ACADGILD

Session 11: Sqoop Flume

Assignment 3

Problem Statement

Create a flume agent that streams data from Twitter and stores in the HDFS.

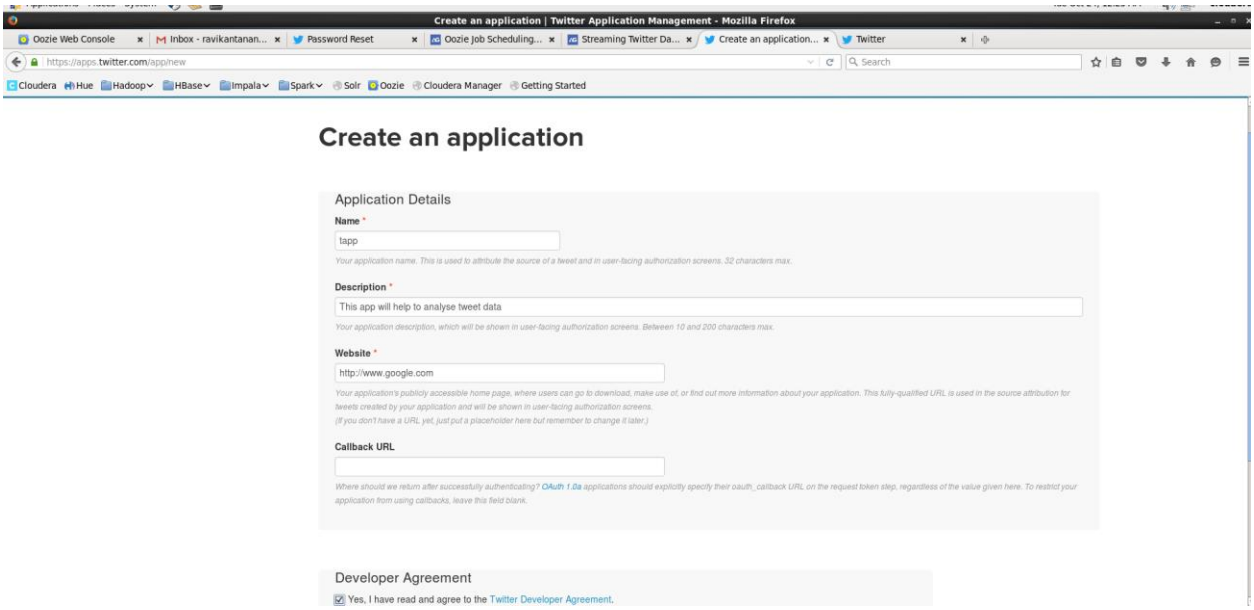
Step 1

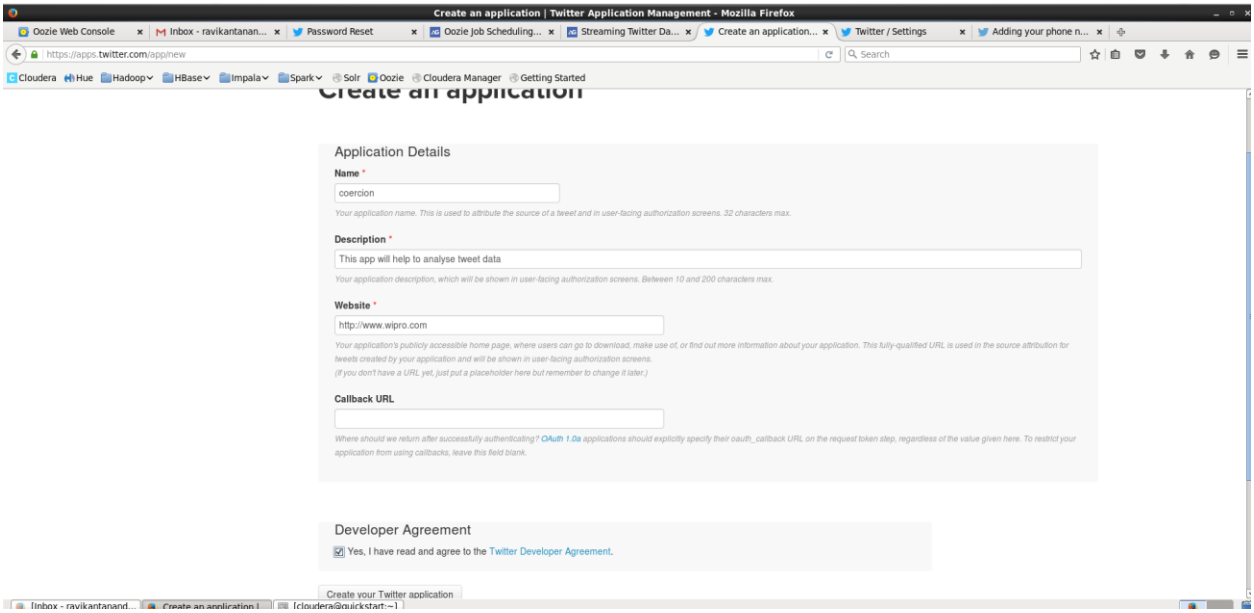
Login to given link https://apps.twitter.com/ and sign in.

