

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`

```
cloudera@cloudera-vm: ~  
File Edit View Search Terminal Help  
cloudera@cloudera-vm:~$ wget http://apache.mirrors.hoobly.com/flume/1.4.0/apache-flume-1.4.0-bin.tar.gz  
  
cloudera@cloudera-vm: ~  
File Edit View Search Terminal Help  
cloudera@cloudera-vm:~$ wget http://apache.mirrors.hoobly.com/flume/1.4.0/apache-flume-1.4.0-bin.tar.gz  
--2013-08-05 02:57:07-- http://apache.mirrors.hoobly.com/flume/1.4.0/apache-flume-1.4.0-bin.tar.gz  
Resolving apache.mirrors.hoobly.com... 66.160.172.98  
Connecting to apache.mirrors.hoobly.com[66.160.172.98]:80... connected.  
HTTP request sent, awaiting response... 200 OK  
Length: 60965956 (58M) [application/x-gzip]  
Saving to: 'apache-flume-1.4.0-bin.tar.gz'  
  
0% [ ] 101,890 11.4K/s eta 89m 17s
```

Check whether the flume tar is present or not:

Command: `ls`

```
cloudera@cloudera-vm: ~  
File Edit View Search Terminal Help  
cloudera@cloudera-vm:~$ ls  
apache-flume-1.4.0-bin.tar.gz Desktop mysql-connector-java-5.0.8.tar.gz oraosch-2.2.0.zip  
cloudera mysql-connector-java-5.0.8 oraloader-2.2.0.x86_64.zip student  
cloudera@cloudera-vm:~$
```

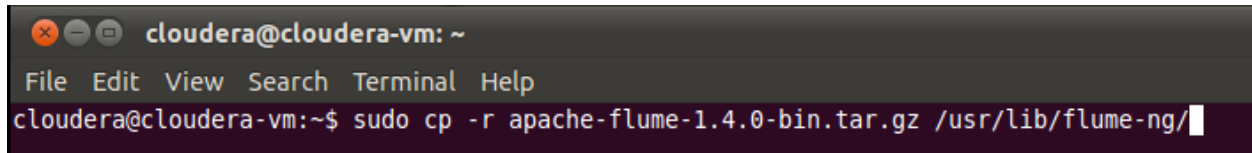
Create flume-ng directory:

Command: `sudo mkdir /usr/lib/flume-ng`

```
cloudera@cloudera-vm: ~  
File Edit View Search Terminal Help  
cloudera@cloudera-vm:~$ sudo mkdir /usr/lib/flume-ng
```

Copy the flume tar to flume-ng directory:

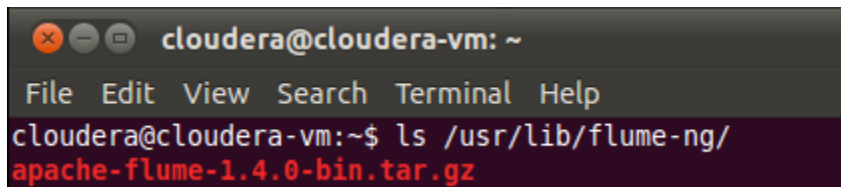
Command: `sudo cp -r apache-flume-1.4.0-bin.tar.gz /usr/lib/flume-ng/`

A terminal window titled 'cloudera@cloudera-vm: ~' with a menu bar (File, Edit, View, Search, Terminal, Help). The command 'sudo cp -r apache-flume-1.4.0-bin.tar.gz /usr/lib/flume-ng/' is entered and executed, with the cursor at the end of the line.

```
cloudera@cloudera-vm: ~  
File Edit View Search Terminal Help  
cloudera@cloudera-vm:~$ sudo cp -r apache-flume-1.4.0-bin.tar.gz /usr/lib/flume-ng/
```

Check whether flume tar is copied or not:

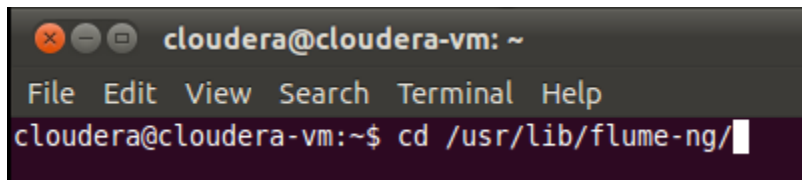
Command: `ls /usr/lib/flume-ng/`

A terminal window titled 'cloudera@cloudera-vm: ~' with a menu bar (File, Edit, View, Search, Terminal, Help). The command 'ls /usr/lib/flume-ng/' is entered and executed, showing the output 'apache-flume-1.4.0-bin.tar.gz' in red text.

```
cloudera@cloudera-vm: ~  
File Edit View Search Terminal Help  
cloudera@cloudera-vm:~$ ls /usr/lib/flume-ng/  
apache-flume-1.4.0-bin.tar.gz
```

Change directory to flume-ng:

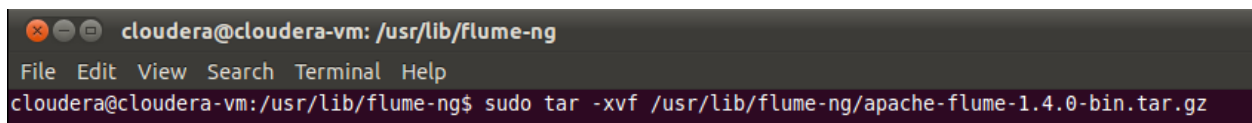
Command: `cd /usr/lib/flume-ng/`

A terminal window titled 'cloudera@cloudera-vm: ~' with a menu bar (File, Edit, View, Search, Terminal, Help). The command 'cd /usr/lib/flume-ng/' is entered and executed, with the cursor at the end of the line.

```
cloudera@cloudera-vm: ~  
File Edit View Search Terminal Help  
cloudera@cloudera-vm:~$ 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`

A terminal window titled 'cloudera@cloudera-vm: /usr/lib/flume-ng' with a menu bar (File, Edit, View, Search, Terminal, Help). The command 'sudo tar -xvf /usr/lib/flume-ng/apache-flume-1.4.0-bin.tar.gz' is entered and executed.

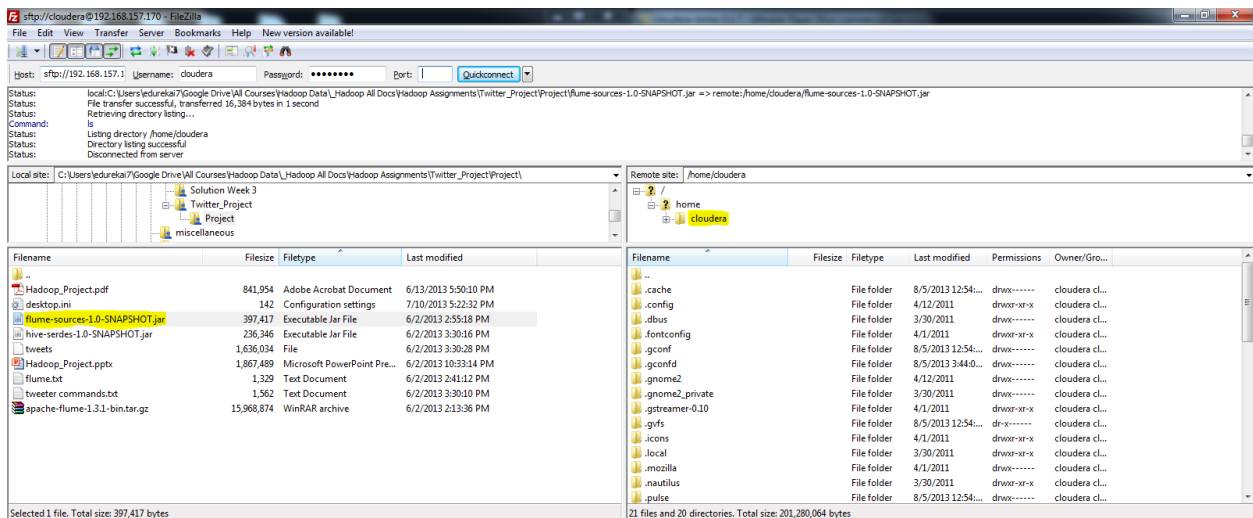
```
cloudera@cloudera-vm: /usr/lib/flume-ng  
File Edit View Search Terminal Help  
cloudera@cloudera-vm:/usr/lib/flume-ng$ sudo tar -xvf /usr/lib/flume-ng/apache-flume-1.4.0-bin.tar.gz
```

Check whether flume files are extracted or not:

Command:ls

