24-12-2017 14:01:17
26461 column=info:amount, timestamp=2024-07-28T16:58:37.962, value=8641225

```

root@ip-172-31-5-189:~

```

29148 column=info:member_id, timestamp=2024-07-28T16:58:39.822, value=548033330364119
29148 column=info:pos_id, timestamp=2024-07-28T16:58:39.822, value=581036828227484
29148 column=info:postcode, timestamp=2024-07-28T16:58:39.822, value=33856
29148 column=info:status, timestamp=2024-07-28T16:58:39.822, value=GENUINE
29148 column=info:transaction_dt, timestamp=2024-07-28T16:58:39.822, value=11-11-2017 00:00:00
29149 column=info:amount, timestamp=2024-07-28T16:58:39.823, value=3546980
29149 column=info:card_id, timestamp=2024-07-28T16:58:39.823, value=347051350857346
29149 column=info:member_id, timestamp=2024-07-28T16:58:39.823, value=548033330364119
29149 column=info:pos_id, timestamp=2024-07-28T16:58:39.823, value=581036828227484
29149 column=info:postcode, timestamp=2024-07-28T16:58:39.823, value=33856
29149 column=info:status, timestamp=2024-07-28T16:58:39.823, value=GENUINE
29149 column=info:transaction_dt, timestamp=2024-07-28T16:58:39.823, value=11-11-2017 00:00:00
2915 column=info:amount, timestamp=2024-07-28T16:58:18.239, value=5747205
2915 column=info:card_id, timestamp=2024-07-28T16:58:18.239, value=4766513674065931
2915 column=info:member_id, timestamp=2024-07-28T16:58:18.239, value=063787588510857
2915 column=info:pos_id, timestamp=2024-07-28T16:58:18.239, value=725999618356193
2915 column=info:postcode, timestamp=2024-07-28T16:58:18.239, value=18929
2915 column=info:status, timestamp=2024-07-28T16:58:18.239, value=GENUINE
2915 column=info:transaction_dt, timestamp=2024-07-28T16:58:18.239, value=09-01-2018 16:15:11
29150 column=info:amount, timestamp=2024-07-28T16:58:39.824, value=9879461
29150 column=info:card_id, timestamp=2024-07-28T16:58:39.824, value=347051350857346
29150 column=info:member_id, timestamp=2024-07-28T16:58:39.824, value=548033330364119
29150 column=info:pos_id, timestamp=2024-07-28T16:58:39.824, value=581036828227484
29150 column=info:postcode, timestamp=2024-07-28T16:58:39.824, value=33856
29150 column=info:status, timestamp=2024-07-28T16:58:39.824, value=GENUINE
29150 column=info:transaction_dt, timestamp=2024-07-28T16:58:39.824, value=11-11-2017 00:00:00
29151 column=info:amount, timestamp=2024-07-28T16:58:39.824, value=939513
29151 column=info:card_id, timestamp=2024-07-28T16:58:39.824, value=347051350857346
29151 column=info:member_id, timestamp=2024-07-28T16:58:39.824, value=548033330364119
29151 column=info:pos_id, timestamp=2024-07-28T16:58:39.824, value=581036828227484
29151 column=info:postcode, timestamp=2024-07-28T16:58:39.824, value=33856
29151 column=info:status, timestamp=2024-07-28T16:58:39.824, value=GENUINE
29151 column=info:transaction_dt, timestamp=2024-07-28T16:58:39.824, value=11-11-2017 00:00:00
29152 column=info:amount, timestamp=2024-07-28T16:58:39.825, value=155915
29152 column=info:card_id, timestamp=2024-07-28T16:58:39.825, value=347051350857346
29152 column=info:member_id, timestamp=2024-07-28T16:58:39.825, value=548033330364119
29152 column=info:pos_id, timestamp=2024-07-28T16:58:39.825, value=581036828227484
29152 column=info:postcode, timestamp=2024-07-28T16:58:39.825, value=33856
29152 column=info:status, timestamp=2024-07-28T16:58:39.825, value=GENUINE
29152 column=info:transaction_dt, timestamp=2024-07-28T16:58:39.825, value=11-11-2017 00:00:00
29153 column=info:amount, timestamp=2024-07-28T16:58:39.825, value=7968836
29153 column=info:card_id, timestamp=2024-07-28T16:58:39.825, value=347051350857346
29153 column=info:member_id, timestamp=2024-07-28T16:58:39.825, value=548033330364119
29153 column=info:pos_id, timestamp=2024-07-28T16:58:39.825, value=581036828227484
29153 column=info:postcode, timestamp=2024-07-28T16:58:39.825, value=33856

```