Step 2

Click on the **Create New App** button.

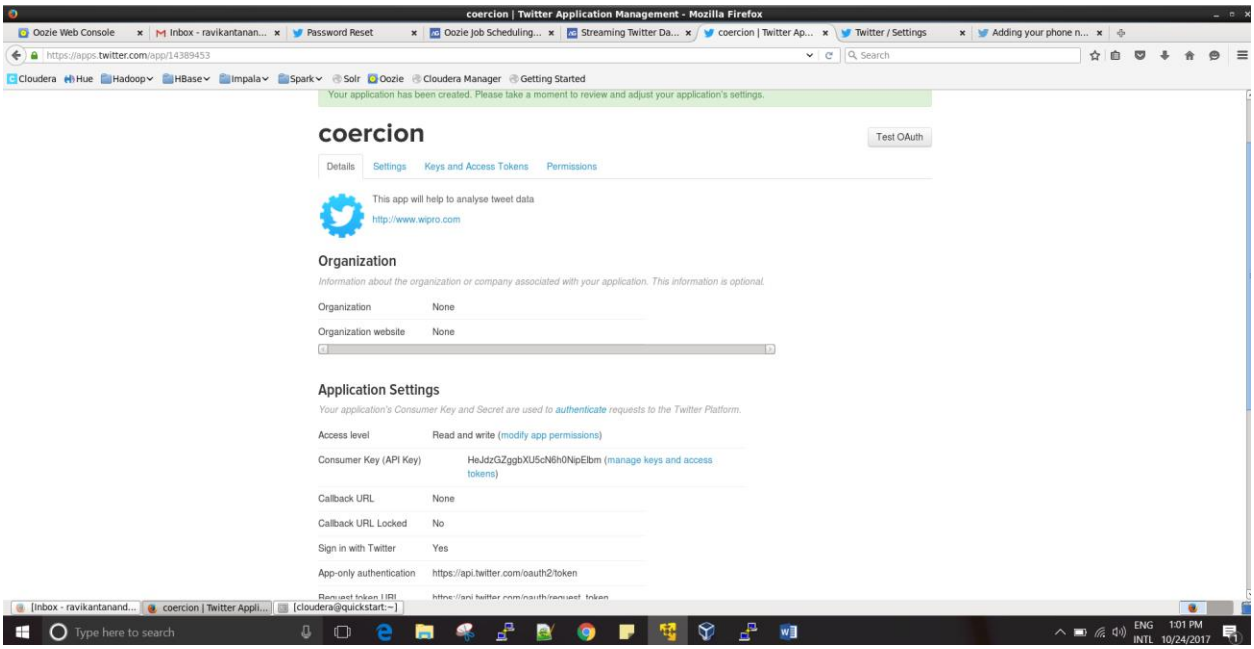
Below given window will pop-up





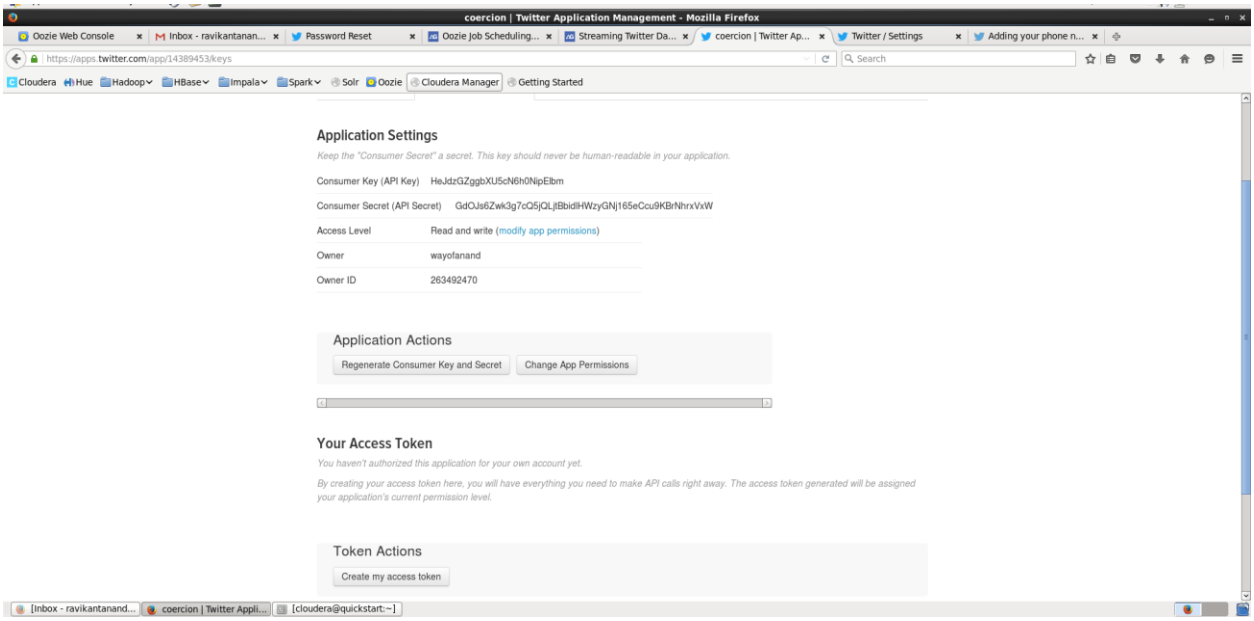
Step 3

Fill in the details, accept the **Developer Agreement** when finished, click on the **Create your Twitter application button** which is at the bottom of the page. If everything goes fine, an App will be created with the given details as shown below.



Step 4

Under **keys and Access Tokens** tab at the bottom of the page, you can observe a button named **Create my access token**. Click on it to generate the access token.

