

Introduction

In this assignment, we are going to stream data from twitter and store it into HDFS and the screen shots are shared.

Problem Statement

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

Prerequisite

To stream data to our database from twitter we should have the following pre-requisites.

- Twitter account
- Hadoop cluster

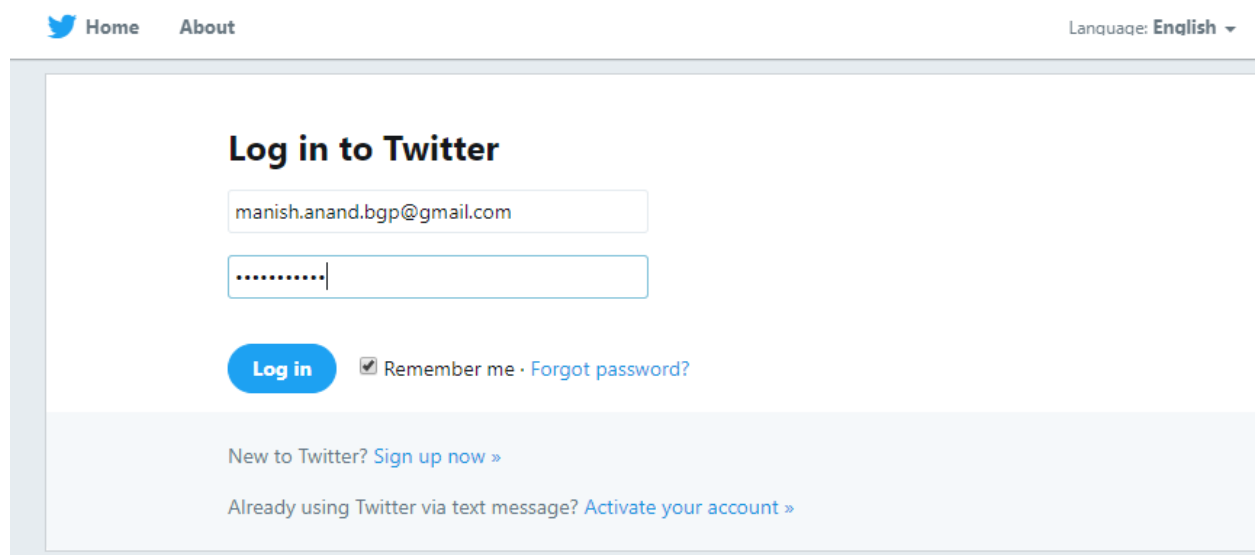
Make sure you have below jars placed in your `$FLUME_HOME/lib/conf` directory:

- twitter4j-core-X.XX.jar
- twitter4j-stream-X.X.X.jar
- twitter4j-media-support-X.X.X.jar

If the above prerequisites are available we can move to our further step.

Step1:

Login to the twitter account,

A screenshot of the Twitter login page. At the top, there is a navigation bar with the Twitter logo, 'Home', 'About', and a language dropdown set to 'English'. The main content area is titled 'Log in to Twitter'. It contains a text input field with the email 'manish.anand.bgp@gmail.com', a password input field with masked characters, a blue 'Log in' button, a checked 'Remember me' checkbox, and a link for 'Forgot password?'. At the bottom, there are two links: 'New to Twitter? Sign up now »' and 'Already using Twitter via text message? Activate your account »'.

Step2:

Go to the following link and click the 'create new app' button.

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

Twitter Apps

You don't currently have any Twitter Apps.

Create New App

Providing necessary details,

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? 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.

Accept the developer agreement and select the 'create your Twitter application' button'

Callback URL

Where should we return after successfully authenticating? [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.

Developer Agreement

☒ Yes, I have read and agree to the [Twitter Developer Agreement](#).

Create your Twitter application

Select the 'Keys and Access Token' tab.

Your application has been created. Please take a moment to review and adjust your application's settings.

Abutokenapp

Test OAuth

Details

Settings

Keys and Access Tokens

Permissions



I want to do analysis in Flume

<https://www.yahoo.com>

Organization

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

Organization None

Organization website None

Copy the consumer key and the consumer secret code, Scroll down further and select the 'create my access token' button.

Now, you will receive a message stating "that you have successfully generated your application access token".

Status

Your application access token has been successfully generated. It may take a moment for changes you've made to reflect.

[Refresh](#) if your changes are not yet indicated.

Copy the Access Token and Access token Secret code.

