Assignment 11.2

Twitter data analysis using flume
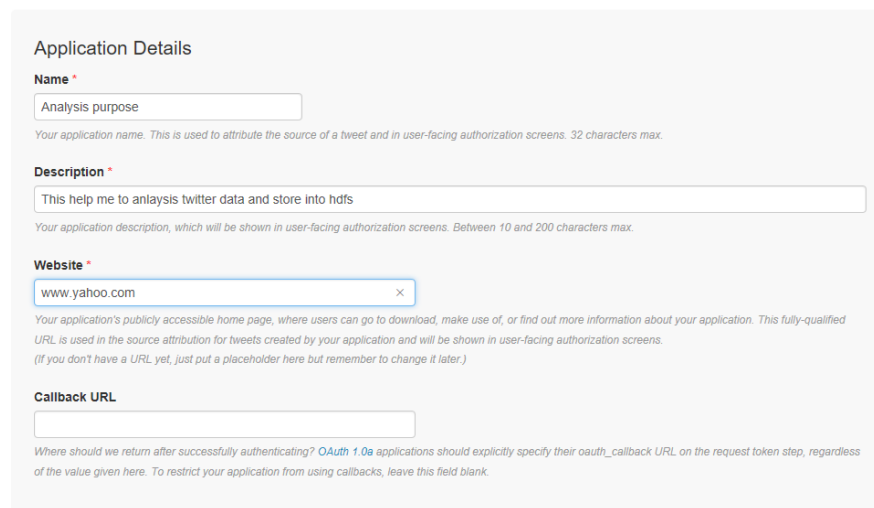
Step1: Login into twitter account ,then head to https://apps.twitter.com/
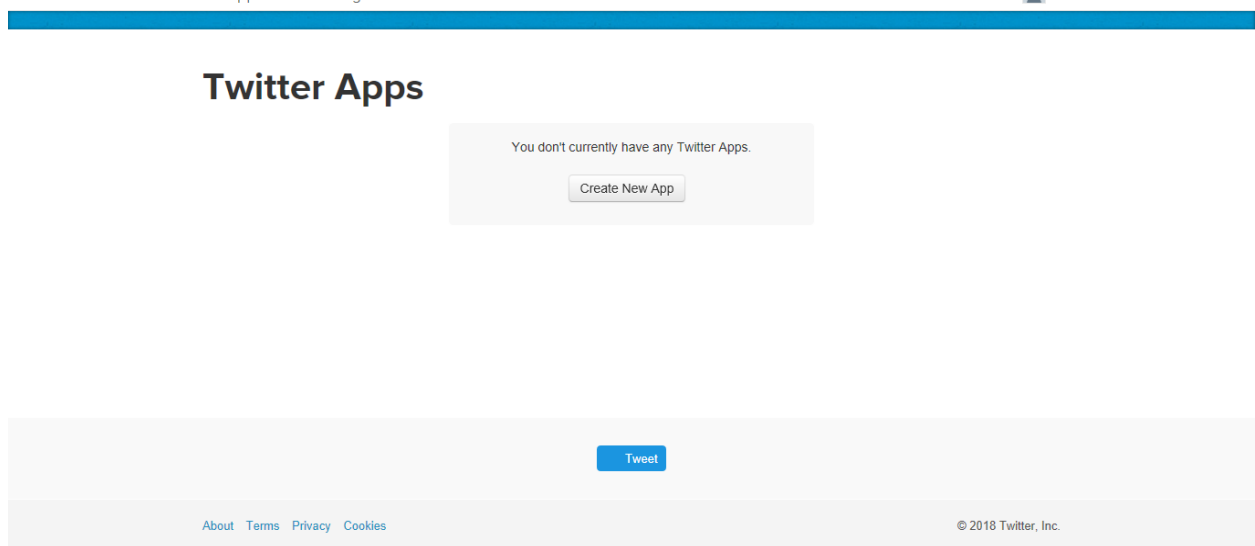
As shown below,

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

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)   XWrKiz8rY7EE5YZt1tmSuXl3fqChmlpvJvWWxQKxQdL6xQaVxk

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-
                              KaMuzXrzF2GrpCMD2rfIsBXW2VNL23O

Access Token Secret   ltEGKbSDsnDupeFZtCXtKBBo0mf1gcC6H56dpRojRuIoe

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-KaMuzXrzF2GrpCMD2rfIsBXW2VNL23O
Acess token scret key:ItEGKbSDsnDupeFZtCXtKBBo0mf1gcC6H56dpRojRuIoe
```

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 libra
ry 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 libra
ry for your platform... using builtin-java classes where applicable
```

Step7: Twiiter configuration file is placed inside /flume/conf directory as shown below,content is shown in next slide
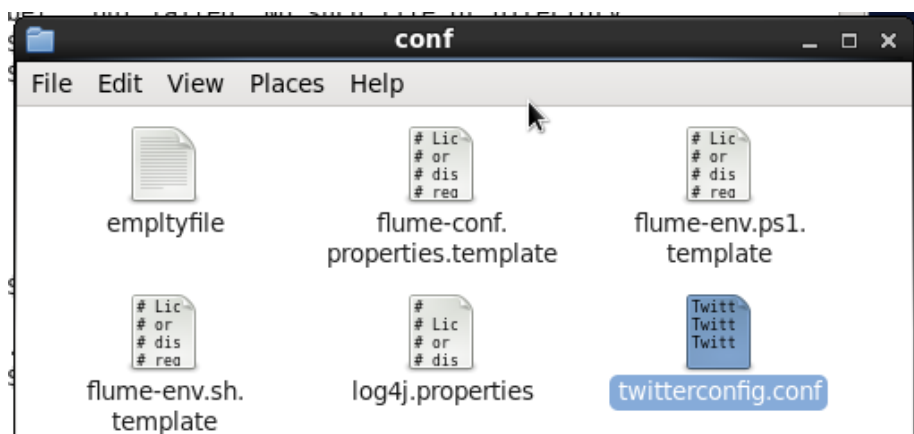
Step8: Access key ,Access Token, Consumer key ,Cosumer Secret are replaced with those are copied previously from our own application,and Keywords can also be changed as per our requirement as shown below

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

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

/user/flume/tweets   Go!

Show 25 ▾ entries   Search:

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	acadgild	supergroup	8 MB	Mar 06 11:42	1	128 MB	FlumeData.1520316136628
-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 feteched inside the files

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