```
cloudera@cloudera-vm: /usr/lib/flume-ng
File Edit View Search Terminal Help
cloudera@cloudera-vm:/usr/lib/flume-ng$ ls
apache-flume-1.4.0-bin  apache-flume-1.4.0-bin.tar.gz
```

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/

```
cloudera@cloudera-vm: /usr/lib/flume-ng
File Edit View Search Terminal Help
cloudera@cloudera-vm:/usr/lib/flume-ng$ sudo mv /home/cloudera/flume-sources-1.0-SNAPSHOT.jar /usr/lib/flume-ng/apache-flume-1.4.0-bin/lib/
```

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: `sudo cp /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/flume-ng/apache-flume-1.4.0-bin/conf/flume-env.sh
```

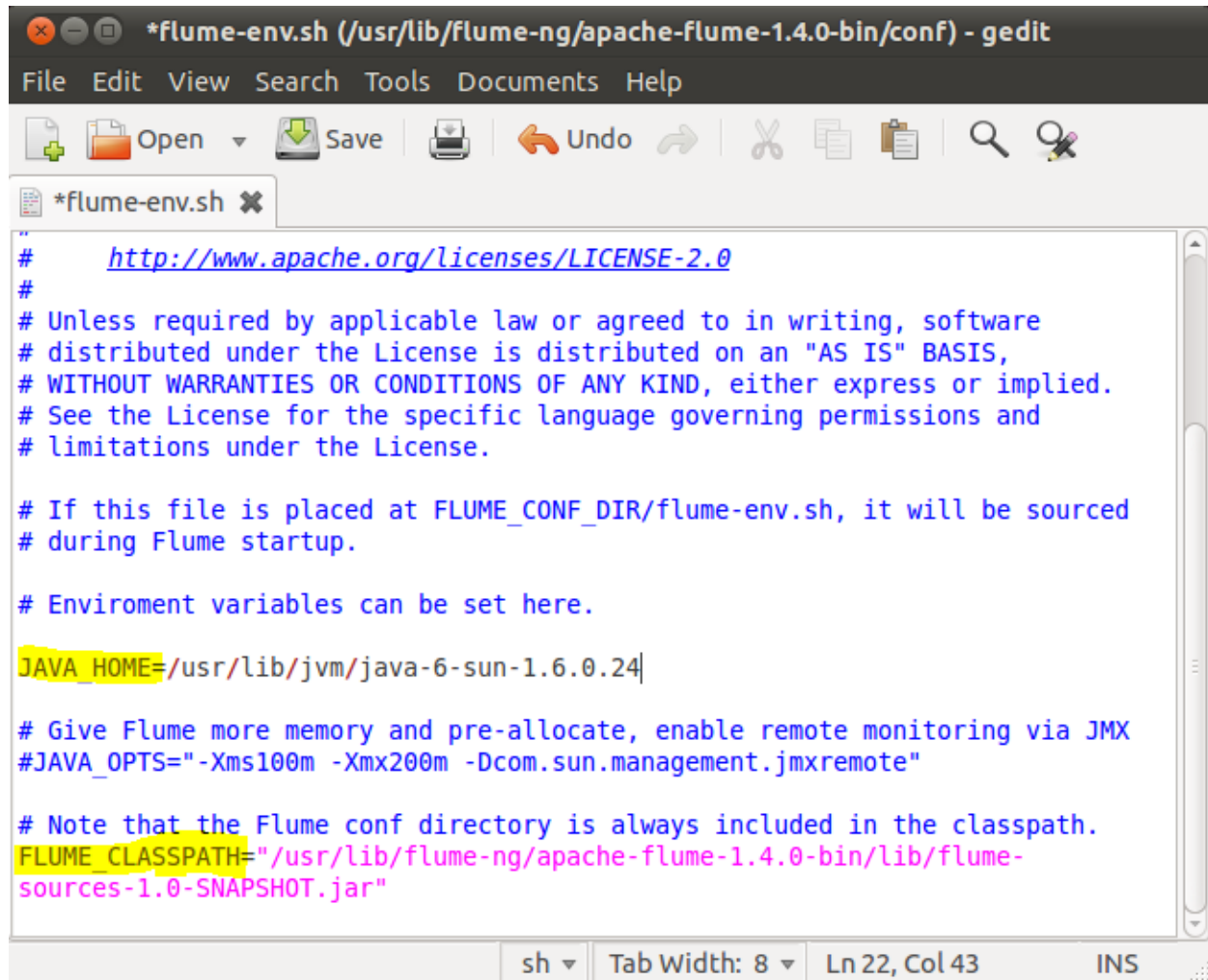
Open flume-env.sh:

Command: `sudo gedit /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

flume-env.sh (/usr/lib/flume-ng/apache-flume-1.4.0-bin/conf) - gedit
File Edit View Search Tools Documents Help
Open Save Undo Redo Cut Copy Paste Find
flume-env.sh x
# Licensed to the Apache Software Foundation (ASF) under one
# or more contributor license agreements. See the NOTICE file
# distributed with this work for additional information
# regarding copyright ownership. The ASF licenses this file
```

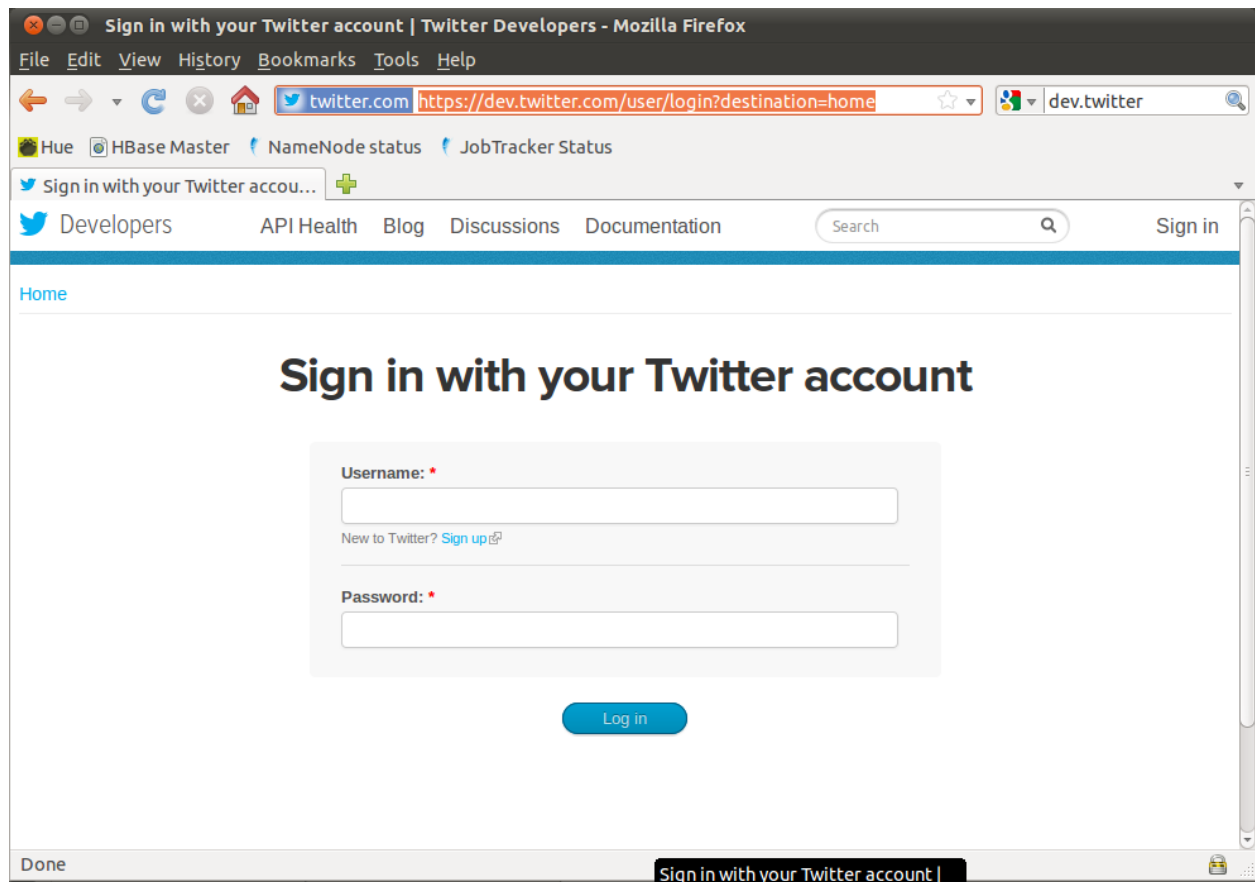
Edit flume-env.sh according to the below snapshot:



```
# http://www.apache.org/licenses/LICENSE-2.0
#
# Unless required by applicable law or agreed to in writing, software
# distributed under the License is distributed on an "AS IS" BASIS,
# WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
# See the License for the specific language governing permissions and
# limitations under the License.
#
# If this file is placed at FLUME_CONF_DIR/flume-env.sh, it will be sourced
# during Flume startup.
#
# Enviroment variables can be set here.
JAVA_HOME=/usr/lib/jvm/java-6-sun-1.6.0.24|
# Give Flume more memory and pre-allocate, enable remote monitoring via JMX
#JAVA_OPTS="-Xms100m -Xmx200m -Dcom.sun.management.jmxremote"
# Note that the Flume conf directory is always included in the classpath.
FLUME_CLASSPATH="/usr/lib/flume-ng/apache-flume-1.4.0-bin/lib/flume-
sources-1.0-SNAPSHOT.jar"
```

Open a Browser and go to the below URL:

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



Enter your Twitter account credentials:

Sign in with your Twitter account


Username: *

New to Twitter? [Sign up](#)

Password: *

Log in

Create a new application:


 Developers


API Health

Blog

Discussions

Documentation

Search 



[Home](#)

My applications

Create a new application

Enter all the details in the application:

Create an application

Application Details

Name: *

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

Description: *

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

Website: *

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.

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

Developer Rules Of The Road

5. While Twitter discourages development in this area, some Services or applications attempt to replicate Twitter's core user experience, typically by accessing the home timeline, account settings, or direct messages API endpoints or [User Streams](#) product. The following additional rules apply to Services or applications that fall within this category.

You must:

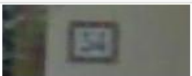

- A. use the Twitter API [as provided by Twitter](#) for functionalities in your Service that are substantially similar to a feature of the Twitter service and present this to your users as the default option. Some examples include media storage and sharing ([pic.twitter.com](#)), trending topics, and suggested user lists.
- B. not pay, or offer to pay, third parties for distribution. This includes offering compensation for downloads (other than transactional fees) or other mechanisms of traffic acquisition.
- C. not arrange for your Service to be pre-installed on any device, promoted as a "zero-rated" service, or marketed as part of a specialized data plan.
- D. not frame or otherwise reproduce significant portions of the Twitter service. You should display Twitter Content from the Twitter API.
- E. display a prominent link or button in your Service that directs new users to Twitter's sign-up functionality.
- F. not use Twitter Content or other data collected from end users to create or maintain a separate status update or social network database or service.




☒ Yes, I agree


By clicking the "I Agree" button, you acknowledge that you have read and understand this agreement and agree to be bound by its terms and conditions.

CAPTCHA

This question is for testing whether you are a human visitor and to prevent automated spam submissions.







[Privacy & Terms](#)

Create your Twitter application



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	G04dAE50aHeqUEh6VZoF0MAC2YMcgkgu8VpWaJam8
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

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	Op8DkYUhQmn8uRt7C71w
Consumer secret	G04dAE50aHeqUEh6VZ0F0MAC2YMcgkgu8VpWaJam8
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

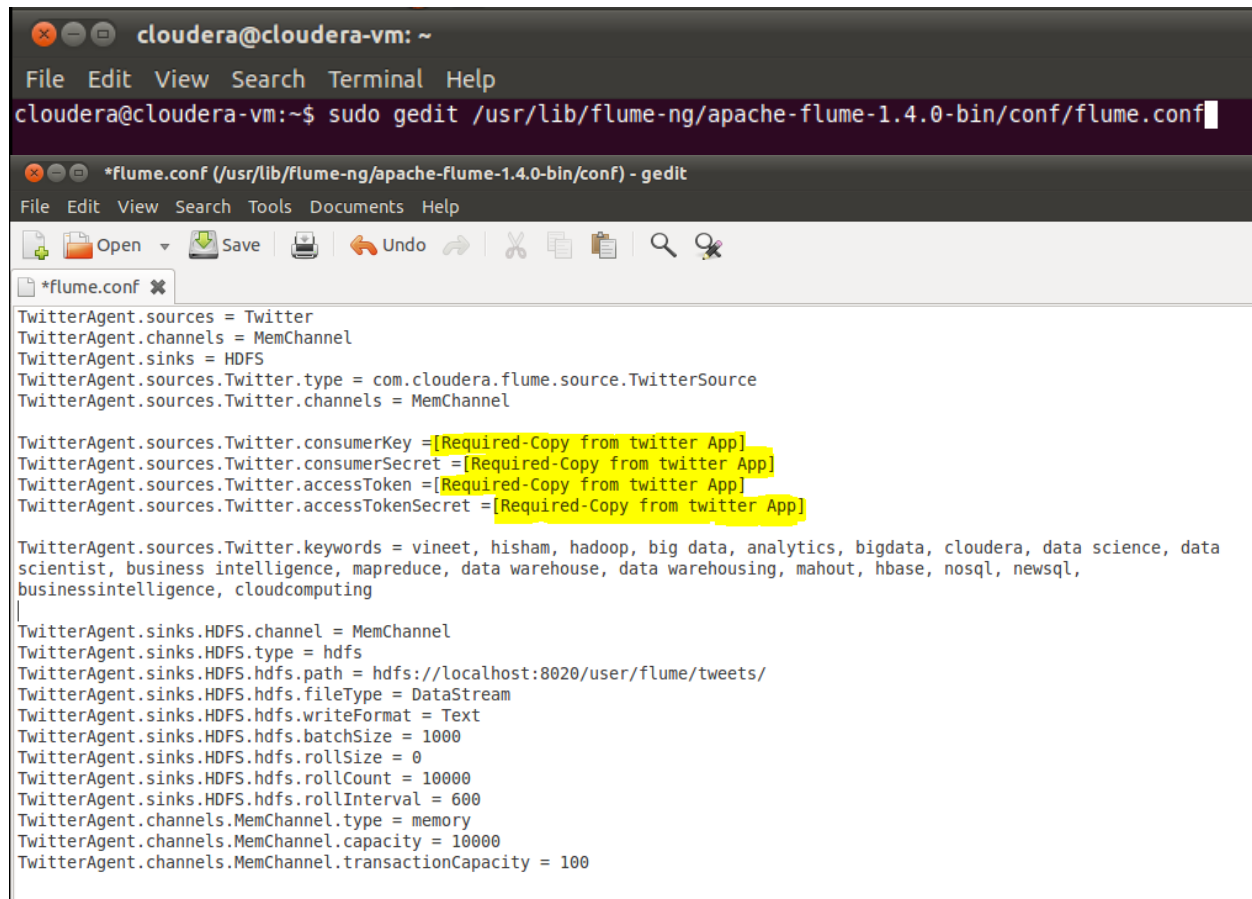
Use the access token string as your "oauth_token" and the access token secret as your "oauth_token_secret" to sign requests with your own Twitter account. Do not share your oauth_token_secret with anyone.

Access token	1635433267-HvStt9EggeUhrxNTTJhc4DKrDymxfHG3oroJpc
Access token secret	gBX9kFeM1K7zSrOIMhUUYfQmvLeuWv7msOEgyhLls
Access level	Read-only

[Recreate my access token](#)

Edit flume.conf:

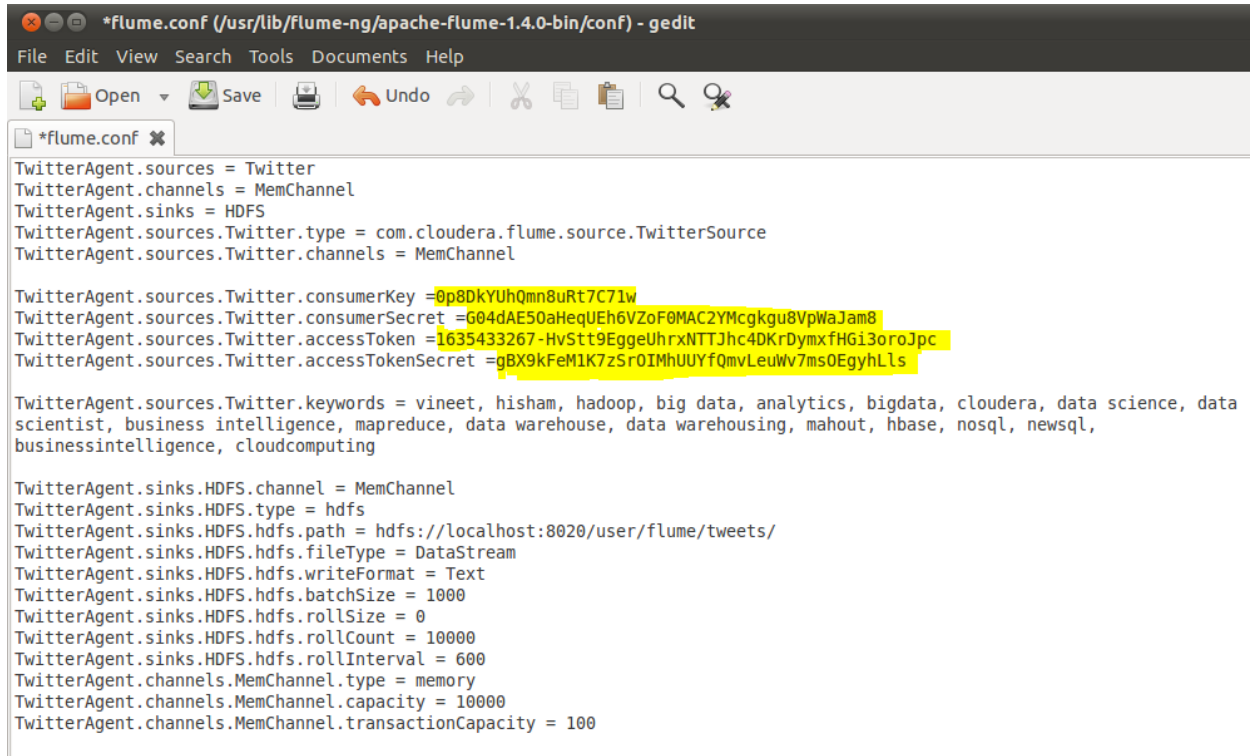
Command: sudo gedit /usr/lib/flume-ng/apache-flume-1.4.0-bin/conf/flume.conf



The screenshot shows a terminal window at the top with the command `sudo gedit /usr/lib/flume-ng/apache-flume-1.4.0-bin/conf/flume.conf` executed. Below the terminal is a gedit editor window titled `*flume.conf (/usr/lib/flume-ng/apache-flume-1.4.0-bin/conf) - gedit`. The editor displays the contents of the `flume.conf` file, which is a configuration file for Apache Flume. The configuration includes settings for a Twitter source, a MemChannel, and an HDFS sink. The Twitter source is configured with a consumer key, consumer secret, access token, and access token secret, all of which are marked as "[Required-Copy from twitter App]". The HDFS sink is configured with a channel of `MemChannel`, a type of `hdfs`, and a path of `hdfs://localhost:8020/user/flume/tweets/`. The sink is also configured with a `DataStream` file type, a `Text` write format, a `batchSize` of `1000`, a `rollSize` of `0`, a `rollCount` of `10000`, and a `rollInterval` of `600`. The channel is configured with a `memory` type, a `capacity` of `10000`, and a `transactionCapacity` of `100`.

