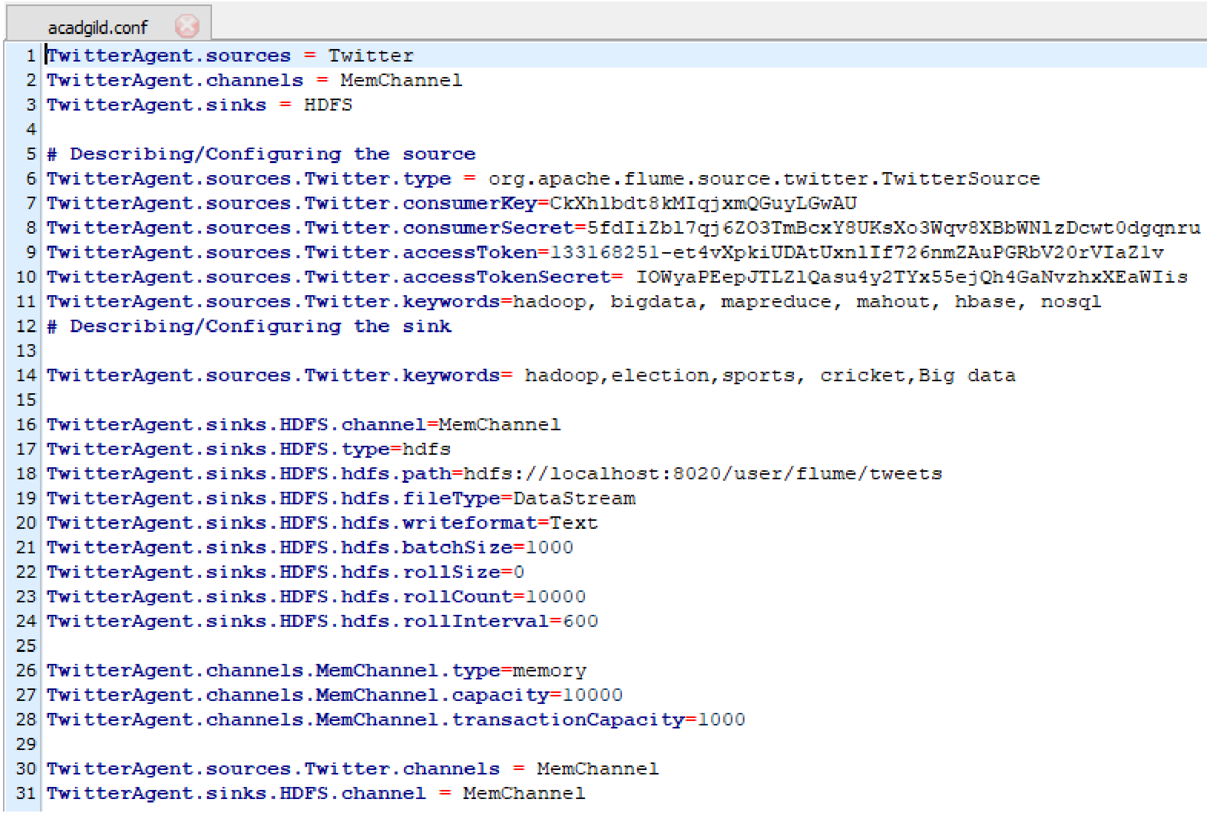
Assignment - 12.1

Task1: Create a flume agent that streams data from Twitter and stores in the HDFS.

Answer1: Below are the steps to be followed to create a Flume agent to stream Twitter data into HDFS: -

