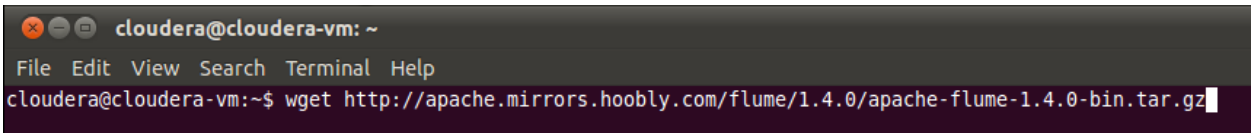
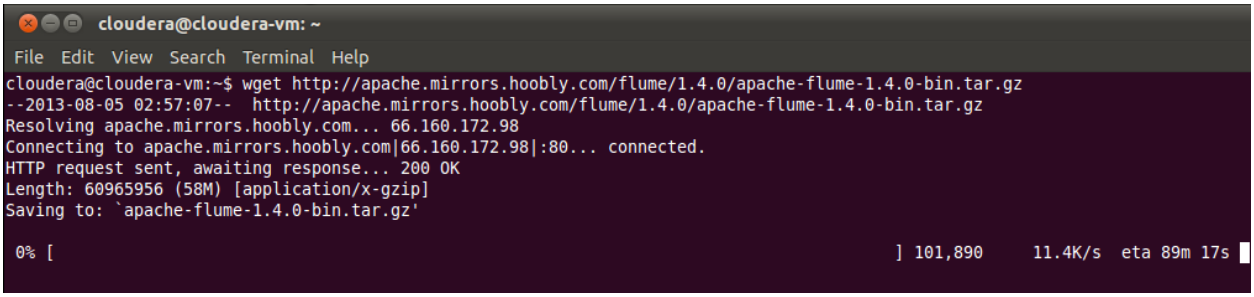
edureka!

Flume Set Up on Cloudera CDH3

Download flume:

Command:wget http://apache.mirrors.hoobly.com/flume/1.4.0/apache-flume-1.4.0-bin.tar.gz



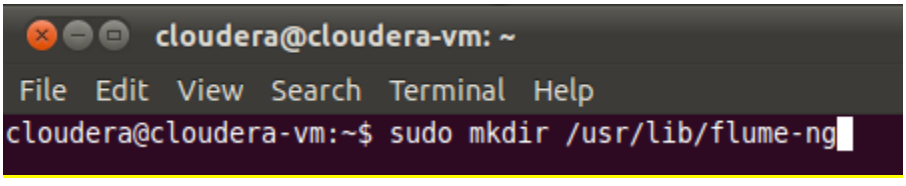


Check whether the flume tar is present or not:

Command:ls

Create flume-ng directory:

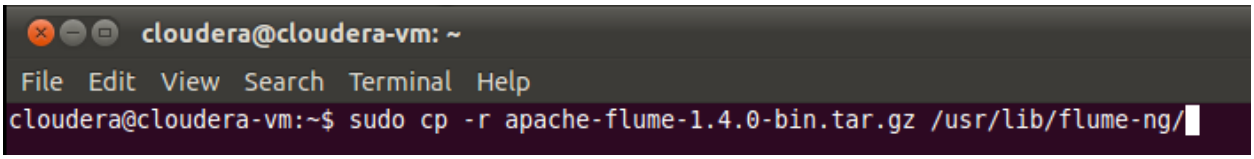
Command:sudomkdir /usr/lib/flume-ng



edureka!

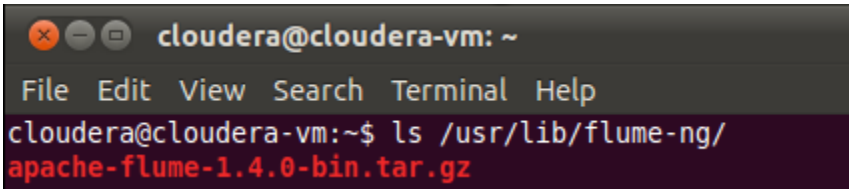
Copy the flume tar to flume-ng directory:

Command:sudocp –r apache-flume-1.4.0-bin.tar.gz /usr/lib/flume-ng/



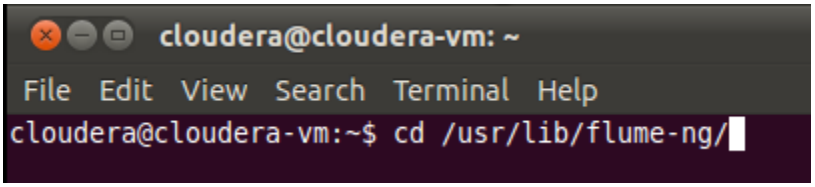
Check whether flume tar is copied or not:

Command:ls /usr/lib/flume-ng/



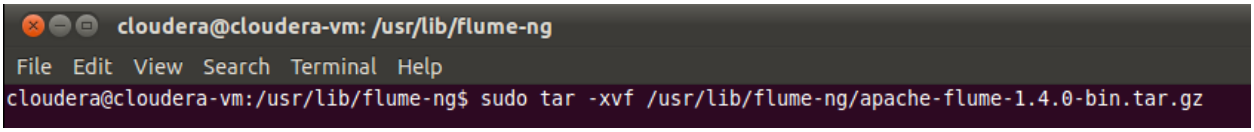
Change directory to flume-ng:

Command: cd /usr/lib/flume-ng/



Extract file from flume tar:

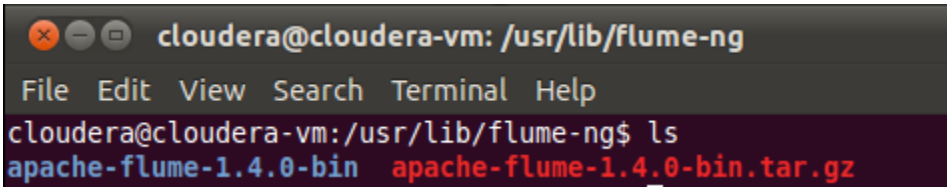
Command:sudo tar –xvf /usr/lib/flume-ng/apache-flume-1.4.0-bin.tar.gz



