Creating Lookup Table

Command to create the Lookup Table

Below is a python file which I used that flavors batch insertion of CSV into hbase table.

I created a dataframe in pyspark which acts as CSV here and use happybase .py file to load data into NoSQL table.

Dataframe Name: look_up_table
Hbase Table: look_up_table

```
import happybase

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

def open_connection():
    connection.open()
def close_connection():
    connection.close()
def list_tables():
 print "fetching all table"
 open_connection()
 tables = connection.tables()
 close_connection()
 print "all tables fetched"
 return tables
```

#check if table already exist and create the 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"
```

 # While loading DF into table, this function gets the create table name

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

#Actual command to create lookup table

```
create_table('look_up_table', {'info' : dict(max_versions=5) })
```

#Function to load data from DF

```
def batch_insert_data(df,tableName):
 print "starting batch insert of events"
 table = get_table(tableName)
 open_connection()
 rows_count=0

rowKey_dict={}
 with table.batch(batch_size=4) as b:
   for row in df.rdd.collect():
    b.put(bytes(row.card_id) , { 'info:card_id':bytes(row.card_id),
                'info:transaction_date':bytes(row.transaction_date),
                'info:score':bytes(row.score),
                'info:postcode':bytes(row.postcode),
                'info:UCL':bytes(row.UCL)})

 print "batch insert done"
 close_connection()
```

#Call function to insert data into Hbase table

```
batch_insert_data(look_up_table,'look_up_table')
```

<Command to see the table created>

1) Login to putty as root user
2) Put up thrift server
3) Give command Hbase shell
4) Give command "list"

Screenshot of the created table

```
hbase(main):001:0> list
TABLE
card_transactions
employee
look_up_table
3 row(s) in 0.3340 seconds

