

Loading Historical Transactions Data into NoSQL Database

Commands to load the past transactions data into NoSQL database

<Commands to load the data>

```
hadoop fs -mkdir /capstone_project
hadoop fs -chown ec2-user:ec2-user /capstone_project
hadoop fs -mkdir /capstone_project/card_transactions
```

```
hadoop distcp s3://creditcapstone/credit/card\_transactions.csv /capstone_project/
card_transactions/
hadoop distcp s3://creditcapstone/creditzip/uszipsv.csv /capstone_project/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>

<Commands to list table and show its count>

```
CREATE DATABASE capstone_project;
```

```
USE capstone_project;
```

1>> card_transactions_ext

```
CREATE EXTERNAL TABLEIF NOT EXISTS card_transactions_ext( `card_id` string, `
member_id` string, `amount` DOUBLE, `postcode` string, `pos_id` string, `tran
saction_dt` string, `status` string) row format delimited fields terminated B
Y ' ' location '/capstone_project/
card_transactions' tblproperties ("skip.header.line.count"="1");
```

2>> card_transactions_hbase

```
CREATE TABLE card_transactions_hbase
(
    `transaction_id` STRING,
    `card_id` STRING,
    `member_id` STRING,
    `amount` DOUBLE,
    `postcode` STRING,
    `pos_id` STRING,
    `transaction_dt` TIMESTAMP,
    `status` STRING
) row format delimited stored BY
'org.apache.hadoop.hive.hbase.HBaseStorageHandler' WITH serdeproperties (
"hbase.columns.mapping"=
":key, card_transactions_family:card_id, card_transactions_family:member_id,
card_transactions_family:amount, card_transactions_family:postcode, card_tran
s a c t i o n s _ f a m i l y : p o s _ i d ,
card_transactions_family:transaction_dt, card_transactions_family:status")
tblproperties ("hbase.table.name"="card_transactions_hive");
```

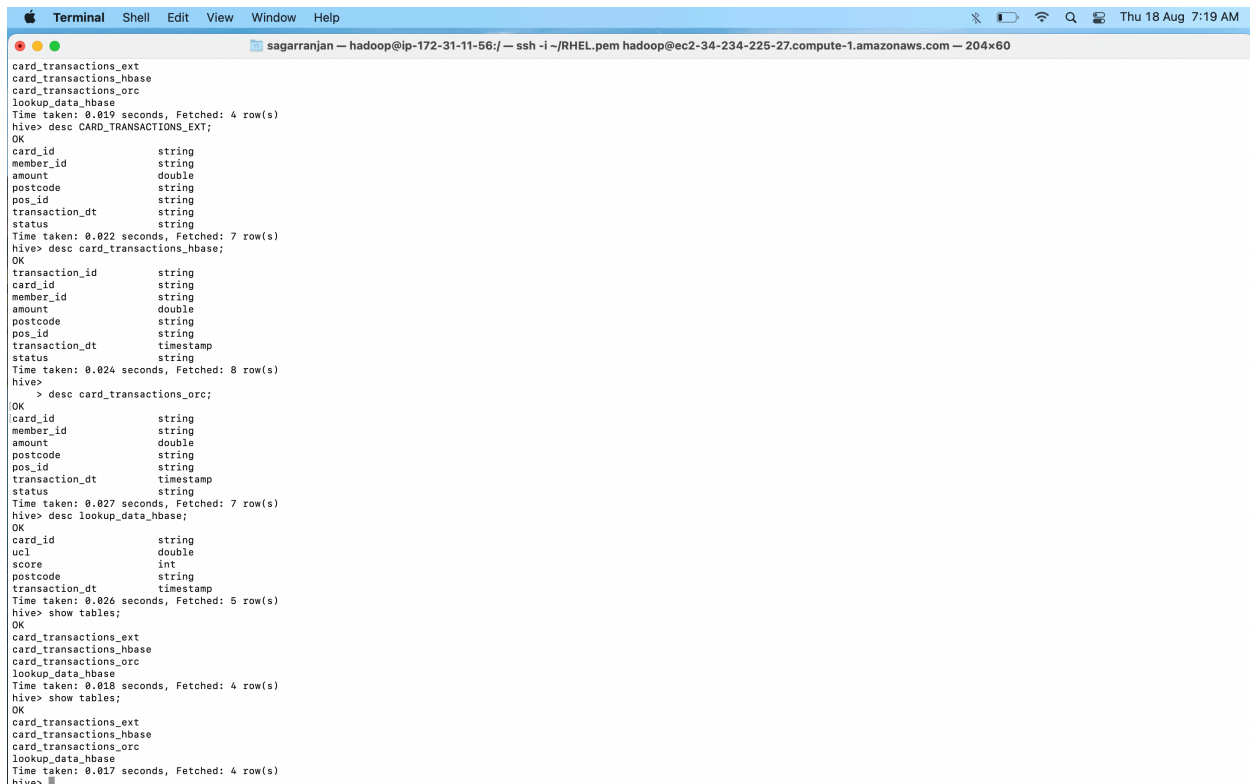
3>> lookup_data_hbase

