

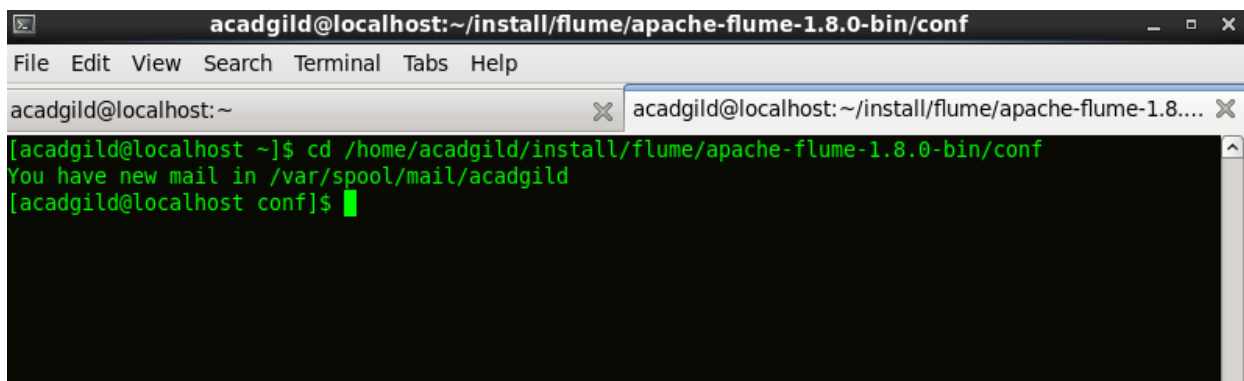
Streaming twitter data using Flume

Steps on twitter :

- 1.Login to twiiter account
- 2.Go to <https://apps.twitter.com/apps> and click on 'create new app' button
- 3.Give your app a name, description and enter a website
4. Accept the developer agreement and select the 'create your Twitter application' button.
5. Select the 'Keys and Access Token' tab. Copy consumer and consumer secret code and keep it.
- 6.Scroll down further to find 'create access token' button. Copy access token and access token secret code

Steps on acadgild VM :

Go to the conf directory inside flume extracted directory, In my case its /home/acadgild/install/flume/apache-flume-1.8.0-bin/conf

A terminal window titled 'acadgild@localhost:~/install/flume/apache-flume-1.8.0-bin/conf'. The terminal shows the user navigating to the conf directory and receiving a mail notification. The prompt changes from '~' to 'conf' after the directory change.

```
acadgild@localhost:~/install/flume/apache-flume-1.8.0-bin/conf
File Edit View Search Terminal Tabs Help
acadgild@localhost:~ acadgild@localhost:~/install/flume/apache-flume-1.8...
[acadgild@localhost ~]$ cd /home/acadgild/install/flume/apache-flume-1.8.0-bin/conf
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost conf]$
```

Note: Make sure you have below jars placed in your lib directory:

1. twitter4j-core-X.XX.jar
2. twitter4j-stream-X.X.X.jar
3. twitter4j-media-support-X.X.X.jar

Run the command `gedit flume.conf` to open up the editor

```
[acadgild@localhost conf]$ gedit flume.conf
```

Copy the flume configuration code and replace it with your own consumer key , consumer secret key, accesstoken and secret accesstoken. Also enter your hdfs path where data from twitter will be dumped