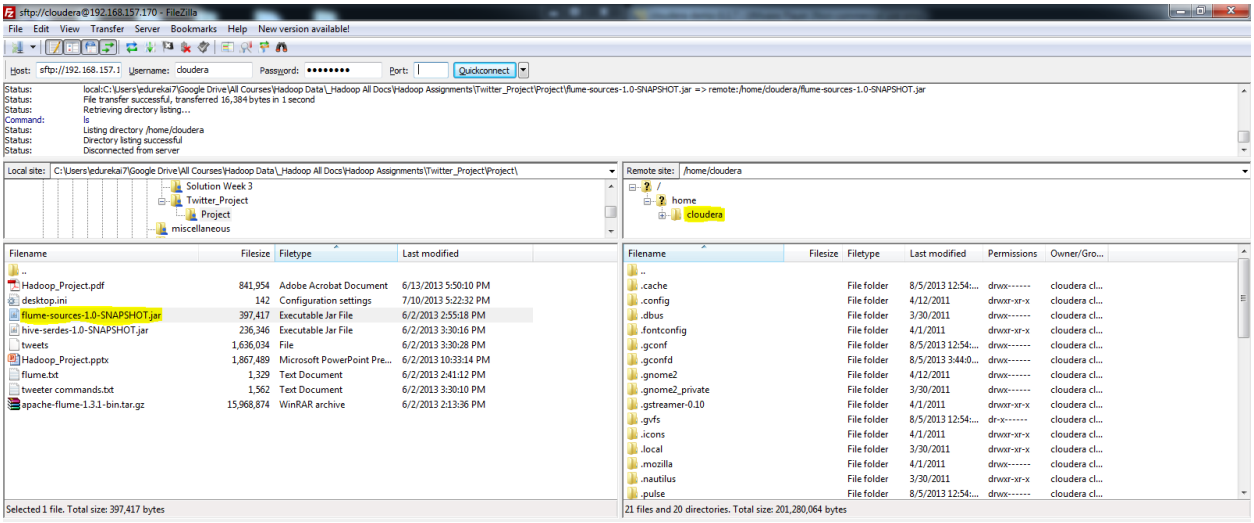
edureka!

Check whether flume files are extracted or not:

Command:ls

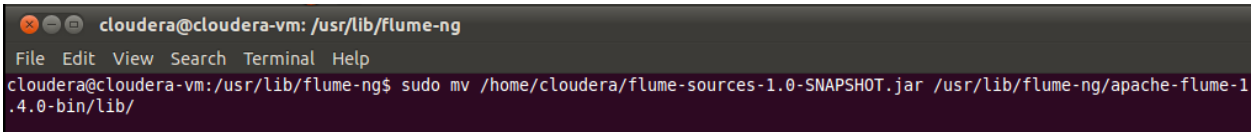


Move flume-sources-1.0-SNAPSHOT.jar to cloudera directory using FTP client from your Windows



Move the file from cloudera directory to lib directory of apache flume:

Command:sudo mv /home/cloudera/flume-sources-1.0-SNAPSHOT.jar /usr/lib/flume-ng/apache-flume-1.4.0-bin/lib/



edureka!

Check whether flume SNAPSHOT has moved to the lib folder of apache flume:

Command:ls /usr/lib/flume-ng/apache-flume-1.4.0-bin/lib/flume-*

```
cloudera@cloudera-vm:/usr/lib/flume-ng$ ls /usr/lib/flume-ng/apache-flume-1.4.0-bin/lib/flume-*
/usr/lib/flume-ng/apache-flume-1.4.0-bin/lib/flume-avro-source-1.4.0.jar
/usr/lib/flume-ng/apache-flume-1.4.0-bin/lib/flume-file-channel-1.4.0.jar
/usr/lib/flume-ng/apache-flume-1.4.0-bin/lib/flume-hdfs-sink-1.4.0.jar
/usr/lib/flume-ng/apache-flume-1.4.0-bin/lib/flume-irc-sink-1.4.0.jar
/usr/lib/flume-ng/apache-flume-1.4.0-bin/lib/flume-jdbc-channel-1.4.0.jar
/usr/lib/flume-ng/apache-flume-1.4.0-bin/lib/flume-jms-source-1.4.0.jar
/usr/lib/flume-ng/apache-flume-1.4.0-bin/lib/flume-ng-configuration-1.4.0.jar
/usr/lib/flume-ng/apache-flume-1.4.0-bin/lib/flume-ng-core-1.4.0.jar
/usr/lib/flume-ng/apache-flume-1.4.0-bin/lib/flume-ng-elasticsearch-sink-1.4.0.jar
/usr/lib/flume-ng/apache-flume-1.4.0-bin/lib/flume-ng-embedded-agent-1.4.0.jar
/usr/lib/flume-ng/apache-flume-1.4.0-bin/lib/flume-ng-hbase-sink-1.4.0.jar
/usr/lib/flume-ng/apache-flume-1.4.0-bin/lib/flume-ng-log4jappender-1.4.0.jar
/usr/lib/flume-ng/apache-flume-1.4.0-bin/lib/flume-ng-morphline-solr-sink-1.4.0.jar
/usr/lib/flume-ng/apache-flume-1.4.0-bin/lib/flume-ng-node-1.4.0.jar
/usr/lib/flume-ng/apache-flume-1.4.0-bin/lib/flume-ng-sdk-1.4.0.jar
/usr/lib/flume-ng/apache-flume-1.4.0-bin/lib/flume-scribe-source-1.4.0.jar
/usr/lib/flume-ng/apache-flume-1.4.0-bin/lib/flume-sources-1.0-SNAPSHOT.jar
/usr/lib/flume-ng/apache-flume-1.4.0-bin/lib/flume-thrift-source-1.4.0.jar
/usr/lib/flume-ng/apache-flume-1.4.0-bin/lib/flume-tools-1.4.0.jar
```

Create flume.env.sh file in the conf directory of apache flume:

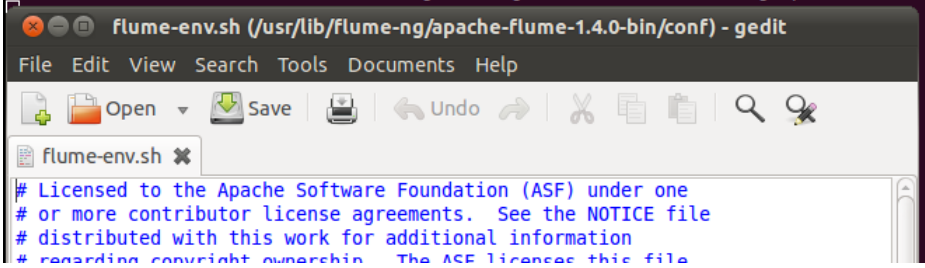
Command:sudocp /usr/lib/flume-ng/apache-flume-1.4.0-bin/conf/flume-env.sh.template /usr/lib/flume-ng/apache-flume-1.4.0-bin/conf/flume-env.sh

```
cloudera@cloudera-vm: /usr/lib/flume-ng
File  Edit  View  Search  Terminal  Help
cloudera@cloudera-vm:/usr/lib/flume-ng$ sudo cp /usr/lib/flume-ng/apache-flume-1.4.0-bin/conf/flume-env.sh.template /usr/lib/f
lume-ng/apache-flume-1.4.0-bin/conf/flume-env.sh
```