```
CREATE TABLE lookup_data_hbase
(
    `card_id`          STRING,
    `ucl`              DOUBLE,
    `score`            INT,
    `postcode`         STRING,
    `transaction_dt`   TIMESTAMP
) stored BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler' WITH
serdeproperties ("hbase.columns.mapping"=
":key,                                lookup_card_family:ucl,
lookup_card_family:score, lookup_transaction_family:postcode, lookup_transact
ion_family:transaction_dt"
) tblproperties ("hbase.table.name" = "lookup_data_hive");
```

4>> card_transactions_orc

```
CREATE TABLE IF NOT EXISTS card_transactions_orc
(
    `card_id`          STRING,
    `member_id`        STRING,
    `amount`           DOUBLE,
    `postcode`         STRING,
    `pos_id`           STRING,
    `transaction_dt`   TIMESTAMP,
    `status`           STRING
) stored AS orc tblproperties ("orc.compress"="SNAPPY");
```

5>> List Tables



```
Terminal Shell Edit View Window Help
sagarranjan — hadoop@ip-172-31-11-56:/ — ssh -i ~/RHEL.pem hadoop@ec2-34-234-225-27.compute-1.amazonaws.com — 204x60

card_transactions_ext
card_transactions_hbase
card_transactions_orc
lookup_data_hbase
Time taken: 0.019 seconds, Fetched: 4 row(s)
hive> desc CARD_TRANSACTIONS_EXT;
OK
card_id          string
member_id        string
amount           double
postcode         string
pos_id           string
transaction_dt    string
status           string
Time taken: 0.022 seconds, Fetched: 7 row(s)
hive> desc card_transactions_hbase;
OK
transaction_id    string
card_id           string
member_id         string
amount            double
postcode          string
pos_id            string
transaction_dt     timestamp
status            string
Time taken: 0.024 seconds, Fetched: 8 row(s)
hive>
> desc card_transactions_orc;
OK
card_id          string
member_id        string
amount           double
postcode         string
pos_id           string
transaction_dt    timestamp
status           string
Time taken: 0.027 seconds, Fetched: 7 row(s)
hive> desc lookup_data_hbase;
OK
card_id          string
ucl              double
score            int
postcode         string
transaction_dt    timestamp
Time taken: 0.026 seconds, Fetched: 5 row(s)
hive> show tables;
OK
card_transactions_ext
card_transactions_hbase
card_transactions_orc
lookup_data_hbase
Time taken: 0.018 seconds, Fetched: 4 row(s)
hive> show tables;
OK
card_transactions_ext
card_transactions_hbase
card_transactions_orc
lookup_data_hbase
Time taken: 0.017 seconds, Fetched: 4 row(s)
hive> █
```

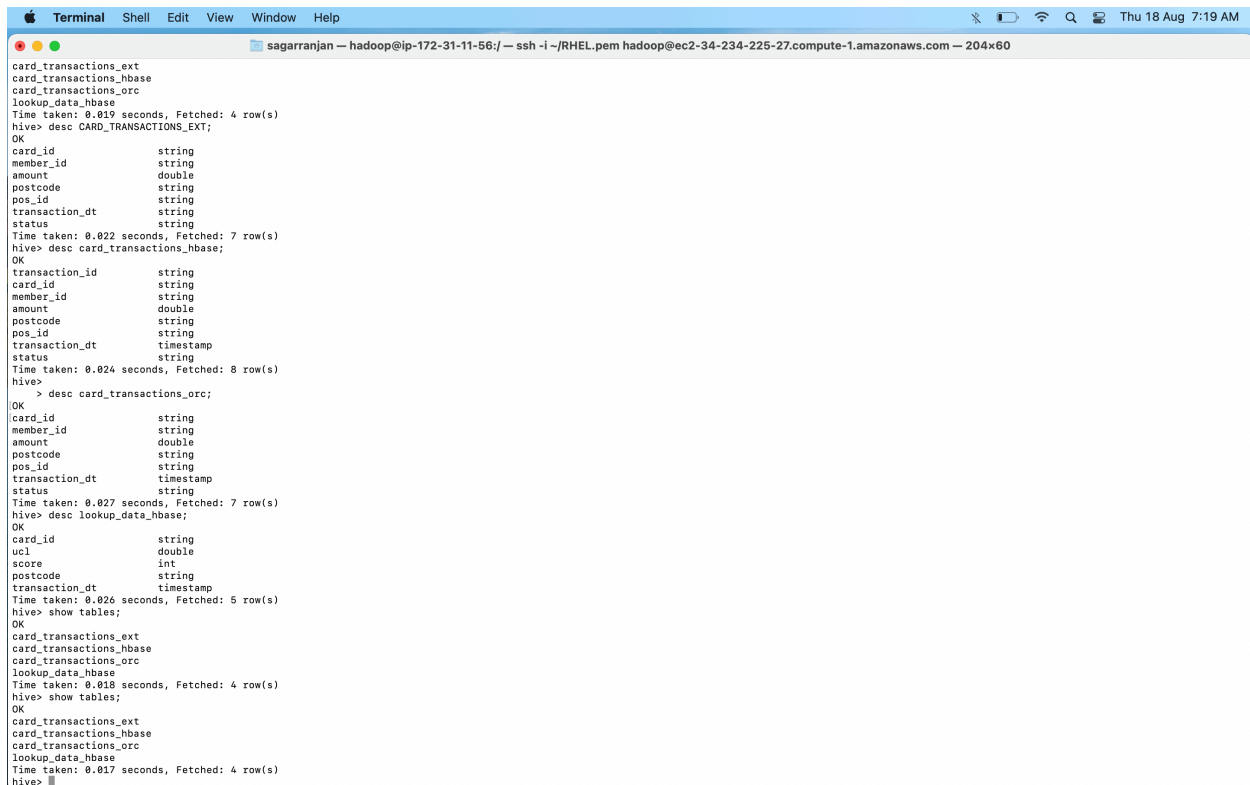
Show TABLES;

6>> Get Count of Rows

