

Assignment 11.3

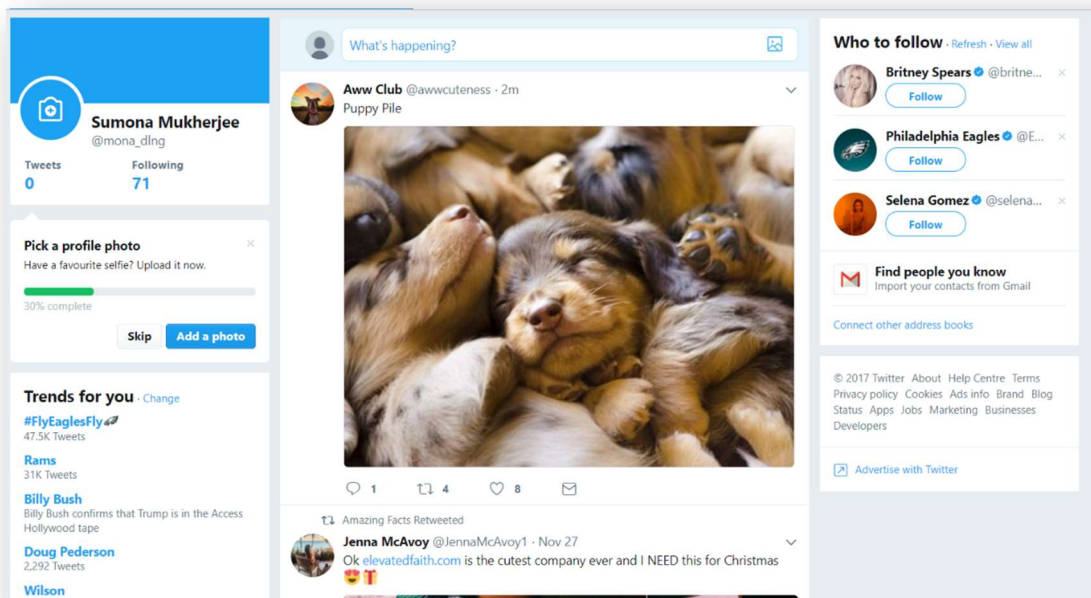
Prerequisites:

- Twitter account
- Hadoop cluster

First we need to download the jar file and extracted and transferred the jar files.

```
[acadgild@chemlabtest sumona-flume]$ ls -l |grep twitter
-rw-rw-r--. 1 acadgild acadgild 14733 Dec  5 14:14 flume-twitter-source-1.6.0.jar
-rw-rw-r--. 1 acadgild acadgild 284077 Dec  5 14:14 twitter4j-core-3.0.3.jar
-rw-rw-r--. 1 acadgild acadgild 27698 Dec  5 14:14 twitter4j-media-support-3.0.3.jar
-rw-rw-r--. 1 acadgild acadgild 56307 Dec  5 14:14 twitter4j-stream-3.0.3.jar
[acadgild@chemlabtest sumona-flume]$
```

Created the Twitter account



Created an application in twitter, application name is **monadIngApp**

Twitter Apps

You don't currently have any Twitter Apps.

Create New App

Create an application

Application Details

Name *

Your application name. This is used to attribute the source of a tweet and in user-facing authorization screens. 32 characters max.

Description *

Your application description, which will be shown in user-facing authorization screens. Between 10 and 200 characters max.

Website *

Your application's publicly accessible home page, where users can go to download, make use of, or find out more information about your application. This fully-qualified URL is used in the source attribution for tweets created by your application and will be shown in user-facing authorization screens.
(If you don't have a URL yet, just put a placeholder here but remember to change it later.)

Callback URL

Where should we return after successfully authenticating? OAuth 1.0a applications should explicitly specify their oauth_callback URL on the request token step, regardless of the value given here. To restrict your application from using callbacks, leave this field blank.

Updated the config file with the **consumer key and accessToken**, uploaded the config file in the location **/home/acadgild/Sumona-flume/acadgild.conf**

```
# Describing/Configuring the source
TwitterAgent.sources.Twitter.type = org.apache.flume.source.twitter.TwitterSource
TwitterAgent.sources.Twitter.consumerKey=FjTX19xyzJN9j1AHu02WKFdh3
TwitterAgent.sources.Twitter.consumerSecret=pExgocnLP68hLjf4sNL6Zz2GLAdw5c2SVZA3FjKXPAQz45ex4i
TwitterAgent.sources.Twitter.accessToken=937627607956201474-nxJm1xXX0F218te40XwsNGXT7koiP0r
TwitterAgent.sources.Twitter.accessTokenSecret=VH2YFGYR760nb8BfyoRfIhh8IpTIqu8J5GP1zVKYZqXMB
TwitterAgent.sources.Twitter.keywords=hadoop, bigdata, mapreduce, mahout, hbase, nosql
# Describing/Configuring the sink
```

