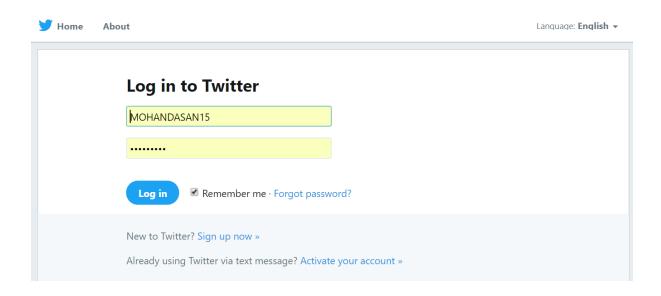Create a flume agent that streams data from Twitter and stores in the HDFS.

🐦 Developer   Use cases   Products   Docs   More   Apply   Apps   🔵

○ User profile

⊘ **Account details**

○ Use case details

○ Terms of service

○ Email verification

🐦 **Why the questions?**

We empower freedom of expression by providing a platform that protects the voices of our users — both on Twitter, and via our developer products. To help verify that all uses of Twitter data comply with our policies, we require additional information from developers signing up to use this service. Providing thorough answers will help us understand your use cases and will help expedite the evaluation of your application. Learn more about our restricted use cases.

Who are you requesting access for?

○ I am requesting access for my organization

I plan to use Twitter's developer platform for projects owned by / in affiliation with a business, organization or institution. Ex: SaaS product, proof of concept, academic research, etc. To enable collaboration, this selection includes additional tools to support team development.

● I am requesting access for my own personal use

I plan to use Twitter's developer platform for projects unaffiliated with an existing business, organization or institution. Ex: Side project, hobby, etc. Personal use accounts do not include team development tools.

Tell us about yourself

Account name

e.g., username, project name, etc.

| DasanAPP |

Primary country of operation

| India ⌄ |

[Continue]

Developer policy and terms   Follow @twitterdev   [Subscribe to developer news]

🐦 Developer   Use cases   Products   Docs   More   Dashboard   DasanAPP ⌄   🔵

Apps / Create an app

🐦 **Understanding apps**

What is an app?   ⌄

Why register an app?   ⌄

Which products require an API key? ⌄

App details

The following app details will be visible to app users and are required to generate the API keys needed to authenticate Twitter developer products.

App name (required) ❓

| DASAN APP1 |

Maximum characters: **32**

Application description (required)

Share a description of your app. This description will be visible to users so this is a good place to tell them what your app does.

| Want to do some analysis on FLUME |

Between 10 and 200 characters

Website URL (required) ❓

| https://www.yahoo.com |

Allow this application to be used to sign in with Twitter Learn more

☐ **Enable Sign in with Twitter**

Callback URLs ⓘ

OAuth 1.0a applications should specify their oauth_callback URL on the request token step, which must match the URLs provided here. To restrict your application from using callbacks, leave these blank.

| https:// or scheme:// | ✕ |

+Add another

Terms of Service URL ⓘ

| https:// |

Privacy policy URL ⓘ

| https:// |

Organization name ⓘ

| |

Organization website URL

| https:// |

Tell us how this app will be used (required)

This field is only visible to Twitter employees. Help us understand how your app will be used. What will it enable you and your customers to do?

Tell us how this app will be used (required)

This field is only visible to Twitter employees. Help us understand how your app will be used. What will it enable you and your customers to do?

Will be basically using the Twitter for analyzing the project on FLUME. This is for a project carried out for personal use only

Cancel   Create

Review our Developer Terms   ✕

As a reminder, you have agreed to our Developer Agreement and Policy.
Please be mindful of the following restricted use cases
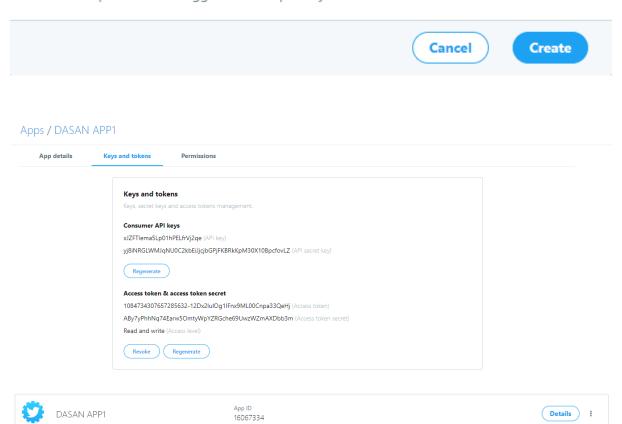
⊗ **Sensitive Information**

Be careful about using Twitter data to derive or infer potentially sensitive characteristics about Twitter users (ie. health, political or religious affiliation, ethnic origin, sexual orientation, and more).

⊗ **Government use and surveillance**

We prohibit the use of Twitter Data and Twitter APIs by any entity for surveillance purposes. Period.

⊗ **Automation**

If your app will be used to post Tweets, follow accounts, or send Direct Messages, you should carefully review the Automation Rules to ensure you comply with our guidelines and never perform bulk, aggressive, or spammy actions.