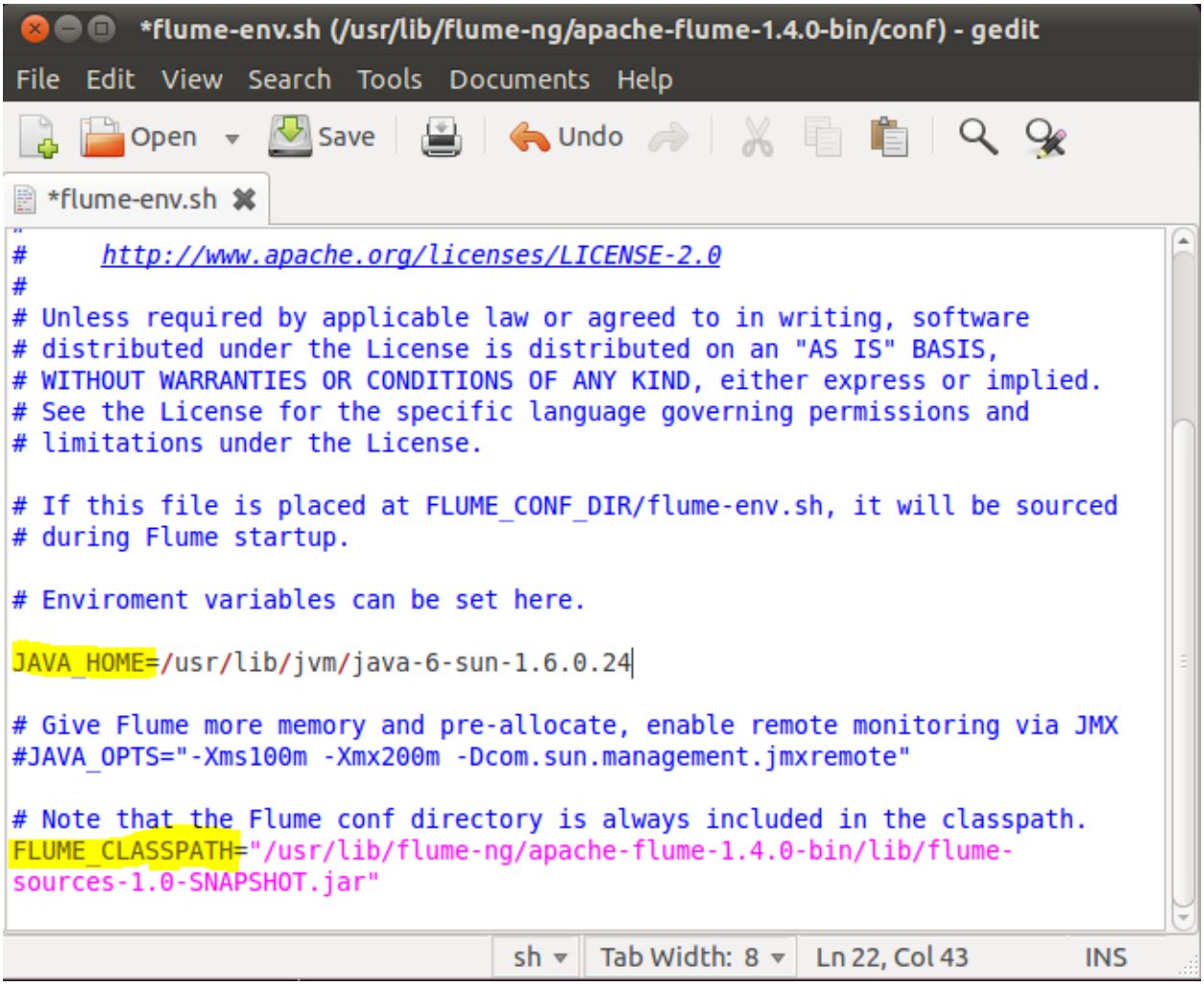
Open flume-env.sh:

Command:sudogedit /usr/lib/flume-ng/apache-flume-1.4.0-bin/conf/flume-env.sh

```
cloudera@cloudera-vm:/usr/lib/flume-ng$ sudo gedit /usr/lib/flume-ng/apache-flume-1.4.0-bin/conf/flume-env.sh
```

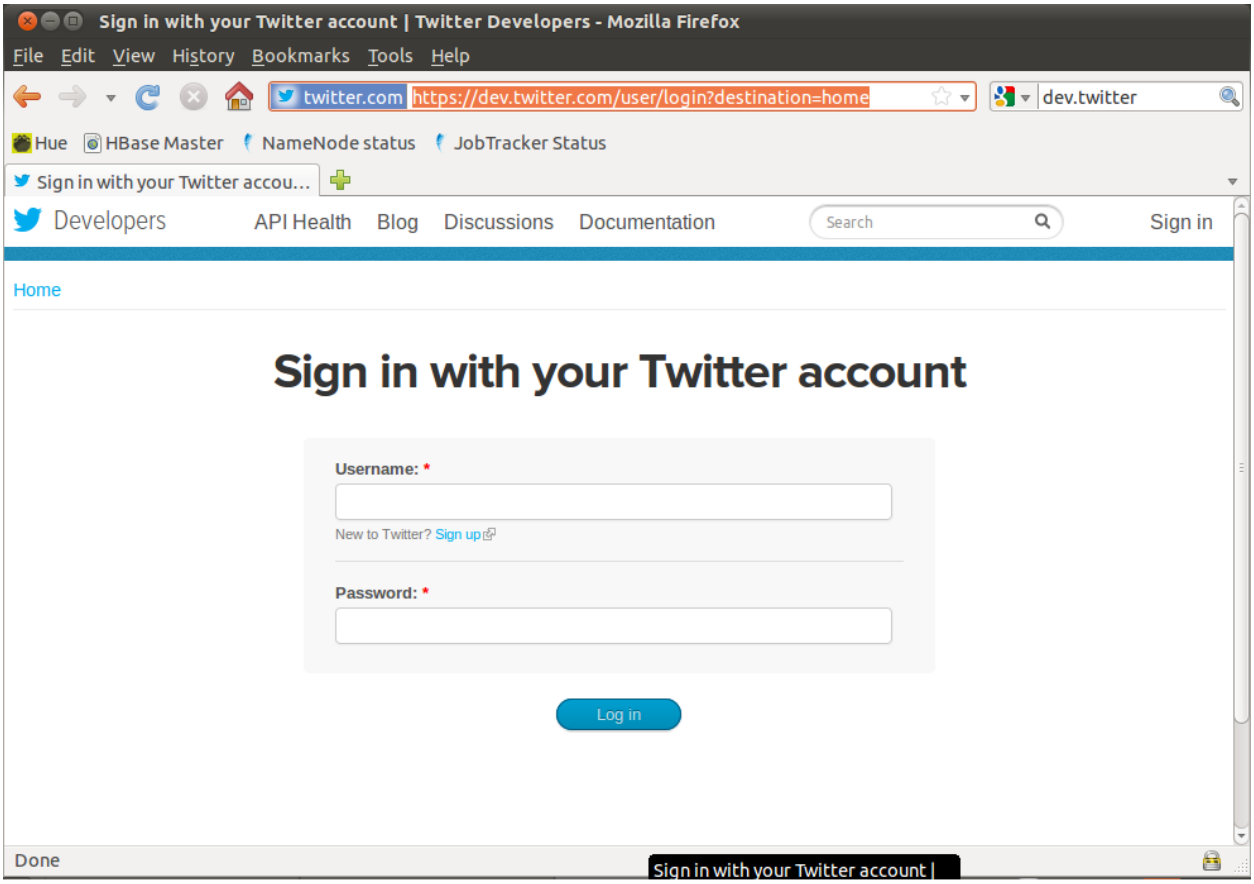


Edit flume-env.sh according to the below snapshot:

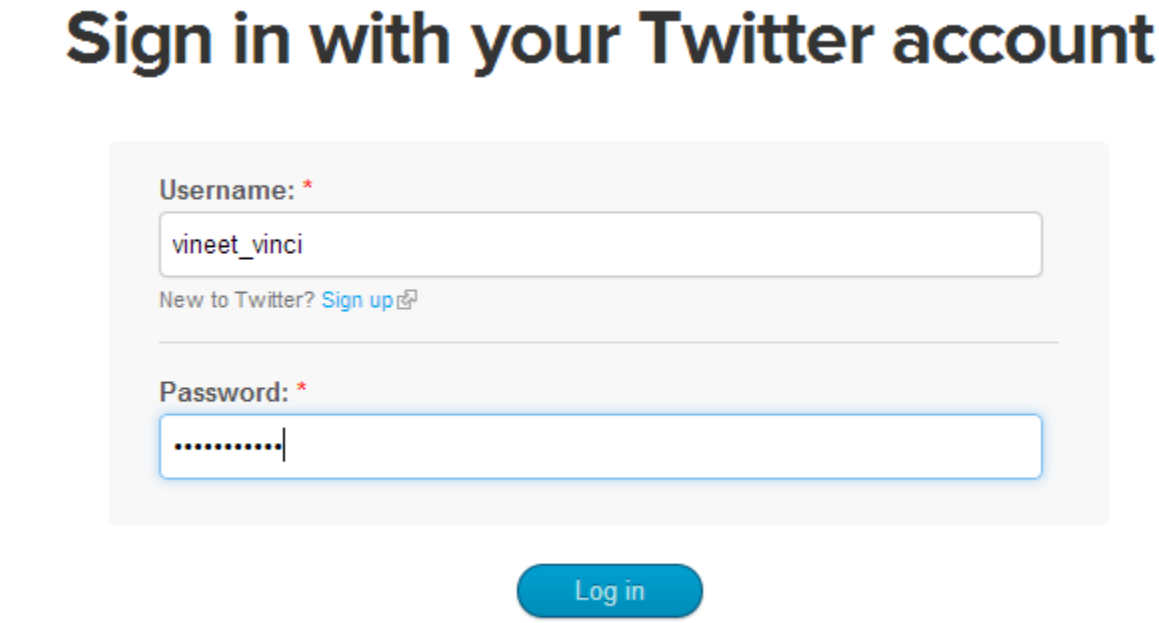


Open a Browser and go to the below URL:

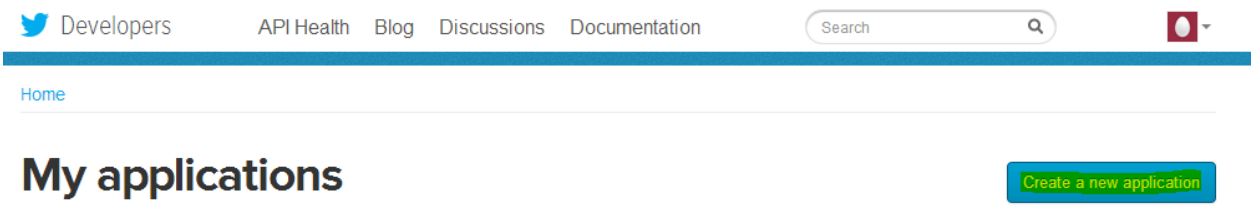
URL:https://dev.twitter.com/user/login?destination=home



Enter your Twitter account credentials:



Create a new application:



Enter all the details in the application:

edureka!

Create an application

Application Details

Name: *

MyApp_Edureka

Your application name. This is used to attribute the source of a tweet and in user-facing authorization screens. 32 characters max.

Description: *

Give a Brief Description

Your application description, which will be shown in user-facing authorization screens. Between 10 and 200 characters max.

Website: *

http://www.yahoo.com/

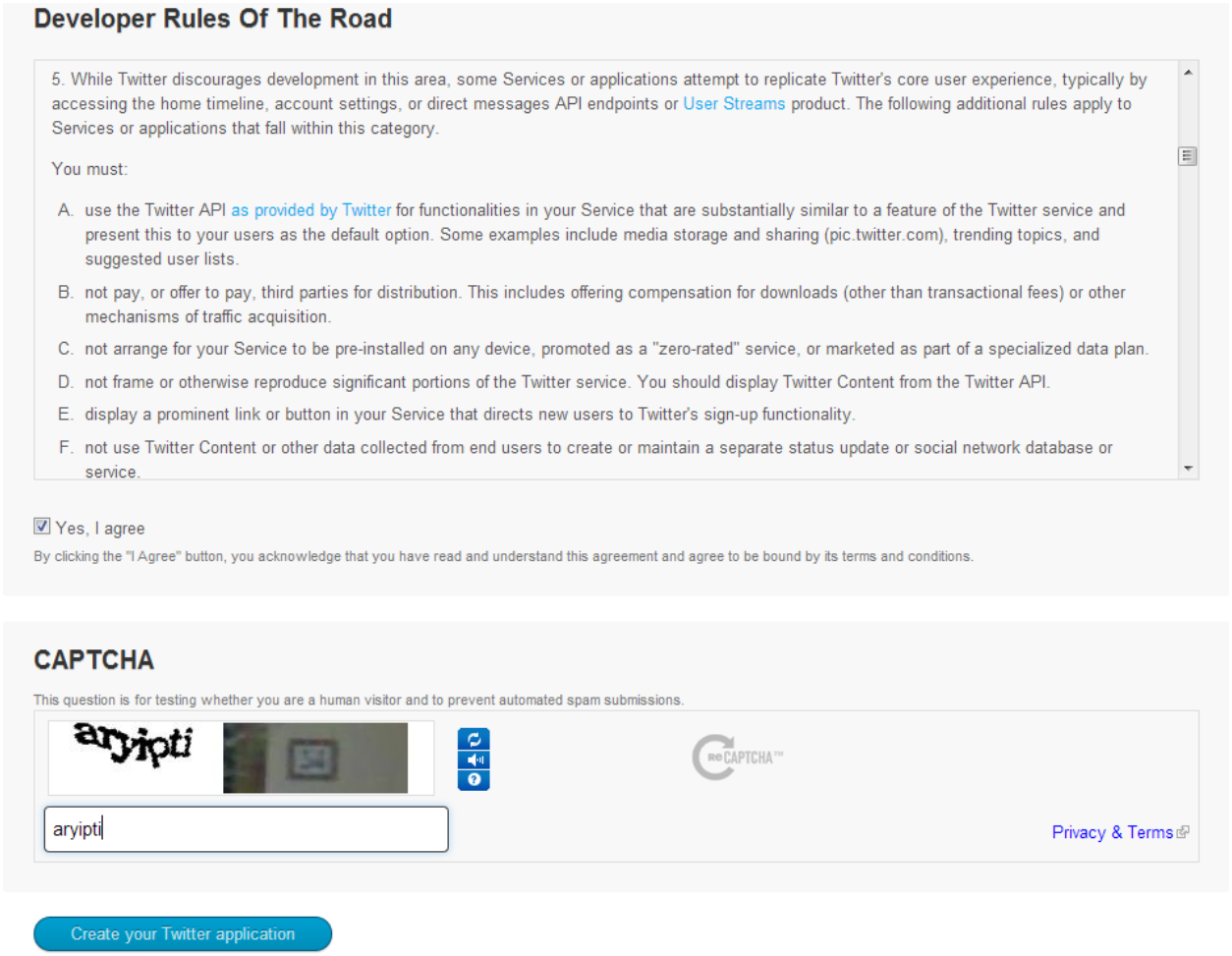
Your application's publicly accessible home page, where users can go to download, make use of, or find out more information about your application. This fully-qualified URL is used in the source attribution for tweets created by your application and will be shown in user-facing authorization screens.
(If you don't have a URL yet, just put a placeholder here but remember to change it later.)

