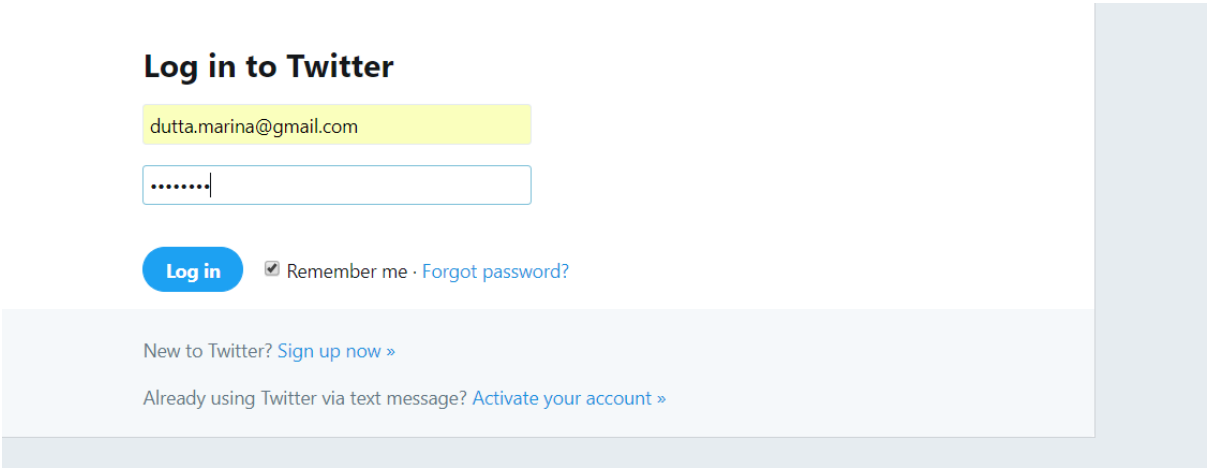
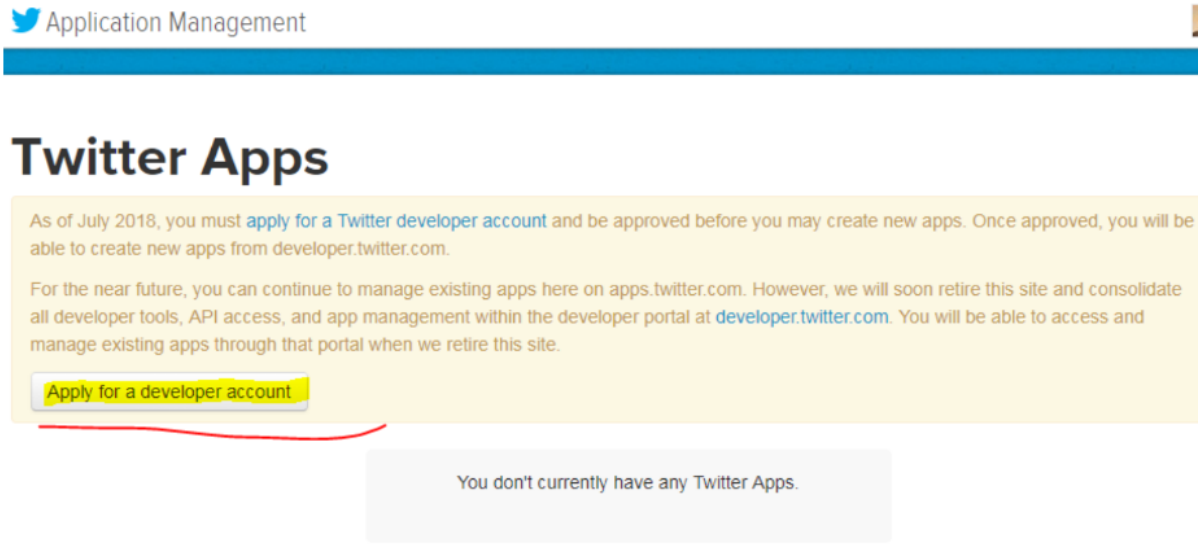