Now we shall start all the Hadoop demons:

```
[acadgild@chemlabtest ~]$ jps
5456 Jps
2594 SecondaryNameNode
2341 NameNode
2437 DataNode
2773 ResourceManager
2879 NodeManager
[acadgild@chemlabtest ~]$
```

Now, lets create a new directory in Hadoop **hadoop fs -mkdir /user/acadgild/flume/tweets/**

```
[acadgild@chemlabtest ~]$ hadoop fs -mkdir /user/acadgild/flume/tweets
17/12/05 14:23:08 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[acadgild@chemlabtest ~]$
```

To fetch the Twitter data, we shall run the below command

flume-ng agent -n TwitterAgent -f /home/acadgild/Sumona-flume/acadgild.conf

This command will start fetching the data from Twitter and stream into the Hadoop file location mentioned in the config file.

```
[acadgild@chemlabtest sumona-flume]$ flume-ng agent -n TwitterAgent -f /home/acadgild/sumona-flume/acadgild.conf
Warning: No configuration directory set! Use --conf <dir> to override.
Info: Including Hadoop libraries found via (/usr/local/hadoop-2.6.0/bin/hadoop) for HDFS access
Info: Excluding /usr/local/hadoop-2.6.0/share/hadoop/common/lib/slf4j-api-1.7.5.jar from classpath
```

```

17/12/05 14:27:19 INFO twitter.TwitterSource: Processed 200 docs
17/12/05 14:27:23 INFO twitter.TwitterSource: Processed 300 docs
17/12/05 14:27:26 INFO twitter.TwitterSource: Processed 400 docs
17/12/05 14:27:29 INFO twitter.TwitterSource: Processed 500 docs
17/12/05 14:27:33 INFO twitter.TwitterSource: Processed 600 docs
17/12/05 14:27:36 INFO twitter.TwitterSource: Processed 700 docs
17/12/05 14:27:39 INFO twitter.TwitterSource: Processed 800 docs
17/12/05 14:27:42 INFO twitter.TwitterSource: Processed 900 docs
17/12/05 14:27:46 INFO twitter.TwitterSource: Processed 1,000 docs
17/12/05 14:27:46 INFO twitter.TwitterSource: Total docs indexed: 1,000, total skipped docs: 0
17/12/05 14:27:46 INFO twitter.TwitterSource: 26 docs/second
17/12/05 14:27:46 INFO twitter.TwitterSource: Run took 38 seconds and processed:
17/12/05 14:27:46 INFO twitter.TwitterSource: 0.007 MB/sec sent to index
17/12/05 14:27:46 INFO twitter.TwitterSource: 0.266 MB text sent to index
17/12/05 14:27:46 INFO twitter.TwitterSource: There were 0 exceptions ignored:
17/12/05 14:27:49 INFO twitter.TwitterSource: Processed 1,100 docs
17/12/05 14:27:53 INFO twitter.TwitterSource: Processed 1,200 docs
17/12/05 14:27:56 INFO twitter.TwitterSource: Processed 1,300 docs
17/12/05 14:27:59 INFO twitter.TwitterSource: Processed 1,400 docs
17/12/05 14:28:03 INFO twitter.TwitterSource: Processed 1,500 docs
17/12/05 14:28:06 INFO twitter.TwitterSource: Processed 1,600 docs
17/12/05 14:28:09 INFO twitter.TwitterSource: Processed 1,700 docs
17/12/05 14:28:12 INFO twitter.TwitterSource: Processed 1,800 docs
17/12/05 14:28:16 INFO twitter.TwitterSource: Processed 1,900 docs
17/12/05 14:28:19 INFO twitter.TwitterSource: Processed 2,000 docs
17/12/05 14:28:19 INFO twitter.TwitterSource: Total docs indexed: 2,000, total skipped docs: 0
17/12/05 14:28:19 INFO twitter.TwitterSource: 28 docs/second
17/12/05 14:28:19 INFO twitter.TwitterSource: Run took 71 seconds and processed:
17/12/05 14:28:19 INFO twitter.TwitterSource: 0.007 MB/sec sent to index
17/12/05 14:28:19 INFO twitter.TwitterSource: 0.532 MB text sent to index
17/12/05 14:28:19 INFO twitter.TwitterSource: There were 0 exceptions ignored:
17/12/05 14:28:22 INFO twitter.TwitterSource: Processed 2,100 docs
17/12/05 14:28:26 INFO twitter.TwitterSource: Processed 2,200 docs
17/12/05 14:28:29 INFO twitter.TwitterSource: Processed 2,300 docs
17/12/05 14:28:32 INFO twitter.TwitterSource: Processed 2,400 docs
17/12/05 14:28:36 INFO twitter.TwitterSource: Processed 2,500 docs
17/12/05 14:28:39 INFO twitter.TwitterSource: Processed 2,600 docs
17/12/05 14:28:42 INFO twitter.TwitterSource: Processed 2,700 docs
17/12/05 14:28:45 INFO twitter.TwitterSource: Processed 2,800 docs
17/12/05 14:28:49 INFO twitter.TwitterSource: Processed 2,900 docs
17/12/05 14:28:52 INFO twitter.TwitterSource: Processed 3,000 docs
17/12/05 14:28:52 INFO twitter.TwitterSource: Total docs indexed: 3,000, total skipped docs: 0
17/12/05 14:28:52 INFO twitter.TwitterSource: 28 docs/second
17/12/05 14:28:52 INFO twitter.TwitterSource: Run took 104 seconds and processed:
17/12/05 14:28:52 INFO twitter.TwitterSource: 0.008 MB/sec sent to index
17/12/05 14:28:52 INFO twitter.TwitterSource: 0.797 MB text sent to index
17/12/05 14:28:52 INFO twitter.TwitterSource: There were 0 exceptions ignored:
17/12/05 14:28:56 INFO twitter.TwitterSource: Processed 3,100 docs
17/12/05 14:28:59 INFO twitter.TwitterSource: Processed 3,200 docs

