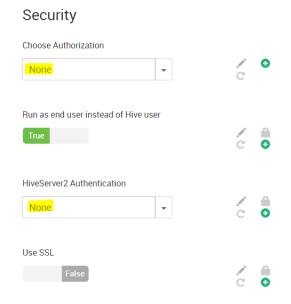
TAN CHANG JUNG

Milestone 3 – Accessing hive data warehouse using Python

1. Set the hive authorization to be PLAIN, which is 'None' in Hortonworks Sandbox.



2. Open 'ODBC Data Source Administration' in Windows, setup the ODBC connection.

The default port for Hive is 10000.

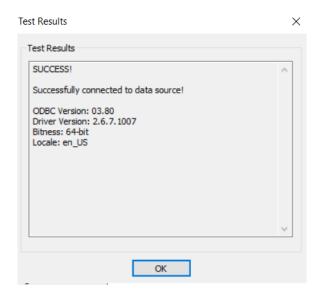
The default credentials for connection is:

User: hive

Password: hive



Then click 'Test' for testing the connection, make sure it shows Success!



TAN CHANG JUNG

3. In Python IDE, use library 'pyodbc' to connect Hive and access Hive database

```
In [6]: import pyodbc
In [7]: import pandas as pd
In [8]: conn = pyodbc.connect(DSN = "hive_connection", autocommit = True, ansi = True)
In [9]: conn
Out[9]: <pyodbc.Connection at 0x150bb90b9f0>
In [10]: db = pd.read_sql("show databases;", conn)
In [11]: print(db)
       database name
             default
0
1
            foodmart
2 information_schema
3
In [12]: bitcoin_table = pd.read_sql("SELECT * FROM bitcoin LIMIT 5", conn)
In [13]: bitcoin table
Out[13]:
  bitcoin.marketdate bitcoin.open ... bitcoin.volume bitcoin.marketcapacity
          2020-05-25
                           8786.11 ...
                                            3.128816e+10
                                                                    1.637605e+11
1
          2020-05-24
                           9212.28 ...
                                            3.251880e+10
                                                                    1.616104e+11
2
          2020-05-23
                           9185.06 ...
                                           2.772787e+10
                                                                    1.693055e+11
3
          2020-05-22
                           9080.33 ... 2.981077e+10
                                                                    1.688076e+11
          2020-05-21
                           9522.74 ...
                                           3.932616e+10
                                                                    1.669480e+11
[5 rows x 7 columns]
```