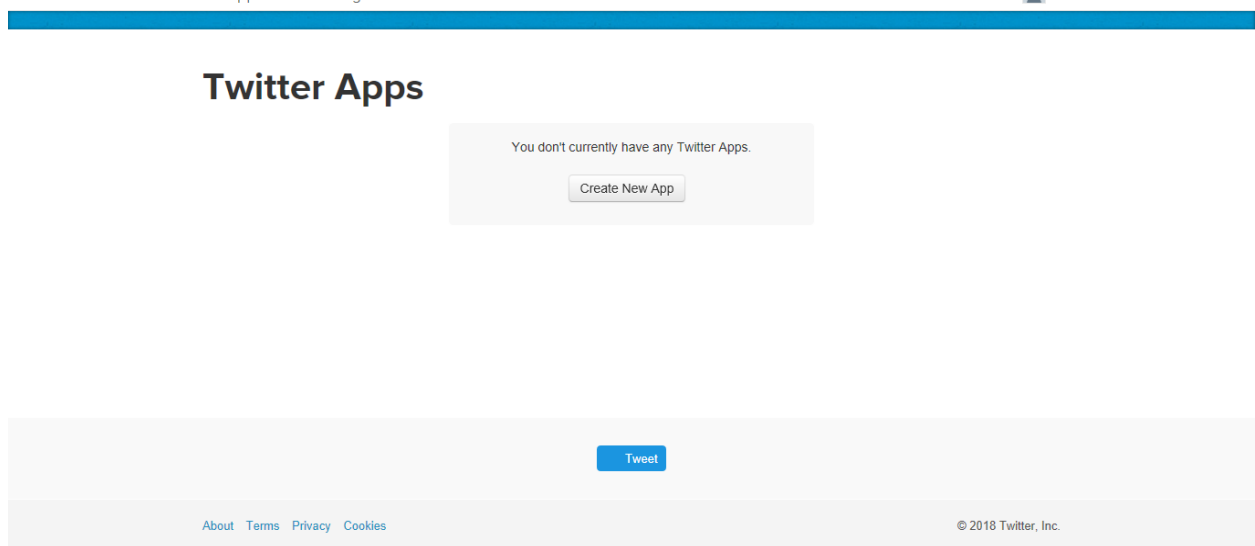


Assignment 11.2

Twitter data analysis using flume

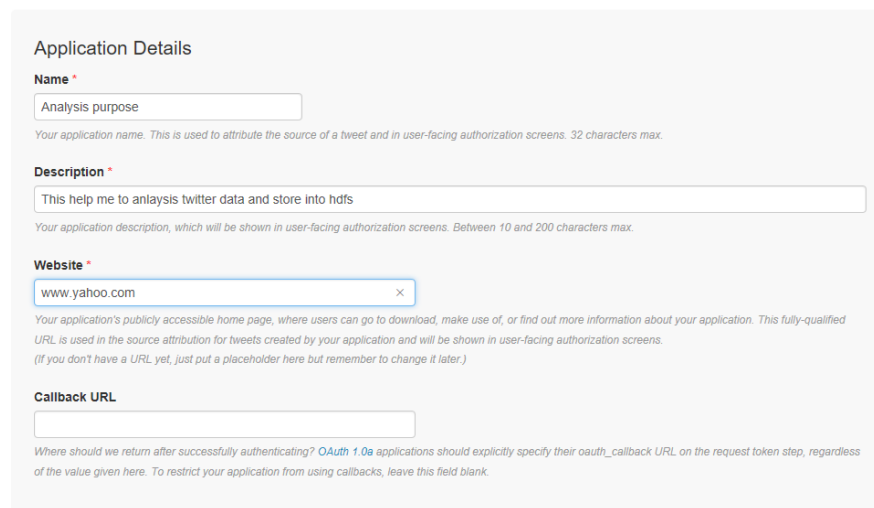
Step1: Login into twitter account ,then head to <https://apps.twitter.com/>

As shown below,



Step2: select “create New App “ option and fill out all details as shown below

Create an application

The screenshot shows the 'Create an application' form. It has a title 'Create an application' and a section 'Application Details'. There are four main fields: 'Name' (with a red asterisk), 'Description' (with a red asterisk), 'Website' (with a red asterisk), and 'Callback URL'. Each field has a text input box and a small 'x' icon to clear the text. Below each field is a small line of explanatory text. The 'Name' field contains 'Analysis purpose'. The 'Description' field contains 'This help me to anlaysis twitter data and store into hdfs'. The 'Website' field contains 'www.yahoo.com'. The 'Callback URL' field is empty.

Step3: Now select Keys and Access Tokens labels as shown below, copy consumer key(API key) ,consumer secret

[Details](#) [Settings](#) [Keys and Access Tokens](#) [Permissions](#)

Application Settings

Keep the "Consumer Secret" a secret. This key should never be human-readable in your application.

