

### Milestone 3 – Accessing hive data warehouse using Python

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

#### Security

Choose Authorization



Run as end user instead of Hive user

☒

HiveServer2 Authentication



Use SSL

☐

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**

## TAN CHANG JUNG

Microsoft Hive ODBC Driver DSN Setup

Data Source Name:

Description:

Host(s):

Port:

Database:

Authentication

Mechanism:

Realm:

Host FQDN:

Service Name:

☒ Canonicalize Principal FQDN  
☐ Delegate Kerberos Credentials

User Name:

Password:

Delegation UID:

Thrift Transport:

v2.6.7.1007 (64 bit)

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

Test Results

Test Results

SUCCESS!

Successfully connected to data source!

ODBC Version: 03.80  
Driver Version: 2.6.7.1007  
Bitness: 64-bit  
Locale: en\_US

## 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", autoccommit = 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
0              default
1             foodmart
2  information_schema
3                  sys
```

```
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
0	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
4	2020-05-21	9522.74	...	3.932616e+10	1.669480e+11

```
[5 rows x 7 columns]
```