

Session 12: Oozie and Flume Assignment

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

1. Check Hadoop daemons are running

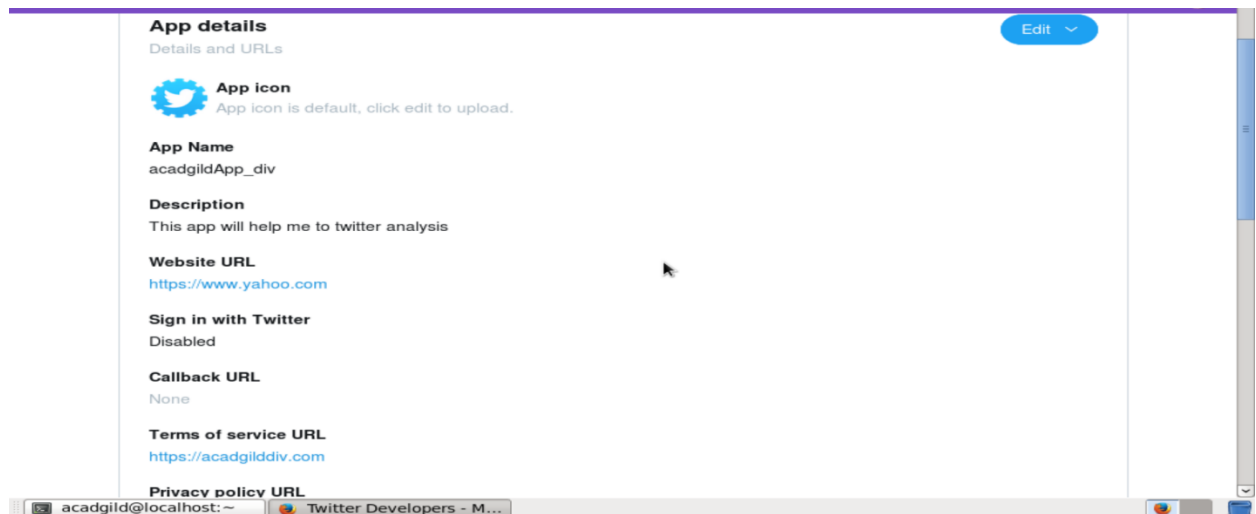
```
[acadgild@localhost ~]$ sudo service mysqld start
Starting mysqld:
[acadgild@localhost ~]$ jps
3297 SecondaryNameNode
3139 DataNode
3043 NameNode
4260 HMaster
4164 HQuorumPeer
4357 HRegionServer
3462 ResourceManager
3594 NodeManager
4703 Jps
[acadgild@localhost ~]$
```



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

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

3. Created app named 'acadgildApp_div'



4. Generated Access Token and Access token Secret code.

Apps / acadgildApp_divCreate an app

App detailsKeys and tokensPermissions

Keys and tokens

Keys, secret keys and access tokens management.

Consumer API keys

mRYe0Lztcks4JObrbmHreQ546 (API key)

4tZiYiStnC0ineOPF9b3h0Y6T2mfQHCGmkhZKYpO9CZLj3M8AC (API secret key)

[Regenerate](#)

Access token & access token secret

525022974-Sm597usulv678tlo7OeXncBRzcBbW2W2gw6HEMy4 (Access token) NEW

N3L8eu9b5cs0P19ADuyim16ba2bdbP8JyB8DdkX993f39 (Access token secret) NEW

Read and write (Access level)

[Revoke](#) [Regenerate](#)

5. Edit bash file.

The screenshot shows a terminal window and a text editor window. The terminal window, titled 'acadgild@localhost:~', shows the command 'sudo gedit .bashrc' being executed. The text editor window, titled '.bashrc (/home/acadgild) - gedit', shows the contents of the .bashrc file. The file contains several lines of code for setting environment variables for HBASE, KAFKA, FLUME, and ZOOKEEPER. The code is as follows:

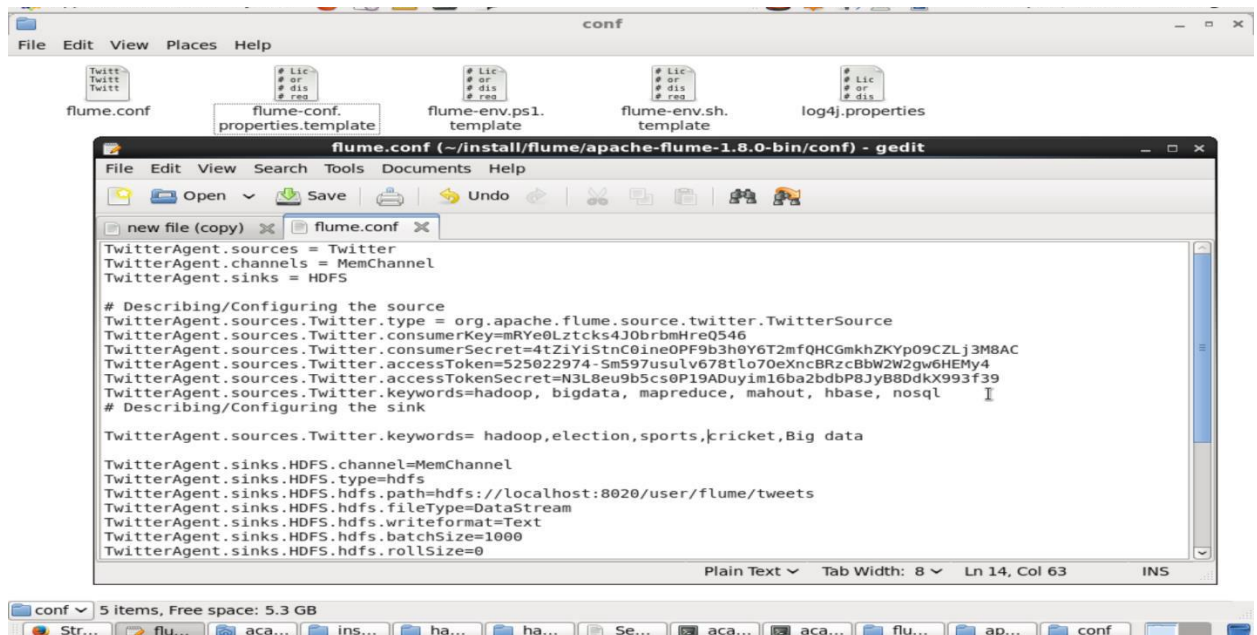
```
# Below 2 lines we have to add for HBASE Installation
export HBASE_HOME=/home/acadgild/install/hbase/hbase-1.2.6
export PATH=$PATH:$HBASE_HOME/bin

# Below 2 lines we have to add for kafka Installation
export KAFKA_HOME=/home/acadgild/install/kafka/kafka_2.12-0.10.1.1
export PATH=$PATH:$KAFKA_HOME/bin

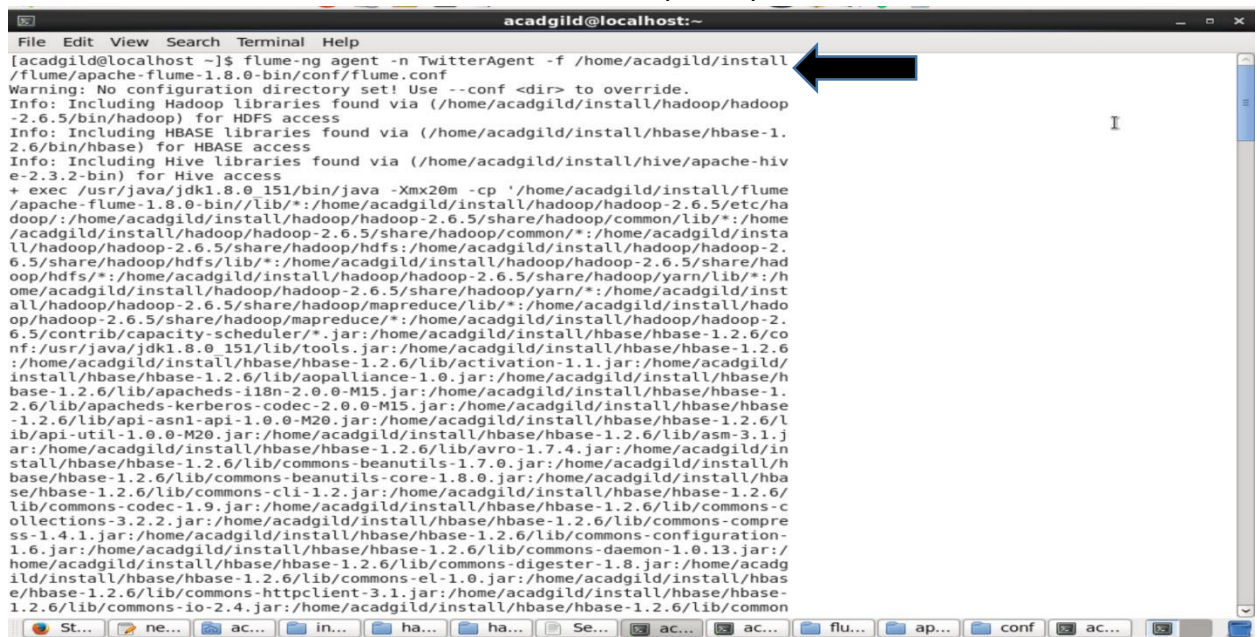
# Below 2 lines we have to add for FLUME Installation
export FLUME_HOME=/home/acadgild/install/flume/apache-flume-1.8.0-bin/
export PATH=$PATH:$FLUME_HOME/bin/

# Below 2 lines we have to add for zookeeper Installation
export ZOOKEEPER_HOME=/home/acadgild/install/zookeeper/zookeeper-3.4.10
export PATH=$PATH:$ZOOKEEPER_HOME/bin
```