```
cloudera@cloudera-vm: ~  
File Edit View Search Terminal Help  
cloudera@cloudera-vm:~$ sudo gedit /usr/lib/flume-ng/apache-flume-1.4.0-bin/conf/flume.conf  
  
*flume.conf (/usr/lib/flume-ng/apache-flume-1.4.0-bin/conf) - gedit  
File Edit View Search Tools Documents Help  
Open Save Undo  
*flume.conf ✕  
TwitterAgent.sources = Twitter  
TwitterAgent.channels = MemChannel  
TwitterAgent.sinks = HDFS  
TwitterAgent.sources.Twitter.type = com.cloudera.flume.source.TwitterSource  
TwitterAgent.sources.Twitter.channels = MemChannel  
  
TwitterAgent.sources.Twitter.consumerKey = [Required-Copy from twitter App]  
TwitterAgent.sources.Twitter.consumerSecret = [Required-Copy from twitter App]  
TwitterAgent.sources.Twitter.accessToken = [Required-Copy from twitter App]  
TwitterAgent.sources.Twitter.accessTokenSecret = [Required-Copy from twitter App]  
  
TwitterAgent.sources.Twitter.keywords = vineet, hisham, hadoop, big data, analytics, bigdata, cloudera, data science, data  
scientist, business intelligence, mapreduce, data warehouse, data warehousing, mahout, hbase, nosql, newsq,  
businessintelligence, cloudcomputing  
  
TwitterAgent.sinks.HDFS.channel = MemChannel  
TwitterAgent.sinks.HDFS.type = hdfs  
TwitterAgent.sinks.HDFS.hdfs.path = hdfs://localhost:8020/user/flume/tweets/  
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 = 100
```

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



```
*flume.conf (/usr/lib/flume-ng/apache-flume-1.4.0-bin/conf) - gedit
File Edit View Search Tools Documents Help

TwitterAgent.sources = Twitter
TwitterAgent.channels = MemChannel
TwitterAgent.sinks = HDFS
TwitterAgent.sources.Twitter.type = com.cloudera.flume.source.TwitterSource
TwitterAgent.sources.Twitter.channels = MemChannel