Step 2:

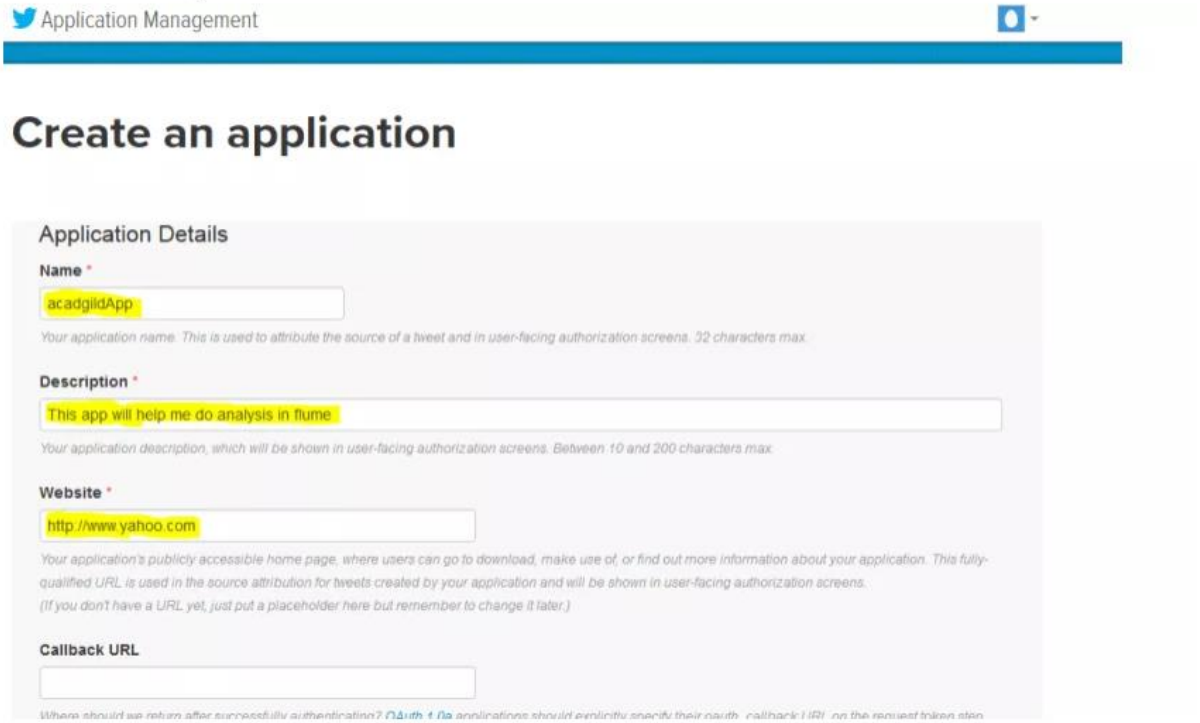
Go to the following link and click the 'Apply for Developer Account' button.

