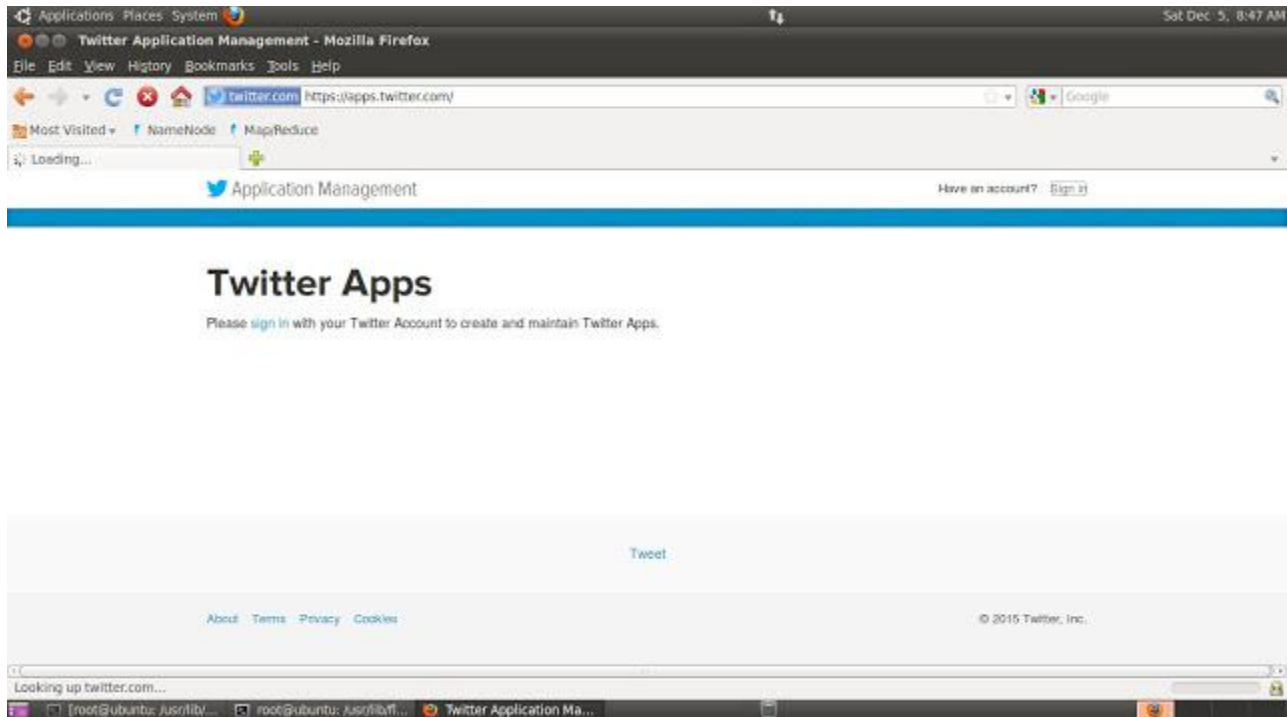


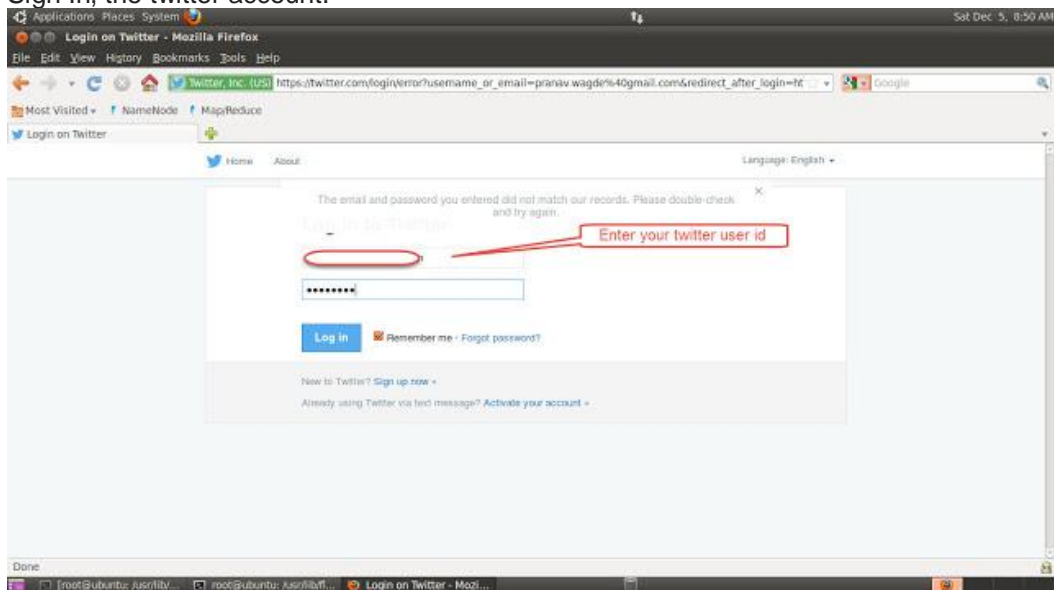
Fetching twitter data using Flume

