Python for HBase Thrift API

# 1.安装thrift环境及python lib for hbase

安装thrift

*$wget* [*http://archive.apache.org/dist/thrift/0.9.0/thrift-0.9.0.tar.gz*](http://archive.apache.org/dist/thrift/0.9.0/thrift-0.9.0.tar.gz)

*$tar –zxvf thrift-0.9.0.tar.gz*

*$./configure --with-boost=/usr/local*

*$make*

*$make install*

安装thrift的pyhon库

*$cd thrift-0.9.0/lib/py*

*$python setup.py install*

生成hbase的连接python lib （gen-py）

*$cd hbase-thrift/src/main/resources/org/apache/hadoop/hbase/thrift/gen-py/hbase*

*$ thrift -gen py hbase.thrift*

在当前目录会生成gen-py目录，如下所示：

*[root@sgserver002 gen-py]# ll*

*hbase*

*\_\_init\_\_.py*

*[root@sgserver002 hbase]# ll*

*constants.py*

*Hbase.py*

*Hbase-remote*

*\_\_init\_\_.py*

*ttypes.py*

将生成的hbase的python库导入到Python中

*$cp –r hbase /usr/lib/python2.6/site-packages*

启动hbase thrift2 server

*$hbase-daemon.sh start thrift2*

测试python连接hbase thrift

*[root@sgserver002 python2.6]# python*

*>>> from thrift.transport import TSocket*

*>>> from thrift.protocol import TBinaryProtocol*

*>>> from hbase import Hbase*

*>>> transport = TSocket.TSocket('sgserver002',9090)*

*>>> protocol = TBinaryProtocol.TBinaryProtocol(transport)*

*>>> client = Hbase.Client(protocol)*

*>>> transport.open()*

*>>> client.getTableNames()*

*['Blog', 'HBaseSamples',, 'test\_info']*

# 2. HBase Thrift2 Example

Thrift2 Example如下所示

*#!/usr/bin/python*

*import sys*

*import time*

*import os*

*from thrift.transport import TTransport*

*from thrift.transport import TSocket*

*from thrift.protocol import TBinaryProtocol*

*from hbase import THBaseService*

*from hbase.ttypes import \**

*print "Thrift2 Demo"*

*print "This demo assumes you hava a table called example with a column family info"*

*host= "sgserver002"*

*port = 9090*

*framed = False*

*socket = TSocket.TSocket(host,port)*

*if framed:*

*transport = TTransport.TFramedTransport(socket)*

*else:*

*transport = TTransport.TBufferedTransport(socket)*

*protocol = TBinaryProtocol.TBinaryProtocol(transport)*

*client = THBaseService.Client(protocol)*

*transport.open()*

*table = "example"*

*put = TPut(row="row2",columnValues=[TColumnValue(family="info",qualifier="q1",value="value2")])*

*print "Putting:",put*

执行命令：

*[hbase@sgserver002 thrift2]$ python Thrift2DemoClient.py*

*Thrift2 Demo*

*This demo assumes you hava a table called example with a column family info*

*Putting: TPut(durability=None, timestamp=None, cellVisibility=None, attributes=None, columnValues=[TColumnValue(value='value2', tags=None, qualifier='q1', family='info', timestamp=None)], row='row2')*

*Getting: TGet(authorizations=None, filterString=None, timestamp=None, maxVersions=None, timeRange=None, attributes=None, columns=None, row='row1')*

*Result: TResult(columnValues=[TColumnValue(value='value1', tags=None, qualifier='qa1', family='info', timestamp=1485029801604)], row='row1')*

执行成功

# 3.开启BCID认证

具体步骤见confluence

http://223.105.0.132:8090/pages/viewpage.action?pageId=24975864

将集群外的一台主机作为客户端，将hadoop和hbase的安装目录复制到客户端主机中，测试如下：

*[hdfs@fys1 hadoop]$ bin/hdfs dfs -ls /*

*ls: Authentication failed! Access id&key should not be empty!*

*[hdfs@fys1 hadoop]$ bin/hdfs dfs -D hadoop.security.bdoc.access.id=admin -D hadoop.security.bdoc.access.key=admin -ls /*

*Found 13 items*

*drwxr-xr-x - hdfs hdfs 0 2016-11-04 05:26 /amshbase*

*drwxrwxrwx - yarn hadoop 0 2017-01-13 04:18 /app-logs*

*drwxr-xr-x - hdfs hdfs 0 2016-12-19 02:52 /apps*

*drwxr-xr-x - hdfs hdfs 0 2016-11-04 03:37 /iothrottle*

*drwxr-xr-x - mapred hdfs 0 2016-11-04 03:36 /mapred*

*drwxrwxrwx - mapred hadoop 0 2016-11-04 03:36 /mr-history*

*drwxrwxrwx - slider hdfs 0 2016-11-04 03:37 /slider*

*drwxrwxrwx - hdfs hdfs 0 2017-01-13 04:18 /user*

**hadoop开启bcid没有问题**

测试hbase使用如下：

*[hdfs@fys1 hbase]$ bin/hbase shell*

*Version 1.1.3-bc1.3.2, re30b5add732ec03353cc4368684123ac2e462b7b, Thu Sep 29 09:55:50 CST 2016*

*hbase(main):001:0> list*

*TABLE*

*ERROR: Connection to sgserver006/10.133.46.143:60000 is closing. Call id=6, waitTime=4*

连接失败，关闭bcid认证后的结果如下：

*hbase(main):001:0> list*

*TABLE*

*Blog*

*HBaseSamples*

*test\_info*

*14 row(s) in 0.2680 seconds*

以上说明bcid开启认证后，不指定bcid是无法访问hbase

# 4.开启bcid后，测试Thrift2 Example

命令如下：

*$ python Thrift2DemoClient.py*

执行结果如下：

*[hdfs@fys1 thrift2]$ python Thrift2DemoClient.py*

*Thrift2 Demo*

*This demo assumes you hava a table called example with a column family info*

*Putting: TPut(durability=None, timestamp=None, cellVisibility=None, attributes=None, columnValues=[TColumnValue(value='value3', tags=None, qualifier='q1', family='info', timestamp=None)], row='row3')*

*Getting: TGet(authorizations=None, filterString=None, timestamp=None, maxVersions=None, timeRange=None, attributes=None, columns=None, row='row3')*

*Result: TResult(columnValues=[TColumnValue(value='value3', tags=None, qualifier='q1', family='info', timestamp=1485050143468)], row='row3')*

程序正常执行， Thrift2-Server没有添加BCID认证

# 参考文献：

安装thrift

http://thrift.apache.org/docs/install/centos

thrift连接hbase的例子

https://wiki.apache.org/hadoop/Hbase/ThriftApi

https://theatr.us/2008/07/19/using-facebook-thrift-with-python-and-hbase/

Thrift API

http://svn.apache.org/viewvc/hbase/trunk/hbase-thrift/src/main/resources/org/apache/hadoop/hbase/thrift/Hbase.thrift?view=markup

# 遇到问题：

问题1：

Thrift 编译时出现的异常如下：

checking for SSL\_ctrl in -lssl... no

configure: error: "Error: libssl required"

解决方法：

export LIBS='-lcrypto -ldl'