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

Give command scan 'look_up_table' to see data inserted into table.

```
5231456036333304                    column=info:transaction_date, timestamp=1607880087970, value=2018-01-22 00:56:57
5232083808576685                    column=info:UCL, timestamp=1607880086427, value=14120434.4
5232083808576685                    column=info:card_id, timestamp=1607880086427, value=5232083808576685
5232083808576685                    column=info:postcode, timestamp=1607880086427, value=17965
5232083808576685                    column=info:score, timestamp=1607880086427, value=566
5232083808576685                    column=info:transaction_date, timestamp=1607880086427, value=2018-01-09 12:44:31
5232271306465150                    column=info:UCL, timestamp=1607880087122, value=10951781.35
5232271306465150                    column=info:card_id, timestamp=1607880087122, value=5232271306465150
5232271306465150                    column=info:postcode, timestamp=1607880087122, value=12920
5232271306465150                    column=info:score, timestamp=1607880087122, value=638
5232271306465150                    column=info:transaction_date, timestamp=1607880087122, value=2018-01-22 16:44:59
5232695950818720                    column=info:UCL, timestamp=1607880087849, value=15220850.52
5232695950818720                    column=info:card_id, timestamp=1607880087849, value=5232695950818720
5232695950818720                    column=info:postcode, timestamp=1607880087849, value=79080
5232695950818720                    column=info:score, timestamp=1607880087849, value=207
5232695950818720                    column=info:transaction_date, timestamp=1607880087849, value=2018-01-29 08:30:32
5239380866598772                    column=info:UCL, timestamp=1607880086358, value=12835247.22
5239380866598772                    column=info:card_id, timestamp=1607880086358, value=5239380866598772
5239380866598772                    column=info:postcode, timestamp=1607880086358, value=72471
5239380866598772                    column=info:score, timestamp=1607880086358, value=440
5239380866598772                    column=info:transaction_date, timestamp=1607880086358, value=2017-12-07 21:44:43
5242841712000086                    column=info:UCL, timestamp=1607880088013, value=15646358.41
5242841712000086                    column=info:card_id, timestamp=1607880088013, value=5242841712000086
5242841712000086                    column=info:postcode, timestamp=1607880088013, value=48821
5242841712000086                    column=info:score, timestamp=1607880088013, value=236
5242841712000086                    column=info:transaction_date, timestamp=1607880088013, value=2018-01-27 10:51:48
5249623960609831                    column=info:UCL, timestamp=1607880087191, value=12497504.76
5249623960609831                    column=info:card_id, timestamp=1607880087191, value=5249623960609831
5249623960609831                    column=info:postcode, timestamp=1607880087191, value=16858
5249623960609831                    column=info:score, timestamp=1607880087191, value=265
5249623960609831                    column=info:transaction_date, timestamp=1607880087191, value=2018-01-28 00:54:29
5252551880815473                    column=info:UCL, timestamp=1607880086480, value=11540779.75
5252551880815473                    column=info:card_id, timestamp=1607880086480, value=5252551880815473
5252551880815473                    column=info:postcode, timestamp=1607880086480, value=39352
5252551880815473                    column=info:score, timestamp=1607880086480, value=449
5252551880815473                    column=info:transaction_date, timestamp=1607880086480, value=2018-02-01 10:14:39
5253084214148600                    column=info:UCL, timestamp=1607880087349, value=13198338.6
5253084214148600                    column=info:card_id, timestamp=1607880087349, value=5253084214148600
5253084214148600                    column=info:postcode, timestamp=1607880087349, value=78054
5253084214148600                    column=info:score, timestamp=1607880087349, value=512
5253084214148600                    column=info:transaction_date, timestamp=1607880087349, value=2018-01-27 10:51:49
5254025009868430                    column=info:UCL, timestamp=1607880087698, value=14556419.87
5254025009868430                    column=info:card_id, timestamp=1607880087698, value=5254025009868430
5254025009868430                    column=info:postcode, timestamp=1607880087698, value=12973
```

```
root@ip-10-0-0-0:~
6591175617713393                    column=info:transaction_date, timestamp=1607880087142, value=2018-01-31 13:10:37
6592184145413632                    column=info:UCL, timestamp=1607880086730, value=13734342.65
6592184145413632                    column=info:card_id, timestamp=1607880086730, value=6592184145413632
6592184145413632                    column=info:postcode, timestamp=1607880086730, value=53186
6592184145413632                    column=info:score, timestamp=1607880086730, value=456
6592184145413632                    column=info:transaction_date, timestamp=1607880086730, value=2018-01-28 00:54:30
6594248319343442                    column=info:UCL, timestamp=1607880086800, value=15065362.77
6594248319343442                    column=info:card_id, timestamp=1607880086800, value=6594248319343442
6594248319343442                    column=info:postcode, timestamp=1607880086800, value=24927
6594248319343442                    column=info:score, timestamp=1607880086800, value=350
6594248319343442                    column=info:transaction_date, timestamp=1607880086800, value=2018-01-31 23:42:38
6595638658736751                    column=info:UCL, timestamp=1607880087351, value=14005069.97
6595638658736751                    column=info:card_id, timestamp=1607880087351, value=6595638658736751
6595638658736751                    column=info:postcode, timestamp=1607880087351, value=68328
6595638658736751                    column=info:score, timestamp=1607880087351, value=310
6595638658736751                    column=info:transaction_date, timestamp=1607880087351, value=2018-01-30 10:50:34
6595814135833988                    column=info:UCL, timestamp=1607880087066, value=14332708.84
6595814135833988                    column=info:card_id, timestamp=1607880087066, value=6595814135833988
6595814135833988                    column=info:postcode, timestamp=1607880087066, value=22508
6595814135833988                    column=info:score, timestamp=1607880087066, value=210
6595814135833988                    column=info:transaction_date, timestamp=1607880087066, value=2018-01-30 02:03:54
6595928469079750                    column=info:UCL, timestamp=1607880087956, value=11824730.01
6595928469079750                    column=info:card_id, timestamp=1607880087956, value=6595928469079750
6595928469079750                    column=info:postcode, timestamp=1607880087956, value=98349
6595928469079750                    column=info:score, timestamp=1607880087956, value=412
6595928469079750                    column=info:transaction_date, timestamp=1607880087956, value=2018-01-24 12:38:22
6597703848279563                    column=info:UCL, timestamp=1607880087391, value=15250624.49
6597703848279563                    column=info:card_id, timestamp=1607880087391, value=6597703848279563
6597703848279563                    column=info:postcode, timestamp=1607880087391, value=95699
6597703848279563                    column=info:score, timestamp=1607880087391, value=218
6597703848279563                    column=info:transaction_date, timestamp=1607880087391, value=2018-01-27 10:51:49
6598830758632447                    column=info:UCL, timestamp=1607880087564, value=12685782.48
6598830758632447                    column=info:card_id, timestamp=1607880087564, value=6598830758632447
6598830758632447                    column=info:postcode, timestamp=1607880087564, value=19421
6598830758632447                    column=info:score, timestamp=1607880087564, value=293
6598830758632447                    column=info:transaction_date, timestamp=1607880087564, value=2018-01-30 00:18:34
6599900931314251                    column=info:UCL, timestamp=1607880087928, value=12487392.07
6599900931314251                    column=info:card_id, timestamp=1607880087928, value=6599900931314251
6599900931314251                    column=info:postcode, timestamp=1607880087928, value=97423
6599900931314251                    column=info:score, timestamp=1607880087928, value=297
6599900931314251                    column=info:transaction_date, timestamp=1607880087928, value=2018-01-31 11:25:16
999 row(s) in 2.5910 seconds
```