Cancel   Create

Apps / DASAN APP1

App details   **Keys and tokens**   Permissions

Keys and tokens
Keys, secret keys and access tokens management.

Consumer API keys
xJZFTIema5Lp01hPELfrVj2qe (API key)

yj8iNRGLWMJqNU0C2kbEiljcjbGPjFKBRkKpM30X10BpcfovLZ (API secret key)

Regenerate

Access token & access token secret
1084734307657285632-12Dx2IuIOg1IFnx9ML00Cnpa33QeHj (Access token)

ABy7yPhhNq74Earw5OmtyWpYZRGche69UwzWZmAXDbb3m (Access token secret)

Read and write (Access level)

Revoke   Regenerate

🐦 DASAN APP1   App ID
                                 16067334   Details   ⋮

/home/acadgild/apache-flume-1.6.0-bin/conf/flume_twitter.conf

Update the newly created file with twitter **api** keys like consumer key, Consumer token, Access token and the access token secret code and with the **key words.**

```
# Describing/Configuring the source
TwitterAgent.sources.Twitter.type = org.apache.flume.source.twitter.TwitterSource
TwitterAgent.sources.Twitter.consumerKey=DCjUjRSucocyREIvZQa6VJ5AP
TwitterAgent.sources.Twitter.consumerSecret=x1D1nQkXJHAhghTztK6519I7U9Taq4WLl8fRqa9UUm5DCwYDVj
TwitterAgent.sources.Twitter.accessToken=797943092-wcNt3mgrbPiHYhEZ2K9RjWvjs3zAlYg1ETi2sOA3
TwitterAgent.sources.Twitter.accessTokenSecret=ohm8hds3X1d2SOJWsOaAu3HlpTjYvSsaI4In3lNVTAJJU
TwitterAgent.sources.Twitter.keywords=hadoop, bigdata, mapreduce, mahout, hbase, nosql
# Describing/Configuring the sink

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

```
-rw-rw-r--. 1 acadgild acadgild 1919 Nov 30
[acadgild@localhost lib]$ jps
3234 NodeManager
2819 DataNode
3125 ResourceManager
4661 Main
2712 NameNode
4315 HMaster
4107 RunJar
6172 Jps
[acadgild@localhost lib]$
```

hadoop dfs –mkdir /user/acadgild/hadoop/tweets

```
Usage: hadoop fs [generic options] -mkdir [-p] <path> ...
[acadgild@localhost lib]$ hadoop dfs -mkdir /user/acadgild/hadoop/tweets
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

Java HotSpot(TM) Client VM warning: You have loaded library /home/acadgild/hadoop-2.7.2/lib/native/libhadoop.so.1.0.0 which m
e stack guard now.
It's highly recommended that you fix the library with 'execstack -c <libfile>', or link it with '-z noexecstack'.
17/11/30 10:03:57 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
[acadgild@localhost lib]$
[acadgild@localhost lib]$
[acadgild@localhost lib]$ hadoop fs -ls /user/acadgild/hadoop
Java HotSpot(TM) Client VM warning: You have loaded library /home/acadgild/hadoop-2.7.2/lib/native/libhadoop.so.1.0.0 which m
e stack guard now.
It's highly recommended that you fix the library with 'execstack -c <libfile>', or link it with '-z noexecstack'.
17/11/30 10:04:19 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
Found 14 items
drwxr-xr-x   - acadgild supergroup          0 2017-10-12 21:15 /user/acadgild/hadoop/InvalidDataMR
drwxr-xr-x   - acadgild supergroup          0 2017-10-12 18:46 /user/acadgild/hadoop/InvalidRecord2
drwxr-xr-x   - acadgild supergroup          0 2017-10-12 17:44 /user/acadgild/hadoop/InvalidRecordsoutput
drwxr-xr-x   - acadgild supergroup          0 2017-10-31 17:26 /user/acadgild/hadoop/OnidaTV
drwxr-xr-x   - acadgild supergroup          0 2017-10-31 17:41 /user/acadgild/hadoop/TV
drwxr-xr-x   - acadgild supergroup          0 2017-10-31 17:49 /user/acadgild/hadoop/TV1
-rw-r--r--   1 acadgild supergroup       1958 2017-10-13 18:56 /user/acadgild/hadoop/WordCount.txt
-rw-r-----   1 acadgild supergroup        237 2017-09-25 11:10 /user/acadgild/hadoop/max-temp.txt
drwxr-xr-x   - acadgild supergroup          0 2017-09-24 14:31 /user/acadgild/hadoop/maxout
-rw-r--r--   1 acadgild supergroup      21007 2017-09-24 14:25 /user/acadgild/hadoop/sample_temperature_dataset.csv
-rw-r--r--   1 acadgild supergroup      26204 2017-11-26 02:06 /user/acadgild/hadoop/student.txt
-rw-r--r--   1 acadgild supergroup       2938 2017-10-31 17:47 /user/acadgild/hadoop/television.txt
drwxr-xr-x   - acadgild supergroup          0 2017-11-30 10:03 /user/acadgild/hadoop/tweets
-rw-r--r--   1 acadgild supergroup        300 2017-09-24 14:16 /user/acadgild/hadoop/word-count.txt
[acadgild@localhost lib]$
```

flume-ng agent -n TwitterAgent -f /home/acadgild/apache-flume-1.6.0-bin/conf/flume_twitter.conf