```
SELECT Count(*) FROM card_transactions_ext;
SELECT Count(*) FROM card_transactions_hbase;
SELECT Count(*) FROM lookup_data_hbase;
SELECT Count(*) FROM card_transactions_orc;
```

Screenshot of the table created

<Attach the screenshot>



```
Terminal Shell Edit View Window Help
sagarranjan — hadoop@ip-172-31-11-56:/ — ssh -i ~/RHEL.pem hadoop@ec2-34-234-225-27.compute-1.amazonaws.com — 204x60

card_transactions_ext
card_transactions_hbase
card_transactions_orc
lookup_data_hbase
Time taken: 0.019 seconds, Fetched: 4 row(s)
hive> desc CARD_TRANSACTIONS_EXT;
OK
card_id          string
member_id       string
amount          double
postcode        string
pos_id          string
transaction_dt   string
status          string
Time taken: 0.022 seconds, Fetched: 7 row(s)
hive> desc card_transactions_hbase;
OK
transaction_id   string
card_id          string
member_id       string
amount          double
postcode        string
pos_id          string
transaction_dt   timestamp
status          string
Time taken: 0.024 seconds, Fetched: 8 row(s)
hive>
> desc card_transactions_orc;
OK
card_id          string
member_id       string
amount          double
postcode        string
pos_id          string
transaction_dt   timestamp
status          string
Time taken: 0.027 seconds, Fetched: 7 row(s)
hive> desc lookup_data_hbase;
OK
card_id          string
ucl              double
score            int
postcode        string
transaction_dt   timestamp
Time taken: 0.026 seconds, Fetched: 5 row(s)
hive> show tables;
OK
card_transactions_ext
card_transactions_hbase
card_transactions_orc
lookup_data_hbase
Time taken: 0.010 seconds, Fetched: 4 row(s)
hive> show tables;
OK
card_transactions_ext
card_transactions_hbase
card_transactions_orc
lookup_data_hbase
Time taken: 0.017 seconds, Fetched: 4 row(s)
hive> █
```