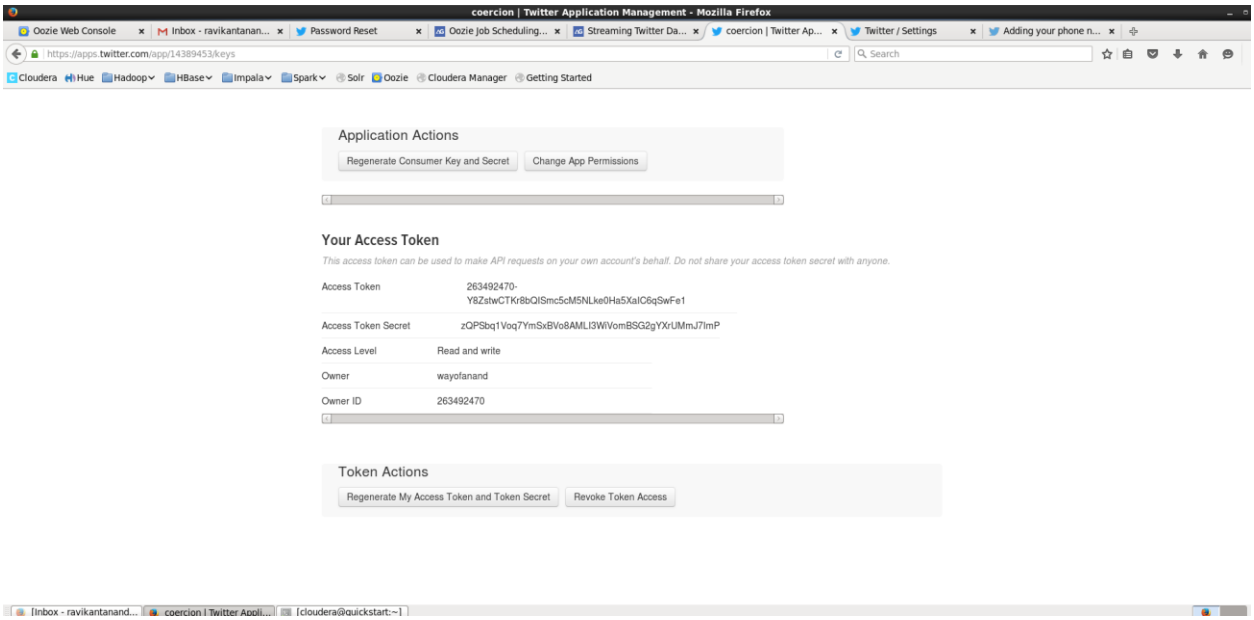


Consumer Key (API Key) HeJdzGZggbXU5cN6h0NipElbm

Consumer Secret (API Secret)
GdOJs6Zwk3g7cQ5jQLjtBbidlHWzyGNj165eCcu9KBrNhrxVxW

Step 5

Finally, click on the **Test OAuth** button which is on the right side top of the page. This will lead to a page which displays your **Consumer key, Consumer secret, Access token,** and **Access token secret**. Copy these details. These are useful to configure the agent in Flume



Your Access Token

This access token can be used to make API requests on your own account's behalf. Do not share your access token secret with anyone.