Setup for Twitter

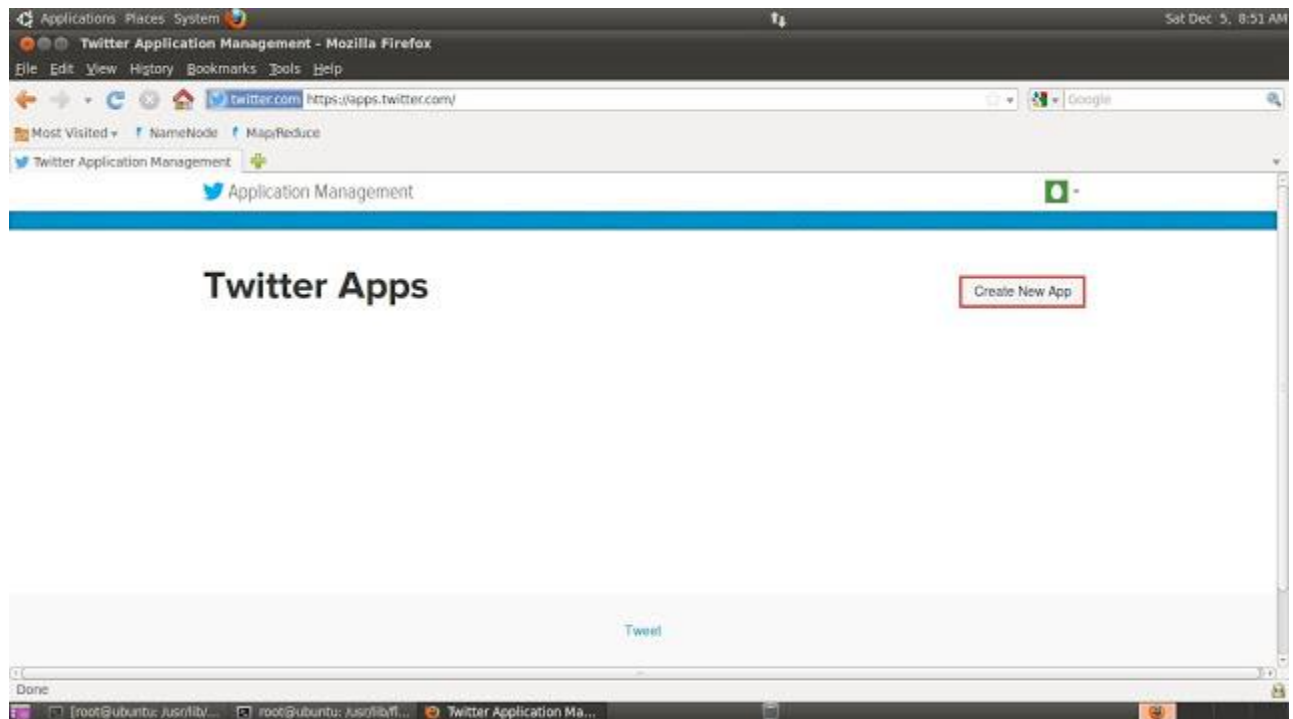
Open <https://apps.twitter.com>



Sign In, the twitter account.

