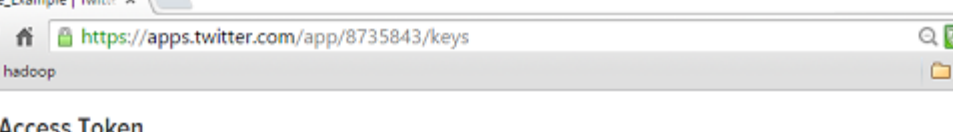


Create a flume agent that streams data from Twitter and stores in the HDFS.

We will create an application and get the tweets from it using the experimental twitter source provided by Apache Flume. We will use the memory channel to buffer these tweets and HDFS sink to push these tweets into the HDFS.



Apply for a Twitter developer account and create a new application there.



Tutor's point

Tp\_Flume\_Example | Twitter x

← → ↺ 🏠 <https://apps.twitter.com/app/8735843/keys> 🔍 📌 ⭐ ⚙

Apps 📁 hadoop 📁 Other bookmarks

## Your Access Token

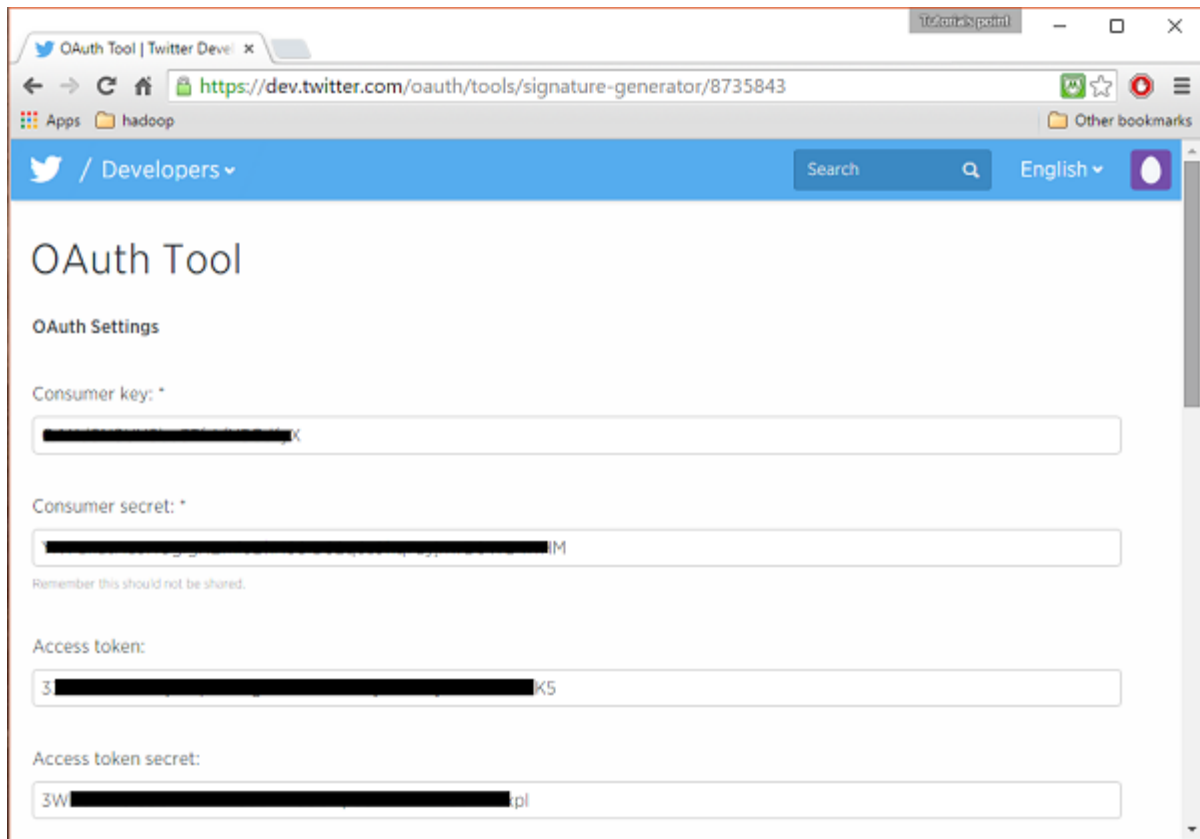
*You haven't authorized this application for your own account yet.*

*By creating your access token here, you will have everything you need to make API calls right away. The access token generated will be assigned your application's current permission level.*

### Token Actions

Create my access token

Click on the **Test OAuth** button which is on the right side top of the page. This leads to a page which displays the **Consumer key**, **Consumer secret**, **Access token**, and **Access token secret**. Copy these details which would be used in configuring the flume agent



## Step#2

The start Hadoop standard demons using start-all.sh command. Use jps to verify that required daemons are running

```
[acadgild.mmisra ~]$  
[acadgild.mmisra ~]$ jps  
28209 Jps  
3780 ResourceManager  
3415 NameNode  
3884 NodeManager  
23196 MainGenericRunner  
3630 SecondaryNameNode
```

## Step#3

Create a directory 'twitter\_data' in HDFS where the twitter data will be stored by the flume agent.

```
[acadgild.mmisra ~]$  
[acadgild.mmisra ~]$ hadoop fs -mkdir /twitter_data
```

## Step#4

Configure the flume agent by writing a new configuration file called **flume\_twitter.conf**.