Callback URL:

Where should we return after successfully authenticating? For @Anywhere applications, only the domain specified in the callback will be used. OAuth 1.0a applications should explicitly specify their `oauth_callback` URL on the request token step, regardless of the value given here. To restrict your application from using callbacks, leave this field blank.

edureka!

Check Yes, I agree and after entering the CAPTCHA click on Create your Twitter application:



edureka!

Click on Create my access token:

OAuth settings

Your application's OAuth settings. Keep the "Consumer secret" a secret. This key should never be human-readable in your application.

Access level	Read-only About the application permission model
Consumer key	0p8DkYUhQmn8uRt7C71w
Consumer secret	G04dAE5OaHeqUEh6VZoF0MAC2YMcgkgu8VpWaJam8
Request token URL	https://api.twitter.com/oauth/request_token
Authorize URL	https://api.twitter.com/oauth/authorize
Access token URL	https://api.twitter.com/oauth/access_token
Callback URL	None
Sign in with Twitter	No

Your access token

It looks like you haven't authorized this application for your own Twitter account yet. For your convenience, we give you the opportunity to create your OAuth access token here, so you can start signing your requests right away. The access token generated will reflect your application's current permission level.

Create my access token

edureka!

Refresh the Page: The highlighted part in the snapshot will be used in flume.conf:

OAuth settings

Your application's OAuth settings. Keep the "Consumer secret" a secret. This key should never be human-readable in your application.

Access level	Read-only About the application permission model
Consumer key	0p8DkYUhQmn8uRt7C71w
Consumer secret	G04dAE5OaHeqUEh6VZoF0MAC2YMcgkgu8VpWaJam8
Request token URL	https://api.twitter.com/oauth/request_token
Authorize URL	https://api.twitter.com/oauth/authorize
Access token URL	https://api.twitter.com/oauth/access_token
Callback URL	None
Sign in with Twitter	No

Your access token

Use the access token string as your "oauth_token" and the access token secret as your "oauth_token_secret" to sign requests with your own Twitter account. Do not share your oauth_token_secret with anyone.

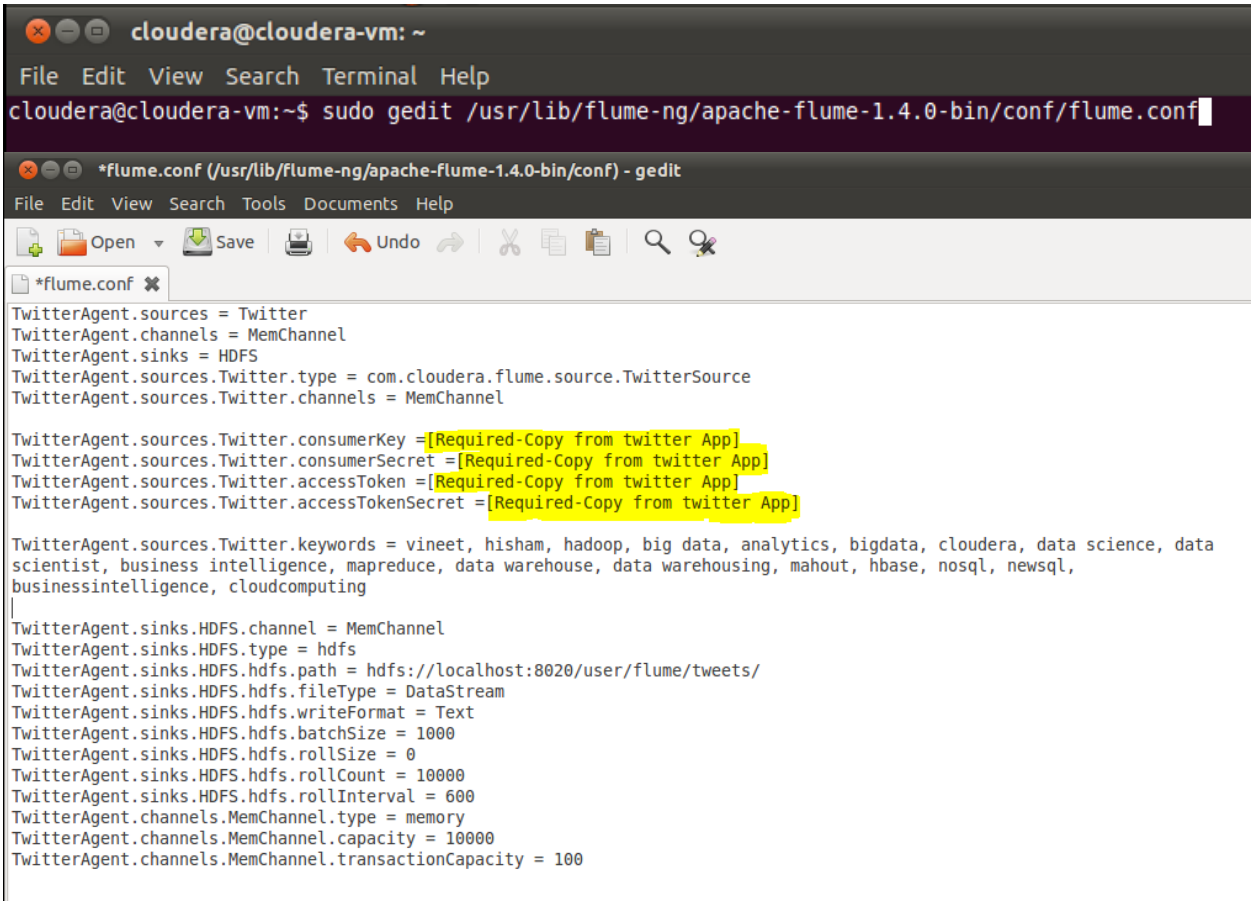
Access token	1635433267-HvStt9EggeUhrxNTTJhc4DKrDymxfHGi3oroJpc
Access token secret	gBX9kFeM1K7zSrOlMhUUYfQmvLeuWv7msOEgyhLIs
Access level	Read-only

Recreate my access token

edureka!