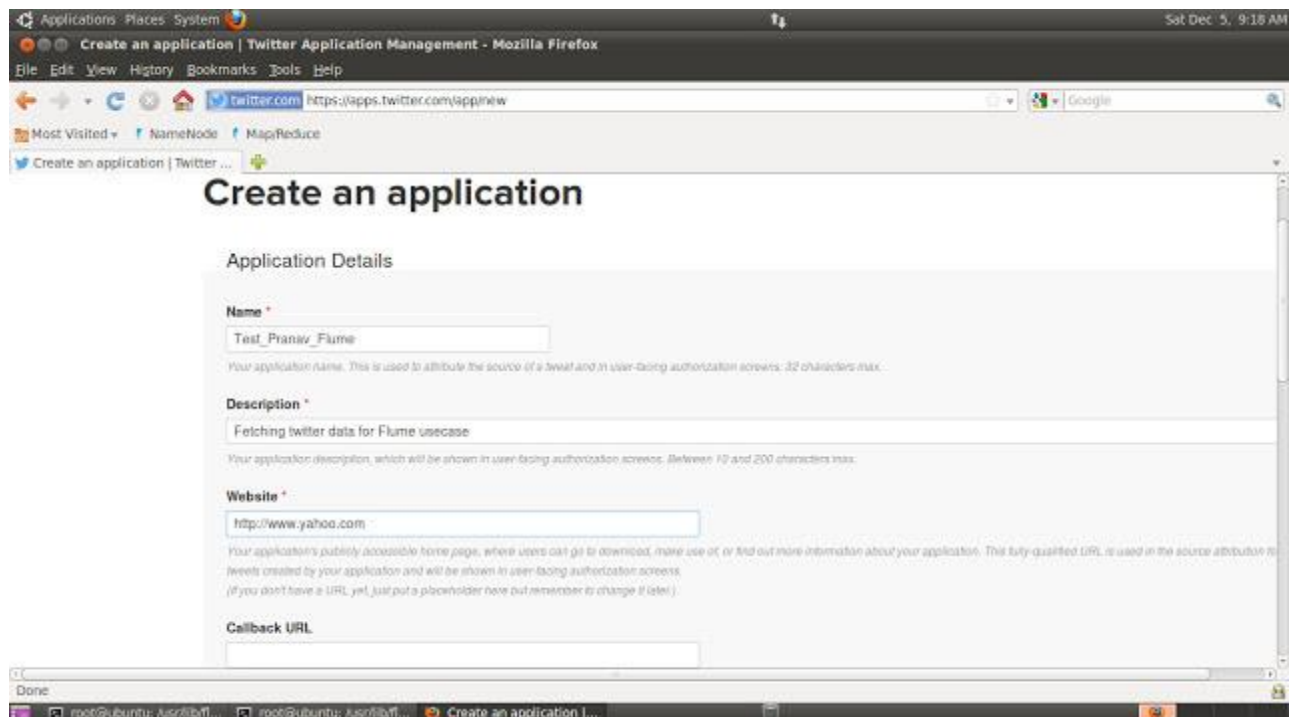


Click on Create New App

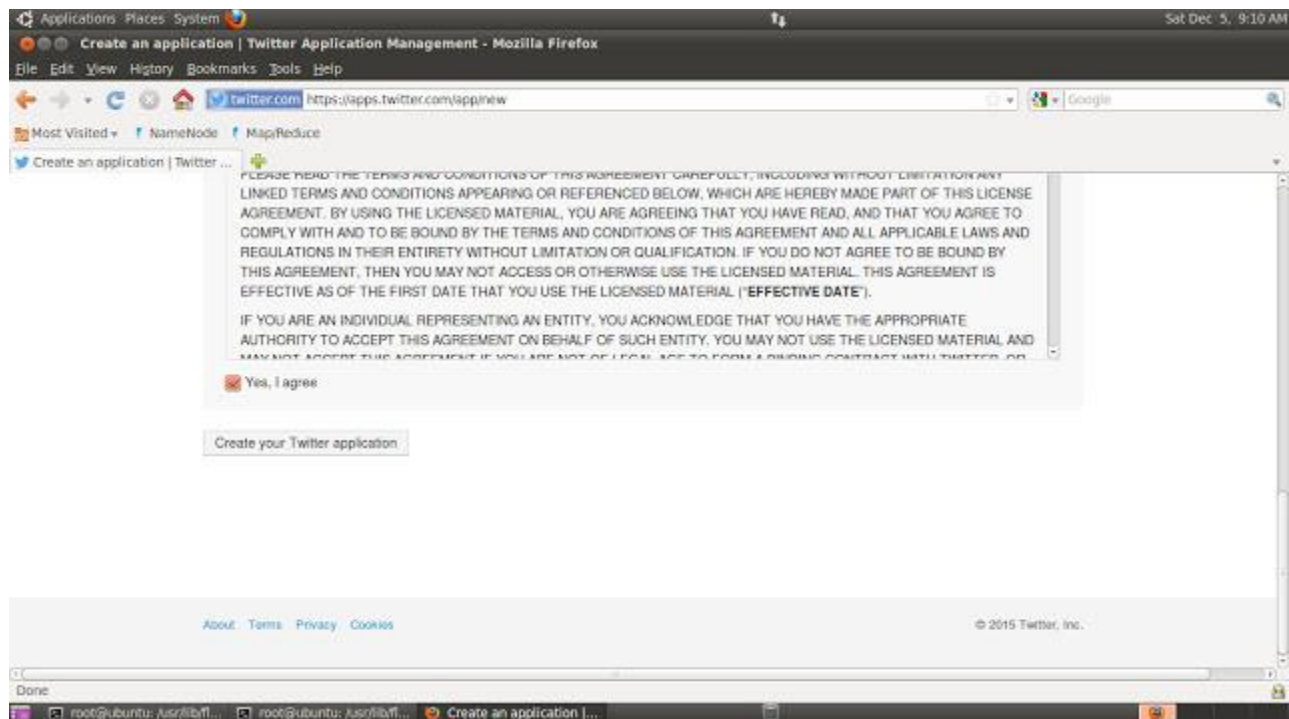


Give the details as shown below. **(Make sure you will give unique name to the application)**

Note: Website should be fully qualified URL.
e.g., instead of google.com, should give http://www.google.com



Tick "Yes, I agree" and click [Create your Twitter application](#).



Application will be created as below

 Application Management

RameshBatch84

[Details](#) [Settings](#) [Keys and Access Tokens](#) [Permissions](#)



flume exmple

<https://www.kellytechnologies.com>

Organization

Information about the organization or company associated with your application. This information is optional.

Organization

Organization website

Application Settings

Your application's Consumer Key and Secret are used to [authenticate](#) requests to the Twitter Platform.

Access level

Click on the created application and select tab “Keys and Access Tokens”.

 Application Management

RameshBatch84

Details Settings **Keys and Access Tokens** Permissions



flume exmaple

<https://www.kellytechnologies.com>

Organization

Information about the organization or company associated with your application. This information is optional.