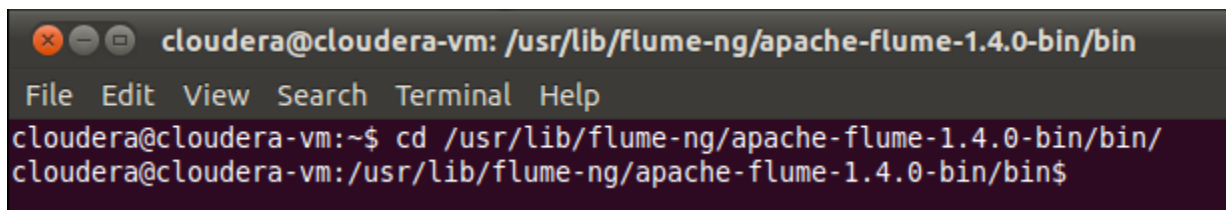
TwitterAgent.sources.Twitter.consumerKey = 0p8DkYUhQmn8uRt7C71w
TwitterAgent.sources.Twitter.consumerSecret = G04dAE50aHeqUEh6VZoF0MAC2YMcgkgu8VpWaJam8
TwitterAgent.sources.Twitter.accessToken = 1635433267-HvStt9EggeUhrxNTTJhc4DKrDymxfHG13oroJpc
TwitterAgent.sources.Twitter.accessTokenSecret = gBX9kFeM1K7zSr0IMhUUYfQmvLeuWv7ms0EgyhLls

TwitterAgent.sources.Twitter.keywords = vineet, hisham, hadoop, big data, analytics, bigdata, cloudera, data science, data
scientist, business intelligence, mapreduce, data warehouse, data warehousing, mahout, hbase, nosql, newsq,
businessintelligence, cloudcomputing

TwitterAgent.sinks.HDFS.channel = MemChannel
TwitterAgent.sinks.HDFS.type = hdfs
TwitterAgent.sinks.HDFS.hdfs.path = hdfs://localhost:8020/user/flume/tweets/
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 = 100
```

Change Directory to the bin folder of apache flume:

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

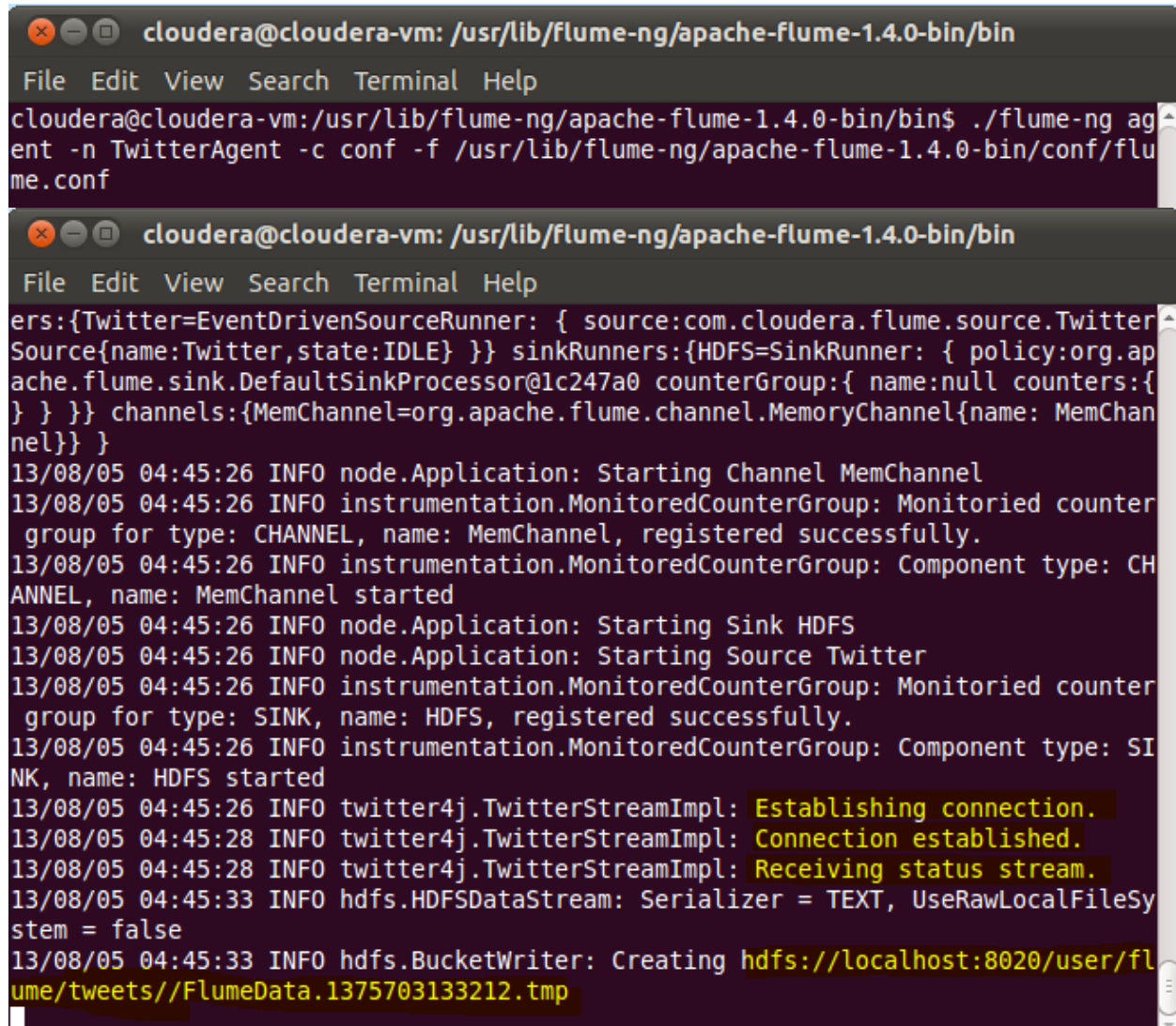


```
cloudera@cloudera-vm: /usr/lib/flume-ng/apache-flume-1.4.0-bin/bin
File Edit View Search Terminal Help

cloudera@cloudera-vm:~$ cd /usr/lib/flume-ng/apache-flume-1.4.0-bin/bin/
cloudera@cloudera-vm:/usr/lib/flume-ng/apache-flume-1.4.0-bin/bin$
```

Start fetching the data from twitter:

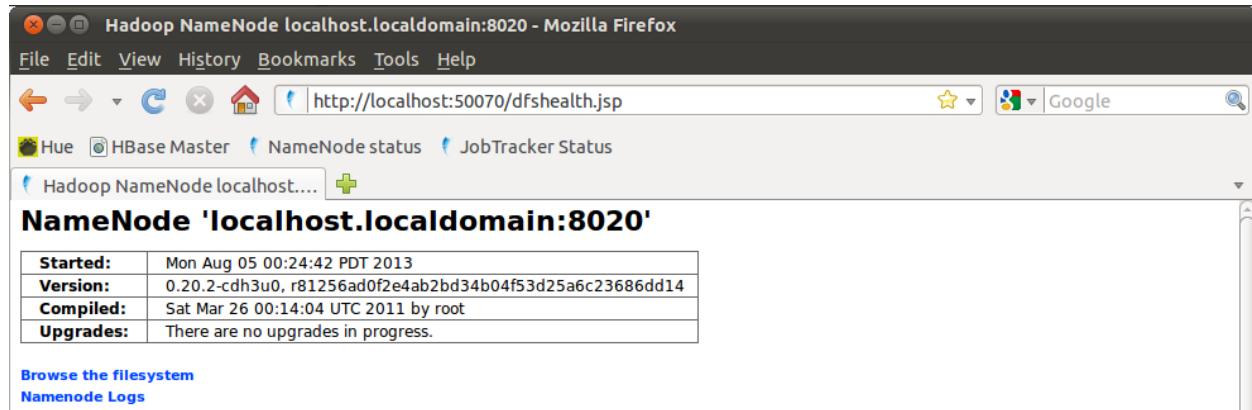
Command: `./flume-ng agent -n TwitterAgent -c conf -f /usr/lib/flume-ng/apache-flume-1.4.0-bin/conf/flume.conf`