The contents of the config file are shown below

```
[acadgild.mmisra ~]$ cat flume-twitter.conf

# Name the source,sink and channels

TwAgent.sources= Twitter
TwAgent.channels= MemChannel
TwAgent.sinks=HDFS

# configure the source
TwAgent.sources.Twitter.type = com.cloudera.flume.source.TwitterSource
TwAgent.sources.Twitter.channels=MemChannel

### provide my keys and tokens
TwAgent.sources.Twitter.consumerKey=bK8UtiI56yTnt6Epn2ruRHBSSqP
TwAgent.sources.Twitter.consumerSecret=8kIosYuaSOPZLrwRrnB6c790EPeoiC80lrTAQE4hWc0qGDf
g092
TwAgent.sources.Twitter.accessToken=899058335-
t2pbxWrwKzovJV78WPcXsU7mVhVwn9EILOWklPuiY0iO
TwAgent.sources.Twitter.accessTokenSecret=peuytXMCozOnPAti8MPtkSlDxTa6U1QxAZ4ydLF6wsUt
y90P

### describe the keywords to get the tweets about
TwAgent.sources.Twitter.keywords= modi, cricket

# configure the Sink

TwAgent.sinks.HDFS.type=hdfs
TwAgent.sinks.HDFS.channel=MemChannel
##### Check this - look at core-site.xml

### provide the hdfs directory name where the data would be stored
TwAgent.sinks.HDFS.hdfs.path=hdfs://localhost:8020/twitter_data
TwAgent.sinks.HDFS.hdfs.fileType=DataStream
TwAgent.sinks.HDFS.hdfs.writeformat=Text
TwAgent.sinks.HDFS.hdfs.batchSize=1000
TwAgent.sinks.HDFS.hdfs.rollSize=0
TwAgent.sinks.HDFS.hdfs.rollCount=10000
TwAgent.sinks.HDFS.hdfs.rollInterval=600

#configure the channel
TwAgent.channels.MemChannel.type=memory
TwAgent.channels.MemChannel.capacity=10000
TwAgent.channels.MemChannel.transactionCapacity=1000

## Bind the source and sink to channel
TwAgent.sources.Twitter.channels=MemChannel
TwAgent.sinks.HDFS.channel=MemChannel
```

## Step#5

Launch the flume agent with the following command and also provide the name of the config file we created in the step#4.

**flume-ng agent -conf \$FLUME\_HOME/conf/ -f flume-twitter.conf -n TwAgent**

```
[acadgild.mmisra ~]$ flume-ng agent --conf $FLUME_HOME/conf/ -f flume-twitter.conf -n TwAgent
Info: Including Hadoop libraries found via (/home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop) for HDFS access
Info: Including HBASE libraries found via (/home/acadgild/install/hbase/hbase-1.2.6/bin/hbase) for HBASE access
Info: Including Hive libraries found via (/home/acadgild/install/hive/apache-hive-2.3.2-bin) for Hive access
+ exec /usr/java/jdk1.8.0_151/bin/java -Xmx20m -cp
'/home/acadgild/install/flume/apache-flume-1.8.0-
bin/conf:/usr/java/jdk1.8.0_151/lib/tools.jar:/home/acadgild/install/hbase/hbase-1.2.6:/home/acadgild/install/hbase/hbase-1.2.6/lib/activation-1.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/aopalliance-1.0.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/apacheds-kerberos-codec-2.0.0-M15.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/api-asn1-api-1.0.0-M20.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/api-util-1.0.0-M20.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/asm-3.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/avro-1.7.4.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-beanutils-1.7.0.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-beanutils-core-1.8.0.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-cli-1.2.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-codec-1.9.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-collections-3.2.2.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-compress-1.4.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-configuration-1.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-daemon-1.0.13.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-digester-1.8.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-el-1.0.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-httpclient-3.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-io-2.4.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-lang-2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-logging-1.2.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-math-2.2.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-math3-3.1.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-net-3.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/disruptor-3.3.0.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/findbugs-annotations-1.3.9-1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/guava-12.0.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/guice-3.0.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/guice-servlet-3.0.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-annotations-
```