Organization None

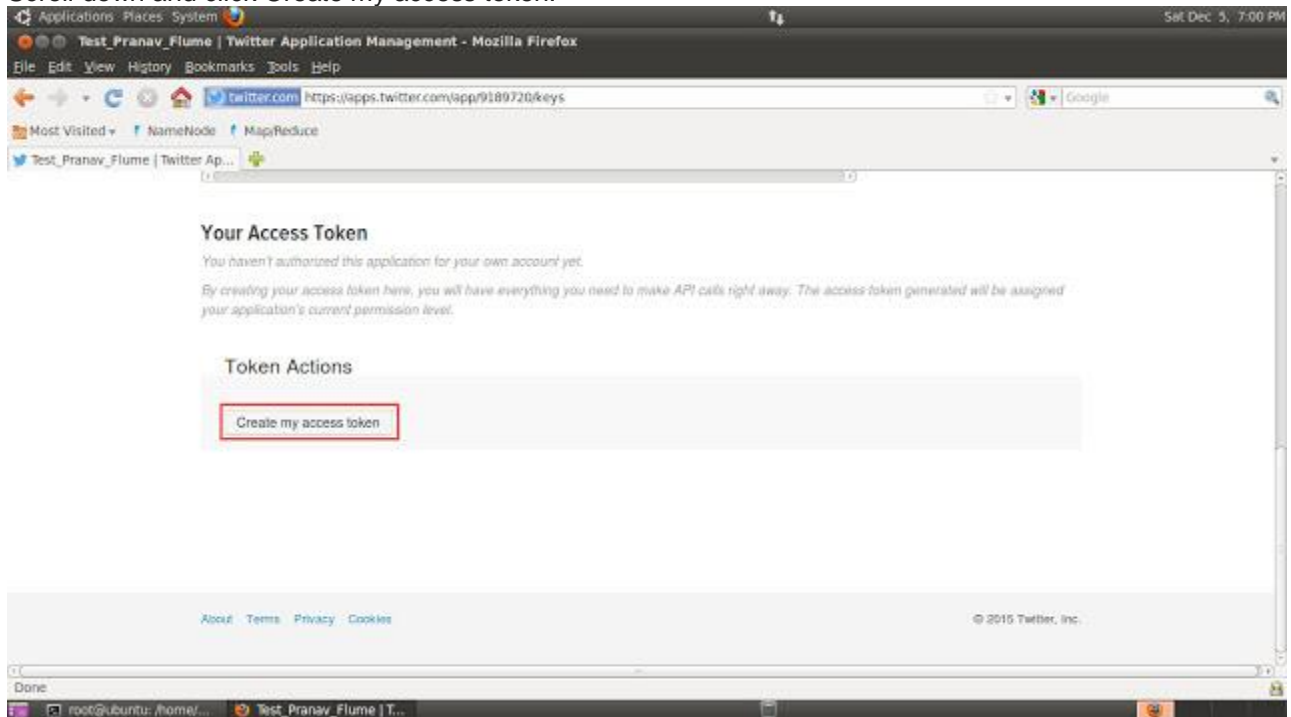
Organization website None

Application Settings

Your application's Consumer Key and Secret are used to [authenticate](#) requests to the Twitter Platform.

Access level Read and write (modify app permissions)

Scroll down and click Create my access token.



If the access token successfully created, you will get below message as Status.

Consumer Key (API Key)	GK290CbplPqBMiM8ZwAQgmY5i
Consumer Secret (API Secret)	URDN1PMwjDjsb5phFo9XJyKU0Y4TubBhu7Q9LJSTgIDdZoukhn
Access Level	Read and write (modify app permissions)
Owner	ramesh1024
Owner ID	277481794

Application Actions
[Regenerate Consumer Key and Secret](#) [Change App Permissions](#)

Your Access Token

This access token can be used to make API requests on your own account's behalf. Do not share your access token secret with anyone.

Access Token	277481794-aMWA5069hTBKFbdtee6mdIOegoKJcP29rh4zkbYw
Access Token Secret	QQ0GJNcCygEzAZmL21hPKi2F0u0NdpCEVgbv11YAL8Cf
Access Level	Read and write
Owner	ramesh1024

These 4 highlighted details in above screenshot, we need as a part of flume configuration file(mentioned in the below screenshot).

We could place the flume.conf file at the \$flume_home/conf/flume.conf

Add the details into flume.conf as below.

ConsumerKey == Consumer Key from Twitter

ConsumerSecret == Consumer Secret from Twitter

accessToken == Access Token from twitter

accessTokenSecret == Access Token Secret from twitter

```
TwitterAgent.sources = Twitter
TwitterAgent.channels = MemChannel
TwitterAgent.sinks = HDFS

# Describing/Configuring the source
TwitterAgent.sources.Twitter.channels = MemChannel
TwitterAgent.sources.Twitter.type = org.apache.flume.source.twitter.TwitterSource
TwitterAgent.sources.Twitter.consumerKey=GK290CbplPqBMiM8ZwAQgmY5i
TwitterAgent.sources.Twitter.consumerSecret=URDN1PMwjDjsb5phFo9XJyKU0Y4TubBhu7Q9LJSTgIDdZoukhn
TwitterAgent.sources.Twitter.accessToken=277481794-aMWA5069hTBKFbdtee6mdIOegoKJcP29rh4zkbYw
TwitterAgent.sources.Twitter.accessTokenSecret= QQ0GJNcCygEzAZmL21hPKi2F0u0NdpCEVgbv11YAL8Cf
TwitterAgent.sources.Twitter.keywords=hadoop,hbase, nosql
# Describing/Configuring the sink

TwitterAgent.sources.Twitter.keywords= hadoop,election,sports, cricket,Big data

TwitterAgent.sinks.HDFS.channel=MemChannel
TwitterAgent.sinks.HDFS.type=hdfs
TwitterAgent.sinks.HDFS.hdfs.path=hdfs://localhost:8020/user/flume/tweets1
TwitterAgent.sinks.HDFS.hdfs.fileType=DataStream
TwitterAgent.sinks.HDFS.hdfs.writeformat=Text
TwitterAgent.sinks.HDFS.hdfs.batchSize=1000
TwitterAgent.sinks.HDFS.hdfs.rollSize=0
TwitterAgent.sinks.HDFS.hdfs.rollCount=10000
TwitterAgent.sinks.HDFS.hdfs.rollInterval=600

TwitterAgent.channels.MemChannel.type=memory
TwitterAgent.channels.MemChannel.capacity=10000
TwitterAgent.channels.MemChannel.transactionCapacity=1000
~
~
~
~
```

