Assignment 11.3

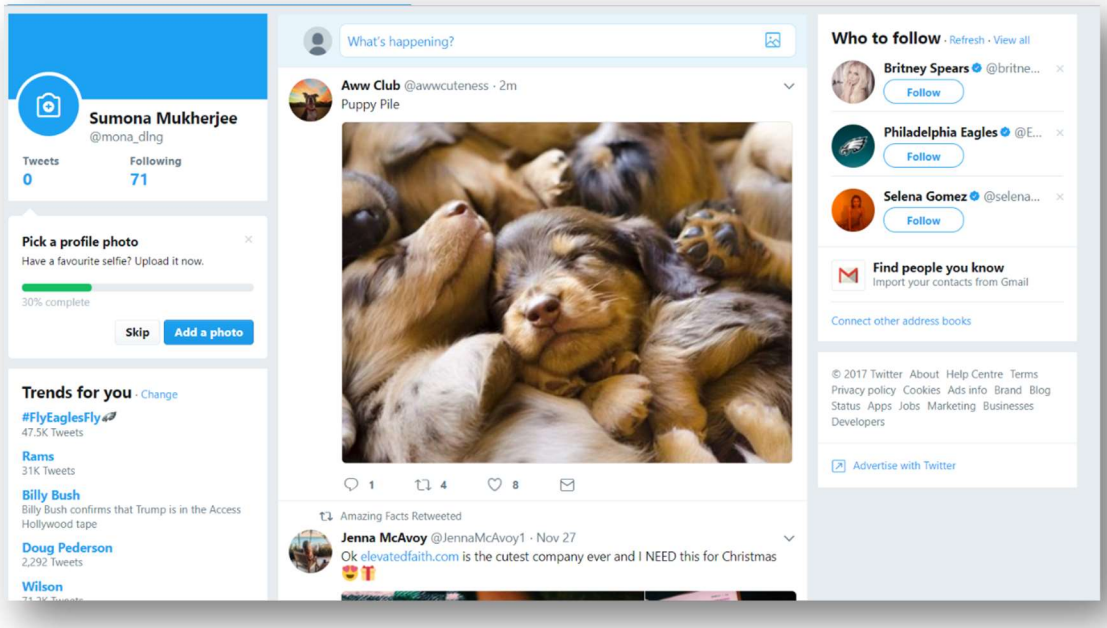
Prerequisites:

- Twitter account
- Hadoop cluster

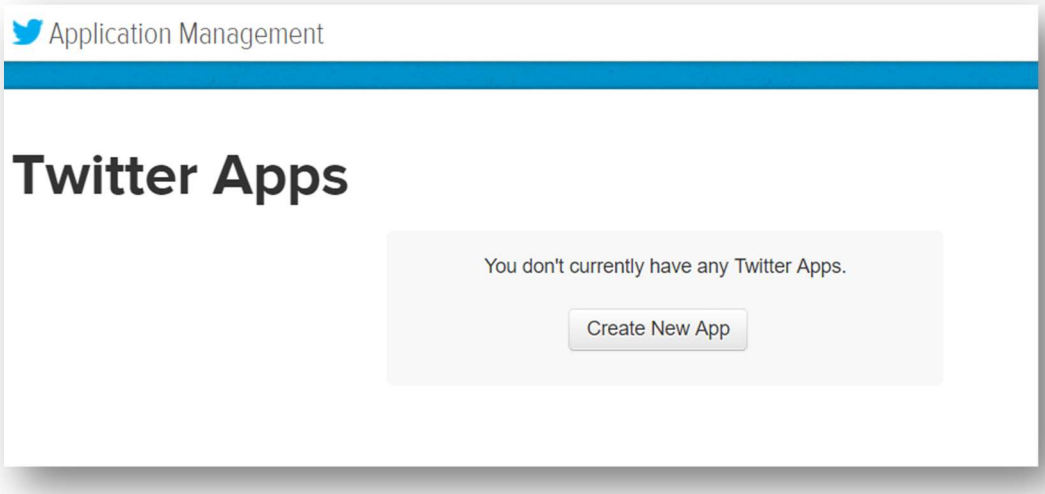
First we need to download the jar file and extracted and transferred the jar files.