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

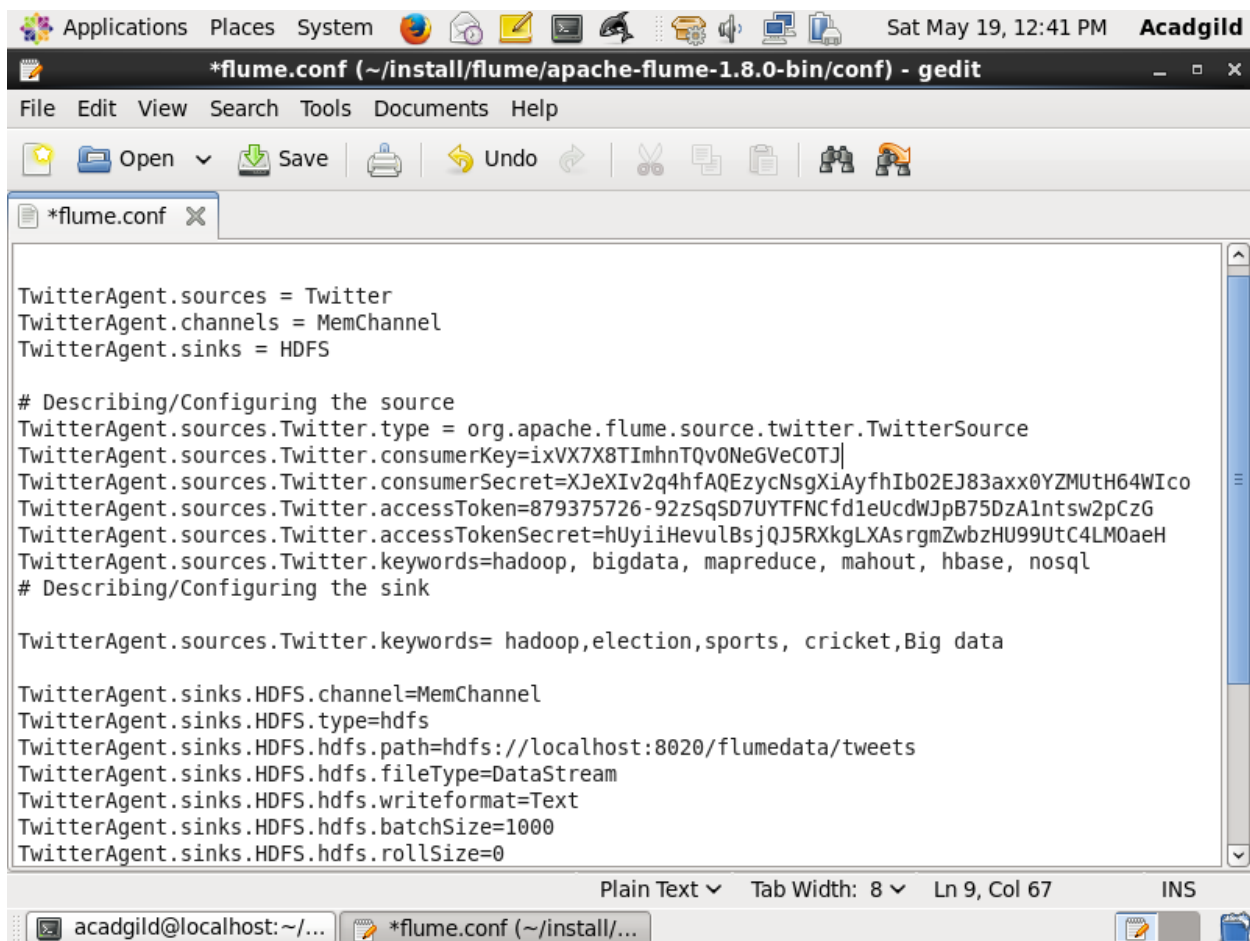
# Describing/Configuring the source
TwitterAgent.sources.Twitter.type =
org.apache.flume.source.twitter.TwitterSource
TwitterAgent.sources.Twitter.consumerKey=uX0TWqkx0okYEjjqLzxIx6mD6
TwitterAgent.sources.Twitter.consumerSecret=rzHIs3TMJnADbZNvdGU7LQUo0kPxPISq3
RGSLfqcBip39X5END
TwitterAgent.sources.Twitter.accessToken=559516596-
yDA9xqOljo4CV32wSnqxs2BXh4RBIRKFxZGSZrPC
TwitterAgent.sources.Twitter.accessTokenSecret=zDxePILZitS5tIWBhre0GWqps0FIj9
OadX8RZb6w8ZCwz
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/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 = MemChan
```



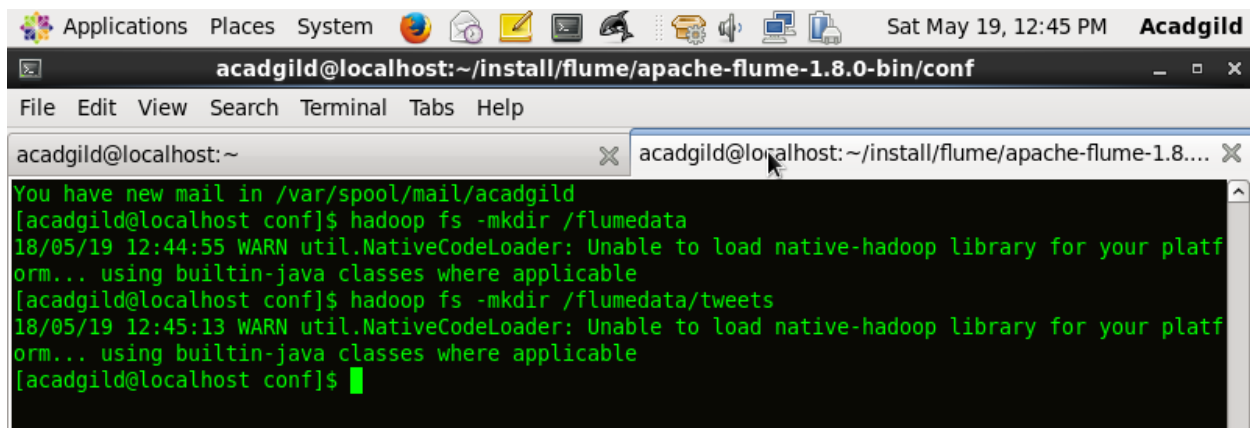
```
TwitterAgent.sources = Twitter
TwitterAgent.channels = MemChannel
TwitterAgent.sinks = HDFS