https://apps.twitter.com/app



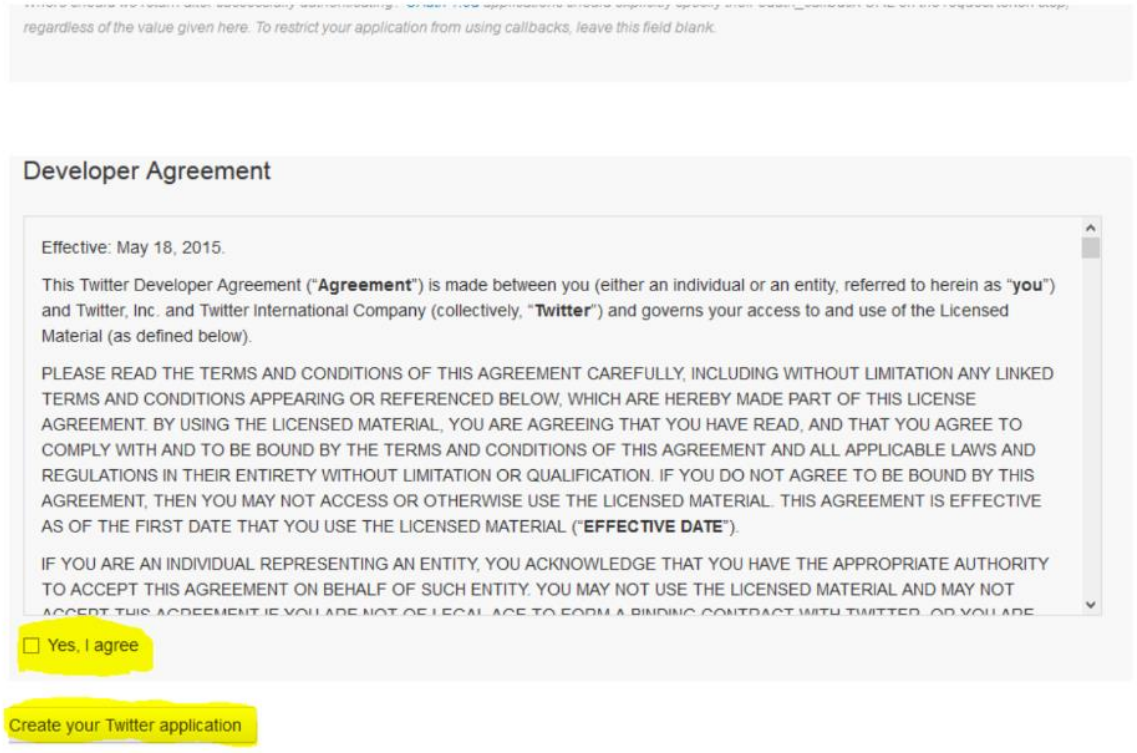
Step 3:

Enter the necessary details.



Step 4:

Accept the developer agreement and select the 'create your Twitter application' button.



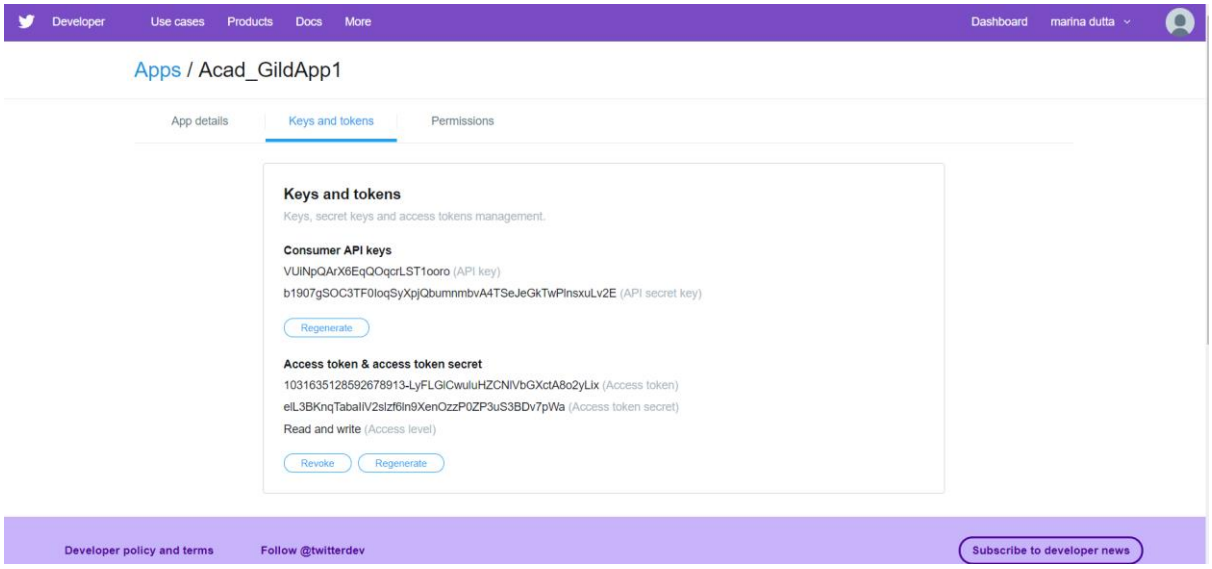
Step 5:

Create the new flume.conf file & copy the Flume configuration code from the below link and paste it in the newly created file flume.conf.