Consumer Key (API Key) DCjUjRSucocyREIvZQa6VJ5AP

Consumer Secret (API Secret) x1D1nQkXJHAghTztK6519I7U9Taq4WLl8fRqa9UUm5DCwYDVj

Access Token 797943092-wcNt3mgrbPiHYhEZ2K9RjWvjs3zAlYg1ETi2sOA3

Access Token Secret ohm8hds3X1d2S0JWsOaAu3HlpTjYvSsaI4In3INVTAJJU

Step 3:

Copy the Flume configuration code from the below link and paste it in the newly created file in the location,

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

<https://drive.google.com/open?id=0B1QaXx7tpw3Sb3U4LW9SWlNidkk>

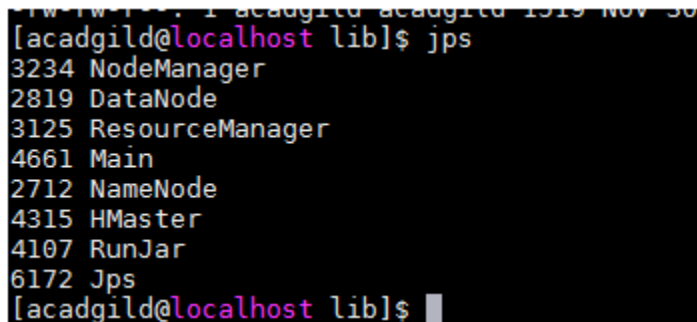
Update the newly created file with twitter **api** keys like consumer key, Consumer token, Access token and the access token secret code and with the **key words**.

```
# Describing/Configuring the source
TwitterAgent.sources.Twitter.type = org.apache.flume.source.twitter.TwitterSource
TwitterAgent.sources.Twitter.consumerKey=DCjUjRSucocyREIvZQa6VJ5AP
TwitterAgent.sources.Twitter.consumerSecret=x1D1nQkXJHAghTztK6519I7U9Taq4WLl8fRqa9UUm5DCwYDVj
TwitterAgent.sources.Twitter.accessToken=797943092-wcNt3mgrbPiHYhEZ2K9RjWvjs3zAlYg1ETi2sOA3
TwitterAgent.sources.Twitter.accessTokenSecret=ohm8hds3X1d2S0JWsOaAu3HlpTjYvSsaI4In3INVTAJJU
TwitterAgent.sources.Twitter.keywords=hadoop, bigdata, mapreduce, mahout, hbase, nosql
# Describing/Configuring the sink

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

Step4:

4.1 start all Hadoop daemons



```
[acadgild@localhost lib]$ jps
3234 NodeManager
2819 DataNode
3125 ResourceManager
4661 Main
2712 NameNode
4315 HMaster
4107 RunJar
6172 Jps
[acadgild@localhost lib]$
```

Step5:

Create a new directory inside HDFS path, where the Twitter tweet data should be stored.

Hadoop dfs -mkdir /user/acadgild/hadoop/tweets

```

[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 might not be compatible with the current JVM.
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 classes where applicable
[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 might not be compatible with the current JVM.
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 classes where applicable
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]$

```

Step6:

For fetching data from Twitter, Use the below command to fetch the twitter tweet data into the HDFS cluster path.

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

```

The above command will start fetching data from Twitter and streams it into the HDFS given path.

```

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 not be compatible with the current JVM.
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 10:13:06 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

```

Once, the tweet data started streaming it into the given HDFS path we can use 'Ctrl+c' command to stop the streaming process.

Step7:

To check the contents of the tweet data we can use the following command:

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

```

[acadgild@localhost lib]s
[acadgild@localhost lib]s 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 the
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":{"doc":{"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"]},"name":"in_re
ply_to_status_id","type":["long","null"]},"name":"media_url_https","type":["string","null"]},"name":"expanded_url","type":["string","null"]}}}}
{"type":"record","name":"Doc","doc":{"doc":{"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"]},"name":"in_re
ply_to_status_id","type":["long","null"]},"name":"media_url_https","type":["string","null"]},"name":"expanded_url","type":["string","null"]}}}}
{"type":"record","name":"Doc","doc":{"doc":{"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"]},"name":"in_re
ply_to_status_id","type":["long","null"]},"name":"media_url_https","type":["string","null"]},"name":"expanded_url","type":["string","null"]}}}}
8461@proud father's Rose Gold800perByDesign(2017-11-30T10:12:31Z)Work a job make a living, invest in yourself make a killing https://t.co/Tis8bUpdial#a href=http://twitter.com/
download/iphone" rel="nofollow">Twitter for iPhone</a>a$93609305887721881700Just trying to get along.#UnionStrong. #USArmyVet #TrueBlueDenverBroncosFan #7is7 #GunSense #ImStillWithHe
r#Resist
Diana
DianaMo(2017-11-30T10:12:31Z)RT @CharlotteAlter: I covered the campaign last year. Over and over men insinuated that women's analysis of HRC's candidacy were "biased."
twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>a$936093058885591040SM(トナー セックスフレンド探してますこはお世話になってます http://goo.gl/wj8mR ) LINE
とかアドレスとか安全に無料で交換できるので是非。 サイトの回し者みたいw
net" rel="nofollow">twittbot.net</a>a$9360930588729630742.5D, 2D 다 파는 슈퍼잡곡. 마블 로키,버키 최애. 데인드만, 빌스카스카드. 2015.12.26
12RT @FFXIV_KOREA: 판이 날 잔다지 14 스들블라드 사건등록 오븐 기념 이벤트 (3주차)
이벤트 기간: 2017.11.28(월) ~ 12.3(일)
참여방법: '리틀의'하단 참여란!
참여 TIP: 판이 날 잔다지 14 공식 트위터 팔로우 하
a href=http://twitter.com" rel="nofollow">Twitter Web Client</a>a$936093058873024512Teas&e1&v=Timmy662266(2017-11-30T10:12:31Z)RT
@1130x1020: SMILESLOVE UP!
1130 11:07
https://t.co/My0A4MjYUu https://t.co/0KfPlAz13g#a href=http://twitter.com/download/android" rel="nofollow">Twitter for Android</a>a$https://pbs.twimg.com/media/DP2WbShVQAEvVwi.jpg
https://twitter.com/1130x1020/status/936069718292676608/photo/1s9360930588898128896sjust call me Arie.
NOMINATED SONG. https://t.co/MvRFVIdQV#a href=http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>a$https://pbs.twimg.com/media/DPX08e_WsAALG6A.jpghttps://tw
itter.com/thug1yfe/status/935507087941128194/video/1s9360930588898099200In Louis Arms And Heart I love @louis tomlinson he is everything to me and Melanie Martinez and zayn malik a
nd troye sivan and Trevor Moran and panic at the disco
Louis is my lifelayda zuniga(2017-11-30T10:12:31Z)RT @heavenlyzishes: When I close my eyes I see this a green eyed bo
y looking at me and holding me in his arms making me feel loved and speci
a href=http://twitter.com/download/android" rel="nofollow">Twitter for Android</a>a$936093063054839808@Oh1
o, USAseventeen.
AbrionneGabrionne pope(2017-11-30T10:12:32Z)I wish there were more genuinely good people in the world. I want to meet someone different, someone countercultural.
a href=http://twit
ter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>a$936093063063187456The official twitter of Joseph Townsend
il Swampyjoey321townsend(2017-11-30T10:12:32Z)@jose a torres21 You should try my mom's
a href=http://twitter.com/download/android" rel="nofollow">Twitter for Android</a>a$
93609306306745549571780:7500:ianryanatalkii Follow Me get my shit bussini
a href=http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>a$936093063075848192
@byron sandersonjovon(2017-11-30T10:12:32Z)RT @wordsporns: Choose people who choose you.
a href=http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>a$936093063
0884024338nQuien diria que te ibas a convertir en mi crush. nena.Nubia Mi CrushNubiaMiCrush(2017-11-30T10:12:32Z)EASY WAY TO GET FREE #FOLLOWERS
#yfp0hR6p75TVrC0m05qCmQ
#kindof 和風 音楽 #HARRY @zarafme @0AMNlookatHIM#a href=http://the.easy.way.to.get.free.followers.everyday27.addmoreinstafollowers.info" rel="nofollow">Instanta S Patel C2d5c/
a$936093063071567872Rhode Island, USA#twitter machine broke@2Pablo Vargasthestoeofshoe(2017-11-30T10:12:32Z)why even try?
a href=http://twitter.com/download/android" rel="nofo
llow">Twitter for Android</a>a$936093063071531008fsslhs // april 16th leeah paige @peachylees(2017-11-30T10:12:32Z)RT @beatsbysteph: Me in Florida: Florida sucks
Me anywhere else: Florida is a year-round tropical paradise with endless opportuniti
a href=http://twitter.com/download/android" rel="nofollow">Twitter for Android</a>a$9360930630
1772857
DM006(2017-11-30T10:12:32Z)(48:السميل) [يملعون] في الريم ولا يملعون https://t.co/bL1vFXF2p7#a href=http://quran.tv" rel="nofollow">Quran.tv</a>a$9360930
63067430914
Querétaro,Méxicogood vibes only
Luisa PantojaFeerPantoja (2017-11-30T10:12:32Z)RT @IMGmodels: #RomeeStrijd takes @VictoriasSecret's #VSFashionShow runway by storm. https://t.co/mD
DNPzZ56R#a href=http://twitter.com/download/android" rel="nofollow">Twitter for Android</a>a$https://pbs.twimg.com/ext_tw_video_thumb/936067353040662528/pu/img/w39dnig3BukZGzm-.jpg-

```

We can observe from the above image that we have successfully fetched twitter data into our HDFS cluster directory using Flume.