Name: Mohd Amirul Shafiq Bin Shafiee
Matric No: WQD180114
Milestone 3: **Processing of data - Accessing hive data warehouse using Python**

Part 1: Configure HiveServer2

- Add proxy user in coresite.xml in hadoop as follows:

```
<property>
        <name>hadoop.proxyuser.amirul.groups</name>
        <value>*</value>
</property>
<property>
        <name>hadoop.proxyuser.amirul.hosts</name>
        <value>*</value>
</property>
```

- Set the hiveserver2 at localhost in hive-site.xml as follows:

```
<property>
        <name>hive.server2.thrift.bind.host</name>
        <value>localhost</value>
</property>
<property>
        <name>hive.server2.thrift.port</name>
        <value>10000</value>
</property>
```

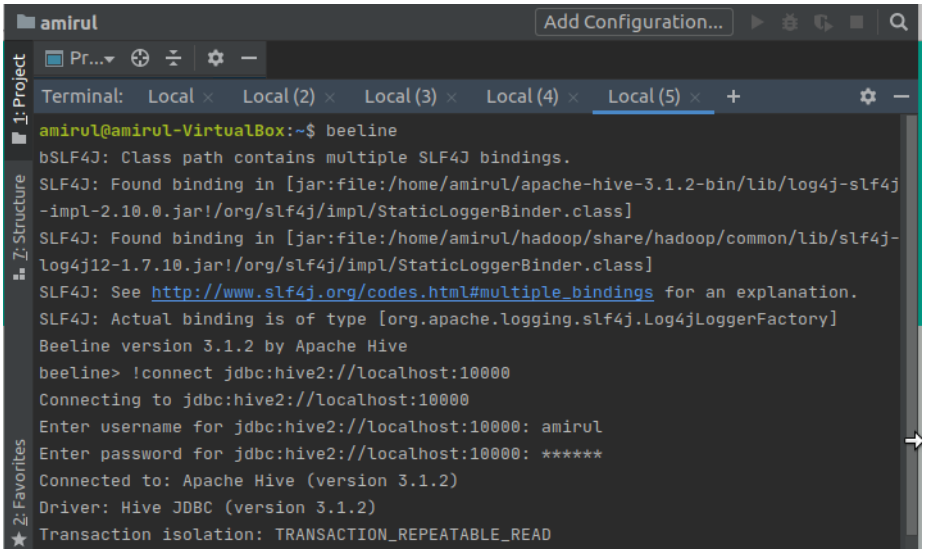
- hadoop fs -chmod 777 /tmp

- Run in terminal: `hiveserver2 &`

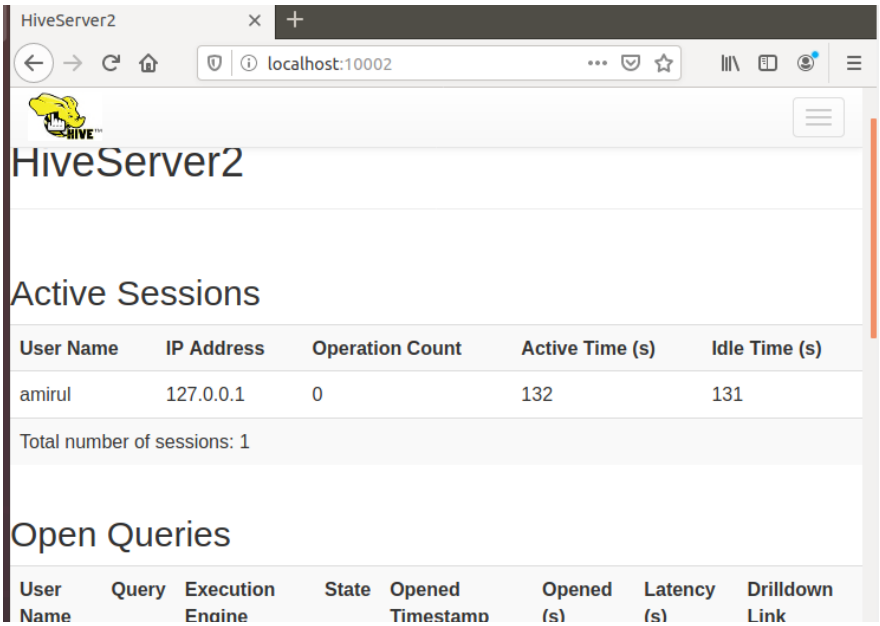
- Test connection to hiveserver2 via beeline:

```
beeline

beeline>!connect jdbc:hive2://localhost:10000
```



- Open http://localhost:10002



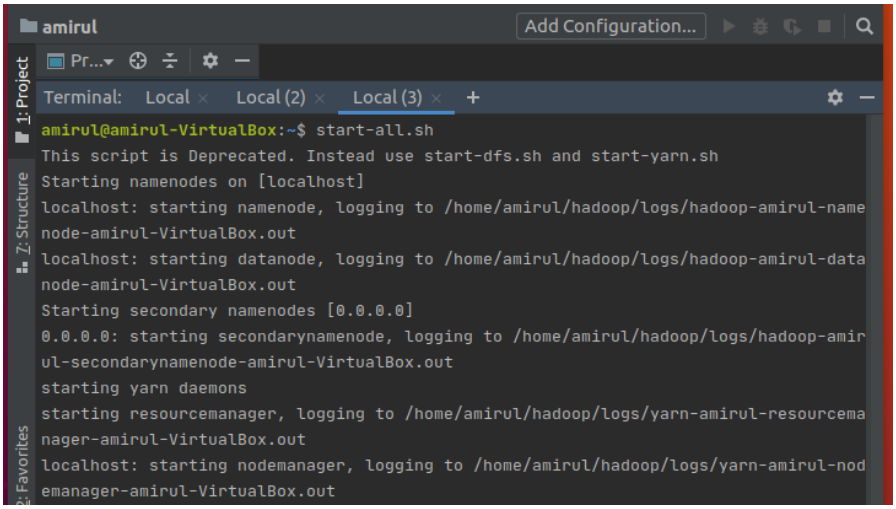
Part 2: Install PyHive and its components