Run the below command to fetch the data from twitter.

```
[acacgild@localhost ~]$ flume-ng agent --conf ./conf/ -f $FLUME_HOME/conf/flume.conf -Dflume.root.logger=DEBUG,console -n TwitterAgent
Info: Including Hadoop libraries found via (/home/acacgild/install/hadoop/hadoop-2.6.5/bin/hadoop) for HDFS access
Info: Including HBase libraries found via (/home/acacgild/install/hbase/hbase-1.2.6/bin/hbase) for HBase access
Info: Including Hive libraries found via (/home/acacgild/install/hive/apache-hive-2.3.2-bin) for Hive access
+ exec /usr/java/jdk1.8.0_151/bin/java -Xmx20m -Dflume.root.logger=DEBUG,console -cp './conf:/home/acacgild/install/flume/apache-flume-2.6.5-bin/lib/*:/home/acacgild/install/hadoop/hadoop-2.6.5/etc/hadoop:/home/acacgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/*:/home/acacgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/*:/home/acacgild/install/hadoop/hadoop-2.6.5/share/hadoop/hdfs:/home/acacgild/install/hadoop/hadoop-2.6.5/share/hadoop/hdfs/lib/*:/home/acacgild/install/hadoop/hadoop-2.6.5/share/hadoop/hdfs/*:/home/acacgild/install/hadoop/hadoop-2.6.5/share/hadoop/yarn/lib/*:/home/acacgild/install/hadoop/hadoop-2.6.5/share/hadoop/yarn/*:/home/acacgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/lib/*:/home/acacgild/install/hadoop/hadoop-2.6.5/share/hadoop/mapreduce/*:/home/acacgild/install/hadoop/hadoop-2.6.5/contrib/capacity-scheduler/*.jar:/home/acacgild/install/hbase/hbase-1.2.6/conf:/usr/java/jdk1.8.0_151/lib/tools.jar:/home/acacgild/install/hbase/hbase-1.2.6:/home/acacgild/install/hbase/hbase-1.2.6/lib/activation-1.1.jar:/home/acacgild/install/hbase/hbase-1.2.6/lib/aopalliance-1.0.jar:/home/acacgild/install/hbase/hbase-1.2.6/lib/apacheds-i18n-2.0.0-M15.jar:/home/acacgild/install/hbase/hbase-1.2.6/lib/apacheds-kerberos-codec-2.0.0-M15.jar:/home/acacgild/install/hbase/hbase-1.2.6/lib/api-asn1-api-1.0.0-M20.jar:/home/acacgild/install/hbase/hbase-1.2.6/lib/api-util-1.0.0-M20.jar:/home/acacgild/install/hbase/hbase-1.2.6/lib/asm-3.1.jar:/home/acacgild/install/hbase/hbase-1.2.6/lib/commons-beanutils-1.7.0.jar:/home/acacgild/install/hbase/hbase-1.2.6/lib/commons-beanutils-core-1.8.0.jar:/home/acacgild/install/hbase/hbase-1.2.6/lib/commons-cli-1.2.jar:/home/acacgild/install/hbase/hbase-1.2.6/lib/commons-codec-1.9.jar:/home/acacgild/install/hbase/hbase-1.2.6/lib/commons-collections-3.2.2.jar:/home/acacgild/install/hbase/hbase-1.2.6/lib/commons-compress-1.4.1.jar:/home/acacgild/install/hbase/hbase-1.2.6/lib/commons-configuration-1.6.jar:/home/acacgild/install/hbase/hbase-1.2.6/lib/commons-daemon-1.0.13.jar:/home/acacgild/install/hbase/hbase-1.2.6/lib/commons-digester-1.8.jar:/
```

If everything goes right, we can see the below output, on HDFS location, mentioned in the flume.conf file.

