

**Solution-**

Apache Flume is a reliable service used for efficiently collecting, aggregating, and moving large amounts of log data.

The next question on your mind would be, “How does Flume do this?”

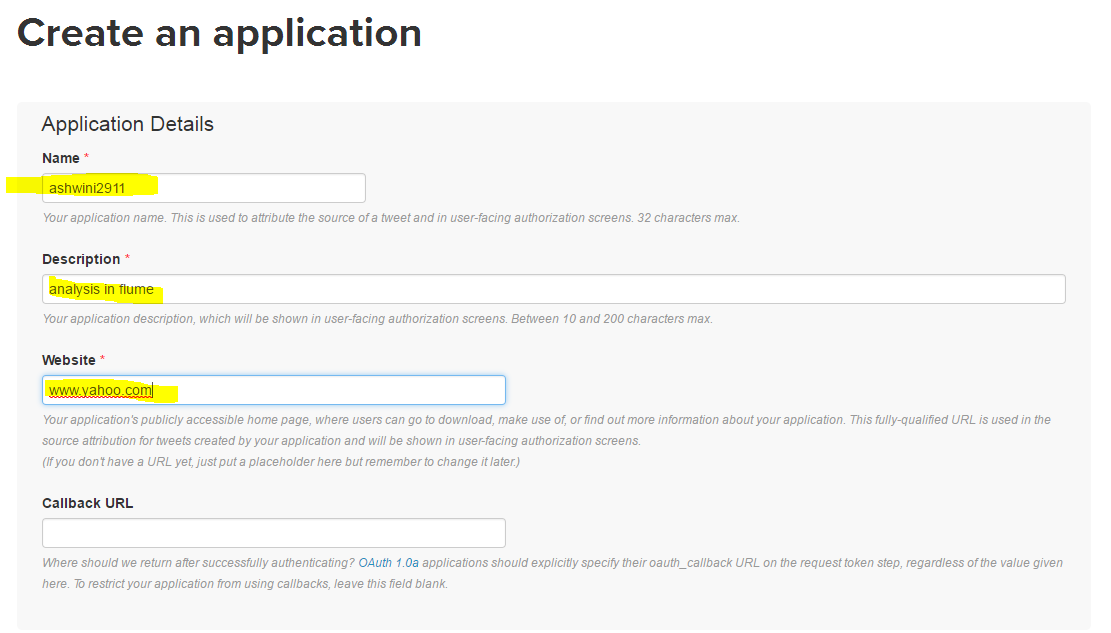
*Read on to find out the answer for your question.*

### **Step-by-step Tutorial: Data Streaming from Twitter to HDFS**

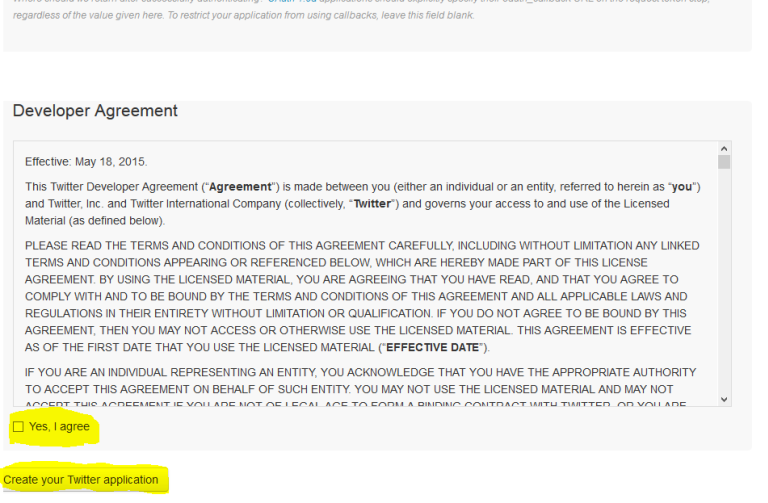
Step 1: Go to the following link and click on ‘create app’.