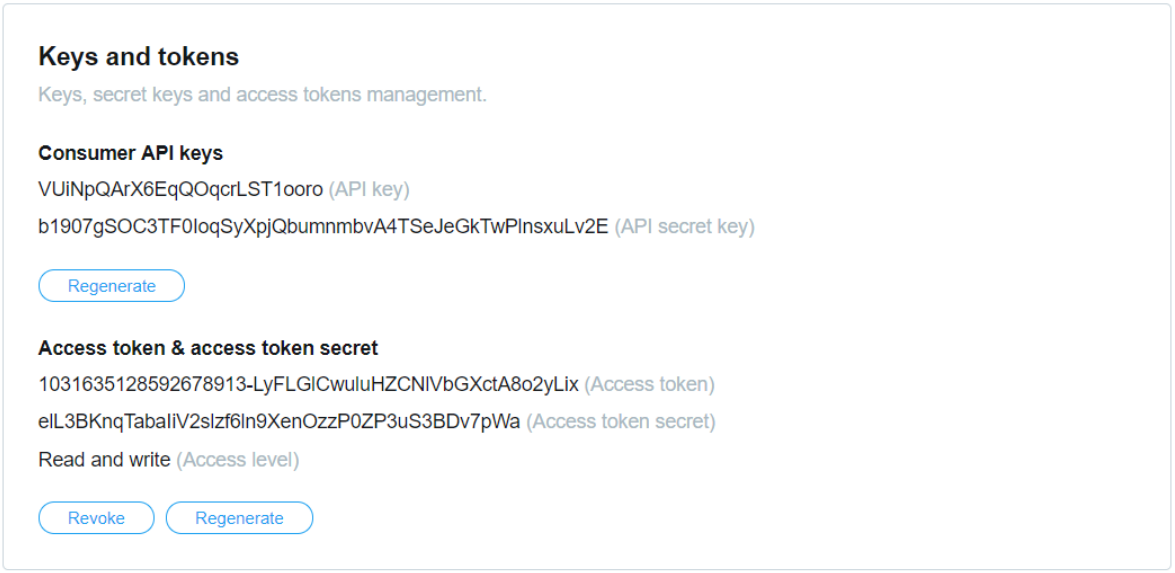
https://drive.google.com/open?id=0B1QaXx7tpw3Sb3U4LW9SWlNidkk

Step 6:

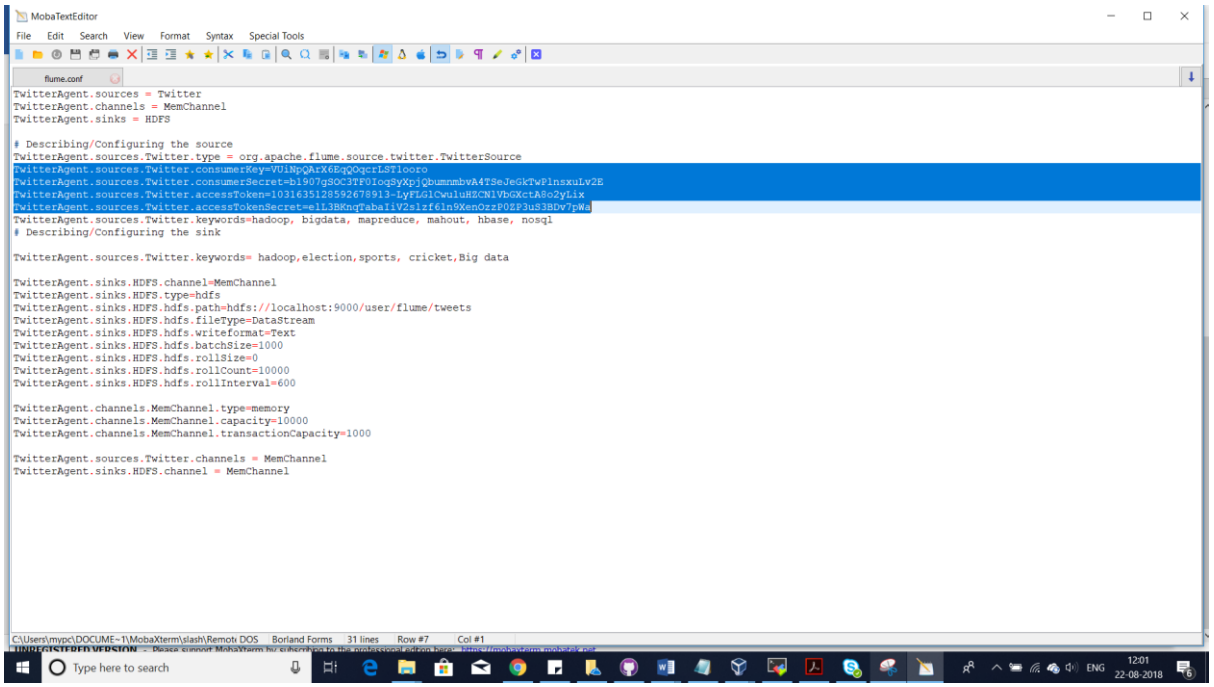
You would receive consumerKey, consumerSecret, accessToken, accessTokenSecret from twitter once developer account is approved.



Copy these four values within flume.config file .



Copy these four values within **flume.conf** file as highlighted below.



Step 7:

Within the same flume.conf file enter the keywords that you want to search the tweets on twitter against the key **TwitterAgent.sources.Twitter.keywords**

TwitterAgent.sources.Twitter.keywords= hadoop, bigdata, mapreduce, mahout, hbase, nosql

Step 8:

Create a new directory tweets which would store tweets stream by flume agent on to

HDFS

```
[acadgild@localhost ~]$ hadoop fs -mkdir -p /hadoopdata/flume/tweets
18/08/22 03:51:17 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -ls /hadoopdata/flume/
18/08/22 03:51:46 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 1 items
drwxr-xr-x   - acadgild supergroup          0 2018-08-22 03:51 /hadoopdata/flume/tweets
[acadgild@localhost ~]$
```

Step 9:

Mention the newly created directory path into the flume.conf as shown

TwitterAgent.sinks.HDFS.hdfs.path=hdfs://localhost:9000/hadoopdata/flume/tweets

```
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=VUiNpQArX6EqQOqcrLST1ooro
TwitterAgent.sources.Twitter.consumerSecret=b1907gSOC3TF0IoqSyXpjQbumnmbvA4TSeJeGkTwPlnsxuLv2E
TwitterAgent.sources.Twitter.accessToken=1031635128592678913-LyFLGlCwuluHZCNlVbGXctA8o2yLix
TwitterAgent.sources.Twitter.accessTokenSecret=elL3BKnqTabaIiV2slzf6ln9XenOzzP0ZP3uS3BDv7pWa
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/hadoopdata/flume/tweets
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
```

We are checking whether all the daemons are started

```
[acadgild@localhost ~]$ jps
3184 NodeManager
4449 Jps
11139 Main
2645 NameNode
23206 org.eclipse.equinox.launcher_1.4.0.v20161219-1356.jar
10840 HMaster
10953 HRegionServer
10777 HQuorumPeer
3082 ResourceManager
2938 SecondaryNameNode
2746 DataNode
8398 RunJar
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

Step 10:

For fetching data from Twitter, Use the below command to fetch the twitter tweet data into the HDFS cluster path.