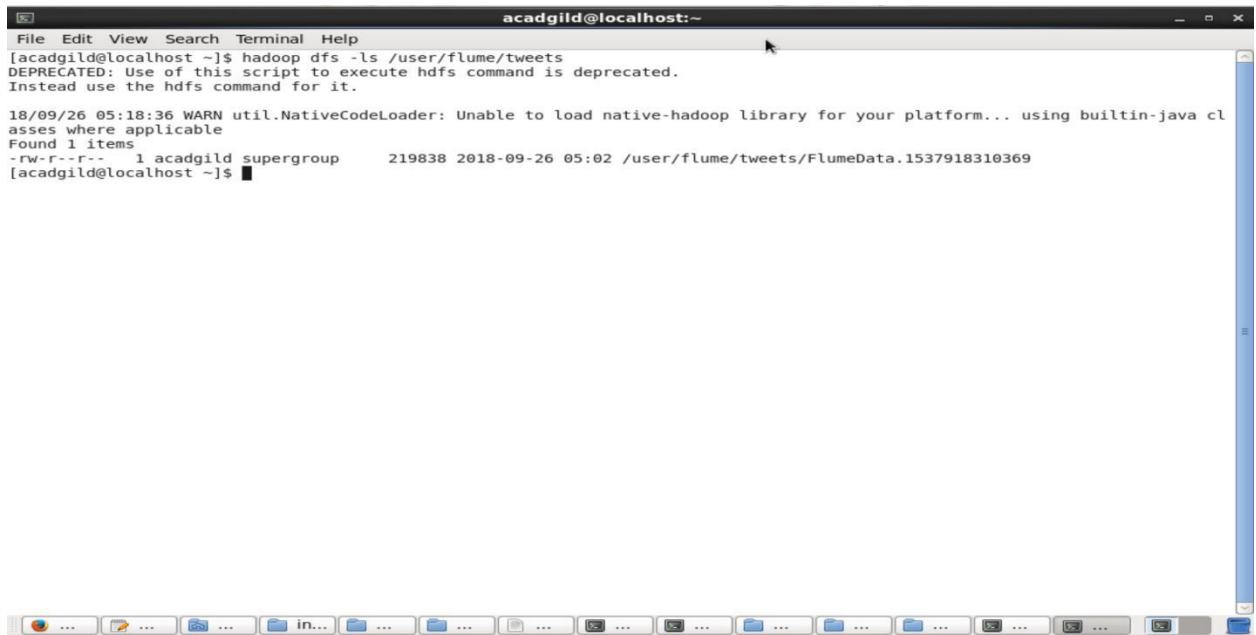
6. Make Flume configuration file with twitter api keys with the keys