Consumer Key (API Key)	rxYCg2mriAkDta7fmegGZCjZh
Consumer Secret (API Secret)	XWrKiz8rY7EE5YZt1tmSuXI3fqChmlpvJvWWxQKxQdL6xQaVxk
Access Level	Read and write (modify app permissions)
Owner	sarwanishabana
Owner ID	970658778277298177

Step4: Scroll down and click on “create my access token”, and we will get as shown below and copy access token and access token secret

Application Actions

[Regenerate Consumer Key and Secret](#) [Change App Permissions](#)

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

Access Token	970658778277298177- KaMuzXrzF2GrpCMD2rflsBXW2VNL23O
Access Token Secret	ItEGKbSDsnDupeFZtCXtKBB0mf1gcC6H56dpRojRuloe
Access Level	Read and write
Owner	sarwanishabana
Owner ID	970658778277298177

Step5: Copy all in any text file as shown below

```
twitter.txt
c-key: rxyCg2mriAkDta7fmegGZCjZh
c-s-k:XWrKiz8rY7EE5YZt1tmSuXI3fqChmIpvJvWwXQKxQdL6xQaVxk

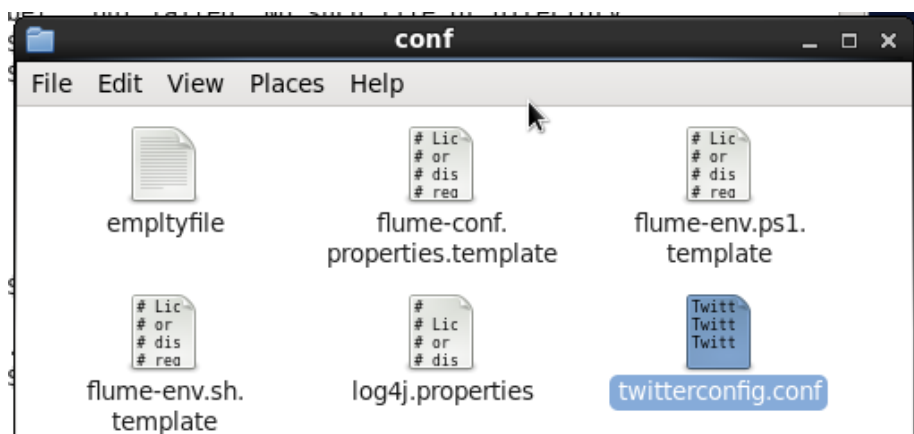
Access token:970658778277298177-KaMuzXrzF2GrpCMD2rfIsBXW2VNL230
Access token scret key:ItEGKbSDsnDupeFZtCXtKBB00mf1gcC6H56dpRojRuIoe|
```

Step6: Open acadgild VM terminal and create tweets directory under flume, flume under user as shown below

```
File Edit View Search Terminal Help

(gedit:8147): Gtk-WARNING **: Attempting to set the permissions of '/root/.local/share/recently-used.xbel', but failed: No such file or directory
[acadgild@localhost ~]$ source .bashrc
[acadgild@localhost ~]$ jps
3237 NameNode
3482 SecondaryNameNode
8365 Jps
3661 ResourceManager
3341 DataNode
3775 NodeManager
[acadgild@localhost ~]$ hadoop fs -mkdir /user/flume
18/03/04 09:23:09 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[acadgild@localhost ~]$ hadoop fs -mkdir /user/flume/tweets
18/03/04 09:23:24 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
```

Step7: Twitter configuration file is placed inside /flume/conf directory as shown below,content is shown in next slide



Step8: Access key ,Access Token, Consumer key ,Cosumer Secret are replaced with those are copied previously from our own application,and Keywords can also be changed as per our requirement as shown below

```
*twitterconfig.conf X
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=rxyCg2mriAkDta7fmegGZCjZh
TwitterAgent.sources.Twitter.consumerSecret=XWrKiz8rY7EE5YZt1tmSuXI3fqChmIpvJvWwXQKxQdL6xQaVxk
TwitterAgent.sources.Twitter.accessToken=970658778277298177-KaMuzXrzF2GrpCMD2rfIsBXW2VNL230
TwitterAgent.sources.Twitter.accessTokenSecret=ItEGKbSDsnDupeFZtCXtKBB00mflgcC6H56dpRojRuIoe
TwitterAgent.sources.Twitter.keywords=acadgild,hadoop
# Describing/Configuring the sink

TwitterAgent.sources.Twitter.keywords= acadgild,hadoop`

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
```

Step9: Checking .bashrc file as shown below,FLUME_HOME ans PATH

```
.bashrc X
export SQOOP_HOME=/home/acadgild/install/sqoop/sqoop-1.4.6.bin_hadoop-2.0.4-
alpha
export PATH=$PATH:$SQOOP_HOME/bin

# Below 2 lines we have to add for HBASE Installation
export HBASE_HOME=/home/acadgild/install/hbase/hbase-
export PATH=$PATH:$HBASE_HOME/bin

# Below 2 lines we have to add for kafka Installation
export KAFKA_HOME=/home/acadgild/install/kafka/kafka_
export PATH=$PATH:$KAFKA_HOME/bin

# Below 2 lines we have to add for FLUME Installation
export FLUME_HOME=/home/acadgild/install/flume/apache
export PATH=$PATH:$FLUME_HOME/bin
```

```
acadgild@localhost:~
File Edit View Search Terminal Help
[acadgild@localhost ~]$ sudo gedit .bashrc
[sudo] password for acadgild:
(gedit:8147): Gtk-WARNING **: Attempting to store changes into `/root/.lo
re/recently-used.xbel', but failed: Failed to create file '/root/.local/s
cently-used.xbel.LTLAFZ': No such file or directory
(gedit:8147): Gtk-WARNING **: Attempting to set the permissions of `/root
/share/recently-used.xbel', but failed: No such file or directory
```

