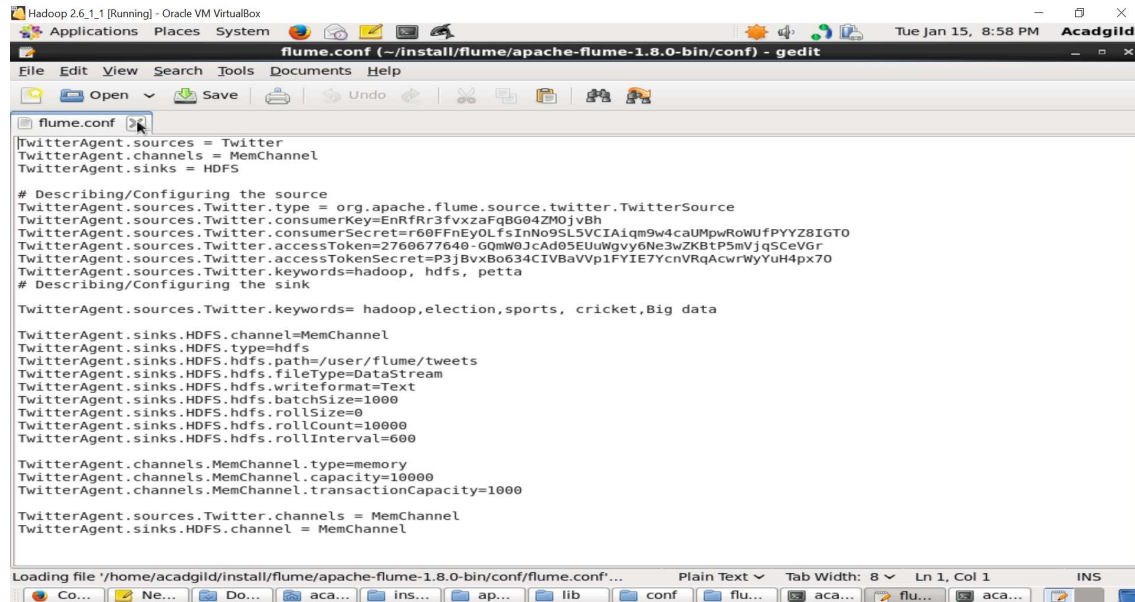


## Vijay's Assignment – Flume

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

Flume.conf file created to use the Access token and API keys created for my login

Search string modified to search for hadoop, hdfs and petta



The screenshot shows a Gedit editor window titled 'flume.conf (~/.install/flume/apache-flume-1.8.0-bin/conf) - gedit'. The editor contains the following configuration for a Twitter source and HDFS sink:

```
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=EnRfRr3fvkzaFqBG04ZM0jvBh
TwitterAgent.sources.Twitter.consumerSecret=r60FFnEy0LfsInNo9SL5VCIA1qm9w4caUMpwRoWUfPYYZ8IGTO
TwitterAgent.sources.Twitter.accessToken=2760677640-GQmW0JcAd05EUuWgy6Ne3wZKBP5mVjq5CevGr
TwitterAgent.sources.Twitter.accessTokenSecret=P3jBvxBo634CIVBaVvp1FYIE7YcnVRqAcwrWYyH4px70
TwitterAgent.sources.Twitter.keywords=hadoop, hdfs, petta
# 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=/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 = MemChannel
```

Command to start the flume agent

```
[acadgild@localhost conf]$ flume-ng agent -n TwitterAgent -f /home/acadgild/install/flume/apache-flume-1.8.0-bin/conf/flume.conf
```

Output

```
[acadgild@localhost conf]$ hadoop fs -ls /user/flume/tweets
19/01/15 20:57:12 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 1 items
-rw-r--r-- 1 acadgild supergroup 186043 2019-01-15 20:56 /user/flume/tweets/FlumeData.1547566006596
```