7. Fetch the twitter tweet data into the HDFS cluster path by flume

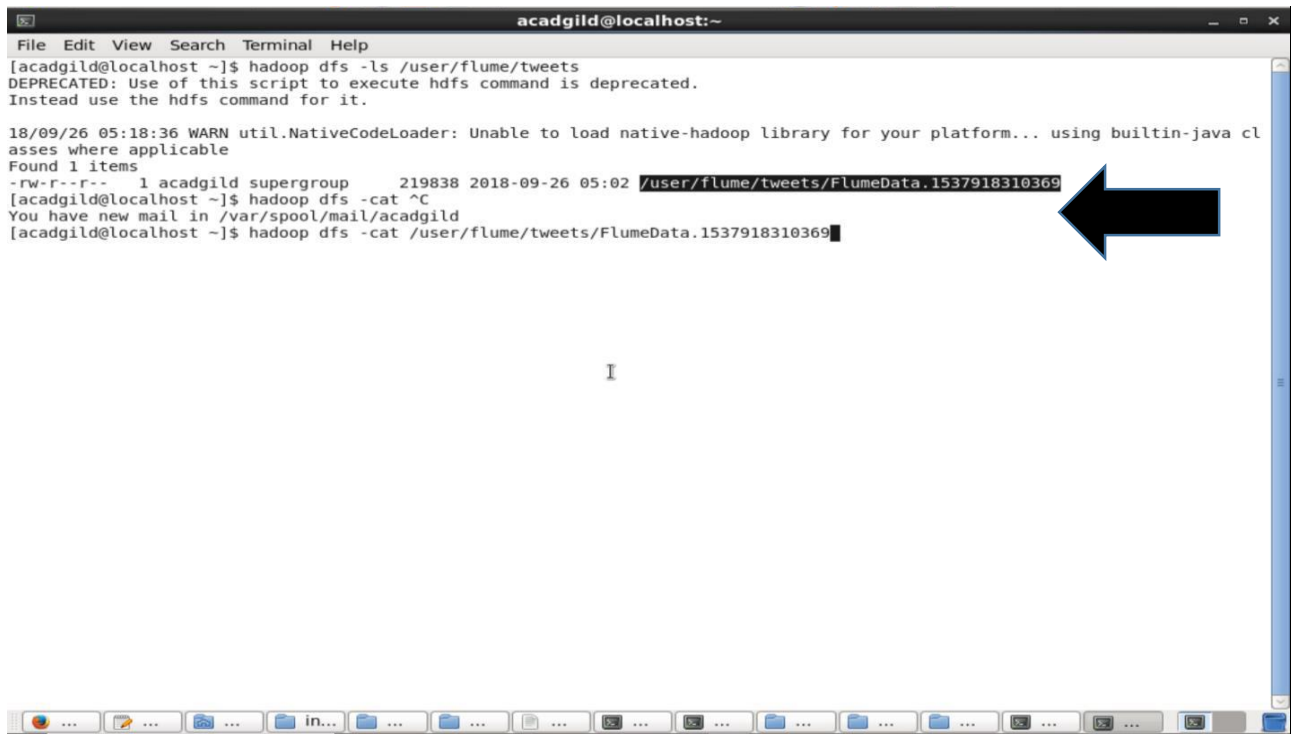


8. Check the contents of the tweet data we can use the following command:
hadoop dfs -ls /user/flume/tweets



```
acadgild@localhost:~  
File Edit View Search Terminal Help  
[acadgild@localhost ~]$ hadoop dfs -ls /user/flume/tweets  
DEPRECATED: Use of this script to execute hdfs command is deprecated.  
Instead use the hdfs command for it.  
  
18/09/26 05:18:36 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl  
asses where applicable  
Found 1 items  
-rw-r--r-- 1 acadgild supergroup 219838 2018-09-26 05:02 /user/flume/tweets/FlumeData.1537918310369  
[acadgild@localhost ~]$
```

9. Display the tweet data inside the /user/flume/tweets/fetched in hdfs using flume



```
acadgild@localhost:~  
File Edit View Search Terminal Help  
[acadgild@localhost ~]$ hadoop dfs -ls /user/flume/tweets  
DEPRECATED: Use of this script to execute hdfs command is deprecated.  
Instead use the hdfs command for it.  
  
18/09/26 05:18:36 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl  
asses where applicable  
Found 1 items  
-rw-r--r-- 1 acadgild supergroup 219838 2018-09-26 05:02 /user/flume/tweets/FlumeData.1537918310369  
[acadgild@localhost ~]$ hadoop dfs -cat ^C  
You have new mail in /var/spool/mail/acadgild  
[acadgild@localhost ~]$ hadoop dfs -cat /user/flume/tweets/FlumeData.1537918310369
```

Twitter data

```
File Edit View Search Terminal Help
i'm funny
i'm ugly

you'll always be laughing at something<?>a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