Access Token 263492470-Y8ZstwCTKr8bQISmc5cM5NLke0Ha5XaIC6qSwFe1

Access Token Secret zQPSbq1Voq7YmSxBVo8AMLI3WiVomBSG2gYXrUMmJ7ImP

Access Level Read and write

Owner wayofanand

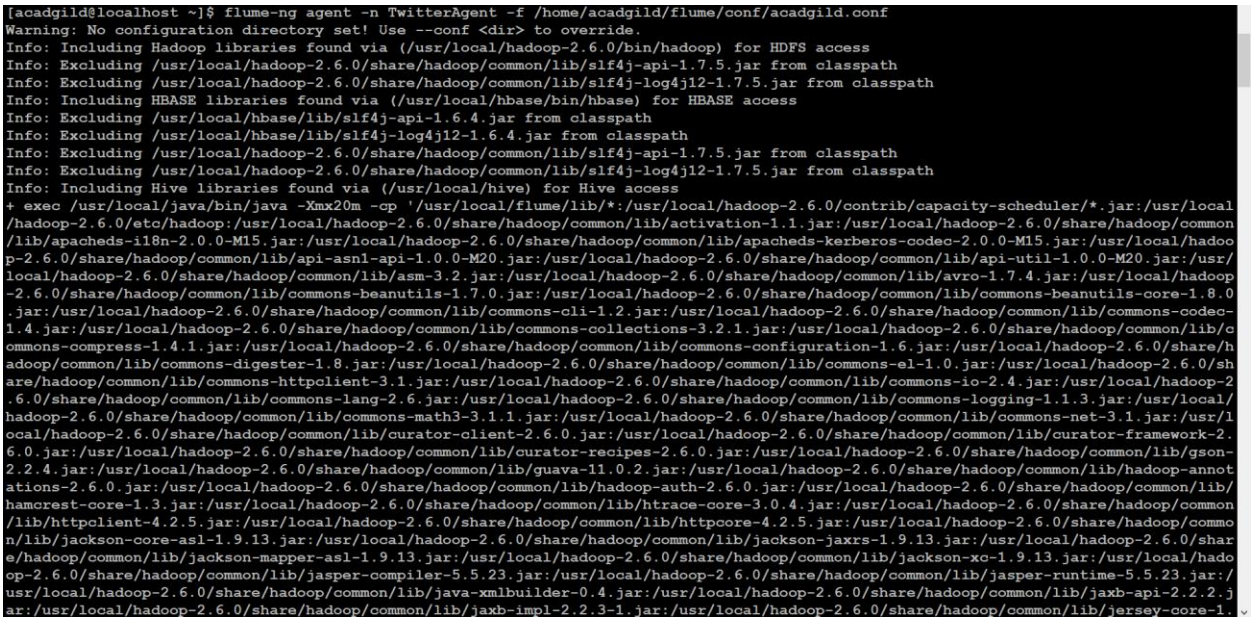
Owner ID 263492470

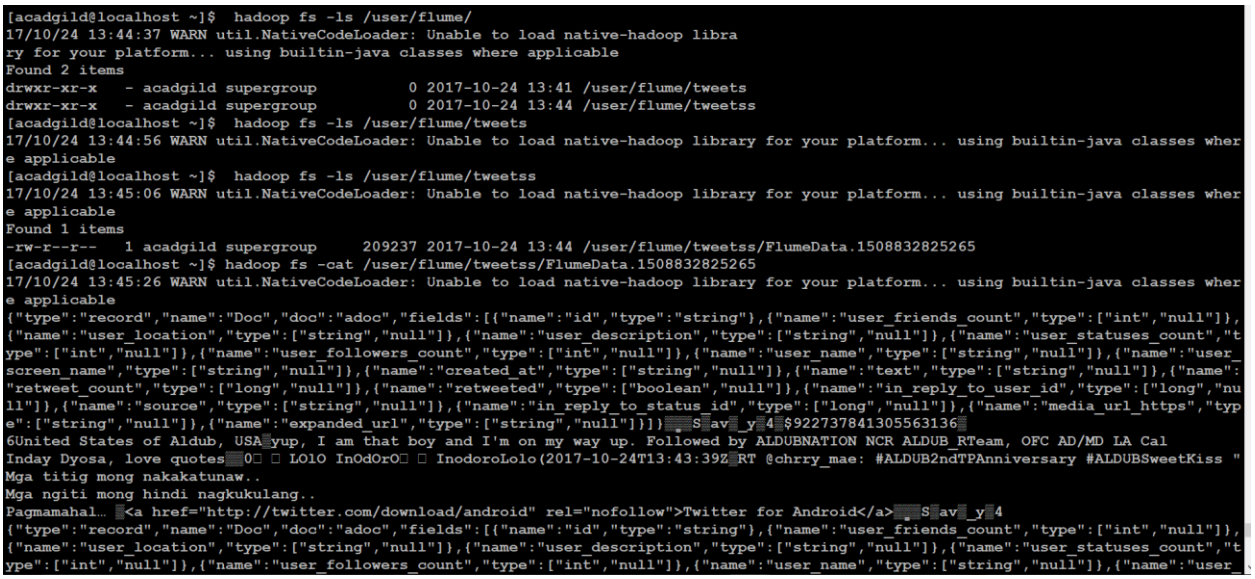
[acadgild@localhost ~]$ hadoop fs -mkdir /user/flume/tweets/