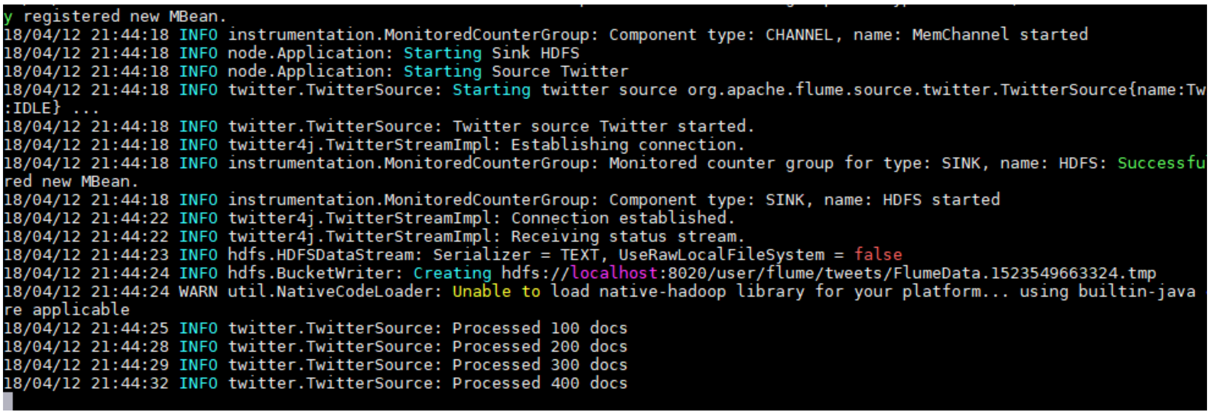
1. Login to Twittwer account.
2. Create a new app in this url - **https://apps.twitter.com/app**
3. Accept the terms and conditions and proceed further.
4. From “Keys and Access Token” tab get the Consumer Key, Consumer Secret, Access Key and Access Secret and hit on “Create my access token”.
5. Now update the keys and access tokens attained in Step4 in the configuration file as mentioned in the Screen shot.



```
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=CkXhlbdt8kMIqjxmQGuyLGwAU
TwitterAgent.sources.Twitter.consumerSecret=5fdIiZbl7qj6ZO3TmBcxY8UKsXo3Wqv8XBbWNlzDcwt0dgqnru
TwitterAgent.sources.Twitter.accessToken=133168251-et4vXpkiUDAtUxnlIf726nmZAuPGRbV20rVIaZlv
TwitterAgent.sources.Twitter.accessTokenSecret= IOWyaPEepJTLZlQasu4y2TYx55ejQh4GaNvzhxXEaWIis
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/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
```

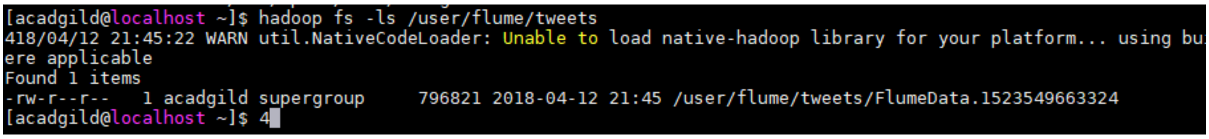
6. Here local host 8020 is updated with a location ‘/user/flume/tweets’ this will be the location where streamed data will be stored in HDFS.
7. Execute the below mentioned command to run Twitter Flume agent
   flume-ng agent -n TwitterAgent -f /home/acadgild/Downloads/apache-flume-1.6.0-bin/conf/acadgild.conf

   The next screenshot shows that the Flume agent is running successfully.



```
y registered new MBean.
18/04/12 21:44:18 INFO instrumentation.MonitoredCounterGroup: Component type: CHANNEL, name: MemChannel started
18/04/12 21:44:18 INFO node.Application: Starting Sink HDFS
18/04/12 21:44:18 INFO node.Application: Starting Source Twitter
18/04/12 21:44:18 INFO twitter.TwitterSource: Starting twitter source org.apache.flume.source.twitter.TwitterSource{name:Tw
:IDLE} ...
18/04/12 21:44:18 INFO twitter.TwitterSource: Twitter source Twitter started.
18/04/12 21:44:18 INFO twitter4j.TwitterStreamImpl: Establishing connection.
18/04/12 21:44:18 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SINK, name: HDFS: Successfu
red new MBean.
18/04/12 21:44:18 INFO instrumentation.MonitoredCounterGroup: Component type: SINK, name: HDFS started
18/04/12 21:44:22 INFO twitter4j.TwitterStreamImpl: Connection established.
18/04/12 21:44:22 INFO twitter4j.TwitterStreamImpl: Receiving status stream.
18/04/12 21:44:23 INFO hdfs.HDFSDataStream: Serializer = TEXT, UseRawLocalFileSystem = false
18/04/12 21:44:24 INFO hdfs.BucketWriter: Creating hdfs://localhost:8020/user/flume/tweets/FlumeData.1523549663324.tmp
18/04/12 21:44:24 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
re applicable
18/04/12 21:44:25 INFO twitter.TwitterSource: Processed 100 docs
18/04/12 21:44:28 INFO twitter.TwitterSource: Processed 200 docs
18/04/12 21:44:29 INFO twitter.TwitterSource: Processed 300 docs
18/04/12 21:44:32 INFO twitter.TwitterSource: Processed 400 docs
```