```
cloudera@cloudera-vm: /usr/lib/flume-ng/apache-flume-1.4.0-bin/bin
File Edit View Search Terminal Help
cloudera@cloudera-vm:/usr/lib/flume-ng/apache-flume-1.4.0-bin/bin$ ./flume-ng ag
ent -n TwitterAgent -c conf -f /usr/lib/flume-ng/apache-flume-1.4.0-bin/conf/fl
me.conf

cloudera@cloudera-vm: /usr/lib/flume-ng/apache-flume-1.4.0-bin/bin
File Edit View Search Terminal Help
ers:{Twitter=EventDrivenSourceRunner: { source:com.cloudera.flume.source.Twitter
Source{name:Twitter,state:IDLE} }} sinkRunners:{HDFS=SinkRunner: { policy:org.ap
ache.flume.sink.DefaultSinkProcessor@1c247a0 counterGroup:{ name:null counters:{
} } }} channels:{MemChannel=org.apache.flume.channel.MemoryChannel{name: MemChan
nel}} }
13/08/05 04:45:26 INFO node.Application: Starting Channel MemChannel
13/08/05 04:45:26 INFO instrumentation.MonitoredCounterGroup: Monitored counter
group for type: CHANNEL, name: MemChannel, registered successfully.
13/08/05 04:45:26 INFO instrumentation.MonitoredCounterGroup: Component type: CH
ANNEL, name: MemChannel started
13/08/05 04:45:26 INFO node.Application: Starting Sink HDFS
13/08/05 04:45:26 INFO node.Application: Starting Source Twitter
13/08/05 04:45:26 INFO instrumentation.MonitoredCounterGroup: Monitored counter
group for type: SINK, name: HDFS, registered successfully.
13/08/05 04:45:26 INFO instrumentation.MonitoredCounterGroup: Component type: SI
NK, name: HDFS started
13/08/05 04:45:26 INFO twitter4j.TwitterStreamImpl: Establishing connection.
13/08/05 04:45:28 INFO twitter4j.TwitterStreamImpl: Connection established.
13/08/05 04:45:28 INFO twitter4j.TwitterStreamImpl: Receiving status stream.
13/08/05 04:45:33 INFO hdfs.HDFSDataStream: Serializer = TEXT, UseRawLocalFileSy
stem = false
13/08/05 04:45:33 INFO hdfs.BucketWriter: Creating hdfs://localhost:8020/user/fl
ume/tweets//FlumeData.1375703133212.tmp
```

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

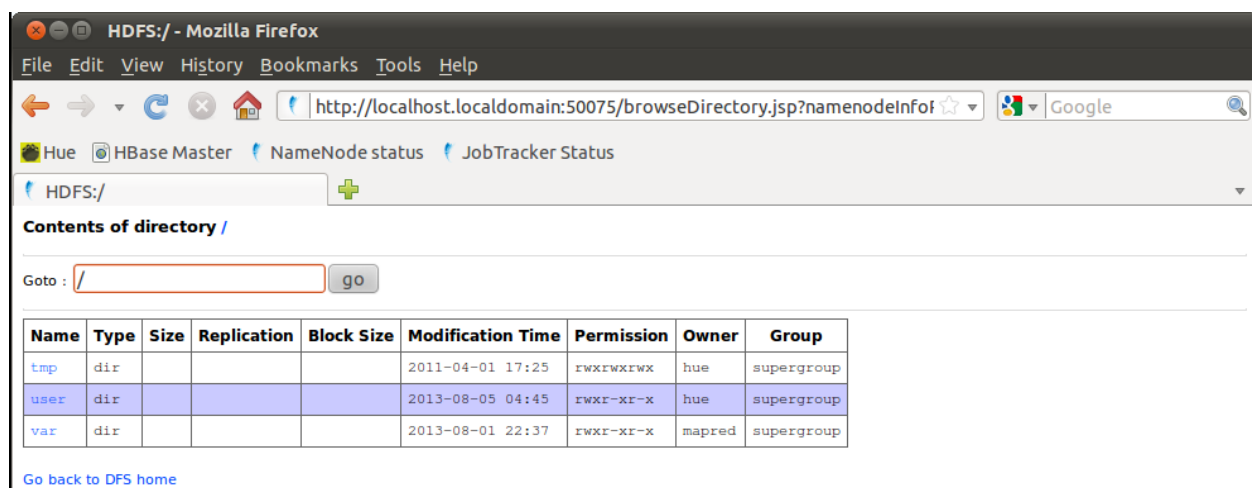


NameNode 'localhost.localdomain:8020'

Started:	Mon Aug 05 00:24:42 PDT 2013
Version:	0.20.2-cdh3u0, r81256ad0f2e4ab2bd34b04f53d25a6c23686dd14