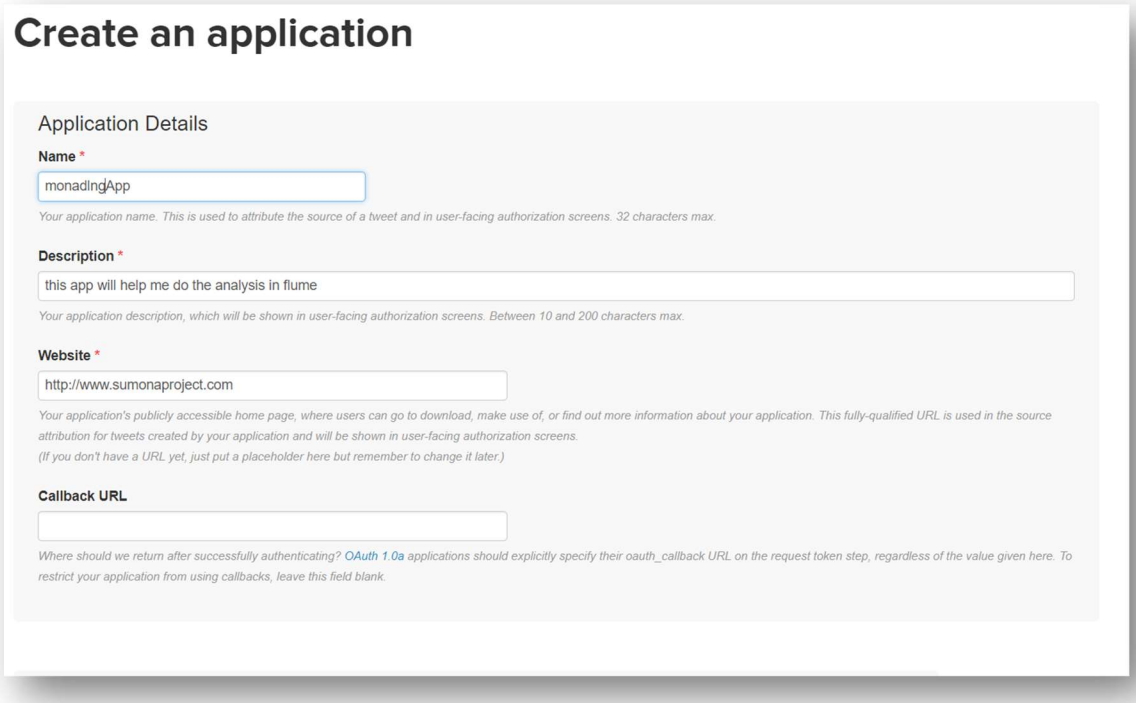
```
[acadgild@chemlabtest sumona-flume]$ ls -l |grep twitter
-rw-rw-r--. 1 acadgild acadgild   14733 Dec  5 14:14 flume-twitter-source-1.6.0.jar
-rw-rw-r--. 1 acadgild acadgild  284077 Dec  5 14:14 twitter4j-core-3.0.3.jar
-rw-rw-r--. 1 acadgild acadgild   27698 Dec  5 14:14 twitter4j-media-support-3.0.3.jar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





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=pExgocnLP6BhLjf4sNL6Zz2GLAdw5c2SVZA3FjKXPAQz45ex4i
TwitterAgent.sources.Twitter.accessToken=937627607956201474-nxJm1xXXOF218te4OXwsNGXT7koiP0r
TwitterAgent.sources.Twitter.accessTokenSecret=VH2YFGYR760nbBbfyoRfIhhBIpTIquBJ5GP1zVKYZqXMb
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/**

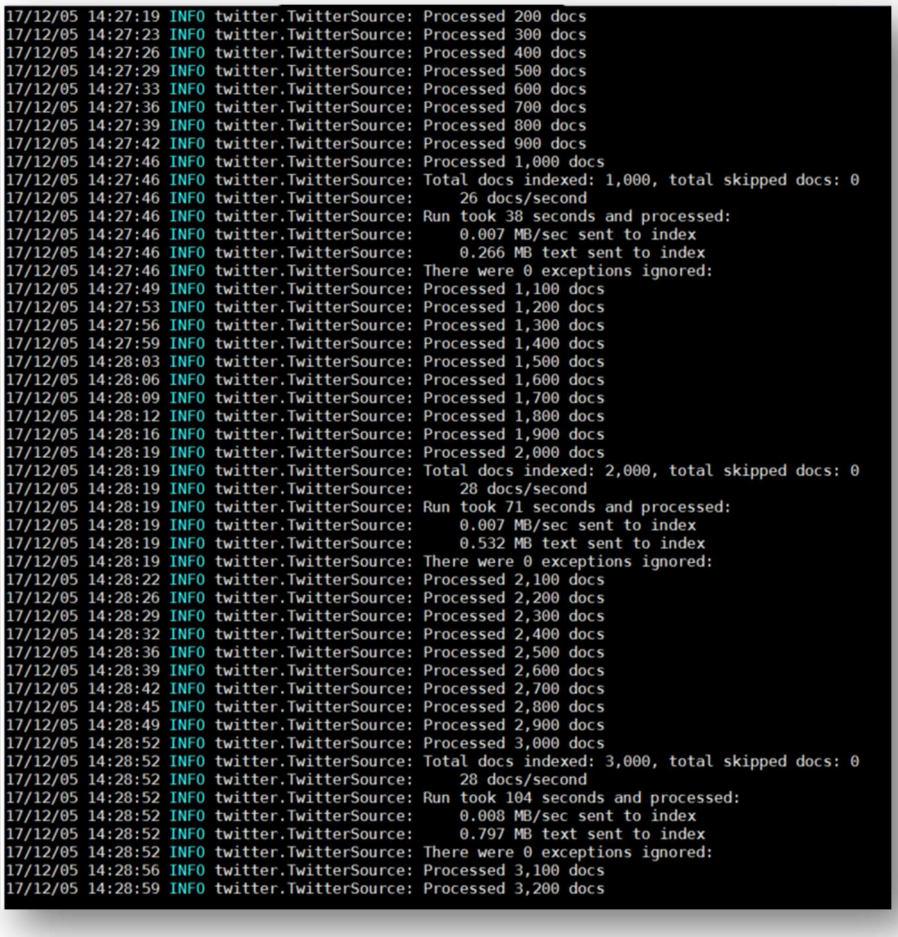
```
[acadgild@chemlabtest ~]$ hadoop fs -mkdir /user/acadgild/flume/tweets
17/12/05 14:23:08 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes
 where applicable
[acadgild@chemlabtest ~]$
```

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
```



Let us check the Hadoop file location and cat the file

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.