python操作hive

1. 安装pyhs2

pyhs2依赖项如下：

gcc-c++

python-devel.x86_64

cyrus-sasl-devel.x86_64

因此pyhs2的安装命令如下：

yum install gcc-c++ python-devel.x86_64 cyrus-sasl-devel.x86_64

安装pip

$ sudo yum install python-pip

pip install pyhs2

此外，可能还需要安装cyrus-sasl-plain：

yum install cyrus-sasl-plain

2. pyhs2示例

```
import pyhs2

with pyhs2.connect(host='192.168.34.130',
port=10000,
authMechanism="PLAIN",
user='hdfs',
password='',
database='default') as conn:
with conn.cursor() as cur:
# Show databases
print cur.getDatabases()
```

```
# Execute query
cur.execute("select * from tags")
# Return column info from query
print cur.getSchema()
# Fetch table results
for i in cur.fetch():
print i
```

参考：

遇到的问题：

按照：

window下执行失败，

报 "thrift.transport.TTransport.TTransportException: Could not start SASL: Error in sasl_client_start (-4) SASL(-4): no mechanism available: Unable to find a callback: 2"

错误

Python通过pyhs2操作Hive

1. 安装pyhs2

1.1 pip安装

1	yum install gcc-c++ cyrus-sasl-devel
2	pip install pyhs2

如果pip没有安装去Github上找到pyhs2和sasl的项目，已安装成功则不用考虑下面。

1.2 pyhs2安装

1	unzip pyhs2-master.zip
2	cd pyhs2-master
3	python setup.py install

1.3 sasl安装

1	unzip python-sasl-master.zip
2	cd python-sasl-master
3	python setup.py install

2. 测试Hive连接

1	# -*- coding: utf-8 -*-
2	# require hive server2
3	# ---
4	# Author: Lich_Amnesia <alwaysxiaop@gmail.com>
5	# Date: 2016-03-09
6	# ---

```python
7    import pyhs2
8    class HiveClient:
9    # create connection to hive server2
10   def __init__(self, db_host, user, password, database, port=10000,
11   authMechanism="PLAIN"):
12   self.conn = pyhs2.connect(host=db_host,

13   port=port,
14   authMechanism=authMechanism,
15   user=user,
16   password=password,
17   database=database,
18   )
19   def query(self, sql):
20   with self.conn.cursor() as cursor:
21   cursor.execute(sql)
22   return cursor.fetch()
23   def close(self):
24   self.conn.close()
25   def main():
26   # 注意authMechanism初始的时候是PLAIN，最好的解决方案是加上认证模
27   块
28   hive_client = HiveClient(db_host='127.0.0.1', port=10000, user='root',
29   password='root123456', database='default', authMechanism='PLAIN')
30   # 测试select语句
31   result = hive_client.query('select * from test limit 10')
32   print result
33   hive_client.close()
34   if __name__ == '__main__':
35   main()
36
37
38
39
40
```