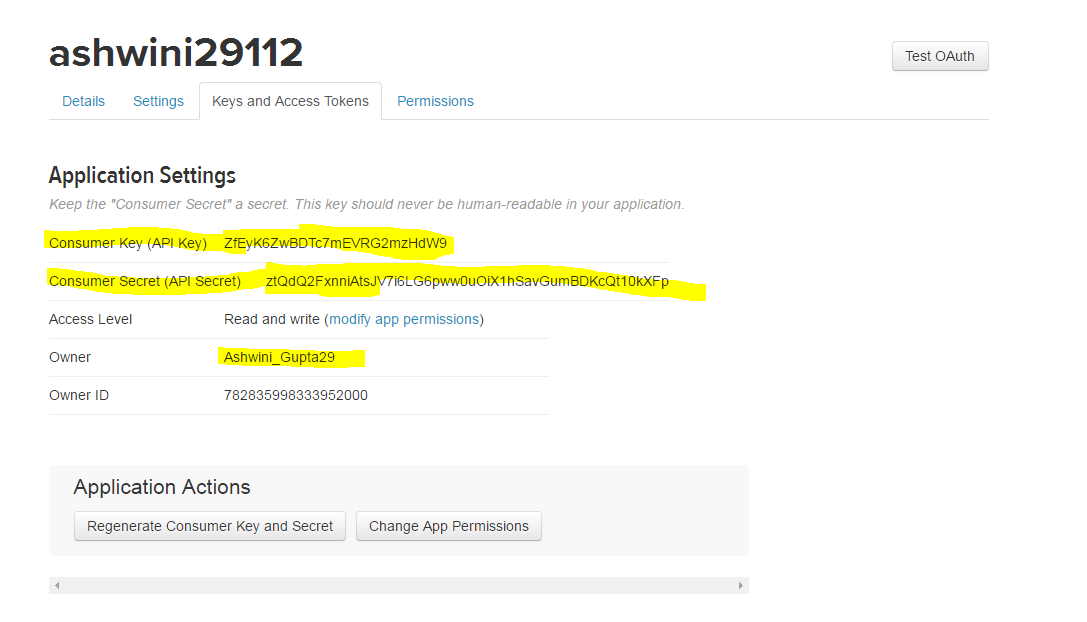
Step 2: Fill in the necessary details.



Step 3: Accept the agreement and click on ‘create your Twitter application’.

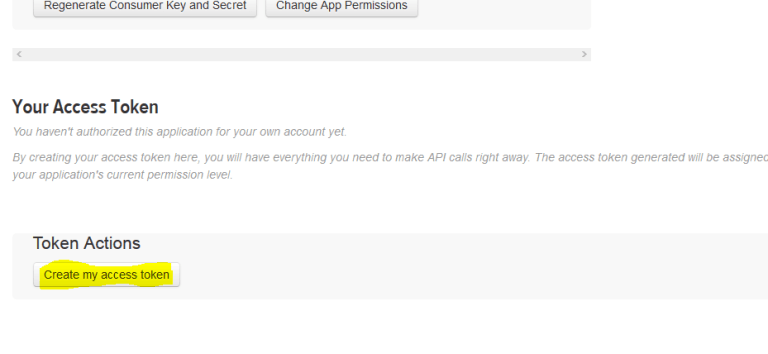


Step 4: Go to ‘Keys and Access Token’ tab.

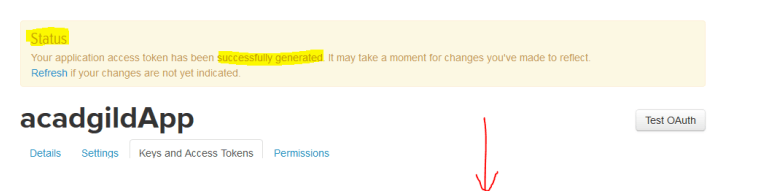


Step 5: Copy the consumer key and the consumer secret.

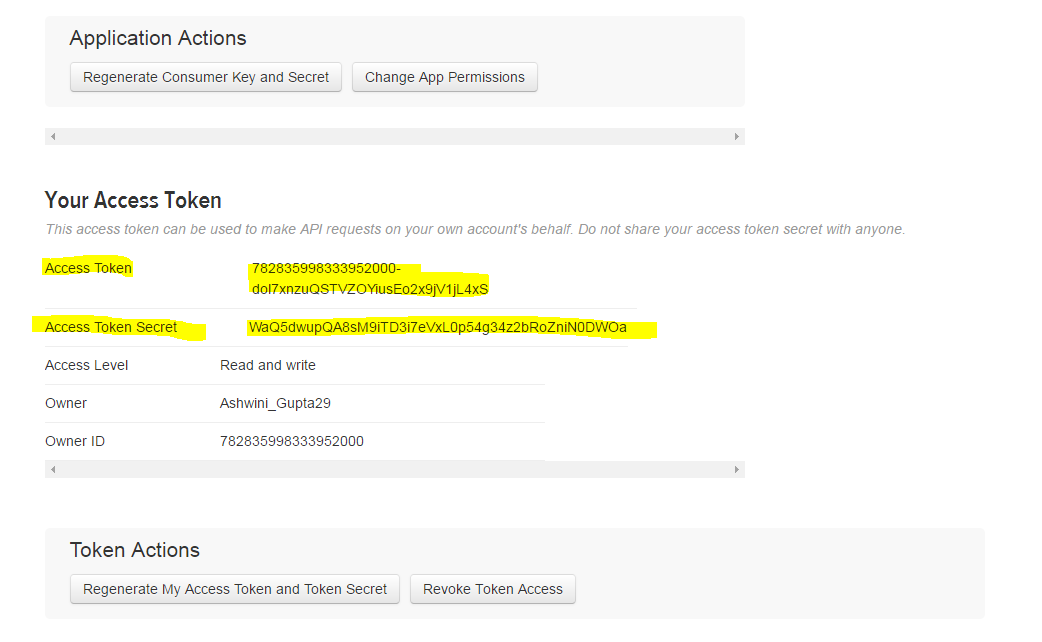
Step 6: Scroll down further and click on ‘create my access token’.