# Describing/Configuring the source
TwitterAgent.sources.Twitter.type = org.apache.flume.source.twitter.TwitterSource
TwitterAgent.sources.Twitter.consumerKey=ixVX7X8TImhnTQv0NeGVeCOTJ|
TwitterAgent.sources.Twitter.consumerSecret=XJeXIV2q4hfAQEzycNsgXiAyfhIb02EJ83axx0YZMUtH64WIco
TwitterAgent.sources.Twitter.accessToken=879375726-92zSqSD7UYTFNCfd1eUcdWJpB75DzA1ntsw2pCzG
TwitterAgent.sources.Twitter.accessTokenSecret=hUyiiHevulBsJQJ5RXkgLXAsrgmZwbzHU99UtC4LM0aeh
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:8020/flumedata/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
```

Note my hdfs path: hdfs ://localhost:8020/flumedata/tweets
Creating flumedata directory inside hdfs and tweets directory in flumedata



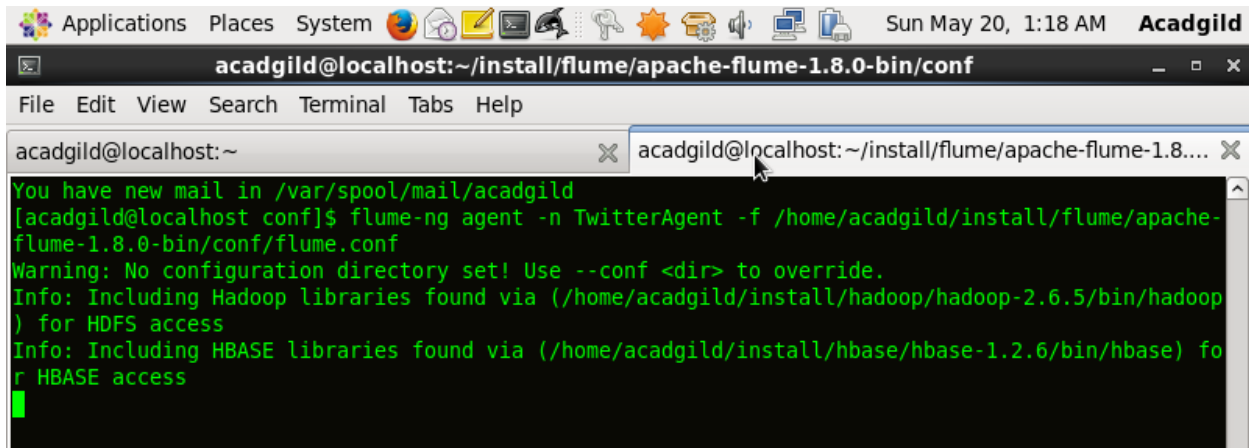
```
acadgild@localhost:~/install/flume/apache-flume-1.8.0-bin/conf
acacgild@localhost:~
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost conf]$ hadoop fs -mkdir /flumedata
18/05/19 12:44:55 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platf
orm... using builtin-java classes where applicable
[acadgild@localhost conf]$ hadoop fs -mkdir /flumedata/tweets
18/05/19 12:45:13 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platf
orm... using builtin-java classes where applicable
[acadgild@localhost conf]$
```

Start all the hadoop daemons run jps to check if all are running

