**ASSIGNMENT 11.3**

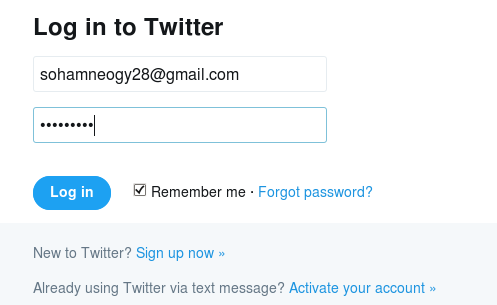
PROBLEM STATEMENT:

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

Step 1:

Login to twitter account:

Output:

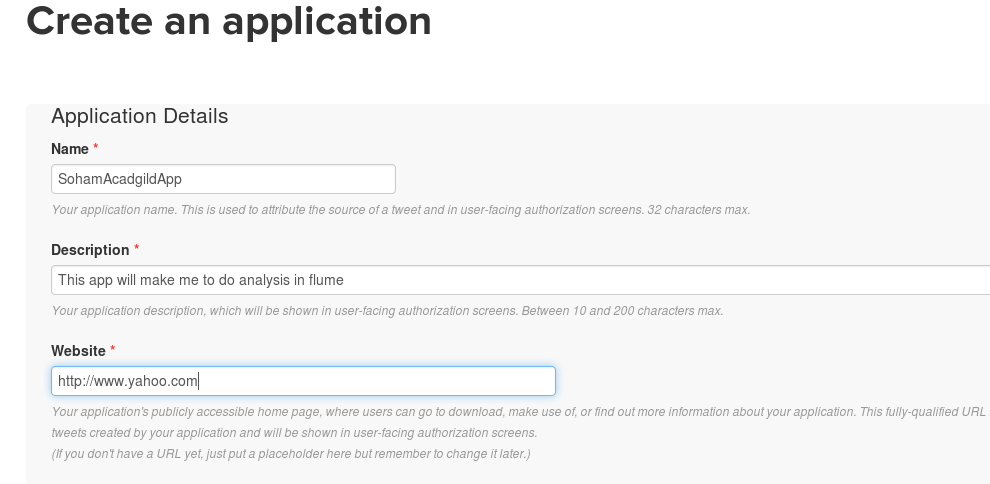


Step 2:

Go to the following link and click the ‘create new app’ button.

<https://apps.twitter.com/app>

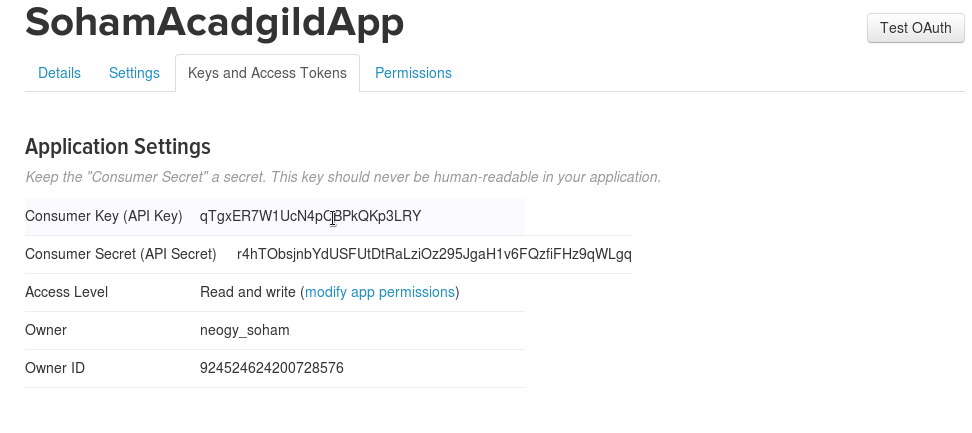
Output:



Step 3:

Select the ‘Keys and Access Token’ tab.

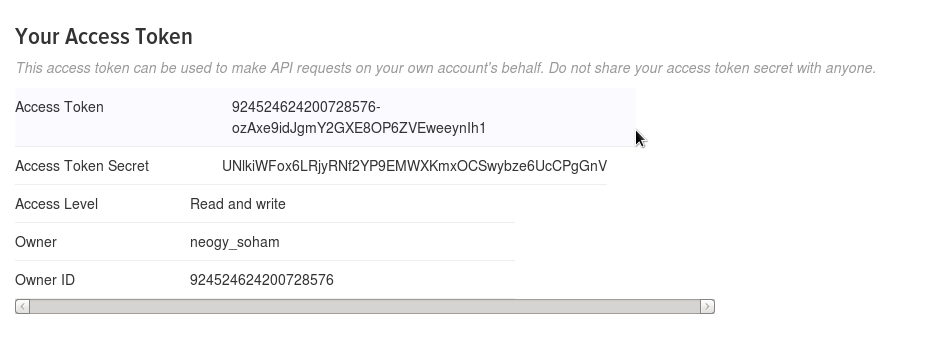
Output:



Step 4:

Scroll down further and select the ‘create my access token’ button.