2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-auth-  
2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-client-  
2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-common-  
2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-hdfs-  
2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-mapreduce-client-app-  
2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-mapreduce-client-common-  
2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-mapreduce-client-core-  
2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-mapreduce-client-  
jobclient-2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-mapreduce-  
client-shuffle-2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-yarn-api-  
2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-yarn-client-  
2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-yarn-common-  
2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hadoop-yarn-server-common-  
2.5.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-annotations-  
1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-annotations-1.2.6-  
tests.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-client-  
1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-common-  
1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-common-1.2.6-  
tests.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-examples-  
1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-external-blockcache-  
1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-hadoop2-compat-  
1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-hadoop-compat-  
1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-it-  
1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-it-1.2.6-  
tests.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-prefix-tree-  
1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-procedure-  
1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-protocol-  
1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-resource-bundle-  
1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-rest-  
1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-server-  
1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-server-1.2.6-  
tests.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-shell-  
1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/hbase-thrift-  
1.2.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/htrace-core-3.1.0-  
incubating.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/httpclient-  
4.2.5.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/httpcore-  
4.4.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jackson-core-asl-  
1.9.13.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jackson-jaxrs-  
1.9.13.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jackson-mapper-asl-  
1.9.13.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jackson-xc-  
1.9.13.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jamon-runtime-  
2.4.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jasper-compiler-  
5.5.23.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jasper-runtime-  
5.5.23.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/javax.inject-  
1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/java-xmlbuilder-  
0.4.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jaxb-api-  
2.2.2.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jaxb-impl-2.2.3-  
1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jcodings-  
1.0.8.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jersey-client-  
1.9.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jersey-core-  
1.9.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jersey-guice-  
1.9.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jersey-json-  
1.9.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jersey-server-  
1.9.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jets3t-  
0.9.0.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jettison-  
1.3.3.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jetty-  
6.1.26.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jetty-sslengine-  
6.1.26.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jetty-util-  
6.1.26.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/joni-  
2.1.2.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jruby-complete-  
1.6.8.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jsch-  
0.1.42.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jsp-2.1-  
6.1.14.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/jsp-api-2.1-

```

6.1.14.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/junit-
4.12.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/leveldbjni-all-
1.8.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/libthrift-
0.9.3.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/log4j-
1.2.17.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/metrics-core-
2.2.0.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/netty-all-
4.0.23.Final.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/paranamer-
2.3.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/protobuf-java-
2.5.0.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/servlet-api-2.5-
6.1.14.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/servlet-api-
2.5.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-api-
1.7.7.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-
1.7.5.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/snappy-java-
1.0.4.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/spymemcached-
2.11.6.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/xmlenc-
0.52.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/xz-
1.0.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/zookeeper-
3.4.6.jar:/home/acadgild/install/hadoop/hadoop-
2.6.5/etc/hadoop:/home/acadgild/install/hadoop/hadoop-
2.6.5/share/hadoop/common/lib/*:/home/acadgild/install/hadoop/hadoop-
2.6.5/share/hadoop/common/*:/home/acadgild/install/hadoop/hadoop-
2.6.5/share/hadoop/hdfs:/home/acadgild/install/hadoop/hadoop-
2.6.5/share/hadoop/hdfs/lib/*:/home/acadgild/install/hadoop/hadoop-
2.6.5/share/hadoop/hdfs/*:/home/acadgild/install/hadoop/hadoop-
2.6.5/share/hadoop/yarn/lib/*:/home/acadgild/install/hadoop/hadoop-
2.6.5/share/hadoop/yarn/*:/home/acadgild/install/hadoop/hadoop-
2.6.5/share/hadoop/mapreduce/lib/*:/home/acadgild/install/hadoop/hadoop-
2.6.5/share/hadoop/mapreduce/*:/home/acadgild/install/hadoop/hadoop-
2.6.5/contrib/capacity-scheduler/*.jar:/home/acadgild/install/hbase/hbase-
1.2.6/conf:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/*' -
Djava.library.path=:/usr/java/packages/lib/amd64:/usr/lib64:/lib64:/lib:/usr/lib:/usr
java/packages/lib/amd64:/usr/lib64:/lib64:/lib:/usr/lib
org.apache.flume.node.Application -f flume-twitter.conf -n TwAgent
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/flume/apache-flume-1.8.0-
bin/lib/slf4j-log4j12-1.6.1.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadgild/install/hadoop/hadoop-
2.6.5/share/hadoop/common/lib/slf4j-log4j12-
1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-
log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadgild/install/hive/apache-hive-2.3.2-
bin/lib/log4j-slf4j-impl-2.6.2.jar!/org/slf4j/impl/StaticLoggerBinder.class]

```

## Step#6

After few minutes the flume agent starts storing the data in HDFS in /twitter\_data directory.

```

[acadgild.mmisra ~]$ hadoop fs -ls /twitter_data
18/08/17 21:14:02 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
your platform... using builtin-java classes where applicable
Found 4 items
-rw-r--r--  1 acadgild supergroup          0 2018-08-17 21:10
/twitter_data/log.1441787158066
-rw-r--r--  1 acadgild supergroup          0 2018-08-17 21:10
/twitter_data/log.1441787158067
-rw-r--r--  1 acadgild supergroup          0 2018-08-17 21:10
/twitter_data/log.1441787158068
-rw-r--r--  1 acadgild supergroup          0 2018-08-17 21:10
/twitter_data/log.1441787158069
You have new mail in /var/spool/mail/acadgild
[acadgild.mmisra ~]$

```