```
6172 Jps
[acadgild@localhost lib]$ flume-ng agent -n TwitterAgent -f /home/acadgild/apache-flume-1.6.0-bin/conf/flume_twitter.conf
Warning: No configuration directory set! Use --conf <dir> to override.
Info: Including Hadoop libraries found via (/home/acadgild/hadoop-2.7.2/bin/hadoop) for HDFS access
```

```
17/11/30 10:12:30 INFO hdfs.HDFSDataStream: Serializer = TEXT, UseRawLocalFileSystem = false
17/11/30 10:12:30 INFO hdfs.BucketWriter: Creating hdfs://localhost:9000/user/acadgild/hadoop/tweets/FlumeData.1512016950366.tmp
Java HotSpot(TM) Client VM warning: You have loaded library /home/acadgild/hadoop-2.7.2/lib/native/libhadoop.so.1.0.0 which might have disabled stack guard. The VM will try to fix th
e stack guard now.
It's highly recommended that you fix the library with 'execstack -c <libfile>', or link it with '-z noexecstack'.
17/11/30 10:12:31 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
17/11/30 10:12:33 INFO twitter.TwitterSource: Processed 100 docs
17/11/30 10:12:35 INFO twitter.TwitterSource: Processed 200 docs
17/11/30 10:12:39 INFO twitter.TwitterSource: Processed 300 docs
17/11/30 10:12:42 INFO twitter.TwitterSource: Processed 400 docs
17/11/30 10:12:44 INFO twitter.TwitterSource: Processed 500 docs
17/11/30 10:12:47 INFO twitter.TwitterSource: Processed 600 docs
17/11/30 10:12:50 INFO twitter.TwitterSource: Processed 700 docs
17/11/30 10:12:53 INFO twitter.TwitterSource: Processed 800 docs
17/11/30 10:12:56 INFO twitter.TwitterSource: Processed 900 docs
17/11/30 10:13:00 INFO twitter.TwitterSource: Processed 1,000 docs
17/11/30 10:13:00 INFO twitter.TwitterSource: Total docs indexed: 1,000, total skipped docs: 0
17/11/30 10:13:00 INFO twitter.TwitterSource:      31 docs/second
17/11/30 10:13:00 INFO twitter.TwitterSource: Run took 32 seconds and processed:
17/11/30 10:13:00 INFO twitter.TwitterSource:      0.008 MB/sec sent to index
17/11/30 10:13:00 INFO twitter.TwitterSource:      0.259 MB text sent to index
17/11/30 10:13:00 INFO twitter.TwitterSource: There were 0 exceptions ignored:
17/11/30 10:13:03 INFO twitter.TwitterSource: Processed 1,100 docs
17/11/30 10:13:05 INFO twitter.TwitterSource: Processed 1,200 docs
17/11/30 10:13:08 INFO twitter.TwitterSource: Processed 1,300 docs
17/11/30 10:13:12 INFO twitter.TwitterSource: Processed 1,400 docs
17/11/30 10:13:15 INFO twitter.TwitterSource: Processed 1,500 docs
```

hadoop fs -cat /user/acadgild/hadoop/tweets/FlumeData.1512016950366

```
[acadgild@localhost lib]$
[acadgild@localhost lib]$ hadoop fs -cat /user/acadgild/hadoop/tweets/FlumeData.1512016950366
Java HotSpot(TM) Client VM warning: You have loaded library /home/acadgild/hadoop-2.7.2/lib/native/libhadoop.so.1.0.0 which might have disabled stack guard. The VM will try to fix th
e stack guard now.
It's highly recommended that you fix the library with 'execstack -c <libfile>', or link it with '-z noexecstack'.
17/11/30 10:18:15 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
{"type":"record","name":"Doc","doc":"adoc","fields":[{"name":"id","type":"string"},{"name":"user_friends_count","type":["int","null"]},{"name":"user_location","type":["string","null"
]},{"name":"user_description","type":["string","null"]},{"name":"user_statuses_count","type":["int","null"]},{"name":"user_followers_count","type":["int","null"]},{"name":"user_name"
,"type":["string","null"]},{"name":"user_screen_name","type":["string","null"]},{"name":"created_at","type":["string","null"]},{"name":"text","type":["string","null"]},{"name":"retwe
et_count","type":["long","null"]},{"name":"retweeted","type":["boolean","null"]},{"name":"in_reply_to_user_id","type":["long","null"]},{"name":"source","type":["string","null"]},{"na
me":"in_reply_to_status_id","type":["long","null"]},{"name":"media_url_https","type":["string","null"]},{"name":"expanded_url","type":["string","null"]}]}
```