Compiled:	Sat Mar 26 00:14:04 UTC 2011 by root
Upgrades:	There are no upgrades in progress.

[Browse the filesystem](#)
[NameNode Logs](#)

Click on user:



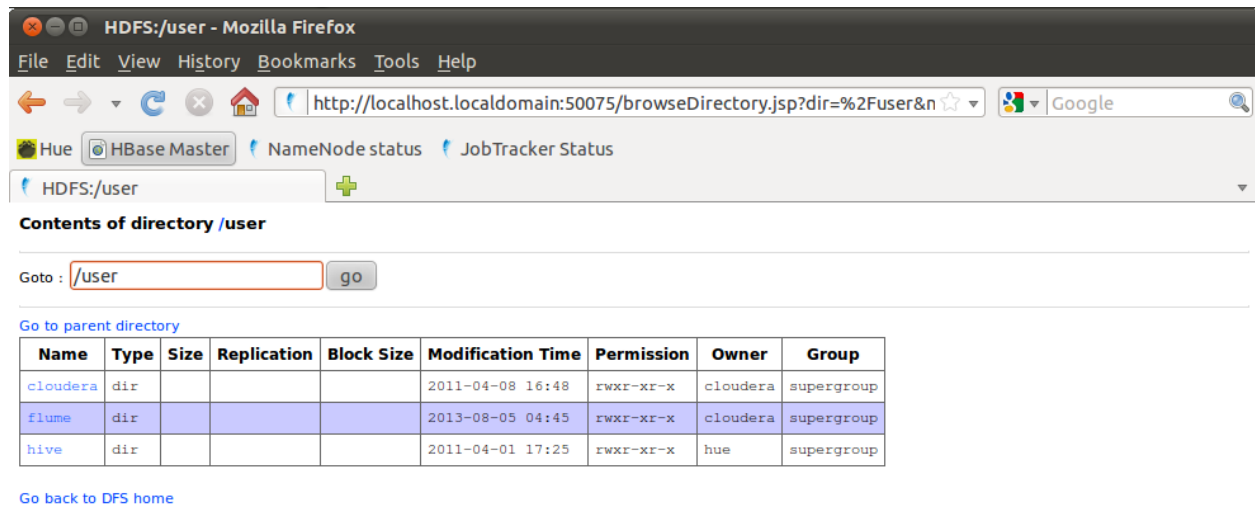
Contents of directory /

Goto : go

Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
tmp	dir				2011-04-01 17:25	rw-rw-rw-	hue	supergroup
user	dir				2013-08-05 04:45	rw-r--r--	hue	supergroup
var	dir				2013-08-01 22:37	rw-r--r--	mapred	supergroup

[Go back to DFS home](#)

Click on flume:



HDFS:/user

Contents of directory /user

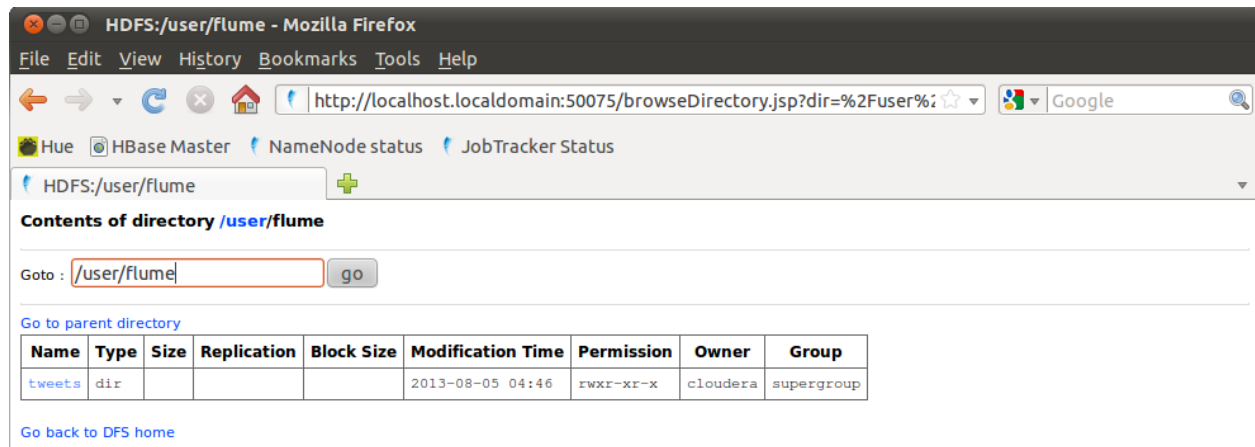
Goto : go

[Go to parent directory](#)

Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
cloudera	dir				2011-04-08 16:48	rw-r--r--	cloudera	supergroup
flume	dir				2013-08-05 04:45	rw-r--r--	cloudera	supergroup
hive	dir				2011-04-01 17:25	rw-r--r--	hue	supergroup

[Go back to DFS home](#)

Click on tweets:



HDFS:/user/flume

Contents of directory /user/flume

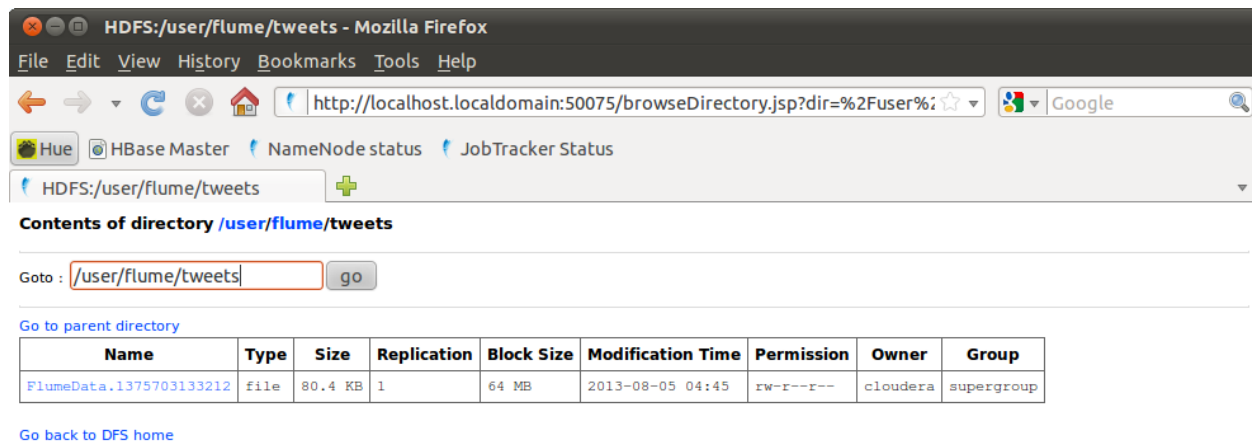
Goto : go

[Go to parent directory](#)

Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
tweets	dir				2013-08-05 04:46	rw-r--r--	cloudera	supergroup

[Go back to DFS home](#)

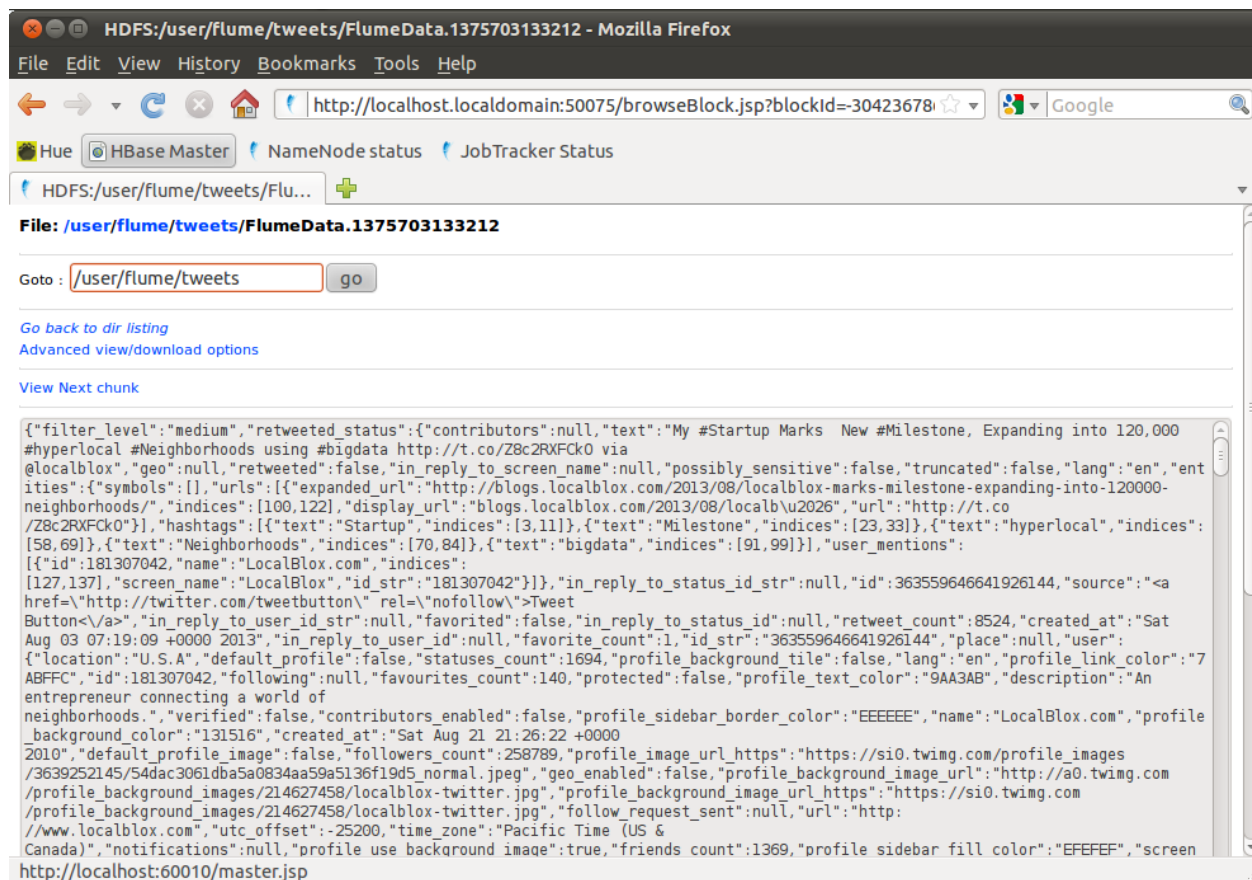
Click on FlumeData file:



The screenshot shows a web browser window titled "HDFS:/user/flume/tweets - Mozilla Firefox". The address bar shows the URL "http://localhost.localdomain:50075/browseDirectory.jsp?dir=%2Fuser%2F". The page displays the contents of the directory "/user/flume/tweets". A table lists the files and directories in the directory.

Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
FlumeData.1375703133212	file	80.4 KB	1	64 MB	2013-08-05 04:45	rw-r--r--	cloudera	supergroup

This is the data that has been downloaded from Twitter:



The screenshot shows a web browser window titled "HDFS:/user/flume/tweets/FlumeData.1375703133212 - Mozilla Firefox". The address bar shows the URL "http://localhost.localdomain:50075/browseBlock.jsp?blockid=30423678". The page displays the contents of the file "FlumeData.1375703133212". The content is a JSON object representing a tweet.