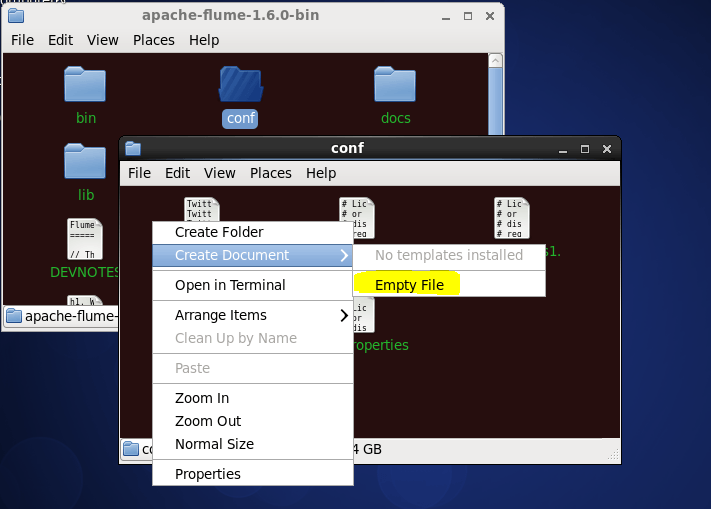
You will now receive a message that says that you have successfully generated your application access token.



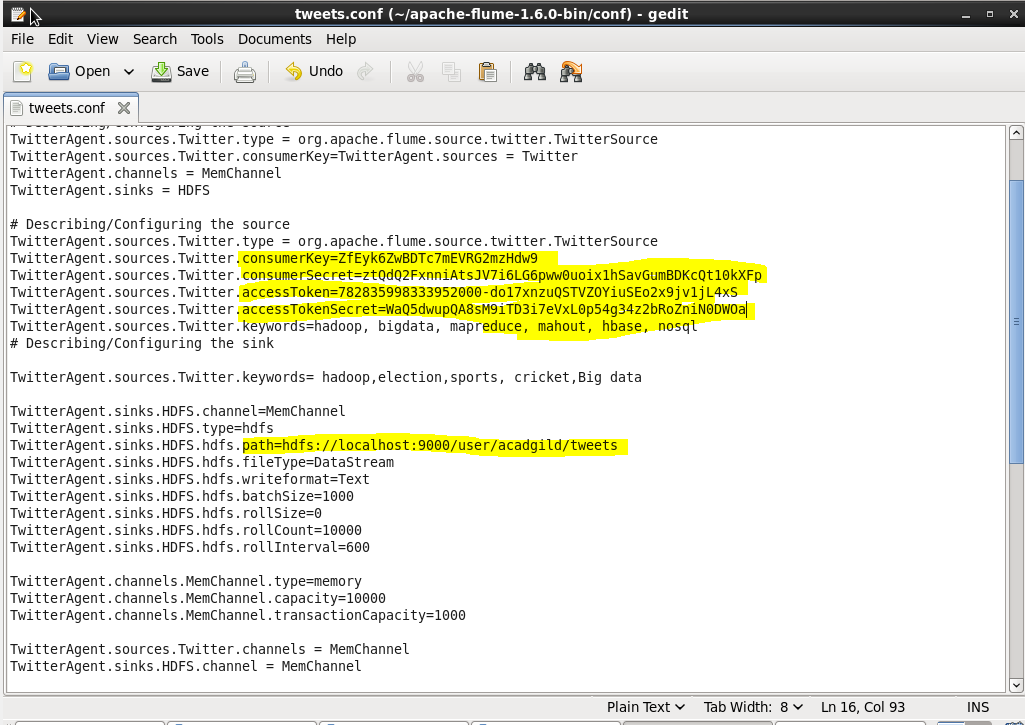
Step 7: Copy the Access Token and Access token Secret.



Step 8: Create a new file inside the ‘conf’ directory inside the Flume-extracted directory.



Step 9: Change the twitter api keys with the keys generated as shown in the step no 6 and step number 8.

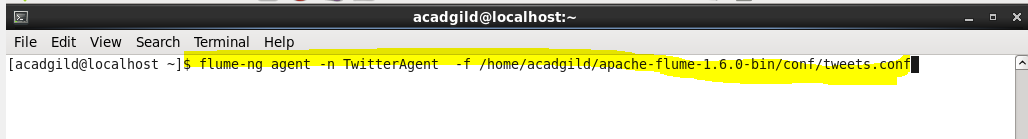


Step 10: Using the below command, create a directory inside HDFS where Twitter data will be stored.

Hadoop dfs –mkdir –p /user/acadgild/tweets

Step 11: For fetching data from Twitter, give the below command in the terminal.

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



This will start fetching data from Twitter and send it to the HDFS.



To stop fetching data, press *‘Ctrl+c’*. This will end the process of fetching the data.

Step 12: To check the contents of the Tweets folder, use the following command:

hadoop dfs –ls /user/flume/tweets