```
[acadgild@localhost conf]$ jps
28643 Jps
9109 NameNode
9544 ResourceManager
9401 SecondaryNameNode
9210 DataNode
9646 NodeManager
29343 HQuorumPeer
10079 JobHistoryServer
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost conf]$
```

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

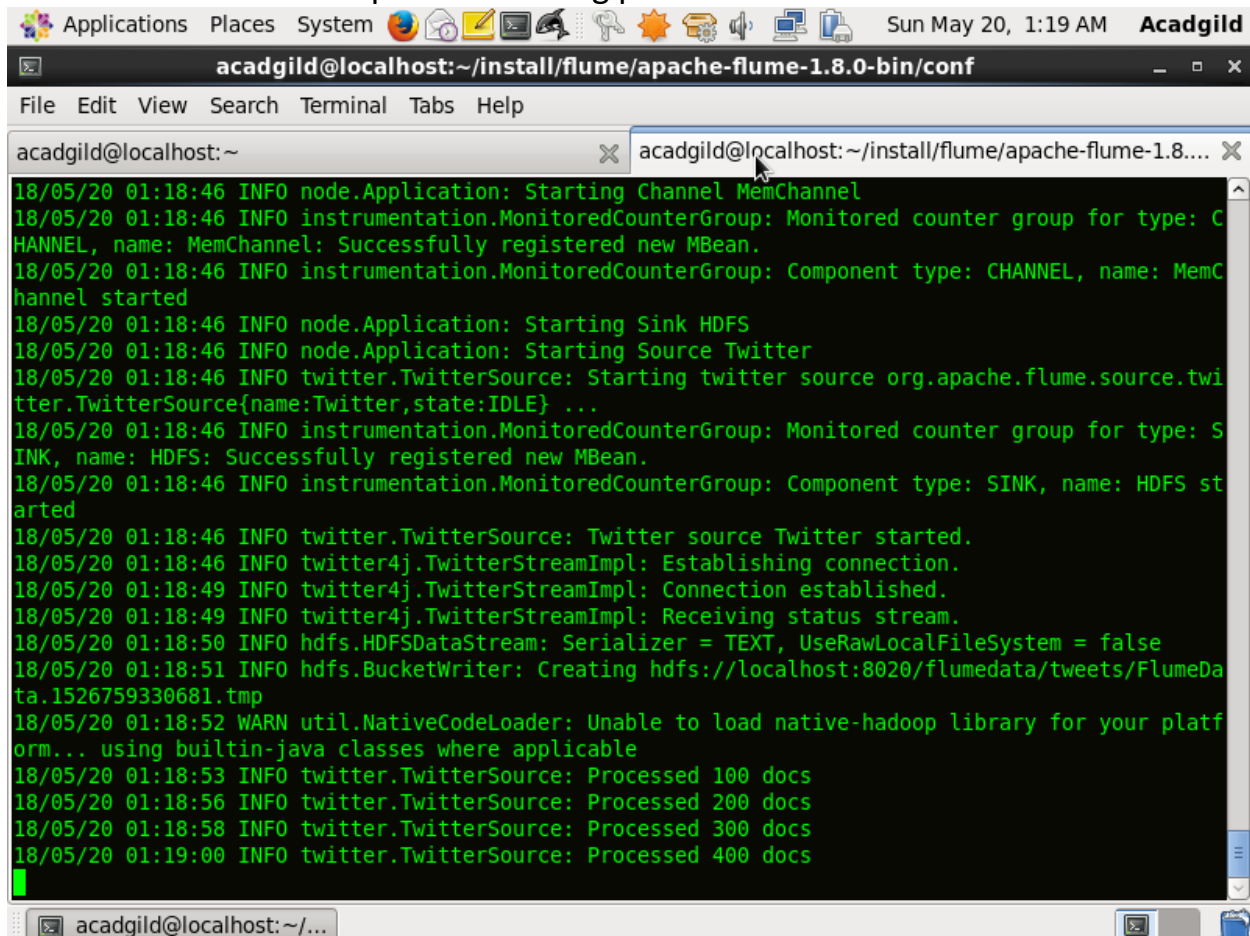
flume-ng agent -n TwitterAgent -f <location of created/edited conf file>



The screenshot shows a terminal window titled "acadgild@localhost: ~/install/flume/apache-flume-1.8.0-bin/conf". The terminal output shows the execution of the command `flume-ng agent -n TwitterAgent -f /home/acadgild/install/flume/apache-flume-1.8.0-bin/conf/flume.conf`. The output includes a warning about the configuration directory and information about including Hadoop and HBase libraries.