HDFS/user/flume/tweet... x Google x +

localhost:5007/browseBlock.jsp?blockId=1073742993&blockSize=986158&genStamp=2173&filename v Search

File: /user/flume/tweets1/FlumeData.1529050723002

Goto : /user/flume/tweets1 go

[Go back to dir listing](#)
[Advanced view/download options](#)
[View Next chunk](#)

Obj avro.schema
{ "type": "record", "name": "Doc", "doc": "adoc", "fields": [{ "name": "id", "type": "string", "name": "user_friends_count", "type": "int", "null": true }, { "name": "user_location", "type": "string", "null": true }, { "name": "user_description", "type": "string", "null": true }, { "name": "user_statuses_count", "type": "int", "null": true }, { "name": "user_followers_count", "type": "int", "null": true }, { "name": "user_name", "type": "string", "null": true }, { "name": "user_screen_name", "type": "string", "null": true }, { "name": "created_at", "type": "string", "null": true }, { "name": "text", "type": "string", "null": true }, { "name": "retweet_count", "type": "long", "null": true }, { "name": "retweeted", "type": "boolean", "null": true }, { "name": "in_reply_to_user_id", "type": "long", "null": true }, { "name": "source", "type": "string", "null": true }, { "name": "in_reply_to_status_id", "type": "long", "null": true }, { "name": "media_url_https", "type": "string", "null": true }, { "name": "expanded_url", "type": "string", "null": true }] }

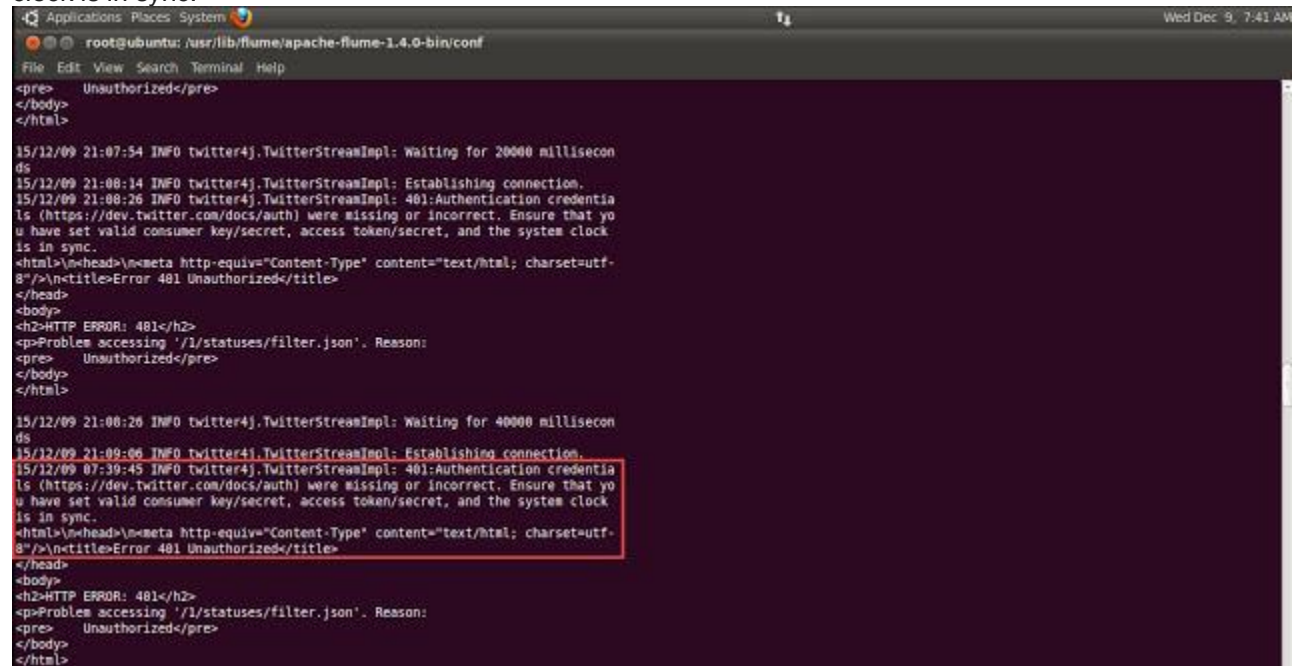
Obj avro.schema
{ "type": "record", "name": "Doc", "doc": "adoc", "fields": [{ "name": "id", "type": "string", "name": "user_friends_count", "type": "int", "null": true }, { "name": "user_location", "type": "string", "null": true }, { "name": "user_description", "type": "string", "null": true }, { "name": "user_statuses_count", "type": "int", "null": true }, { "name": "user_followers_count", "type": "int", "null": true }, { "name": "user_name", "type": "string", "null": true }, { "name": "user_screen_name", "type": "string", "null": true }, { "name": "created_at", "type": "string", "null": true }, { "name": "text", "type": "string", "null": true }, { "name": "retweet_count", "type": "long", "null": true }, { "name": "retweeted", "type": "boolean", "null": true }, { "name": "in_reply_to_user_id", "type": "long", "null": true }, { "name": "source", "type": "string", "null": true }, { "name": "in_reply_to_status_id", "type": "long", "null": true }, { "name": "media_url_https", "type": "string", "null": true }, { "name": "expanded_url", "type": "string", "null": true }] }