- gcc: sudo apt-get install gcc
- Python developer package : sudo apt-get python3.8-devel.x86_64
- libsasl2-dev: sudo apt-get install libsasl2-dev
- sasl: pip3 install sasl
- thrift sasl : pip3 install thrift_sasl

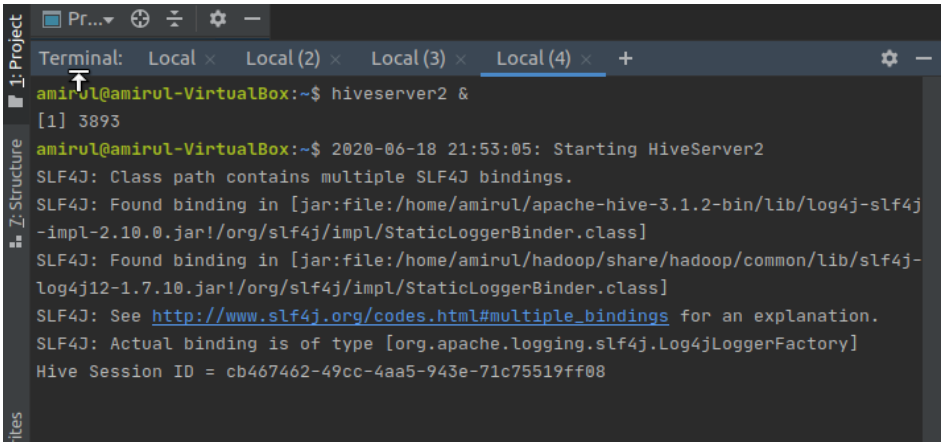
Part 3: Access Hive through Python

<u>At command prompt</u>

- Start HDFS

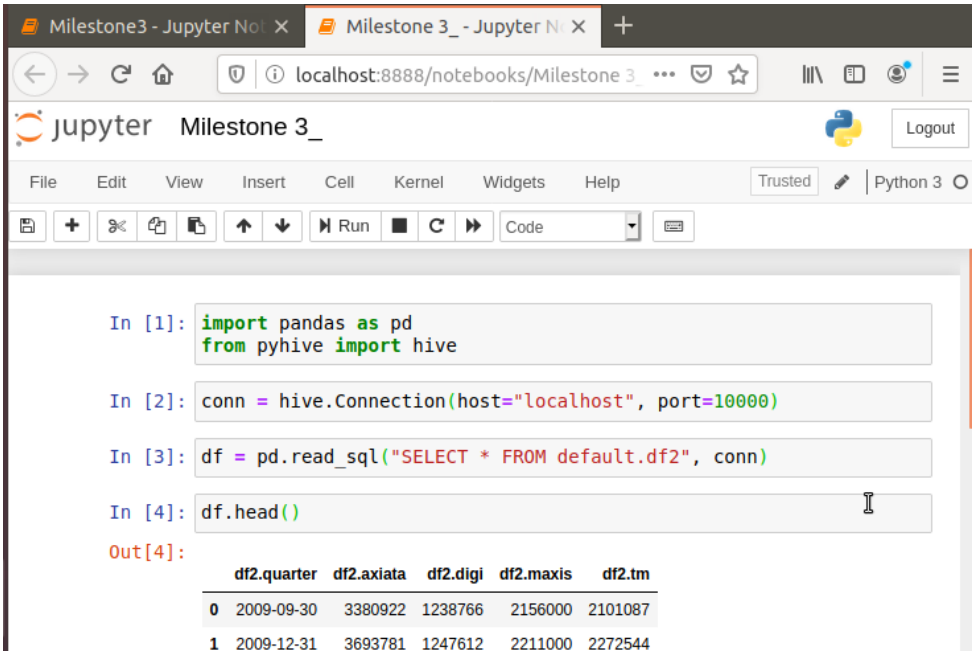


- Initiate hiveserver2 &



<u>At Jupyter Notebook</u>

- Import pyhive
- Define hive connection
- Read data file (df2) and make sure it is correct



Milestone3 - Jupyter Not | Milestone 3_ - Jupyter N

localhost:8888/notebooks/Milestone 3_

jupyter Milestone 3_

Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

In [5]:
```
df.describe()
```

Out[5]:

	df2.axiata	df2.digi	df2.maxis	df2.tm
count	4.300000e+01	4.300000e+01	4.300000e+01	4.300000e+01
mean	4.982791e+06	1.586184e+06	2.214674e+06	2.694206e+06
std	8.641711e+05	1.388962e+05	9.578068e+04	3.345451e+05
min	3.380922e+06	1.238766e+06	2.065000e+06	2.101087e+06
25%	4.344400e+06	1.547072e+06	2.154000e+06	2.404248e+06
50%	4.747318e+06	1.618345e+06	2.214000e+06	2.778922e+06
75%	5.874018e+06	1.674696e+06	2.245000e+06	2.943220e+06
max	6.267007e+06	1.798623e+06	2.590000e+06	3.237032e+06