```
acacgild@localhost:~
You have new mail in /var/spool/mail/acacgild
[acacgild@localhost conf]$ flume-ng agent -n TwitterAgent -f /home/acacgild/install/flume/apache-
flume-1.8.0-bin/conf/flume.conf
Warning: No configuration directory set! Use --conf <dir> to override.
Info: Including Hadoop libraries found via (/home/acacgild/install/hadoop/hadoop-2.6.5/bin/hadoop
) for HDFS access
Info: Including HBASE libraries found via (/home/acacgild/install/hbase/hbase-1.2.6/bin/hbase) fo
r HBASE access
```

Once, the tweet data started streaming it into the given HDFS path we can use 'Ctrl+c' command to stop the streaming process.

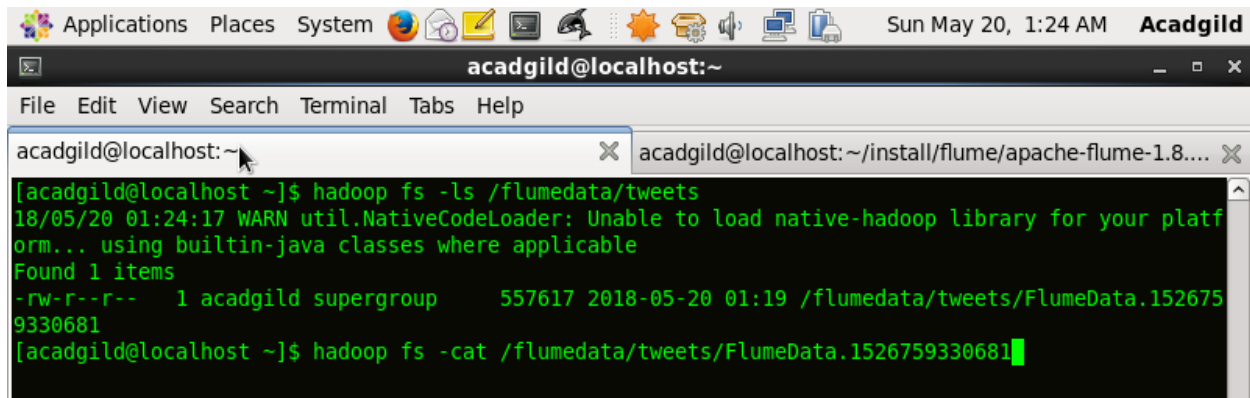


```
acadgild@localhost:~/install/flume/apache-flume-1.8.0-bin/conf
File Edit View Search Terminal Tabs Help

acadgild@localhost:~
acadgild@localhost:~/install/flume/apache-flume-1.8...

