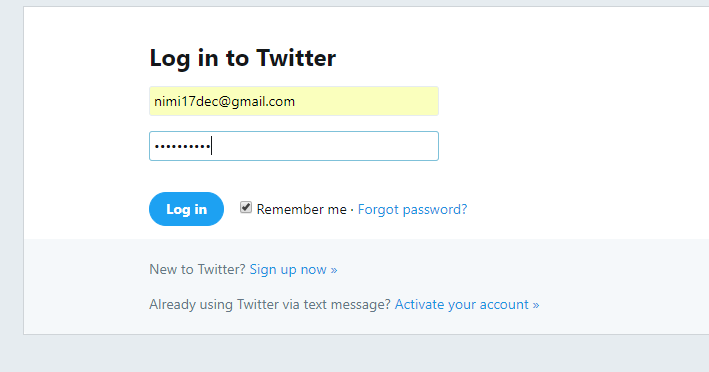
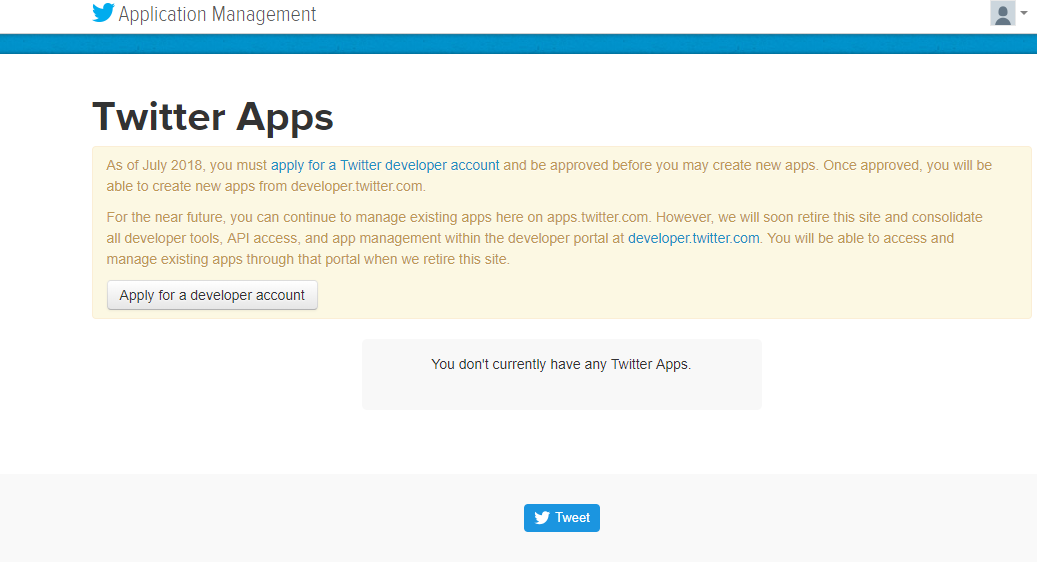
**Create a flume agent that streams data from Twitter and stores in the HDFS.**

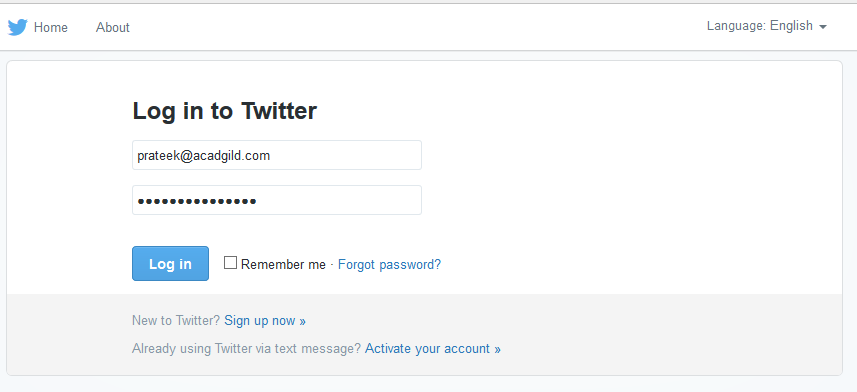
Note : As of now, twitter as blocked to create a developer account.





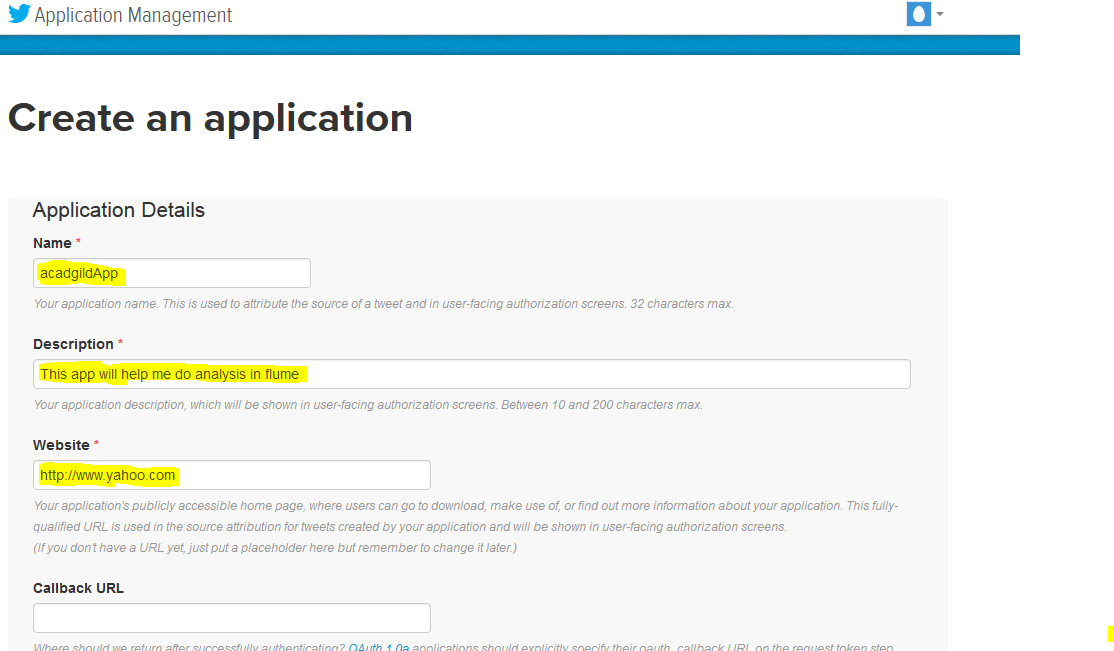
**So as per Acadgild’s support team :** they mentioned to follow and learn the procedure shared in the blog and share the same steps in the assignment solution link in the git hub for further evaluation.

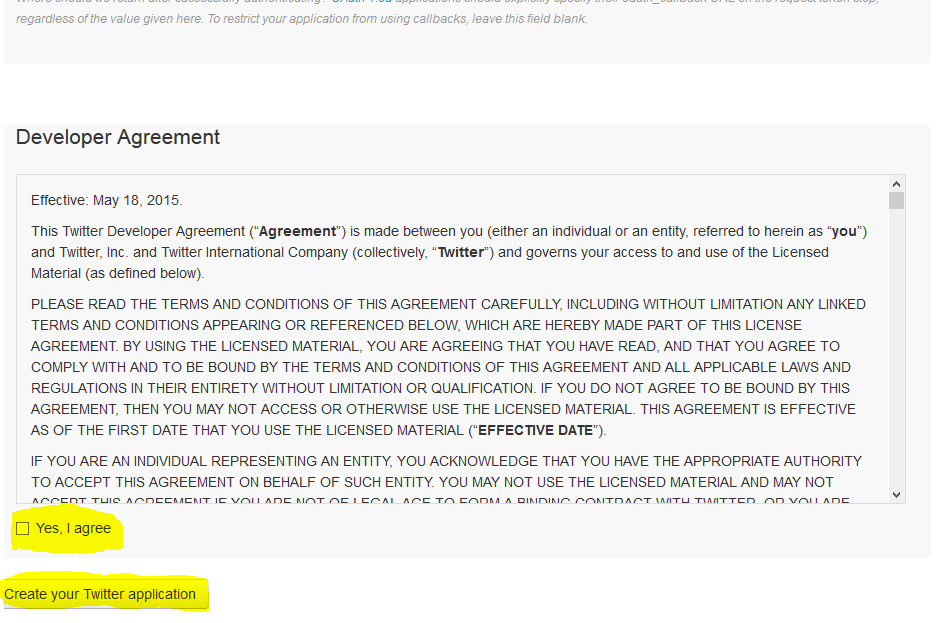
**Step 1:**  
Login to  the twitter account



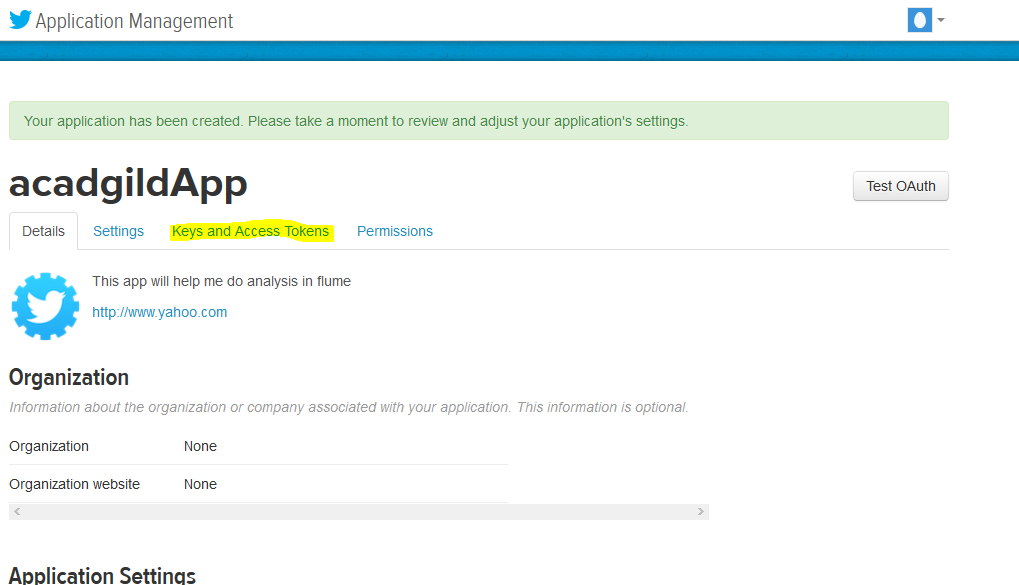
**Step 2:**

**Go to the following link and click the  ‘create new app’ button.**  
[**https://apps.twitter.com/app**](https://apps.twitter.com/app)  
  
  
**Step 3:**  
**Enter the necessary details.**



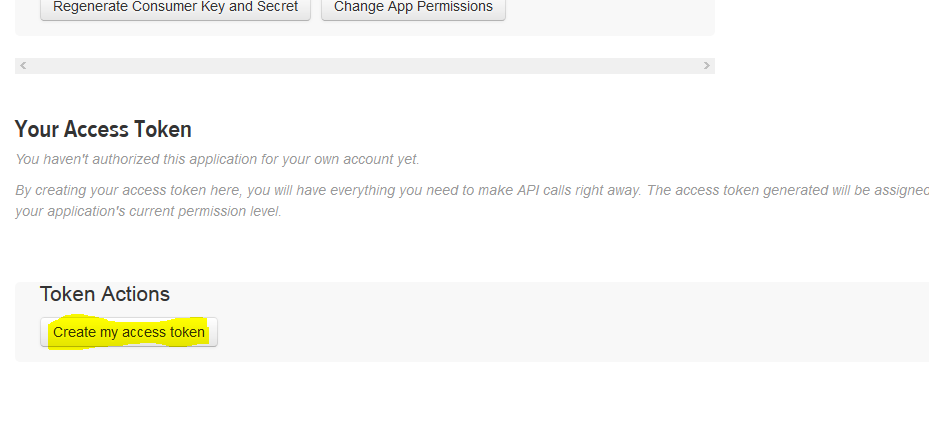
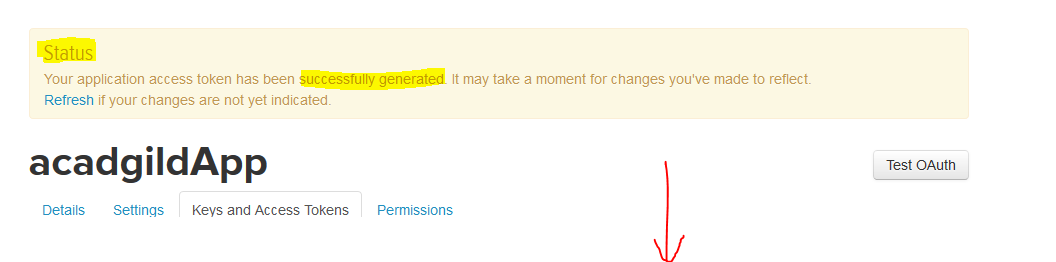
**Step 4:**  
Accept the developer agreement and select the ‘create your Twitter application’ button.  


**Step 5:**  
Select the ‘Keys and Access Token’ tab.

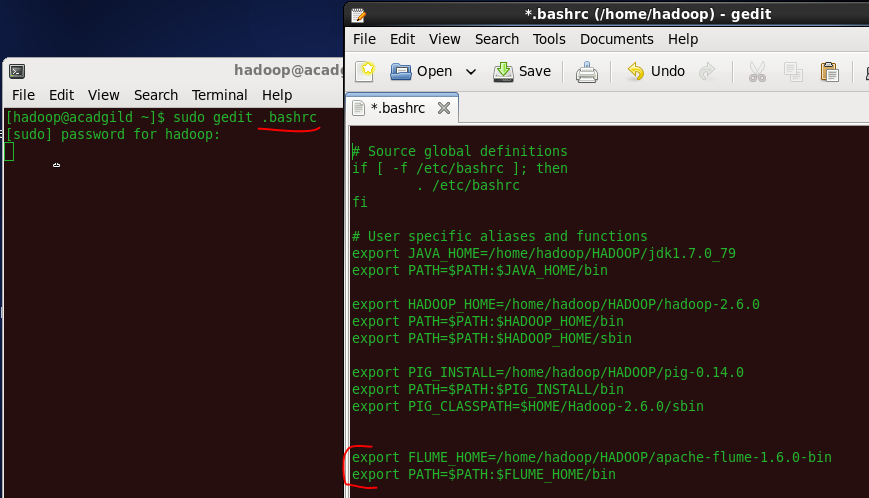


**Step 6:**  
Copy the consumer key and the consumer secret code.

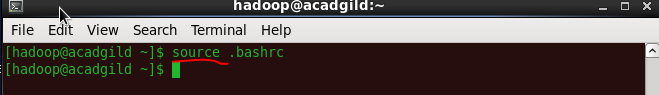
**Step 7:**  
Scroll down further and select the ‘create my access token’ button.

  
Now, you will receive a message stating “that you have successfully generated your application access token”.  


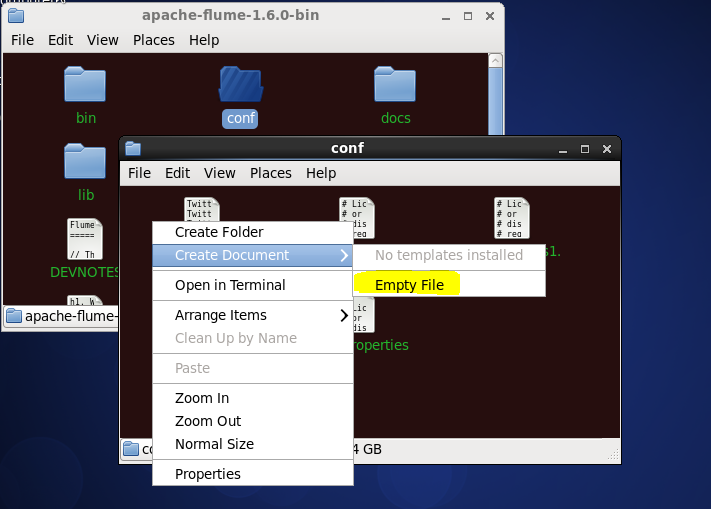
**Step 8:**  
Copy the Access Token and Access token Secret code.

Follow Step 9 and Step 10 to install Apache flume  
**Step 9:**Download flume tar file from below link and extract it.  
[**https://drive.google.com/drive/u/0/folders/0B1QaXx7tpw3SWkMwVFBkc3djNFk**](https://drive.google.com/drive/u/0/folders/0B1QaXx7tpw3SWkMwVFBkc3djNFk)  
Right click on the downloaded flume tar file and select the option as Extract Here to untar the flume directory and update the path of extracted flume directory in the .bashrc file as mentioned in the below image.  
**NOTE: keep the path same as where the extracted file exists.**  


After setting the path of flume directory, save and close the .bashrc file.  And then in the terminal type the below command to update the .bashrc file.

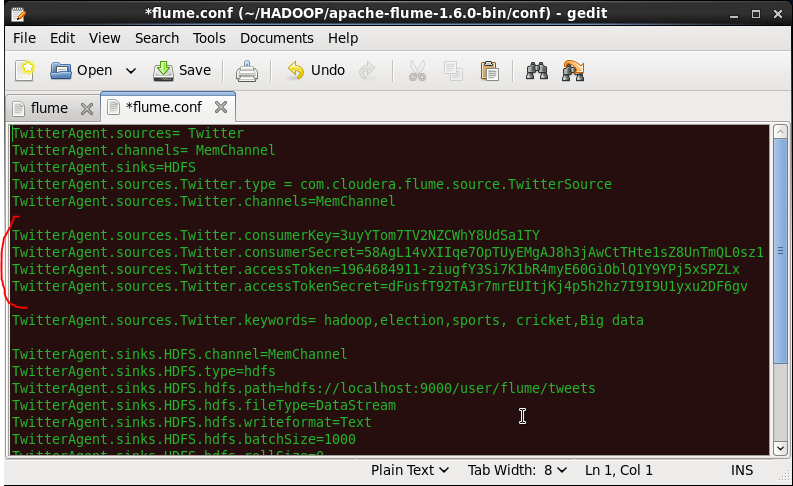


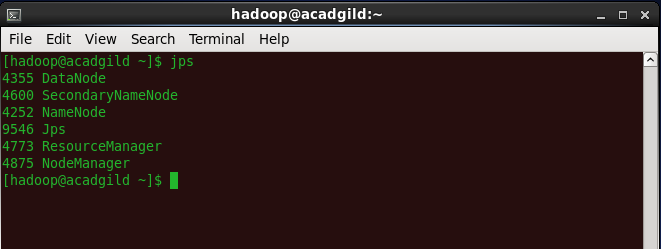
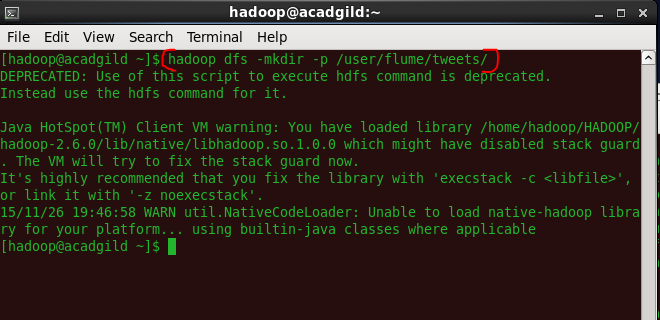
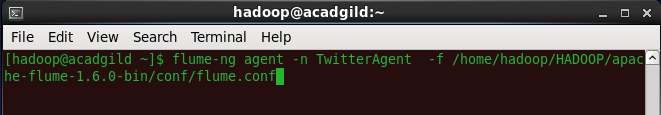
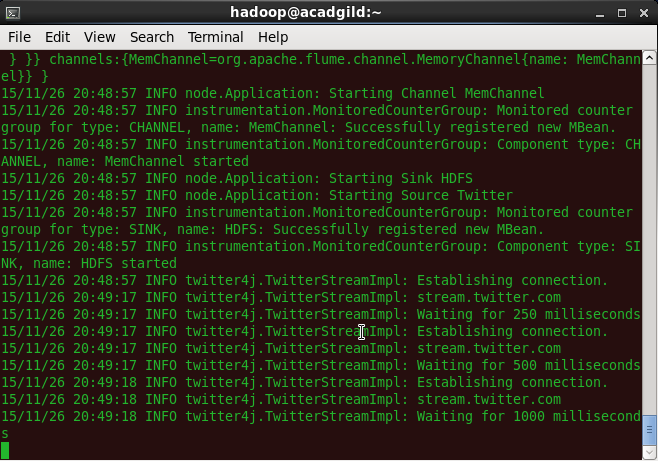
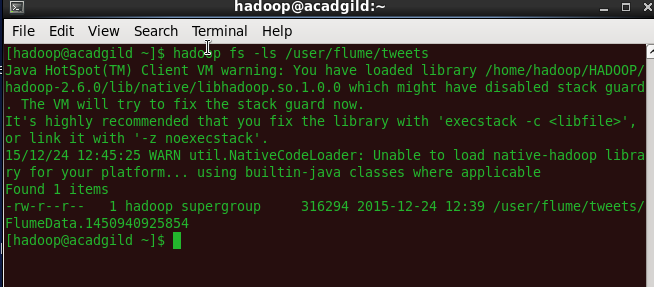
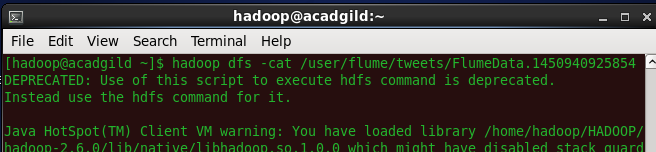
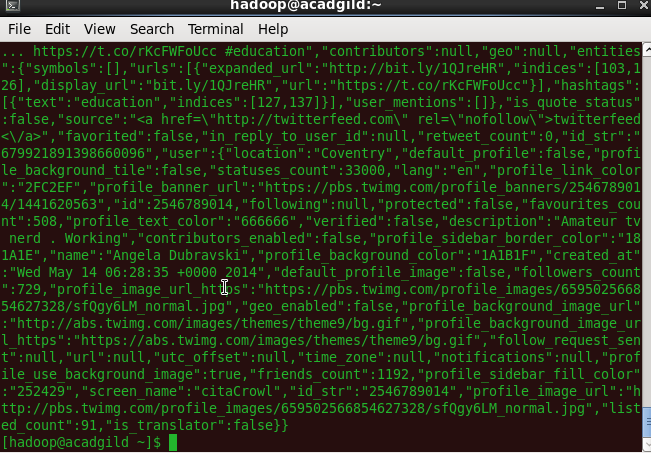
**Step 10:**  
Create a new file inside the conf directory inside the Flume-extracted directory.



Note: Make sure you have below jars placed in your $FLUME\_HOME/lib directory:

1. twitter4j-core-X.XX.jar
2. twitter4j-stream-X.X.X.jar
3. twitter4j-media-support-X.X.X.jar

**Step 11:**  
Copy the Flume configuration code from the below link and paste it in the newly created file.  
<https://drive.google.com/open?id=0B1QaXx7tpw3Sb3U4LW9SWlNidkk>  
**Step 12:**  
Change the twitter api keys with the keys generated as shown in the step no 6 and step number 8.  


**Step 13:**  
We have to decide which keywords tweet data to be collected from the twitter application. So, you can change the keywords in the TwitterAgent.sources.Twitter.keywords command.   
In our example, we are fetching tweet data related to Hadoop, election, sports, cricket and Big data.  
**Step 14:**  
Open a new terminal and start all the Hadoop daemons, before running the flume command to fetch the twitter data.   
Use the ‘jps’ command to see the running Hadoop daemons.  
****  
**Step 15:**  
Create a new directory inside HDFS path, where the Twitter tweet data should be stored.  
**Hadoop dfs –mkdir –p /user/flume/tweets**  
****  
**Step 16:**  
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>**  
  
The above command will start fetching data from Twitter and steams it into the HDFS given path.  
  
Once, the tweet data started streaming it into the given HDFS path we can use ‘Ctrl+c’ command to stop the streaming process.   
**Step 17:**  
To check the contents of the tweet data we can use the following command:  
**hadoop dfs –ls /user/flume/tweets**  
  
**Step 18:**  
We can use the *‘cat’* command to display the tweet  data inside the /user/flume/tweets/FlumeData.145\* path.  
**hadoop dfs –cat /us er/flume/tweets/<flumeData file name>**  
  
  
We can observe from the above image that we have successfully fetched twitter data into our HDFS cluster directory.  Once the tweets have been successfully stored in your database, you can manipulate the tweet data to fit the needs of our future projects.