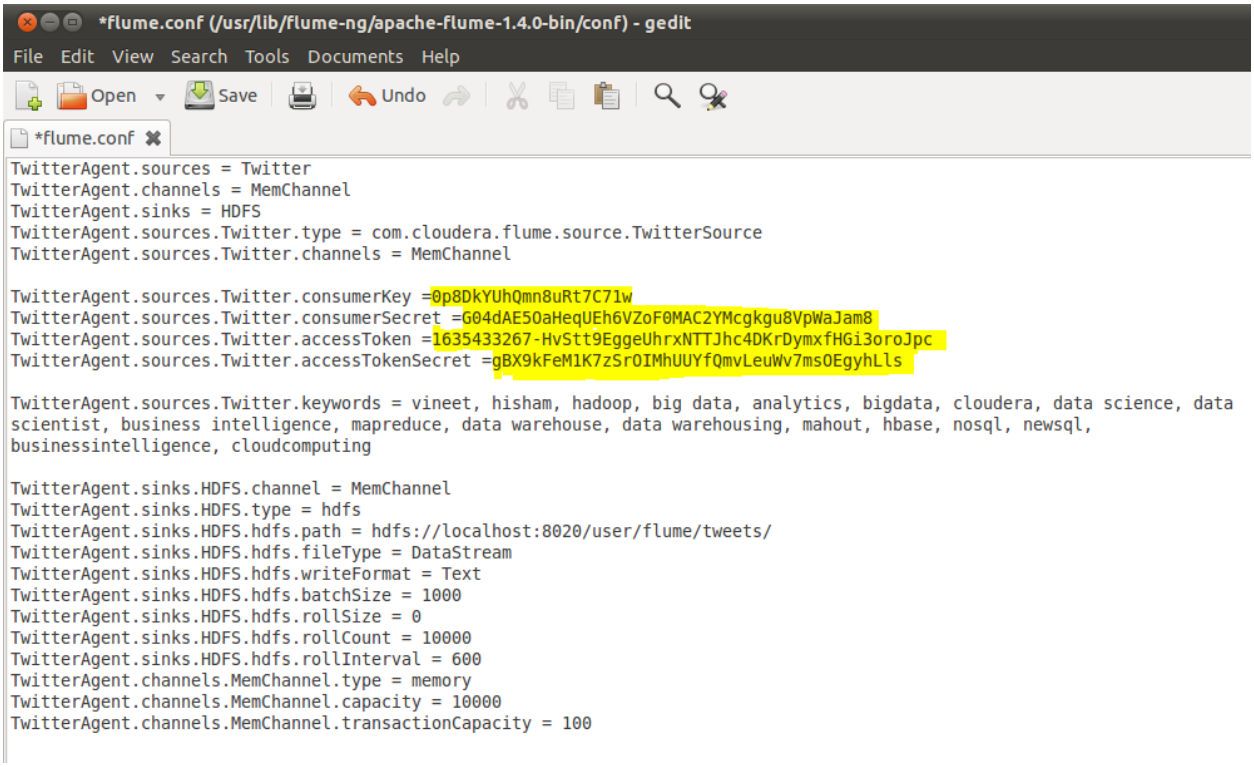
Edit flume.conf:

Command:sudo gedit /usr/lib/flume-ng/apache-flume-1.4.0-bin/conf/flume.conf



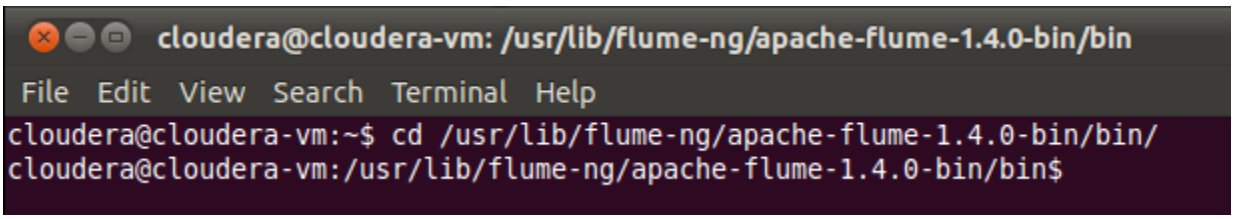
edureka!

After editing flume.conf will look like the below snapshot:



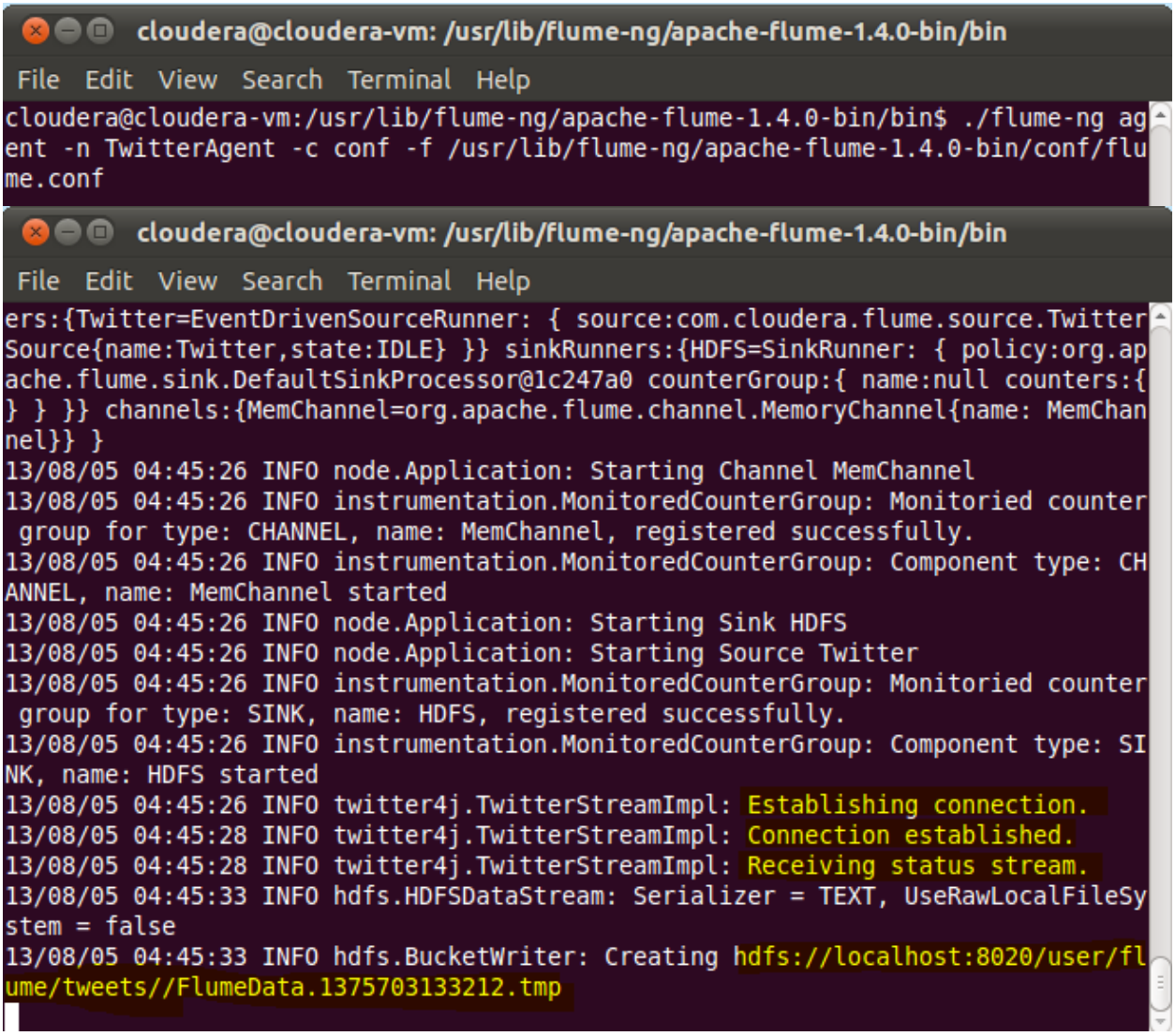
Change Directory to the bin folder of apache flume:

Command:cd /usr/lib/flume-ng/apache-flume-1.4.0-bin/bin/



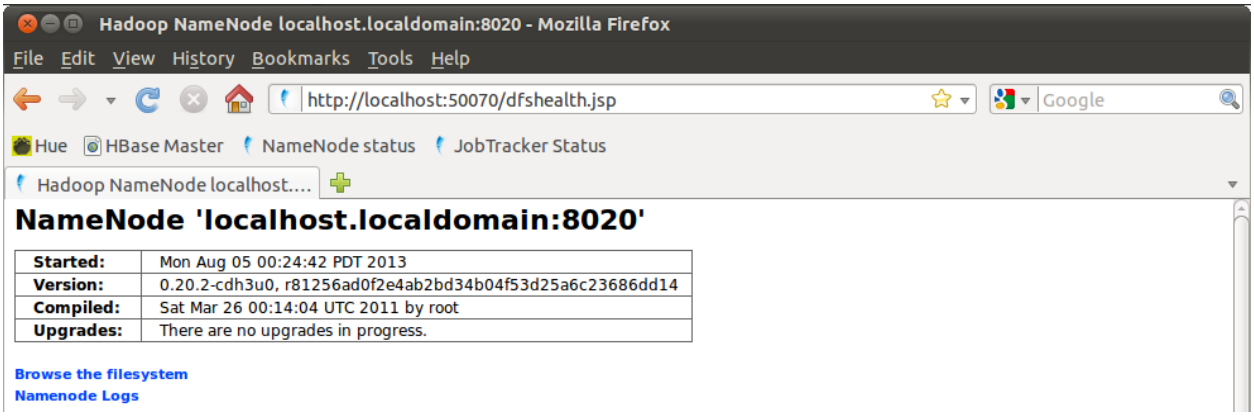
Start fetching the data from twitter:

Command: ./flume-ng agent –n TwitterAgent –c conf –f /usr/lib/flume-ng/apache-flume-1.4.0-bin/conf/flume.conf

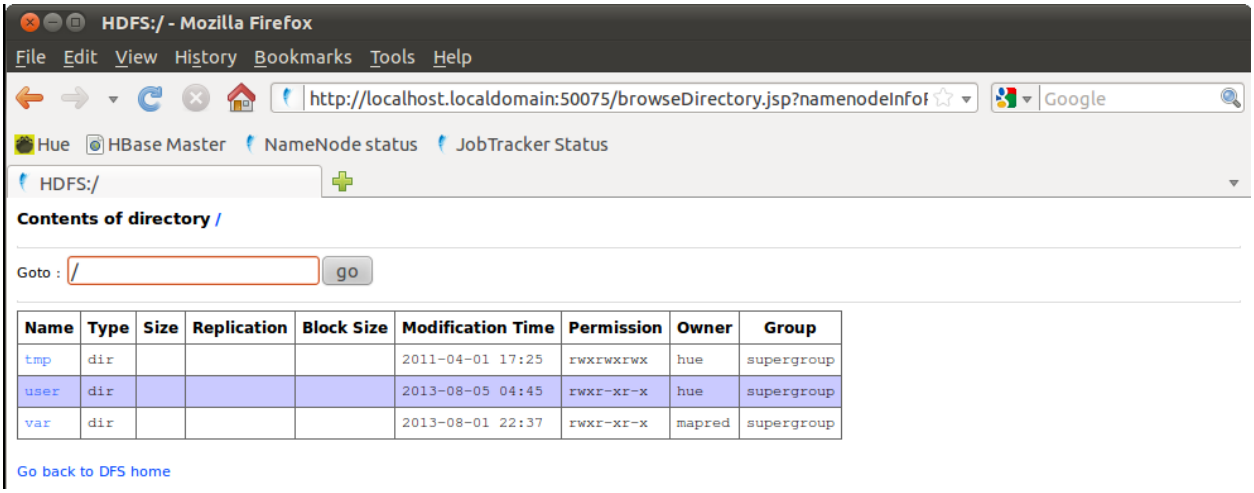


edureka!

Open the Browser and click on NameNode status and then click on Browse the filesystem:

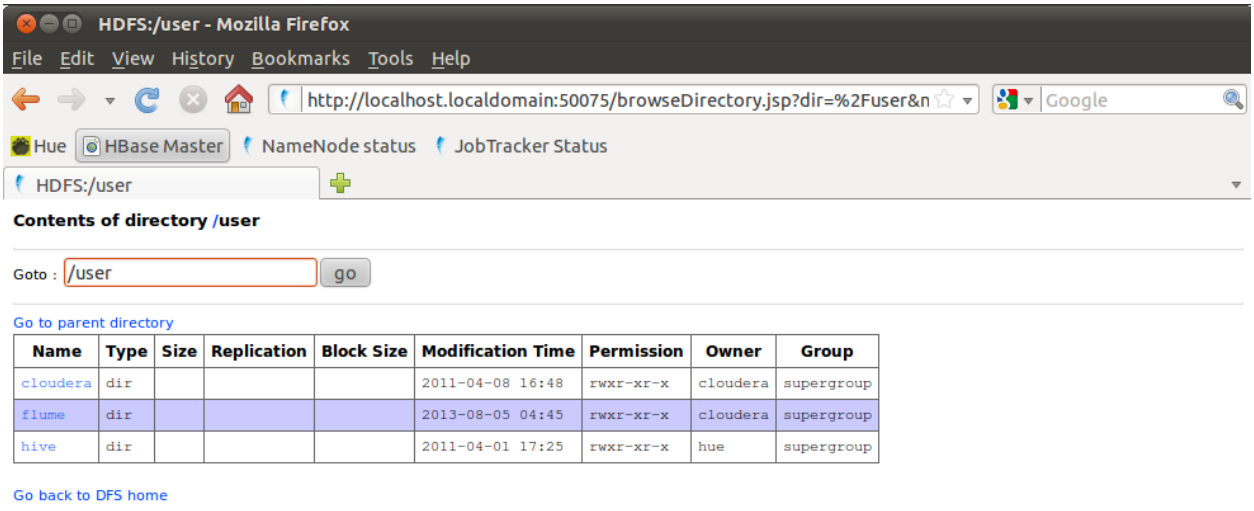


Click on user:

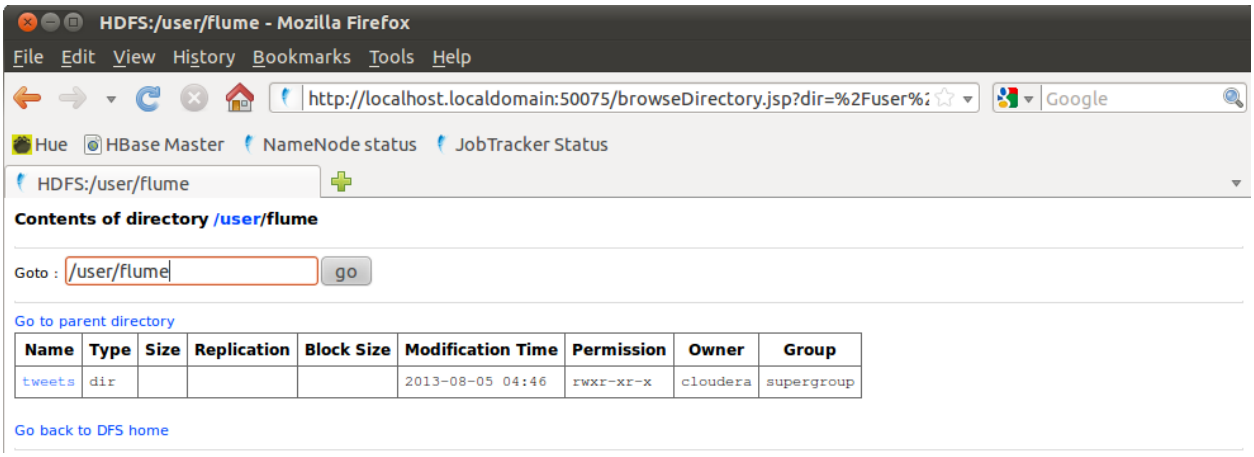


edureka!

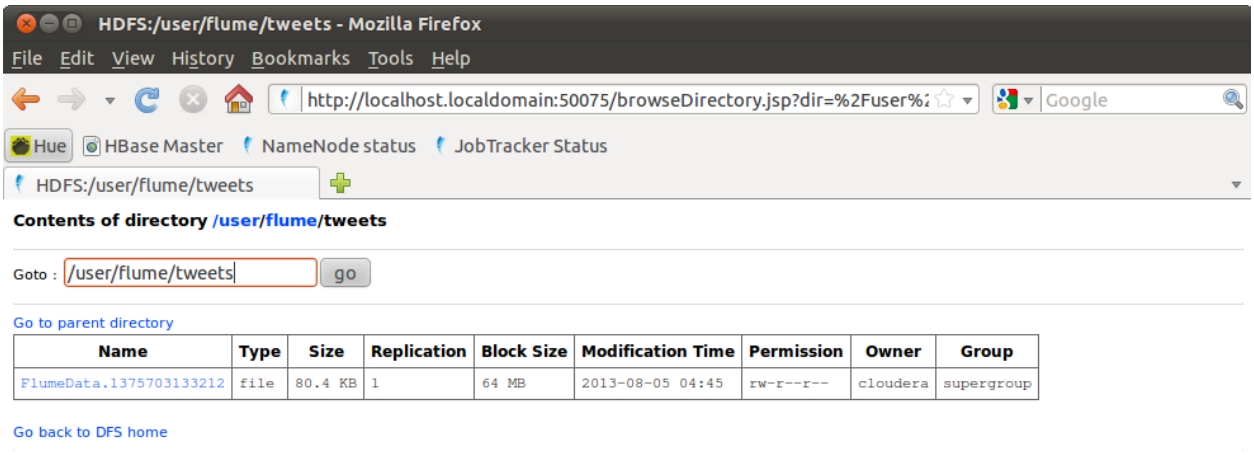
Click on flume:



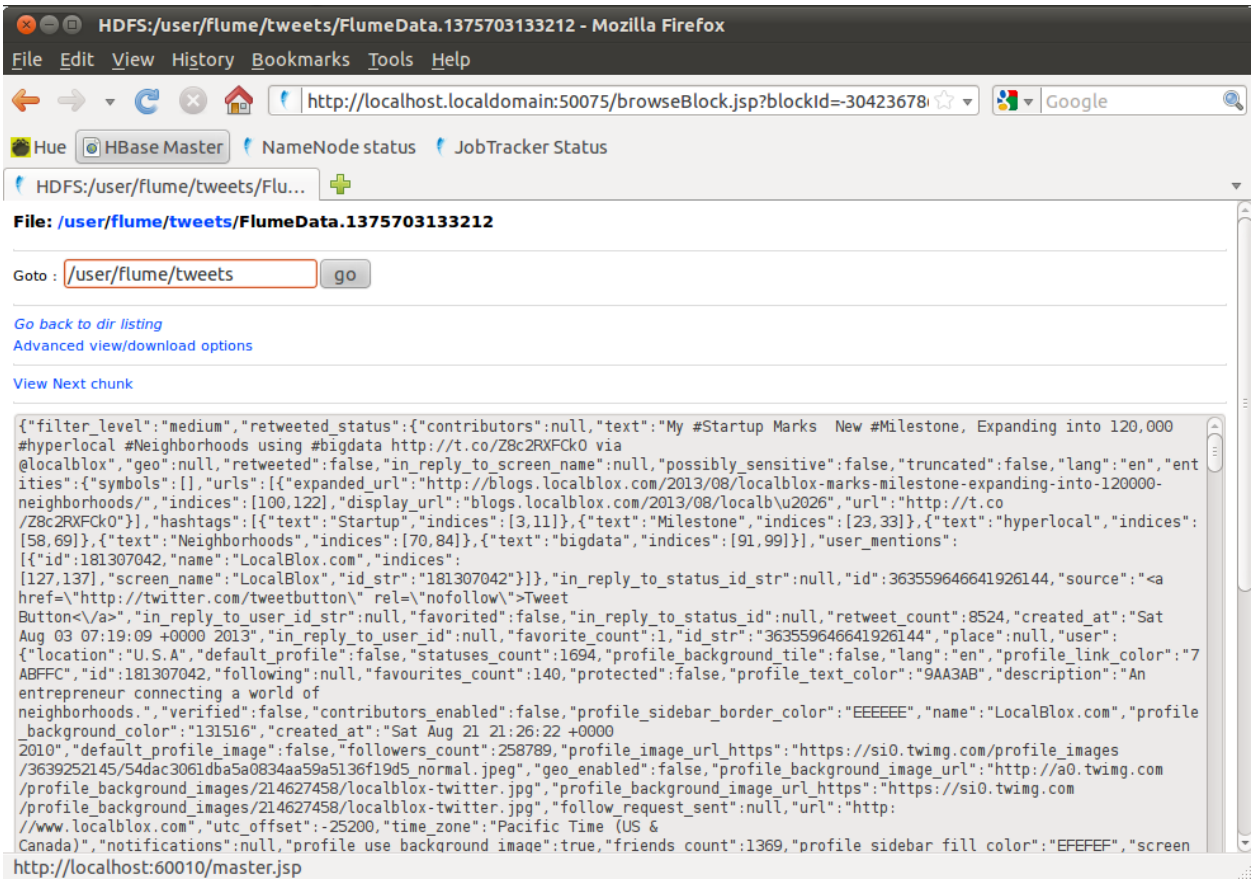
Click on tweets:



Click on FlumeData file:



This is the data that has been downloaded from Twitter:



----------**The End**----------