

Exact Steps to run the Code provided by me

This document comes into place as approach followed to achieve capstone project i.e., credit card fraud detection is not same as guidelines given in step 7.

There's code written in single file and here are steps to follow to execute.

- 1) **Open Putty and connect to ec2-user.**
- 2) **Login as root user.**
- 3) **Give Command Hbase Shell and give command list. Note down table names for card transaction and look up.**
These are required in driver.py to connect, insert and update data.
- 4) **Exit to come back to root user.**
- 5) **Create a directory capstone and place driver.py along with geomap.py & dao.py(in db) and uszipsv.csv**

```
total 1000
-rw-r--r-- 1 root root 1108 Jan 4 14:58 1
drwxr-xr-x 2 root root 38 Jan 4 15:36 db
-rw-r--r-- 1 root root 0 Jan 4 16:49 driver.py
drwxr-xr-x 2 root root 22 Jan 4 15:52 rules
-rw-r--r-- 1 root root 3243 Jan 2 15:04 src.zip
-rw-r--r-- 1 root root 1013345 Jan 2 15:02 uszipsv.csv
```

- 6) Few corrections were required in "DB" files. Dao.py had to be replaced self.host to 'localhost' to make it work. Similarly zipcode csv given in resources list is spelt differently in geomap as uszipsv.csv. Hence make this correction in geomap.py. Lastly, uszipsv.csv only has 5 columns hence, geomap.py has columns from 'A' to 'E'.
Whilst zipcodes csv given in resource list has 6 columns. Hence modifying geomap.py to add 'F' after 'E' to make driver run error free.
- 7) **Ensure thrift server is up now and running.**
- 8) **Give command - python -c 'import happybase'**
This is to make sure happybase is imported already.
- 9) **In root user, give following commands one after other.**
 - a. **wget https://ds-spark-sql-kafka-jar.s3.amazonaws.com/spark-sql-kafka-0-10_2.11-2.3.0.jar**
to create jar file that's required to run the spark code
 - b. **export SPARK_KAFKA_VERSION=0.10**
 - c. **spark2-submit --jars spark-sql-kafka-0-10_2.11-2.3.0.jar --files uszipsv.csv driver.py**
to execute the driver program.
- 10) **Program runs and executes all rules , shall insert rows into look up table when transaction is found genuine. Updates rows to card transaction table with post code, status of transaction & transaction date.**
- 11) **At end, logon to Hbase Shell, list tables and scan card_transactions table to see count of rows after execution.**

```

Current count: 20000, row: 27999
Current count: 21000, row: 28899
Current count: 22000, row: 29799
Current count: 23000, row: 30698
Current count: 24000, row: 31598
Current count: 25000, row: 32498
Current count: 26000, row: 33398
Current count: 27000, row: 341724964458347.210778177559185.12-06-2018152638.2021-01-04171328.398477
Current count: 28000, row: 346618652451637.540752175696215.29-04-2018005259.2021-01-04171400.227023
Current count: 29000, row: 35264
Current count: 30000, row: 36164
Current count: 31000, row: 370582035866789.433646648625434.08-07-2018034337.2021-01-04171349.489639
Current count: 32000, row: 375806375521605.880937166605469.26-05-2018130045.2021-01-04171430.733012
Current count: 33000, row: 38176
Current count: 34000, row: 39076
Current count: 35000, row: 39977
Current count: 36000, row: 40768
Current count: 37000, row: 41560
Current count: 38000, row: 42387
Current count: 39000, row: 4318541450654035.496612742732167.12-02-2018145807.2021-01-04171356.009418
Current count: 40000, row: 43999
Current count: 41000, row: 44784
Current count: 42000, row: 45546
Current count: 43000, row: 46306
Current count: 44000, row: 47134
Current count: 45000, row: 47925
Current count: 46000, row: 48730
Current count: 47000, row: 49500
Current count: 48000, row: 50351
Current count: 49000, row: 5120
Current count: 50000, row: 51888
Current count: 51000, row: 5257502990314019.205172644364018.14-07-2018070014.2021-01-04171327.867742
Current count: 52000, row: 53290
Current count: 53000, row: 5620
Current count: 54000, row: 6211
Current count: 55000, row: 6478888441720966.273246841077378.06-10-2018212851.2021-01-04171333.585477
Current count: 56000, row: 6968
Current count: 57000, row: 7868
Current count: 58000, row: 8768
Current count: 59000, row: 9668
59367 row(s) in 3.8140 seconds

=> 59367
base(main):003:0> █

```

```

Batch: 0
-----+-----+-----+-----+-----+-----+-----+
|card_id|member_id|amount|pos_id|postcode|transaction_dt_ts|status|
+-----+-----+-----+-----+-----+-----+-----+
|348702330256514|37495066290|4380912|248063406800722|96774|2017-12-31 08:24:29|GENUINE|
|348702330256514|37495066290|6703385|786562777140812|84758|2017-12-31 04:15:03|FRAUD|
|348702330256514|37495066290|7454328|466952571393508|93645|2017-12-31 09:56:42|GENUINE|
|348702330256514|37495066290|4013428|45845320330319|15868|2017-12-31 05:38:54|GENUINE|
|348702330256514|37495066290|5495353|545499621965697|79033|2017-12-31 21:51:54|GENUINE|
|348702330256514|37495066290|3966214|369266342272501|22832|2017-12-31 03:52:51|GENUINE|
|348702330256514|37495066290|1753644|9475029292671|17923|2017-12-31 00:11:30|FRAUD|
|348702330256514|37495066290|1692115|27647525195860|55708|2017-12-31 17:02:39|GENUINE|
|5189563368503974|117826301530|9222134|525701337355194|64002|2017-12-31 20:22:10|GENUINE|
|5189563368503974|117826301530|4133848|182031383443115|26346|2017-12-31 01:52:32|FRAUD|
|5189563368503974|117826301530|8938921|799748246411019|76934|2017-12-31 05:20:53|FRAUD|
|5189563368503974|117826301530|1786366|131276818071265|63431|2017-12-31 14:29:38|GENUINE|
|5189563368503974|117826301530|9142237|564240259678903|50635|2017-12-31 19:37:19|GENUINE|
|5407073344486464|1147922084344|6885448|887913906711117|59031|2017-12-31 07:53:53|FRAUD|
|5407073344486464|1147922084344|4028209|116266051118182|80118|2017-12-31 01:06:50|FRAUD|
|5407073344486464|1147922084344|3858369|896105817613325|53820|2017-12-31 17:37:26|GENUINE|
|5407073344486464|1147922084344|9307733|729374116016479|14898|2017-12-31 04:50:16|FRAUD|
|5407073344486464|1147922084344|4011296|543373367319647|44028|2017-12-31 13:09:34|GENUINE|
|5407073344486464|1147922084344|9492531|211980095659371|49453|2017-12-31 14:12:26|GENUINE|
|5407073344486464|1147922084344|7550074|345533088112099|15030|2017-12-31 02:34:52|FRAUD|
+-----+-----+-----+-----+-----+-----+-----+
only showing top 20 rows

```