{"type":"record","name":"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":"retweet_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_reply_to_status_id","type":["long","null"]}, {"name":"media_url_https","type":["string","null"]}, {"name":"expanded_url","type":["string","null"]}]}
@SenBillNelson: Los venezolanos siguen sufriendo a manos del régimen corrupto de Maduro. Insto a mis colegas a que prueben nuestro proy...
MARSIndia013(2018-09-26T05:01:59Z)RT @SenBillNelson: Los venezolanos siguen sufriendo a manos del régimen corrupto de Maduro. Insto a mis colegas a que prueben nuestro proy...
Lite+@104473127028922776A6Republic of the PhilippinesToo much wondering makes me depressedpsukhsychedelic3am(2018-09-26T05:01:59Z)URTWTheDailyWords : https://t.co/P1wtJl3f3n@a href="https://about.twitter.com/products/tweetdeck" rel="nofollow">TweetDeck</a>https://pbs.twimg.com/media/Dnm7x3hU4AAuo_p.jpghttps://twitter.com/TheDailyWords/status/104307126781975040/photo/1&1044731270272634880West Virginia, USA I fly a starshipBanner StatenSteaksauccyy(2018-09-26T05:01:59Z-17 tell you something you may want to hear, but i'll lie.@a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>@104473127027674726albaniavillalbaablv31(2018-09-26T05:01:59Z)RT @GOPChairwoman: Beto O'Rourke wants to:
* abolish ICE & decriminalize unauthorized border crossings
* have govt take over health care
*a href="http://twitter.com/download/android" rel="nofollow">Twitter for Android</a>@1044731270276816896Junior0lin(2018-09-26T05:01:59Z)RPabloNews #MelhorClipeTVZIndestrutivel@a href="http://twitter.com/download/android" rel="nofollow">Twitter for Android</a>@10447312702976081936東京古いビールが好き。社会人
学部の頃、自転車漕いだりボール蹴ったり。
プロフィール画像：「サイドミラーに映ったモノは意外と近くにいるので、気をつけましょう」@UtahKentaro HAYASHIKent Hayashi(2018-09-26T05:01:59Z)RT @ICRRpr: 東京大宇宙線研究所は10月1日から、若手研究者の育成支援のため、スーパーカミオカンデのジグソー
```