18/05/20 01:18:46 INFO node.Application: Starting Channel MemChannel
18/05/20 01:18:46 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: C
HANNEL, name: MemChannel: Successfully registered new MBean.
18/05/20 01:18:46 INFO instrumentation.MonitoredCounterGroup: Component type: CHANNEL, name: MemC
hannel started
18/05/20 01:18:46 INFO node.Application: Starting Sink HDFS
18/05/20 01:18:46 INFO node.Application: Starting Source Twitter
18/05/20 01:18:46 INFO twitter.TwitterSource: Starting twitter source org.apache.flume.source.twi
tter.TwitterSource{name=Twitter,state=IDLE} ...
18/05/20 01:18:46 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: S
INK, name: HDFS: Successfully registered new MBean.
18/05/20 01:18:46 INFO instrumentation.MonitoredCounterGroup: Component type: SINK, name: HDFS st
arted
18/05/20 01:18:46 INFO twitter.TwitterSource: Twitter source Twitter started.
18/05/20 01:18:46 INFO twitter4j.TwitterStreamImpl: Establishing connection.
18/05/20 01:18:49 INFO twitter4j.TwitterStreamImpl: Connection established.
18/05/20 01:18:49 INFO twitter4j.TwitterStreamImpl: Receiving status stream.
18/05/20 01:18:50 INFO hdfs.HDFSDataStream: Serializer = TEXT, UseRawLocalFileSystem = false
18/05/20 01:18:51 INFO hdfs.BucketWriter: Creating hdfs://localhost:8020/flumedata/tweets/FlumeDa
ta.1526759330681.tmp
18/05/20 01:18:52 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platf
orm... using builtin-java classes where applicable
18/05/20 01:18:53 INFO twitter.TwitterSource: Processed 100 docs
18/05/20 01:18:56 INFO twitter.TwitterSource: Processed 200 docs
18/05/20 01:18:58 INFO twitter.TwitterSource: Processed 300 docs
18/05/20 01:19:00 INFO twitter.TwitterSource: Processed 400 docs
```

Lets check the content of /flumedata/tweets

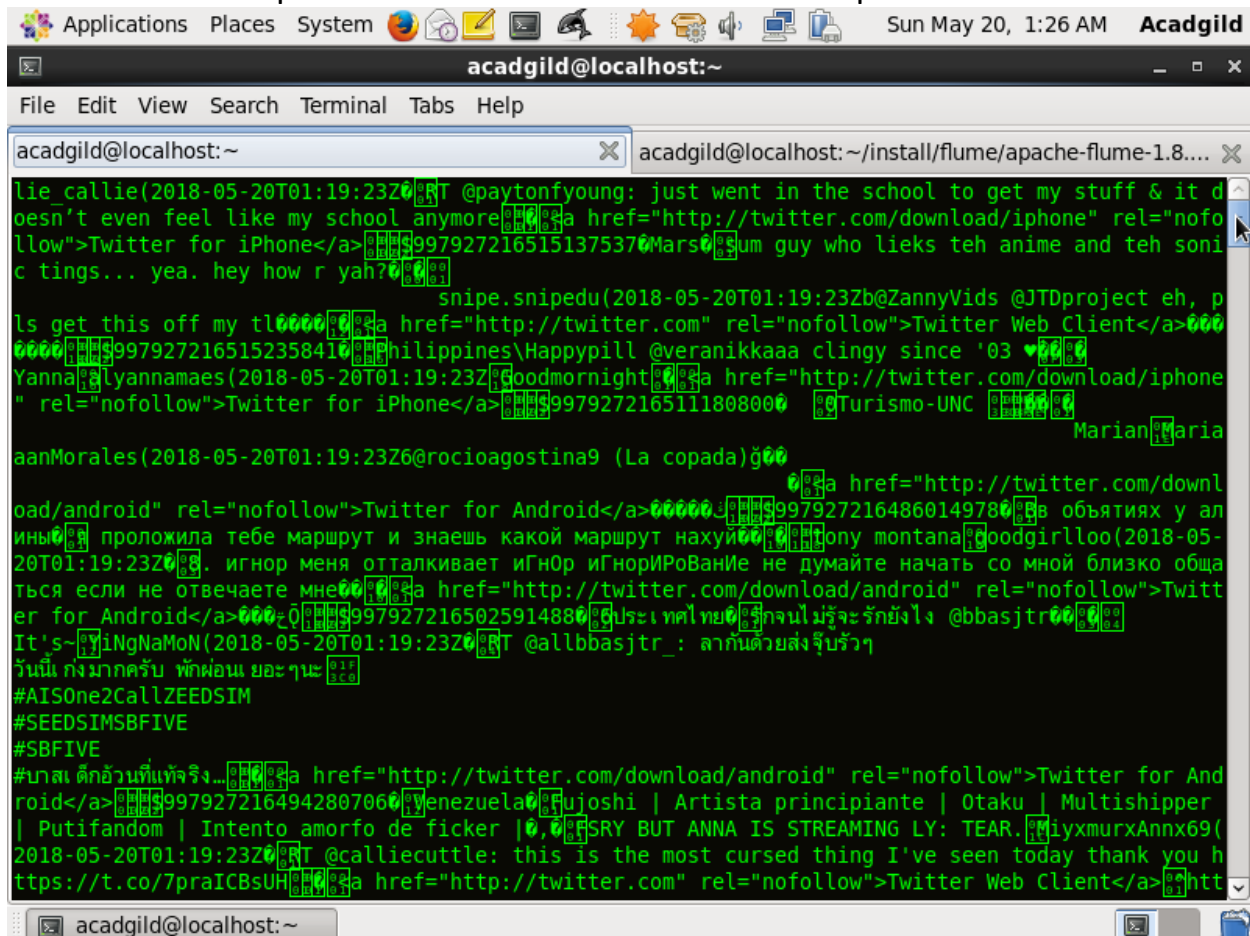