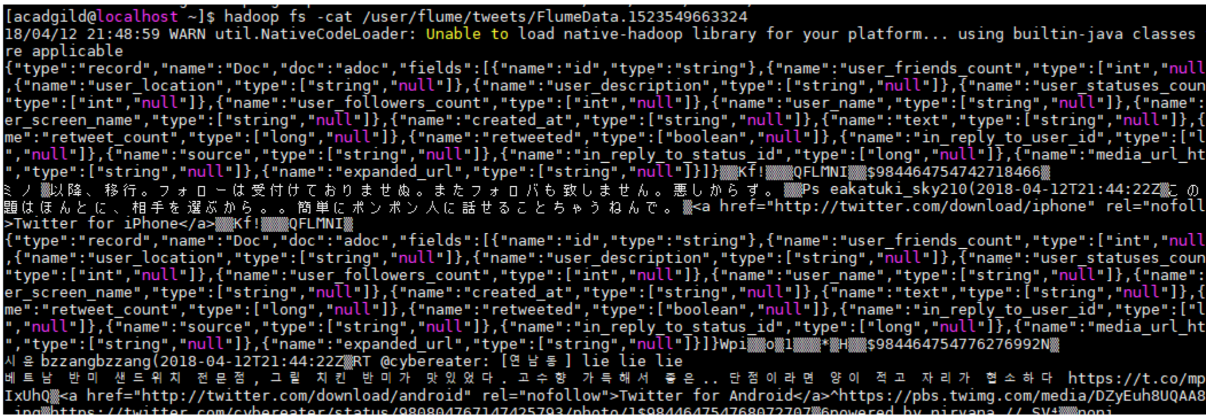
8. **In order to check the data getting streamed we have to check the file in mentioned HDFS location in configuration file.**



```
[acadgild@localhost ~]$ hadoop fs -ls /user/flume/tweets
418/04/12 21:45:22 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using bu
ere applicable
Found 1 items
-rw-r--r--   1 acadgild supergroup     796821 2018-04-12 21:45 /user/flume/tweets/FlumeData.1523549663324
[acadgild@localhost ~]$ 4
```

9. **Use cat command to see the data in the streamed file.**

   hadoop fs -cat /user/flume/tweets/FlumeData.1523549663324



```
[acadgild@localhost ~]$ hadoop fs -cat /user/flume/tweets/FlumeData.1523549663324
18/04/12 21:48:59 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes
re applicable
{"type":"record","name":"Doc","doc":"adoc","fields":[{"name":"id","type":"string"},{"name":"user_friends_count","type":["int","null
,{"name":"user_location","type":["string","null"]},{"name":"user_description","type":["string","null"]},{"name":"user_statuses_coun
"type":["int","null"]},{"name":"user_followers_count","type":["int","null"]},{"name":"user_name","type":["string","null"]},{"name":
er_screen_name","type":["string","null"]},{"name":"created_at","type":["string","null"]},{"name":"text","type":["string","null"]},{
me":"retweet_count","type":["long","null"]},{"name":"retweeted","type":["boolean","null"]},{"name":"in_reply_to_user_id","type":["l
","null"]},{"name":"source","type":["string","null"]},{"name":"in_reply_to_status_id","type":["long","null"]},{"name":"media_url_ht
","type":["string","null"]},{"name":"expanded_url","type":["string","null"]}]}▒▒Kf!▒▒▒QFLMNI▒▒$984464754742718466▒
ミノ▒以降、移行。フォローは受付けておりませぬ。またフォロバも致しません。悪しからず。▒▒Ps eakatuki_sky210(2018-04-12T21:44:22Z▒この
題はほんとに、相手を選ぶから。。簡単にポンポン人に話せることちゃうねんで。▒<a href="http://twitter.com/download/iphone" rel="nofoll
>Twitter for iPhone</a>▒▒Kf!▒▒▒QFLMNI▒
{"type":"record","name":"Doc","doc":"adoc","fields":[{"name":"id","type":"string"},{"name":"user_friends_count","type":["int","null
,{"name":"user_location","type":["string","null"]},{"name":"user_description","type":["string","null"]},{"name":"user_statuses_coun
"type":["int","null"]},{"name":"user_followers_count","type":["int","null"]},{"name":"user_name","type":["string","null"]},{"name":
er_screen_name","type":["string","null"]},{"name":"created_at","type":["string","null"]},{"name":"text","type":["string","null"]},{
me":"retweet_count","type":["long","null"]},{"name":"retweeted","type":["boolean","null"]},{"name":"in_reply_to_user_id","type":["l
","null"]},{"name":"source","type":["string","null"]},{"name":"in_reply_to_status_id","type":["long","null"]},{"name":"media_url_ht
","type":["string","null"]},{"name":"expanded_url","type":["string","null"]}]}Wpi▒▒o▒1▒▒▒*▒H▒▒$984464754776276992N▒
시 윤 bzzangbzzang(2018-04-12T21:44:22Z▒RT @cybereater: [연 남 동 ] lie lie lie
베트남 반미 샌드위치 전문점, 그릴 치킨 반미가 맛있었다. 고수향 가득해서 좋은.. 단점이라면 양이 적고 자리가 협소하다 https://t.co/mp
IxUhQ▒<a href="http://twitter.com/download/android" rel="nofollow">Twitter for Android</a>^https://pbs.twimg.com/media/DZyEuh8UQAA8
 ing▒https://twitter.com/cybereater/status/980804767147425793/photo/1$984464754768072707▒6powered by nirvana // SV*▒▒noni
```