Output:



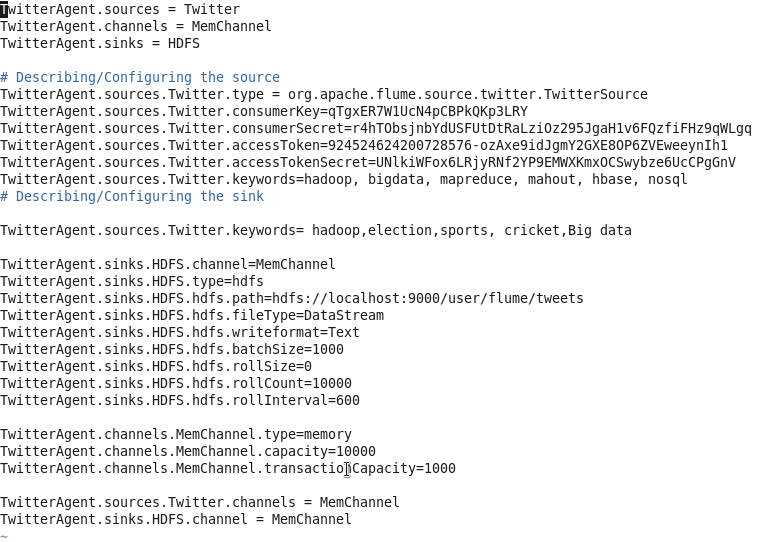
Step 5:

Copy the Flume configuration code from the given link and paste it in the newly created file.

https://drive.google.com/open?id=0B1QaXx7tpw3Sb3U4LW9SWlNidkk

Change the twitter api keys with the keys generated

Output:

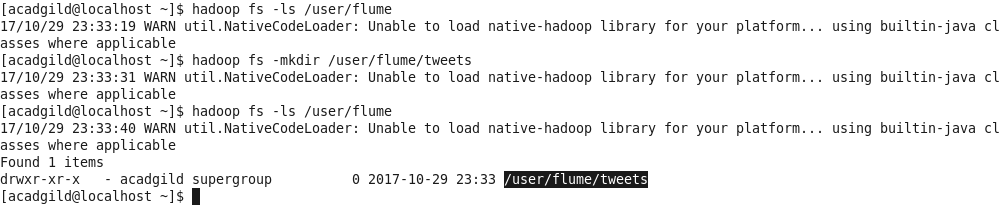


Step 6:

Open the hdfs file system and run the servers.

Create a new directory in hdfs :

Output:



Step 7:

For fetching data from Twitter, Use the below command to fetch the twitter tweet data into the HDFS cluster path.

**flume-ng agent -n TwitterAgent -f <location of created/edited conf file>**

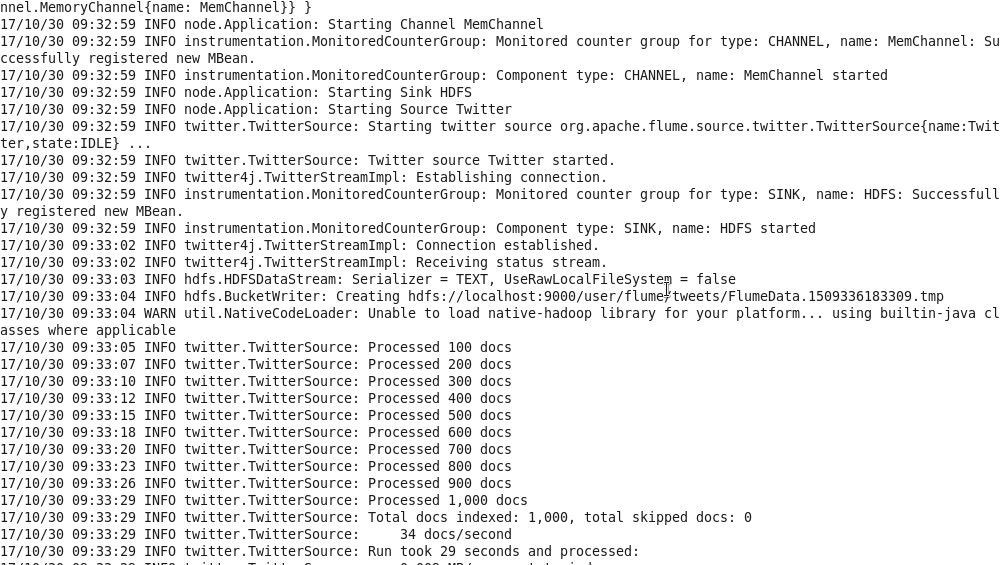
Output:



Step 8:

The above command will start fetching data from Twitter and steams it into the HDFS given path.

Output:

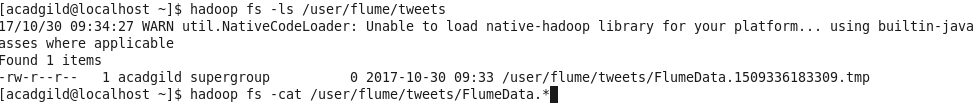


Step 9:

To check the contents of the tweet data we can use the following command:

**hadoop fs –ls /user/flume/tweets**

Output:



Step 10:

We use ‘cat’ command to display the tweet data inside the /user/flume/tweets/FlumeData.\*

Output:

