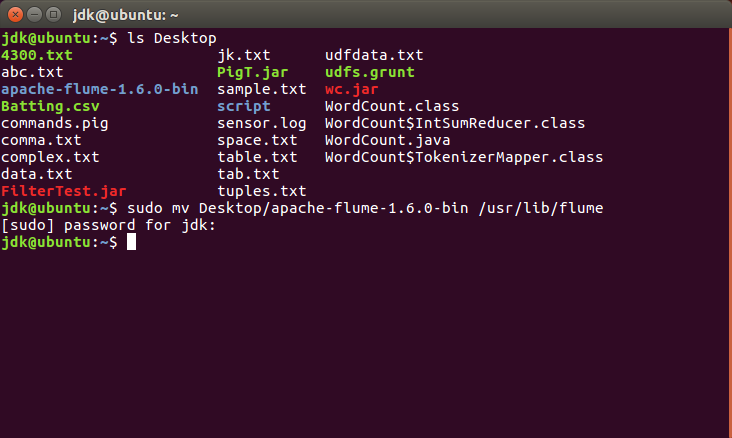
**INSTALL FLUME IN YOUR SYSTEM**

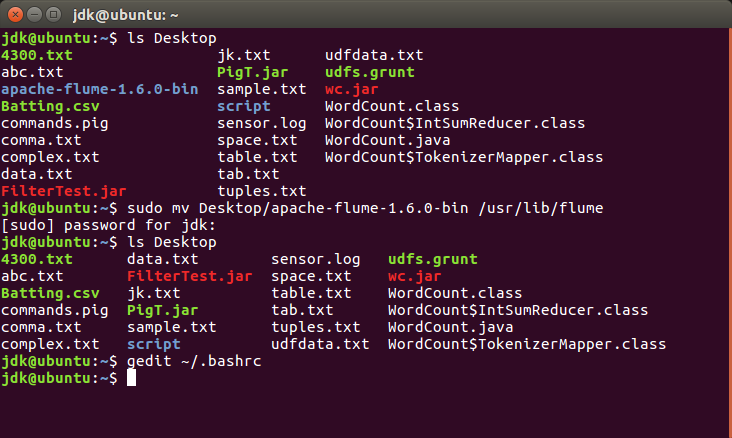
**AND EXTRACT TWITTER DATA**

We start with the Live streaming of data as follows:

1. First we will install flume in the system. Copy the tar file of the apache flume and then extract it on the desktop. Now we move this extracted file from the desktop to the destination **/usr/lib/flume** using the **sudo mv** command.



1. To check if the file has been moved, we use the command **ls.** Now open the bashrc file as shown.



1. We add the following lines in this file for configuration purposes:

**export FLUME\_HOME=/usr/lib/flume**

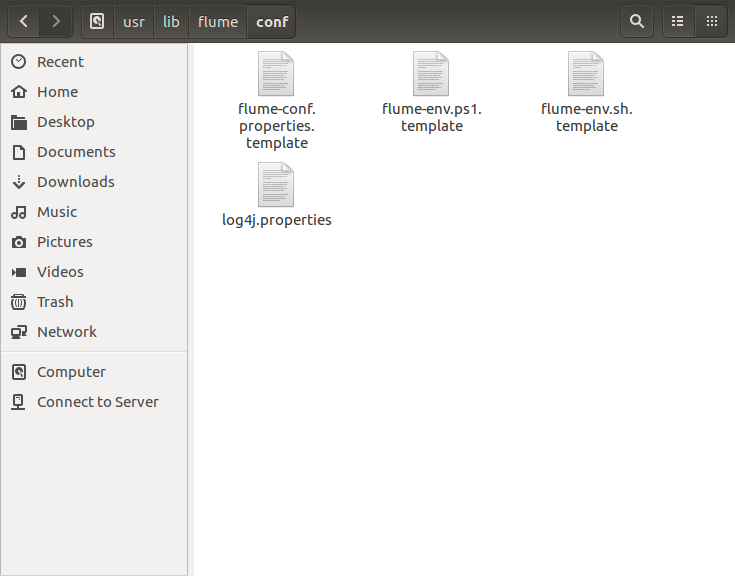
**export FLUME\_CONF\_DIR=$FLUME\_HOME/conf**

**export FLUME\_CLASSPATH=$FLUME\_CONF\_DIR**

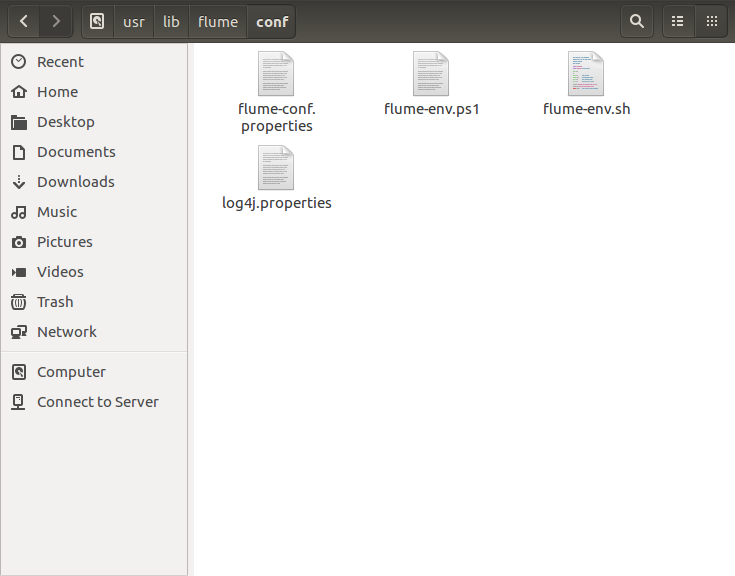
**export PATH=$PATH:$FLUME\_HOME/bin**



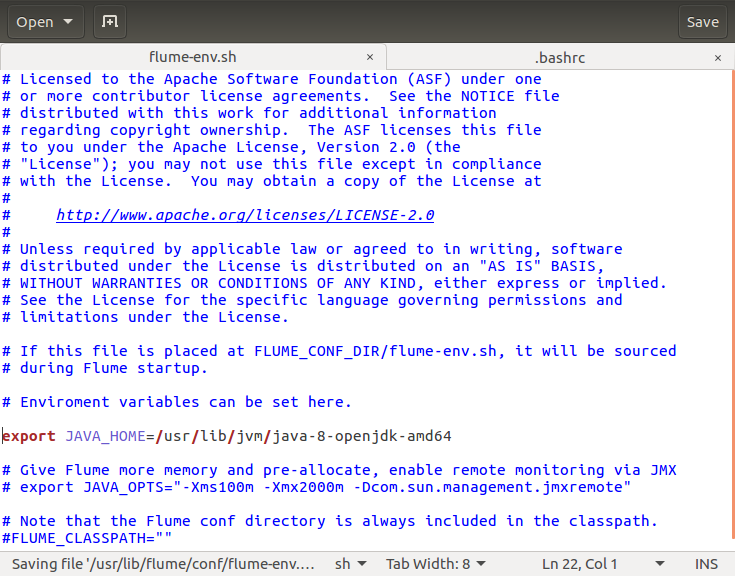
1. Now, we change the names of 3 files **flume-conf.properties.template, flume-env.ps1.template** and **flume-env.sh.template** renaming them and removing the word template from their name.



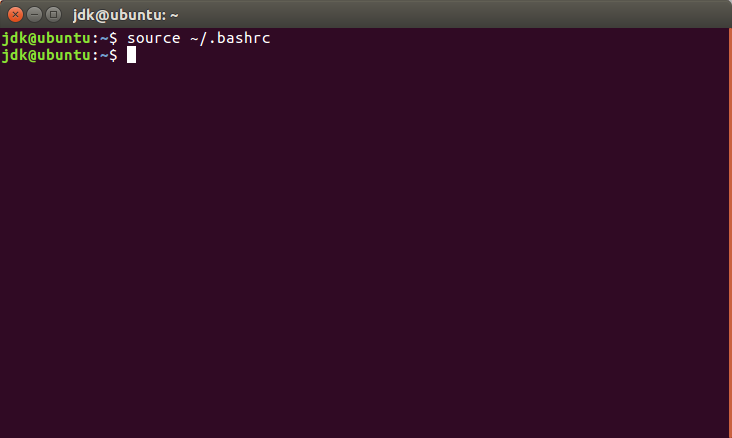
1. This is the output of the same.



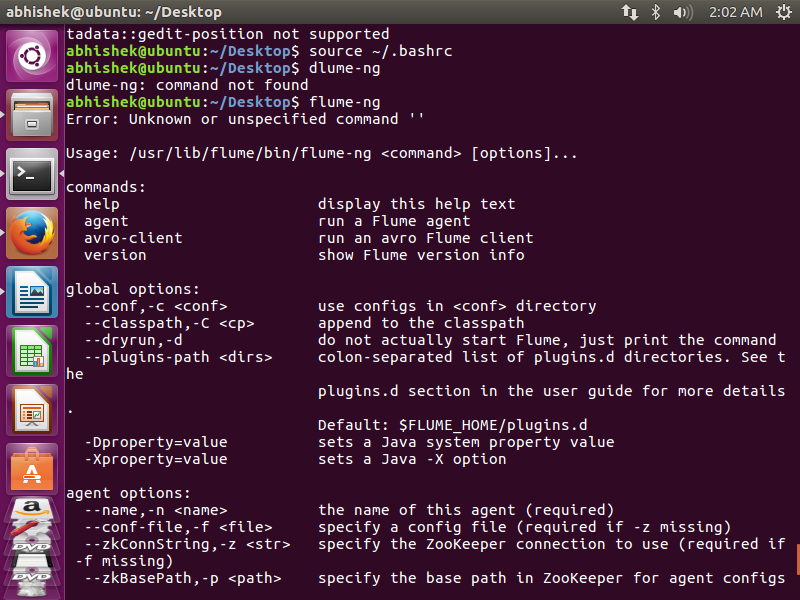
1. Now, we open the **flume-env.sh** file and add the JAVA\_HOME PATH in it.



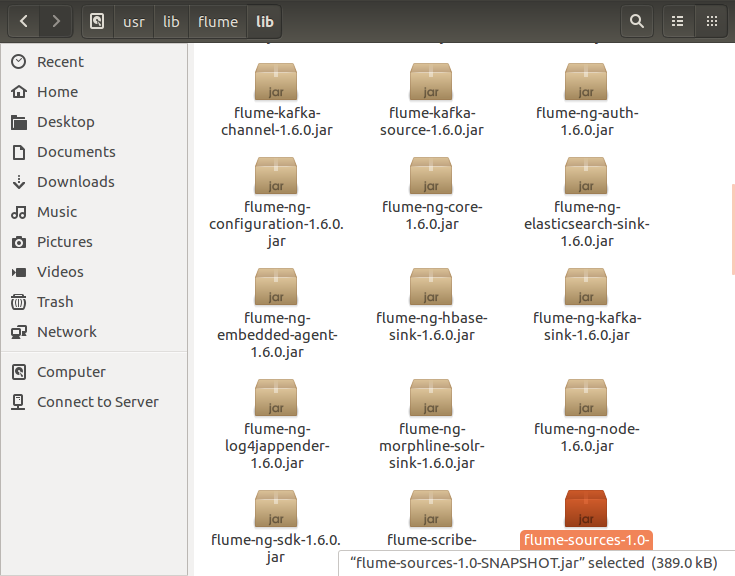
1. Now we permanently save the bashrc file.



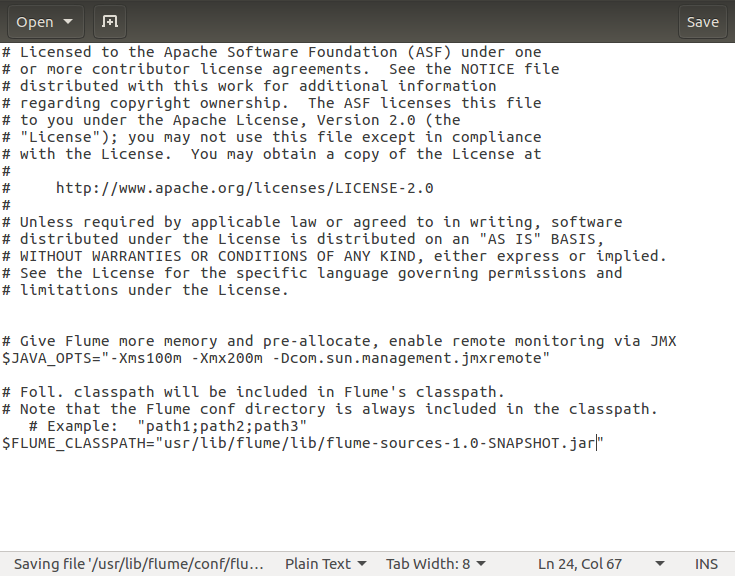
1. We use the **flume-ng** command. If the help desk opens then flume has been successfully installed.



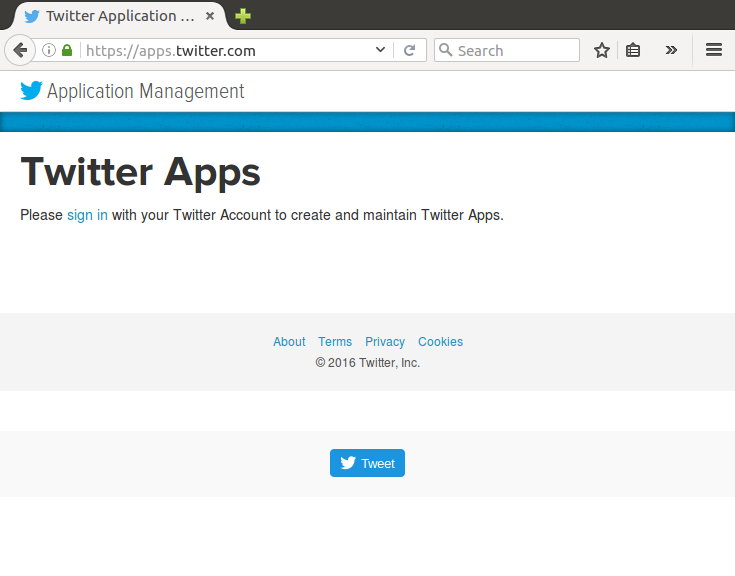
1. Now we move to **/usr/libflume/lib** and copy a jar file named **flume-sources-1.0-SNAPSHOT.jar** into this directory.



1. After that we open the **flume-env.ps1** file and add the **FLUME\_CLASSPATH** in it.

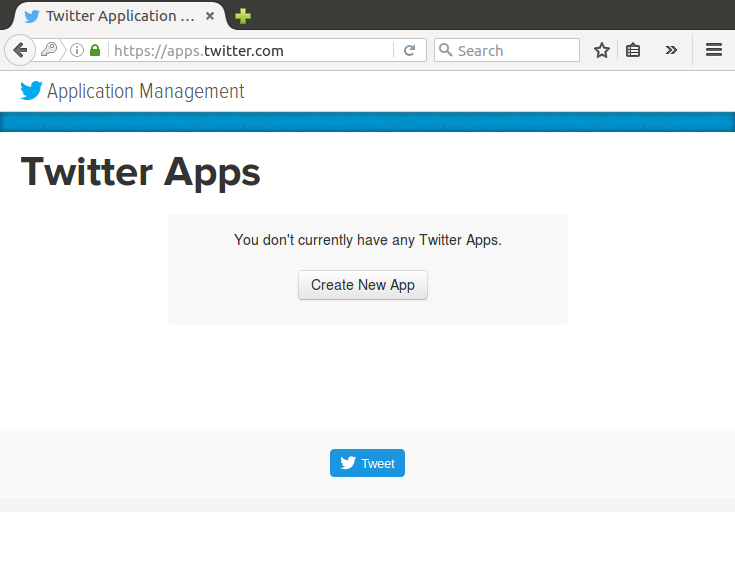


1. Now we open the website **dev.twitter.com/apps** in the Mozilla Firefox Browser.

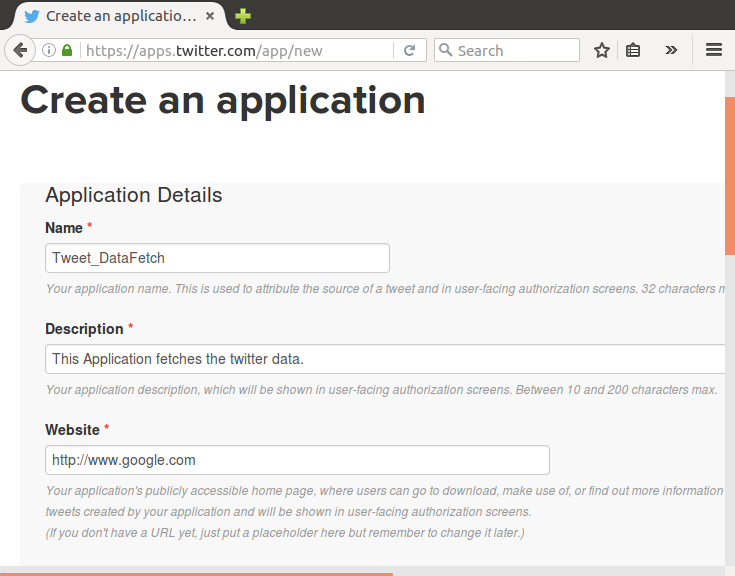


1. We will now see the website suggesting us to sign in. So, we sign into our twitter account.

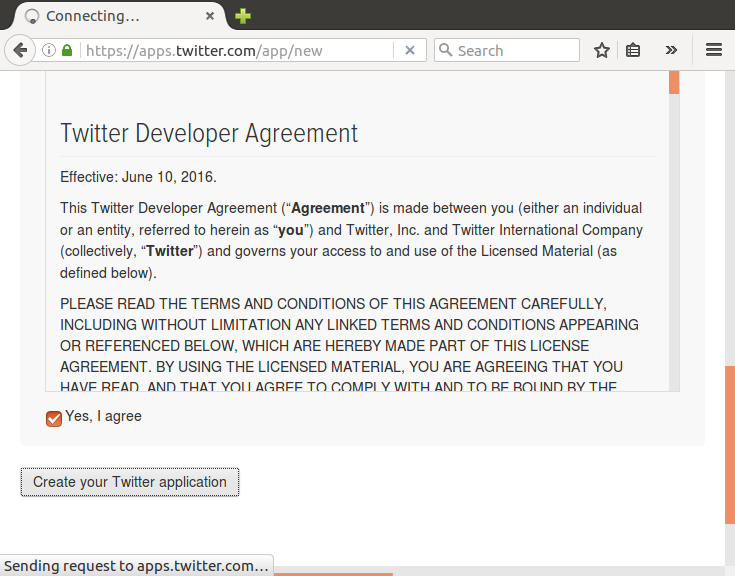
Click on **Create New App.**



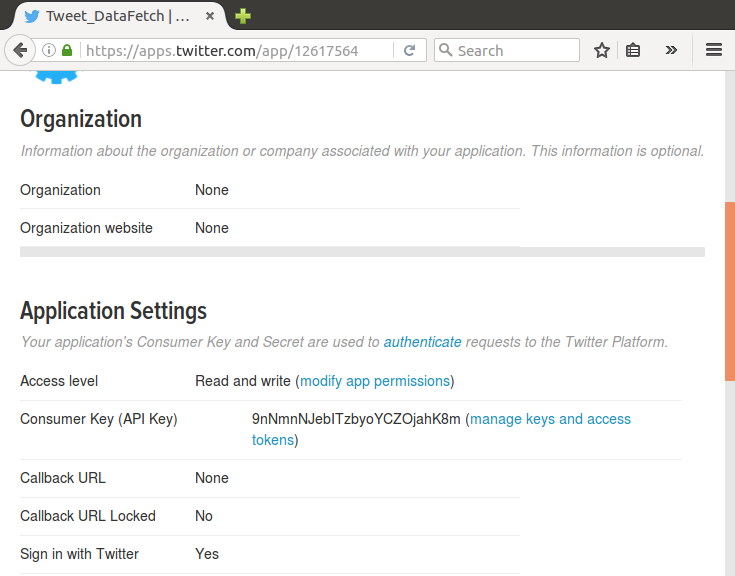
1. Fill in all the required fields to make the application and use the website as **google.com**.



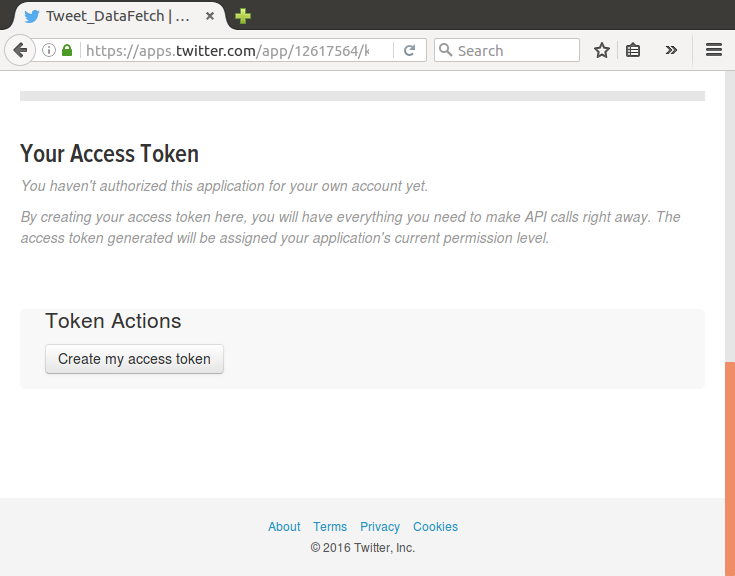
1. Now, scroll down and tick the option **Yes, I agree** and then click **Create your Twitter application.**



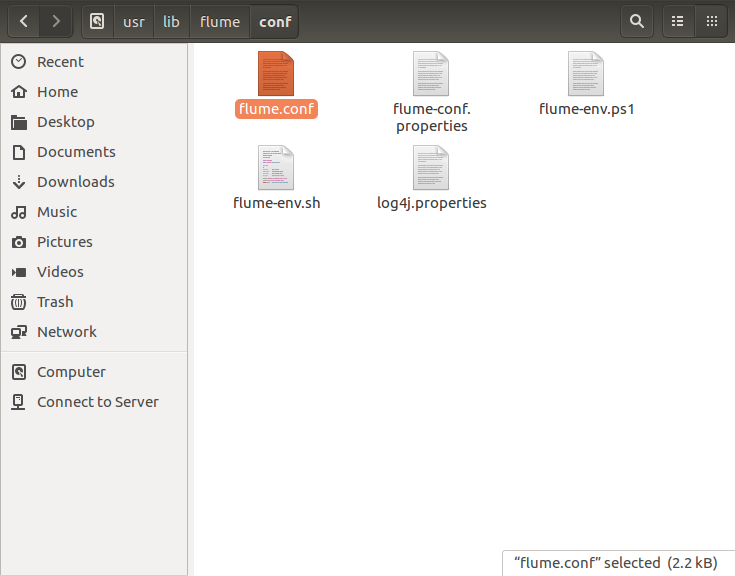
1. Click on **manage keys and access tokens.**



1. Now click on **Create my access token.**



1. Now, we open **flume.conf** file in the directory **/usr/lib/flume/conf** and then change the following keys in the file. These keys will be obtained from the page above.

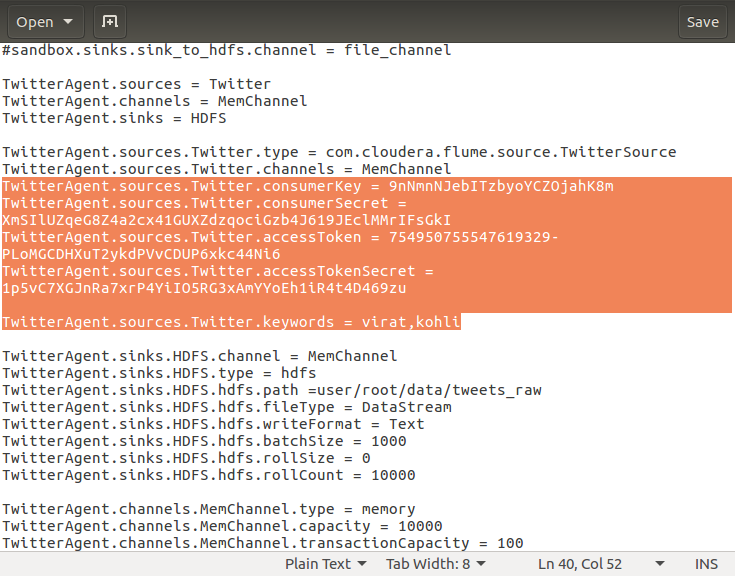


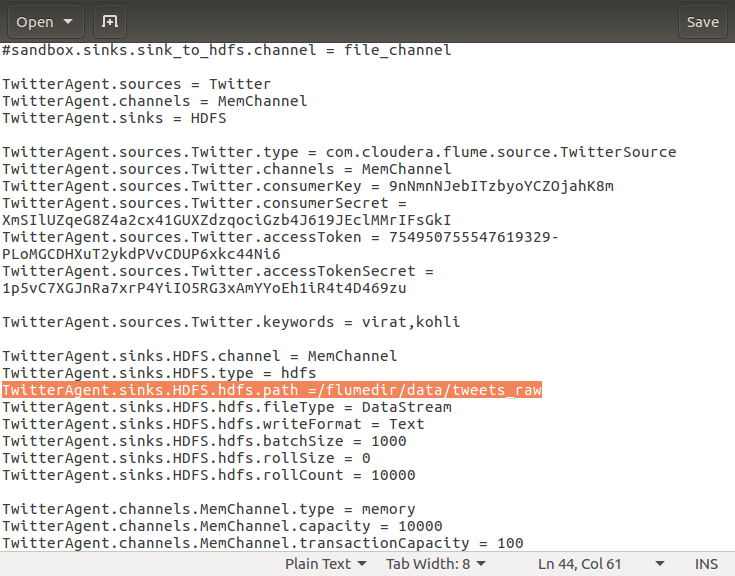
1. These are the keys which we will change in the flume.conf file:

**Access Token, Access Token Secret, Consumer Key (API Key), Consumer Secret (API Secret)**



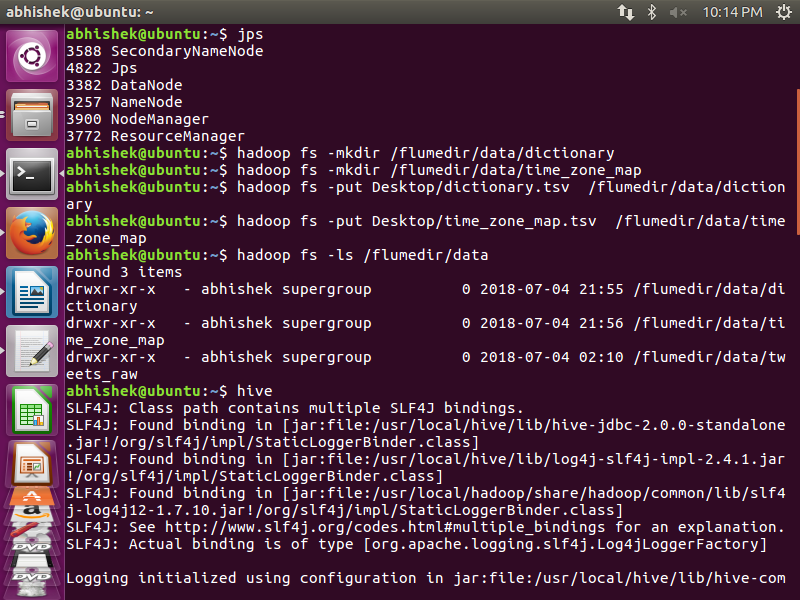
1. The keys are changed in this. Also add the keywords that we want to extract from twitter.





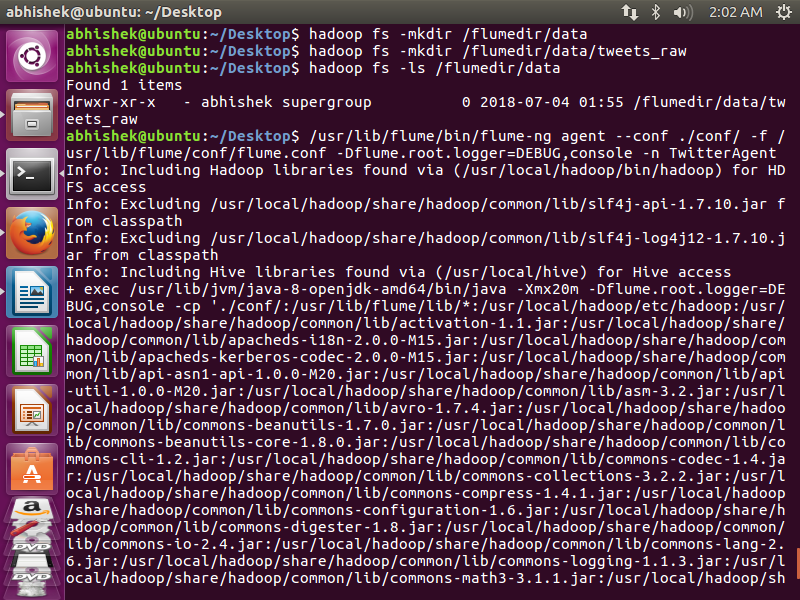
**DATA EXTRACTION FROM TWITTER:**

1. Start all the services using the **start-all.sh** command.
2. Now, make the following new directories:
3. Then we copy the files **dictionary.tsv** and **time\_zone\_map.tsv** to their respective directories:

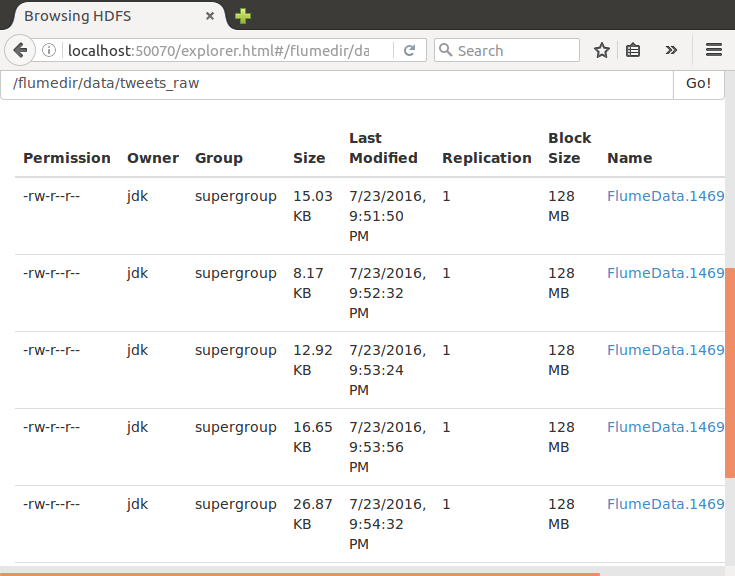


1. We will now start the flume agent using the following command:

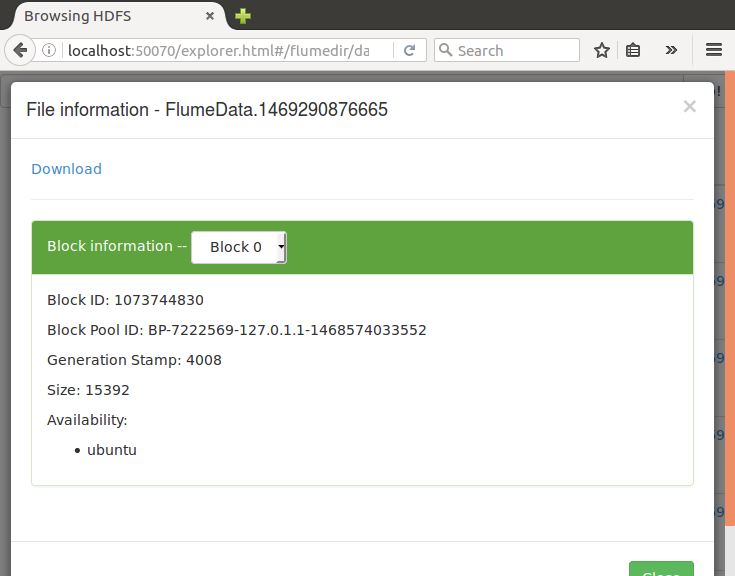
**/usr/lib/flume/bin/flume-ng agent --conf ./conf/ -f /usr/lib/flume/conf/flume.conf -Dflume.root.logger=DEBUG,console -n TwitterAgent**



1. This is the list of twitter data extracted which contains the keyword as specified in the conf file.



1. We can check the files by downloading them and seeing the tweets relating to the keyword.



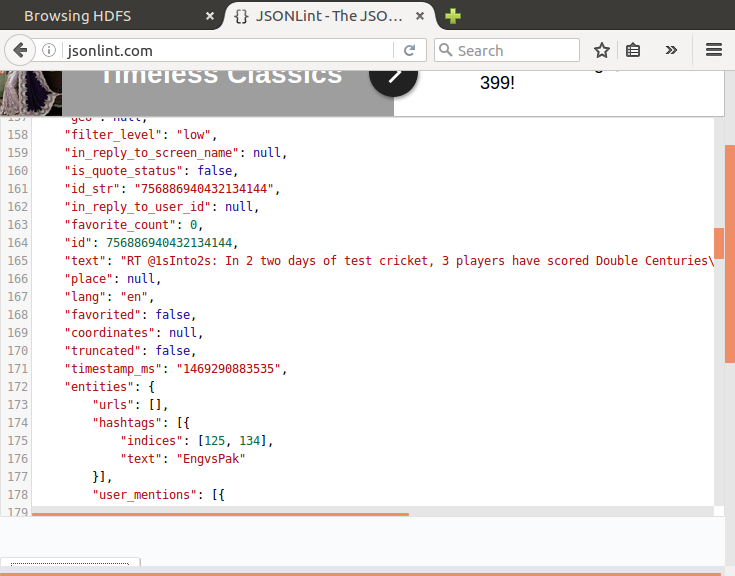
1. The file downloaded will be like this:



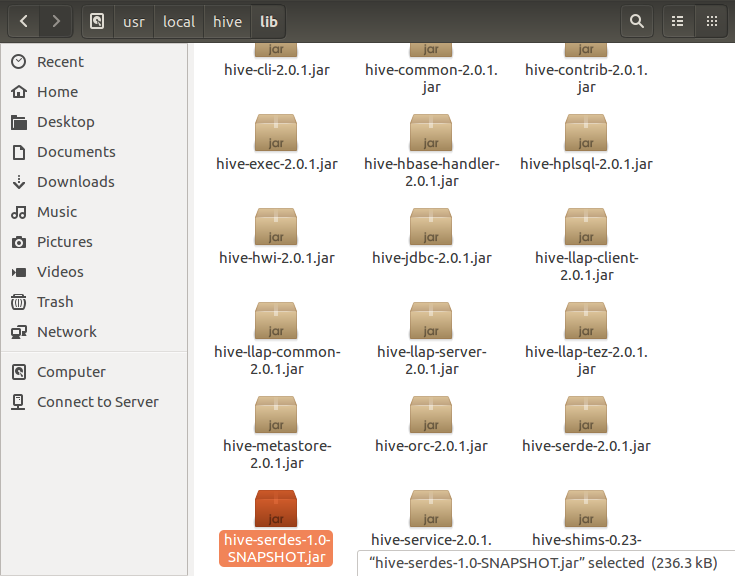
1. If we clean this downloaded file using the **JSON Validator** we will see the tweets in a human readable format as shown:



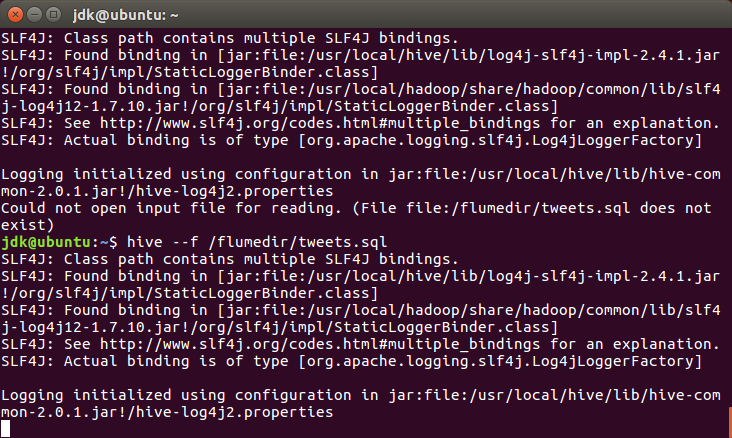
Then click on Validate JSON. The output that will become is as follows:

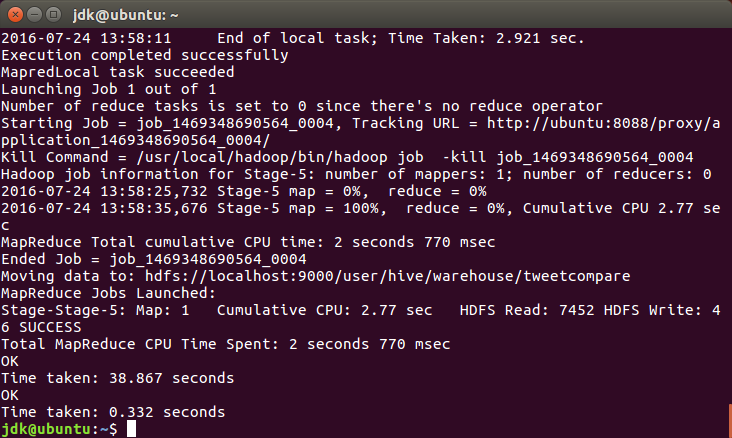


1. Now we copy the **hive-serdes-1.0-SNAPSHOT.jar** AND **jason-1.3.10-sources.jar** file into the directory **/usr/local/hive/lib**. This will be used by the hive shell to extract the clean data from the downloaded data into the hive table.(Assuming Apache Hive is already installed in your system)



1. Now we create a file **tweets.sql** which is on the Desktop.
2. Now run the tweets.sql file using the hive command.(Assuming Apache Hive is already installed in your system.)



1. After running the script, we receive this output as SUCCESS.
2. Now we look into all the created tables in the **hive shell** and **default database**.