[acadgild@localhost ~]$ hadoop fs -ls /user/flume/

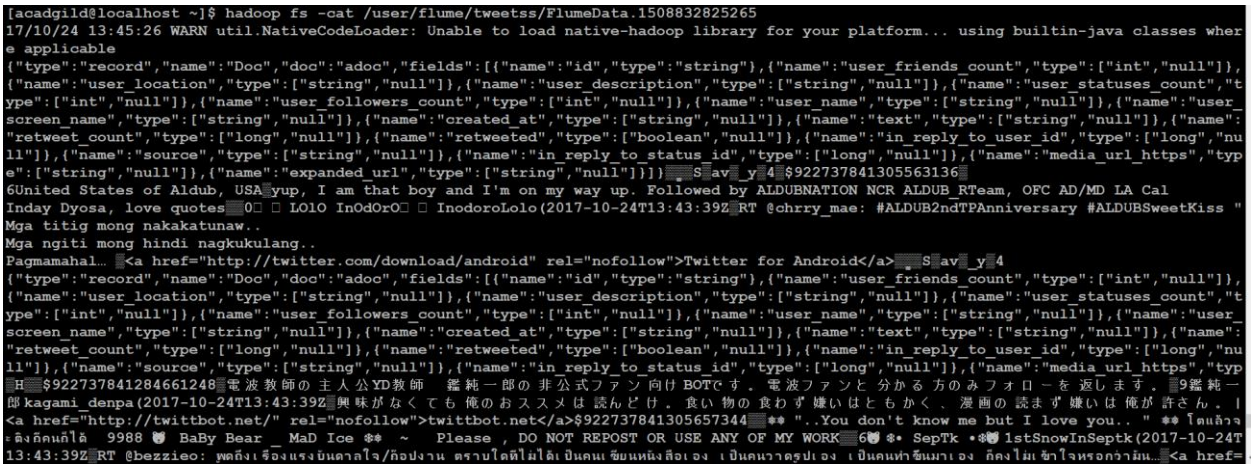
[acadgild@localhost ~]$ hadoop fs -ls /user/flume/