```
acadgild@localhost:~
File Edit View Search Terminal Tabs Help

acadgild@localhost:~
acadgild@localhost:~/install/flume/apache-flume-1.8...

[acadgild@localhost ~]$ hadoop fs -ls /flumedata/tweets
18/05/20 01:24:17 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platf
orm... using builtin-java classes where applicable
Found 1 items
-rw-r--r--  1 acadgild supergroup    557617 2018-05-20 01:19 /flumedata/tweets/FlumeData.152675
9330681
[acadgild@localhost ~]$ hadoop fs -cat /flumedata/tweets/FlumeData.1526759330681
```

Screenshot of output of some data which has been dumped



```
acacgild@localhost:~
File Edit View Search Terminal Tabs Help

acacgild@localhost:~ acacgild@localhost: ~/install/flume/apache-flume-1.8...
lie callie(2018-05-20T01:19:23Z) @paytonfyoung: just went in the school to get my stuff & it d
oesn't even feel like my school anymore href="http://twitter.com/download/iphone" rel="nofo
llow">Twitter for iPhone</a> 99979272165151375370 Mars0 sum guy who lieks teh anime and teh soni
c tings... yea. hey how r yah?
snipe.snipedu(2018-05-20T01:19:23Z) @ZannyVids @JTDproject eh, p
ls get this off my tl href="http://twitter.com" rel="nofollow">Twitter Web Client</a>
99979272165152358410 Philippines\Happypill @veranikkaaa clingy since '03
Yanna Alyannamaes(2018-05-20T01:19:23Z) @goodmornight href="http://twitter.com/download/iphone"
rel="nofollow">Twitter for iPhone</a> 9997927216511808000 Turismo-UNC
Marian Maria
aanMorales(2018-05-20T01:19:23Z) @rocioagostina9 (La copada)
href="http://twitter.com/downl
oad/android" rel="nofollow">Twitter for Android</a> 99979272164860149780
В объятиях у ал
ины проложила тебе маршрут и знаешь какой маршрут нахуй
ony montana @oodgirlloo(2018-05-
20T01:19:23Z). игнор меня отталкивает игнор игнорИРование не думайте начать со мной близко обща
ться если не отвечаете мне href="http://twitter.com/download/android" rel="nofollow">Twitt
er for Android</a> 99979272165025914880
ประเทศไทย งานไม่รู้จักยัง @bbasjtr
It's- ingNaMoN(2018-05-20T01:19:23Z) @allbbasjtr_ ลากันด้วยส่งจีบรัวๆ
วันนั้น คงมากครับ พักผ่อน ยะยะ
#AISOne2CallZEEDSIM
#SEEDSIMSBFIVE
#SBFIVE
#บาลเด็กอ้วนที่แท้จริง href="http://twitter.com/download/android" rel="nofollow">Twitter for And
roid</a> 99979272164942807060 Venezuela @Fujoshi | Artista principiante | Otaku | Multishipper
| Putifandom | Intento amorfo de flicker | SRY BUT ANNA IS STREAMING LY: TEAR. MiyxmurxAnnx69(
2018-05-20T01:19:23Z) @calliecuttle: this is the most cursed thing I've seen today thank you h
ttps://t.co/7praICBsUH href="http://twitter.com" rel="nofollow">Twitter Web Client</a> htt
```