```

Let us check the Hadoop file location and cat the file


```
[acadgild@chemlabtest ~]$ hadoop fs -ls /user/acadgild/flume/tweets
17/12/05 14:32:09 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 0 2017-12-05 14:31 /user/acadgild/flume/tweets/FlumeData.1512464504275.tmp
[acadgild@chemlabtest ~]$ hadoop fs -cat /user/acadgild/flume/tweets/FlumeData.1512464504275.tmp
17/12/05 14:32:36 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where 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 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"]}]}
```

51r \$93797023227799040 한국여과 ZVIXX Leo 정격은 / 마이스타그램 4

리 xx tw00(2017-12-05T14:31:44Z) 난で日本語にするの？

死にたあ https://t.co/RB93cv0VpkTwitter for iPhonehttps://pbs.twimg.com/media/DQRW3m0UEAIpnk4.jpghttps://twitter.com/xx_tw00/status/93797023227799040/photo/1

51r

```

{"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"]}]}
```

70976Planaltina, Brasília(Chama No Probleminha)nenémthamirescaval17(2017-12-05T14:31:45Z)Nem to acreditando ainda que clonara m a placa do carro da minha mãeTwitter for Android\$937970236430790473am عبد الرحمنNadoo0000000(2017-12-05T14:31:45Z)RT @Y0ez8TFH0FjRW8wمعاذ الله استعاد لا مع ل ذلك النوع من الا استعاد يسمى أ كفاء نام من أسTwitter for iPhone\$937970236418084864East LondonIn all honesty I don't know what the f... I'm doing but asikho laphoXZM2eak israel(2017-12-05T14:31:45Z)RT @Anele: Do not go home to see family over the holidays if it damages your mental health. #tipsfortheholidayTwitter for iPhone\$93797023639709696p(مفعو)للا علا نان.خامTwitter for iPhone\$937970236418084864الاستفسارلتابعي الحسابال يعلن للموصلين وطلب النقل والمندوبينTwitter for iPhone\$937970236418084864DalekoinRiyadh(2017-12-05T14:31:45Z)إلىدليلكم في الربامTwitter for iPhone\$937970236418084864ابن مدرسه اهليه كويسه بشمال الربامTwitter for iPhone\$937970236418084864الربام مدراس#лочка с присыпкой в виде боли2штолицStolzAkira(2017-12-05T14:31:45Z)RT @twibbley: Однажды у человека не принявшего гитлера в Художку спросили каково это совершать самую большую ошибку а он ответил...Twitter for Android\$937970236426297345TLIVE:28.快感昼夜 29.0SIRISファンミ毎日いろんな推しのおかげとノリで生きてます。ONE PIECE Free! いきものがかり 三浦春馬 青伊東歌詞太郎 イトラカシ un:c ウォルビスカーター XYZ アニメ 声優 ソシャゲ i7 バ！ A3! シャニライ シノアリス 戦ブラなど多趣味マン Sens i:ちゃんぬるちゃん ゆきの*aniu_no(2017-12-05T14:31:45Z)三角かわいTwitter for iPhone\$937970236418084864

You will notice a file **FlumeData.1512464504275.tmp** has been created in **/user/acadgild/flume/acadgild**

Once we cat the file **/user/acadgild/flume/tweets/FlumeData.1512464504275.tmp**, we will find the Twitter data.