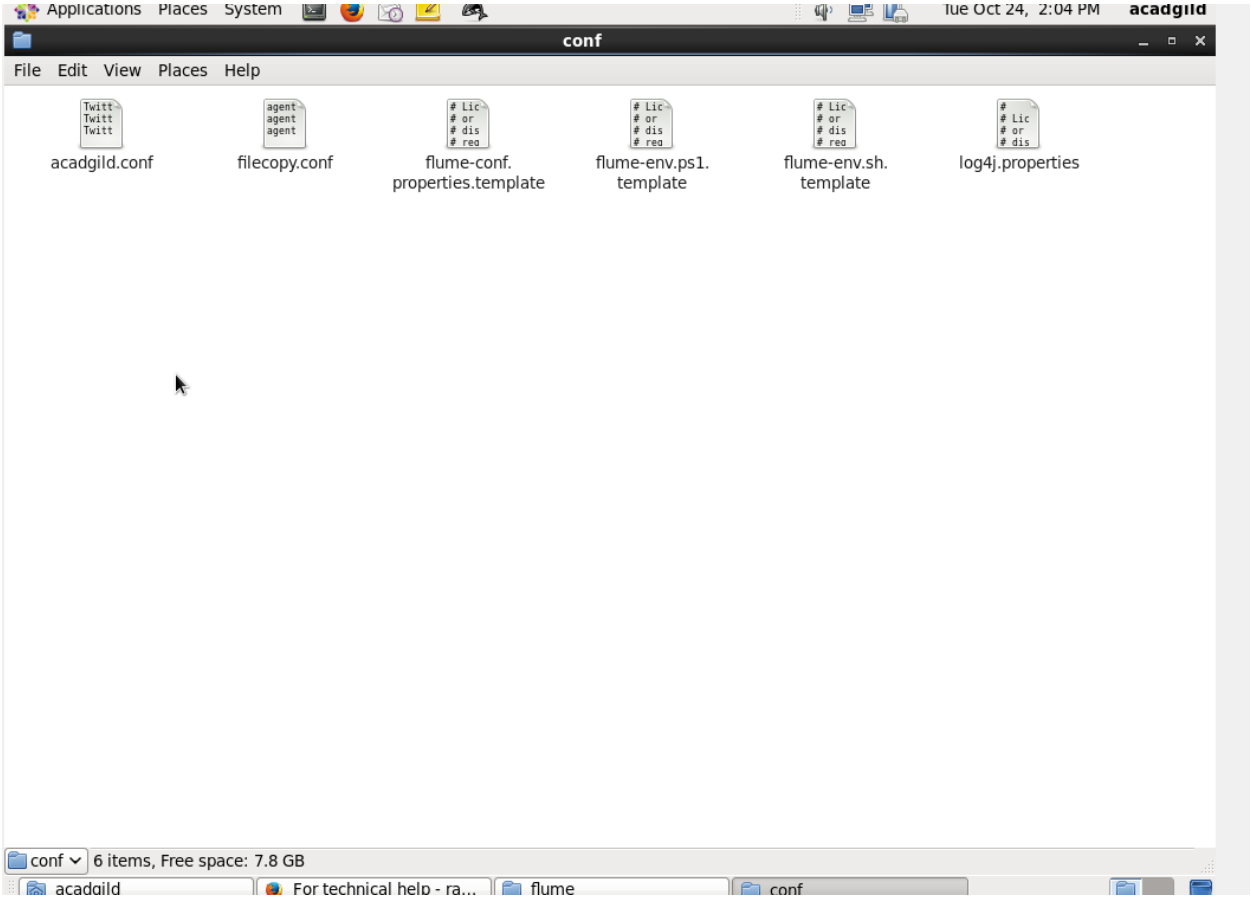
[acadgild@localhost ~]$ flume-ng agent -n TwitterAgent -f /home/acadgild/flume/conf/acadgild.conf