Step10: Using below Flume statement successfully records from twitter are analysis for the given keywords

```
acadgild@localhost:~  
File Edit View Search Terminal Help  
[acadgild@localhost ~]$ flume-ng agent -n TwitterAgent -f /home/acadgild/install  
/flume/apache-flume-1.8.0-bin/conf/twitterconfig.conf
```

Step11: The below snapshot shows number of documents are processed to find desired keywords, by pressing Ctrl+c we can stop the process and can view results

```
acadgild@localhost:~  
File Edit View Search Terminal Help  
18/03/06 11:33:14 INFO twitter4j.TwitterStreamImpl: Connection established.  
18/03/06 11:33:14 INFO twitter4j.TwitterStreamImpl: Receiving status stream.  
18/03/06 11:33:17 INFO twitter.TwitterSource: Processed 1,500 docs  
18/03/06 11:33:21 INFO twitter.TwitterSource: Processed 1,600 docs  
18/03/06 11:33:25 INFO twitter.TwitterSource: Processed 1,700 docs  
18/03/06 11:33:29 INFO twitter.TwitterSource: Processed 1,800 docs  
18/03/06 11:33:33 INFO twitter.TwitterSource: Processed 1,900 docs  
18/03/06 11:33:35 INFO twitter.TwitterSource: Processed 2,000 docs  
18/03/06 11:33:35 INFO twitter.TwitterSource: Total docs indexed: 2,000, total s  
kipped docs: 0  
18/03/06 11:33:35 INFO twitter.TwitterSource: 23 docs/second  
18/03/06 11:33:35 INFO twitter.TwitterSource: Run took 84 seconds and processed:  
18/03/06 11:33:35 INFO twitter.TwitterSource: 0.006 MB/sec sent to index  
18/03/06 11:33:35 INFO twitter.TwitterSource: 0.529 MB text sent to index  
18/03/06 11:33:35 INFO twitter.TwitterSource: There were 0 exceptions ignored:  
18/03/06 11:33:39 INFO twitter.TwitterSource: Processed 2,100 docs  
18/03/06 11:33:41 INFO twitter.TwitterSource: Processed 2,200 docs  
18/03/06 11:33:43 INFO twitter.TwitterSource: Processed 2,300 docs  
18/03/06 11:33:45 INFO twitter.TwitterSource: Processed 2,400 docs  
18/03/06 11:33:48 INFO twitter.TwitterSource: Processed 2,500 docs
```

Step12: We can see the result is stored in files in the Browse Directory

Browse Directory

Show 25 entries

Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
<input type="checkbox"/>	-rw-r--r--	acadgild	supergroup	8 MB	Mar 06 11:42	1	128 MB	FlumeData.1520316136628
<input type="checkbox"/>	-rw-r--r--	acadgild	supergroup	1.71 MB	Mar 06 11:44	1	128 MB	FlumeData.1520316748933

Showing 1 to 2 of 2 entries

Previous

1

Next

Step13: The below is the twitter data got fetched inside the files

```

acadgild@localhost:~
File Edit View Search Terminal Help

//twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>{"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_count","type":["int","null"]}, {"name":"user_followers_count","type":["int","null"]}, {"name":"user_name","type":["string","null"]}, {"name":"user_screen_name","type":["string","null"]}, {"name":"created_at","type":["string","null"]}, {"name":"text","type":["string","null"]}, {"name":"retweet_count","type":["long","null"]}, {"name":"retweeted","type":["boolean","null"]}, {"name":"in_reply_to_user_id","type":["long","null"]}, {"name":"source","type":["string","null"]}, {"name":"in_reply_to_status_id","type":["long","null"]}, {"name":"media_url_https","type":["string","null"]}, {"name":"expanded_url","type":["string","null"]}}]
00U01000p0m 00h0009709327430632898560진우의 보조개 박정수 ♥ 김희철 위너+박인서 >> 강김송이 ♥ 나의 빛은 김진우 #연정 Girl Group BNK48 ← ส่วนมากไร้สาระ

4กขม.#위 너만xvitaminz91(2018-03-06T13:32:59ZRT @NuaRa0r_ : ก็มีความน่าเอ็นดูอยู่ละ เครื่องดนตรี บบิ https://t.co/P38ZFzew5a href="https://twitter.com/NuaRa0r_/status/970925487563288576/photo/1$9709327430716579840Toy Rabbits Records【運営公式】 みんな嬉しい！楽しい！美味

conf "emptyfile" selected (0 bytes)

```

These are all steps we need to follow to analysis the realtime twitter data.