writing in wappad
หลวงปู่ฝั้น อาจาโรน: ๐๐๐๐๐๐
[marvel fan]
16 years old

Error Rectifications

While executing the Flume command, user may get the following error.

Error says that Ensure that you have set the valid consumer key/secret, access token/secret and system clock is in sync.



```
root@ubuntu: /usr/lib/flume/apache-flume-1.4.0-bin/conf
File Edit View Search Terminal Help

<pre> Unauthorized</pre>
</body>
</html>

15/12/09 21:07:54 INFO twitter4j.TwitterStreamImpl: Waiting for 20000 millisecond
15/12/09 21:08:14 INFO twitter4j.TwitterStreamImpl: Establishing connection.
15/12/09 21:08:26 INFO twitter4j.TwitterStreamImpl: 401: Authentication credentials
(https://dev.twitter.com/docs/auth) were missing or incorrect. Ensure that yo
u have set valid consumer key/secret, access token/secret, and the system clock
is in sync.
<html>\n<head>\n<meta http-equiv="Content-Type" content="text/html; charset=utf-
8"/>\n<title>Error 401 Unauthorized</title>
</head>
<body>
<h2>HTTP ERROR: 401</h2>
<p>Problem accessing '/1/statuses/filter.json'. Reason:
<pre> Unauthorized</pre>
</body>
</html>

15/12/09 21:08:26 INFO twitter4j.TwitterStreamImpl: Waiting for 40000 millisecond
15/12/09 21:09:06 INFO twitter4j.TwitterStreamImpl: Establishing connection.
15/12/09 21:09:45 INFO twitter4j.TwitterStreamImpl: 401: Authentication credentials
(https://dev.twitter.com/docs/auth) were missing or incorrect. Ensure that yo
u have set valid consumer key/secret, access token/secret, and the system clock
is in sync.
<html>\n<head>\n<meta http-equiv="Content-Type" content="text/html; charset=utf-
8"/>\n<title>Error 401 Unauthorized</title>
</head>
<body>
<h2>HTTP ERROR: 401</h2>
<p>Problem accessing '/1/statuses/filter.json'. Reason:
<pre> Unauthorized</pre>
</body>
</html>
```

Resolution 1: First of all, check the access key/secret and access tokens/secret are correct as per twitter's application values.

Resolution 2: There may be a chance that the Host OS and the guest OS has difference in timezone, e.g., My Host OS Windows 8 has time set as per India (IST) and Guest OS has US time zone.

To resolve the timezone problem, do the following.

From root user login

Stop ntp service

\$ service ntp stop => This will stop the ntp service

\$ ntpdate ntp.ubuntu.com => This will update the Guest OS time same as Host OS

\$ service ntp start => This will start the ntp service

Re-execute the Flume command...