[acadgild@localhost ~]$ hadoop fs -cat /user/flume/tweetss/FlumeData.1508832825265




Applications Places System
Tue Oct 24, 2:04 PM acadgild
conf
File Edit View Places Help
acadgild.conf
filecopy.conf
flume-conf.
properties.template
flume-env.ps1.
template
flume-env.sh.
template
log4j.properties
conf 6 items, Free space: 7.8 GB
acadgild
For technical help - ra...
flume
conf

acadgild.conf

```
TwitterAgent.sources = Twitter
TwitterAgent.channels = MemChannel
TwitterAgent.sinks = HDFS

# Describing/Configuring the source
TwitterAgent.sources.Twitter.type = org.apache.flume.source.twitter.TwitterSource
TwitterAgent.sources.Twitter.consumerKey=HeJdzGZggbXU5cN6h0NipElbm
TwitterAgent.sources.Twitter.consumerSecret=GdOJs6Zwk3g7cQ5jQLjtBbidlHWzyGNj165eCcu9KBrNhrxVxW
TwitterAgent.sources.Twitter.accessToken=263492470-Y8ZstwCTKr8bQISmc5cM5NLke0Ha5XaIC6qSwFe1
TwitterAgent.sources.Twitter.accessTokenSecret=zQPSbq1Voq7YmSxBVo8AMLI3WiVomBSG2gYXrUMmJ7ImP
TwitterAgent.sources.Twitter.keywords=hadoop, bigdata, mapreduce, mahout, hbase, nosql
# Describing/Configuring the sink

TwitterAgent.sources.Twitter.keywords= hadoop,election,sports, cricket,Big data

TwitterAgent.sinks.HDFS.channel=MemChannel
TwitterAgent.sinks.HDFS.type=hdfs
TwitterAgent.sinks.HDFS.hdfs.path=hdfs://localhost:9000/user/flume/tweetss
TwitterAgent.sinks.HDFS.hdfs.fileType=DataStream
TwitterAgent.sinks.HDFS.hdfs.writeformat=Text
TwitterAgent.sinks.HDFS.hdfs.batchSize=1000
TwitterAgent.sinks.HDFS.hdfs.rollSize=0
TwitterAgent.sinks.HDFS.hdfs.rollCount=10000
TwitterAgent.sinks.HDFS.hdfs.rollInterval=600

TwitterAgent.channels.MemChannel.type=memory
TwitterAgent.channels.MemChannel.capacity=10000
TwitterAgent.channels.MemChannel.transactionCapacity=1000

TwitterAgent.sources.Twitter.channels = MemChannel
TwitterAgent.sinks.HDFS.channel = MemChannel
```