```
[acadgild@localhost ~]$ flume-ng agent -n TwitterAgent -f/home/acadgild/flume.conf
Warning: No configuration directory set! Use --conf <dir> to override.
Info: Including Hadoop libraries found via (/home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop) for HDFS access
Info: Including HBASE libraries found via (/home/acadgild/install/hbase/hbase-1.2.6/bin/hbase) for HBASE access
Info: Including Hive libraries found via (/home/acadgild/install/hive/apache-hive-2.3.2-bin) for Hive access
+ exec /usr/java/jdk1.8.0_151/bin/java -Xmx20m -cp '/home/acadgild/install/flume/apache-flume-1.8.0-bin/lib/*:/home/acadgild/install/hadoop/hadoop-2.6.5/etc/ha
doop/:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/*:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/*:/home/acadgild/inst
all/hadoop/hadoop-2.6.5/share/hadoop/hdfs:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/hdfs/lib/*:/home/acadgild/install/hadoop/hadoop-2.6.5/share/h
adoop/hdfs/*:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/yarn/lib/*:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/yarn/*:/home/acadgild/i
nstall/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/lib/*:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/*:/home/acadgild/install/hadoop/hadoo
p-2.6.5/contrib/capacity-scheduler/*.jar:/home/acadgild/install/hbase/hbase-1.2.6/conf:/usr/java/jdk1.8.0_151/lib/tools.jar:/home/acadgild/install/hbase/hbase-
1.2.6:/home/acadgild/install/hbase/hbase-1.2.6/lib/activation-1.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/aopalliance-1.0.jar:/home/acadgild/install/h
base/hbase-1.2.6/lib/apacheds-i18n-2.0.0-M15.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/apacheds-kerberos-codec-2.0.0-M15.jar:/home/acadgild/install/hbas
e/hbase-1.2.6/lib/api-asn1-api-1.0.0-M20.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/api-util-1.0.0-M20.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/a
sm-3.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/avro-1.7.4.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-beanutils-1.7.0.jar:/home/acadgild/
install/hbase/hbase-1.2.6/lib/commons-beanutils-core-1.8.0.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-cli-1.2.jar:/home/acadgild/install/hbase/hb
ase-1.2.6/lib/commons-codec-1.9.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-collections-3.2.2.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/com
mons-compress-1.4.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-configuration-1.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-daemon-
1.0.13.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-digester-1.8.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-el-1.0.jar:/home/acadgild
/install/hbase/hbase-1.2.6/lib/commons-httpclient-3.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-io-2.4.jar:/home/acadgild/install/hbase/hbase-1.
2.6/lib/commons-lang-2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-logging-1.2.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-math-2.2
.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-math3-3.1.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-net-3.1.jar:/home/acadgild/insta
ll/hbase/hbase-1.2.6/lib/disruptor-3.3.0.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/findbugs-annotations-1.3.9-1.jar:/home/acadgild/install/hbase/hbase-1
.2.6/lib/guava-12.0.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/guice-3.0.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/guice-servlet-3.0.jar:/home/a
cadgild/install/hbase/hbase-1.2.6/lib/hadoop-annotations-2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-auth-2.5.1.jar:/home/acadgild/install/hb
ase/hbase-1.2.6/lib/hadoop-client-2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-common-2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/h
adoop-hdfs-2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-mapreduce-client-app-2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-map
reduce-client-common-2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-mapreduce-client-core-2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib
/hadoop-mapreduce-client-jobclient-2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-mapreduce-client-shuffle-2.5.1.jar:/home/acadgild/install/hbas
e/hbase-1.2.6/lib/hadoop-yarn-api-2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-yarn-client-2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/
lib/hadoop-yarn-common-2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-yarn-server-common-2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/
hbase-annotations-1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-annotations-1.2.6-tests.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-c
lient-1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-common-1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-common-1.2.6-tests.jar:/
home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-examples-1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-external-blockcache-1.2.6.jar:/home/aca
dgild/install/hbase/hbase-1.2.6/lib/hbase-hadoop2-compat-1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-hadoop-compat-1.2.6.jar:/home/acadgild/in
```

The streaming starts

To stop streaming press ctrl c.

Step 11:

To check the contents of the tweet go to the output directory at hdfs:

hadoop fs -ls /hadoopdata/flume/tweets