```
{
  "filter_level": "medium",
  "retweeted_status": {
    "contributors": null,
    "text": "My #Startup Marks New #Milestone, Expanding into 120,000 #hyperlocal #Neighborhoods using #bigdata http://t.co/Z8c2RXfCk0 via @localblox",
    "geo": null,
    "retweeted": false,
    "in_reply_to_screen_name": null,
    "possibly_sensitive": false,
    "truncated": false,
    "lang": "en",
    "entities": {
      "symbols": [],
      "urls": [
        {
          "expanded_url": "http://blogs.localblox.com/2013/08/localblox-marks-milestone-expanding-into-120000-neighborhoods/",
          "indices": [100, 122],
          "display_url": "blogs.localblox.com/2013/08/localblox-marks-milestone-expanding-into-120000-neighborhoods/"
        },
        {
          "expanded_url": "http://t.co/Z8c2RXfCk0",
          "indices": [123, 145],
          "display_url": "t.co/Z8c2RXfCk0"
        }
      ],
      "hashtags": [
        {
          "text": "Startup",
          "indices": [3, 11]
        },
        {
          "text": "Milestone",
          "indices": [23, 33]
        },
        {
          "text": "hyperlocal",
          "indices": [58, 69]
        },
        {
          "text": "Neighborhoods",
          "indices": [70, 84]
        },
        {
          "text": "bigdata",
          "indices": [91, 99]
        }
      ],
      "user_mentions": [
        {
          "id": 181307042,
          "name": "LocalBlox.com",
          "indices": [127, 137],
          "screen_name": "LocalBlox",
          "id_str": "181307042"
        }
      ]
    },
    "in_reply_to_status_id_str": null,
    "id": 363559646641926144,
    "source": "<a href='\"http://twitter.com/tweetbutton\"' rel='\"nofollow\"'>Tweet Button</a>",
    "in_reply_to_user_id_str": null,
    "favorited": false,
    "in_reply_to_status_id": null,
    "retweet_count": 8524,
    "created_at": "Sat Aug 03 07:19:09 +0000 2013",
    "in_reply_to_user_id": null,
    "favorite_count": 1,
    "id_str": "363559646641926144",
    "place": null,
    "user": {
      "location": "U.S.A",
      "default_profile": false,
      "statuses_count": 1694,
      "profile_background_tile": false,
      "lang": "en",
      "profile_link_color": "7ABFFC",
      "id": 181307042,
      "following": null,
      "favourites_count": 140,
      "protected": false,
      "profile_text_color": "9AA3AB",
      "description": "An entrepreneur connecting a world of neighborhoods.",
      "verified": false,
      "contributors_enabled": false,
      "profile_sidebar_border_color": "EEEEEE",
      "name": "LocalBlox.com",
      "profile_background_color": "131516",
      "created_at": "Sat Aug 21 21:26:22 +0000 2010",
      "default_profile_image": false,
      "followers_count": 258789,
      "profile_image_url_https": "https://si0.twimg.com/profile_images/3639252145/54dac3061dba0834aa59a5136f19d5_normal.jpeg",
      "geo_enabled": false,
      "profile_background_image_url": "http://a0.twimg.com/profile_background_images/214627458/localblox-twitter.jpg",
      "profile_background_image_url_https": "https://si0.twimg.com/profile_background_images/214627458/localblox-twitter.jpg",
      "follow_request_sent": null,
      "url": "http://www.localblox.com",
      "utc_offset": -25200,
      "time_zone": "Pacific Time (US & Canada)",
      "notifications": null,
      "profile_use_background_image": true,
      "friends count": 1369,
      "profile_sidebar_fill_color": "E6E6FA",
      "screen_name": "LocalBlox"
    }
  },
  "text": "My #Startup Marks New #Milestone, Expanding into 120,000 #hyperlocal #Neighborhoods using #bigdata http://t.co/Z8c2RXfCk0 via @localblox",
  "geo": null,
  "retweeted": false,
  "in_reply_to_screen_name": null,
  "possibly_sensitive": false,
  "truncated": false,
  "lang": "en",
  "entities": {
    "symbols": [],
    "urls": [
      {
        "expanded_url": "http://blogs.localblox.com/2013/08/localblox-marks-milestone-expanding-into-120000-neighborhoods/",
        "indices": [100, 122],
        "display_url": "blogs.localblox.com/2013/08/localblox-marks-milestone-expanding-into-120000-neighborhoods/"
      },
      {
        "expanded_url": "http://t.co/Z8c2RXfCk0",
        "indices": [123, 145],
        "display_url": "t.co/Z8c2RXfCk0"
      }
    ],
    "hashtags": [
      {
        "text": "Startup",
        "indices": [3, 11]
      },
      {
        "text": "Milestone",
        "indices": [23, 33]
      },
      {
        "text": "hyperlocal",
        "indices": [58, 69]
      },
      {
        "text": "Neighborhoods",
        "indices": [70, 84]
      },
      {
        "text": "bigdata",
        "indices": [91, 99]
      }
    ],
    "user_mentions": [
      {
        "id": 181307042,
        "name": "LocalBlox.com",
        "indices": [127, 137],
        "screen_name": "LocalBlox",
        "id_str": "181307042"
      }
    ]
  },
  "in_reply_to_status_id_str": null,
  "id": 363559646641926144,
  "source": "<a href='\"http://twitter.com/tweetbutton\"' rel='\"nofollow\"'>Tweet Button</a>",
  "in_reply_to_user_id_str": null,
  "favorited": false,
  "in_reply_to_status_id": null,
  "retweet_count": 8524,
  "created_at": "Sat Aug 03 07:19:09 +0000 2013",
  "in_reply_to_user_id": null,
  "favorite_count": 1,
  "id_str": "363559646641926144",
  "place": null,
  "user": {
    "location": "U.S.A",
    "default_profile": false,
    "statuses_count": 1694,
    "profile_background_tile": false,
    "lang": "en",
    "profile_link_color": "7ABFFC",
    "id": 181307042,
    "following": null,
    "favourites_count": 140,
    "protected": false,
    "profile_text_color": "9AA3AB",
    "description": "An entrepreneur connecting a world of neighborhoods.",
    "verified": false,
    "contributors_enabled": false,
    "profile_sidebar_border_color": "EEEEEE",
    "name": "LocalBlox.com",
    "profile_background_color": "131516",
    "created_at": "Sat Aug 21 21:26:22 +0000 2010",
    "default_profile_image": false,
    "followers_count": 258789,
    "profile_image_url_https": "https://si0.twimg.com/profile_images/3639252145/54dac3061dba0834aa59a5136f19d5_normal.jpeg",
    "geo_enabled": false,
    "profile_background_image_url": "http://a0.twimg.com/profile_background_images/214627458/localblox-twitter.jpg",
    "profile_background_image_url_https": "https://si0.twimg.com/profile_background_images/214627458/localblox-twitter.jpg",
    "follow_request_sent": null,
    "url": "http://www.localblox.com",
    "utc_offset": -25200,
    "time_zone": "Pacific Time (US & Canada)",
    "notifications": null,
    "profile use background image": true,
    "friends count": 1369,
    "profile sidebar fill color": "E6E6FA",
    "screen_name": "LocalBlox"
  }
}
```

-----The End-----