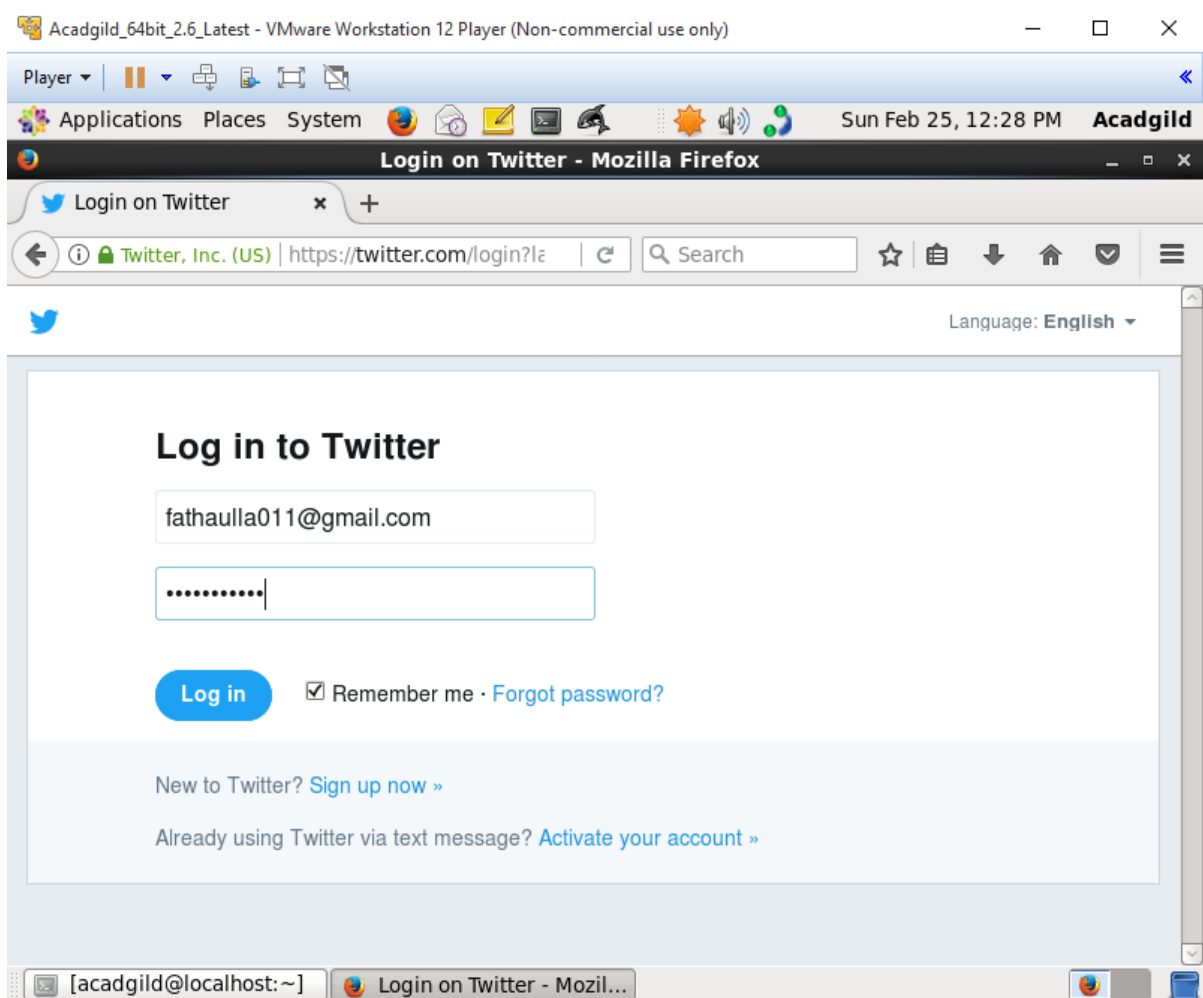# Assignment 12

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

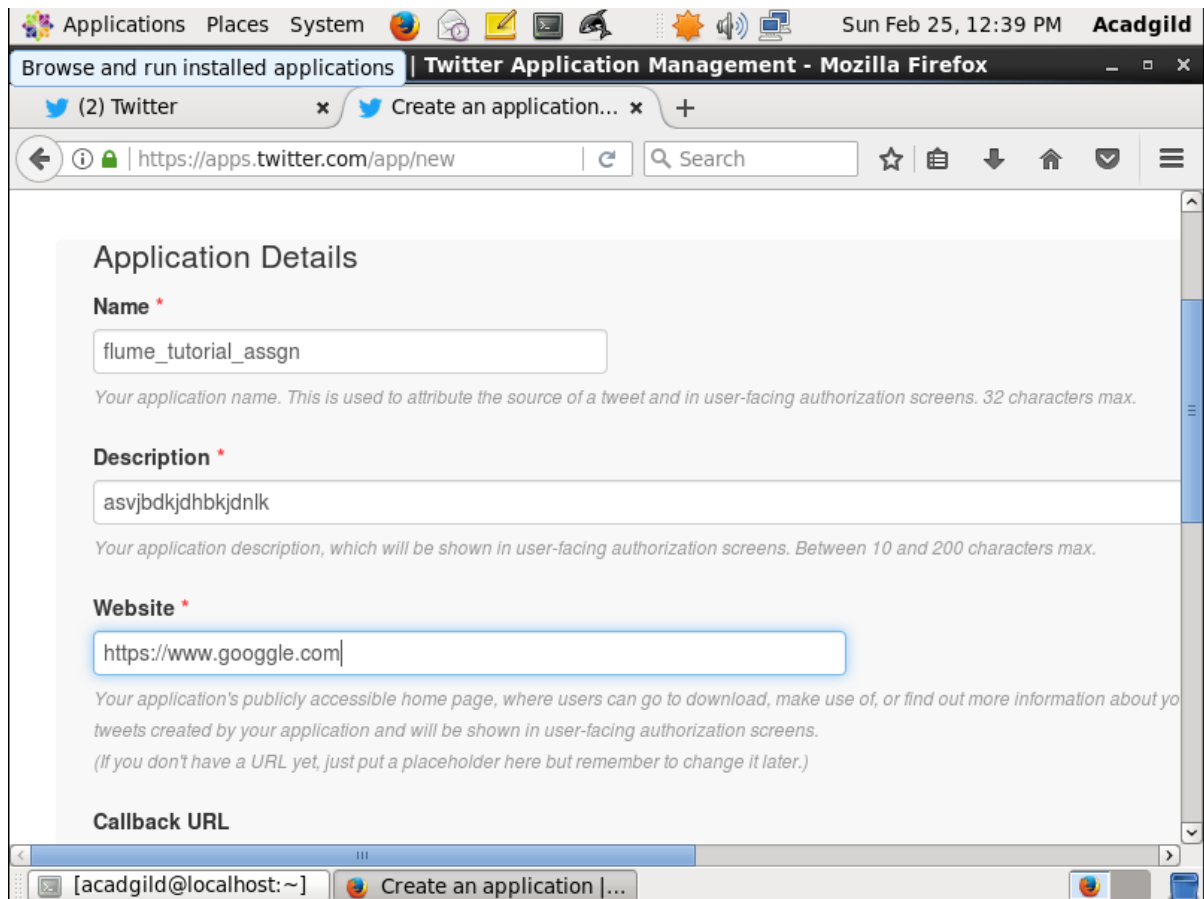## Step1:
Login to the twitter account

Step2:

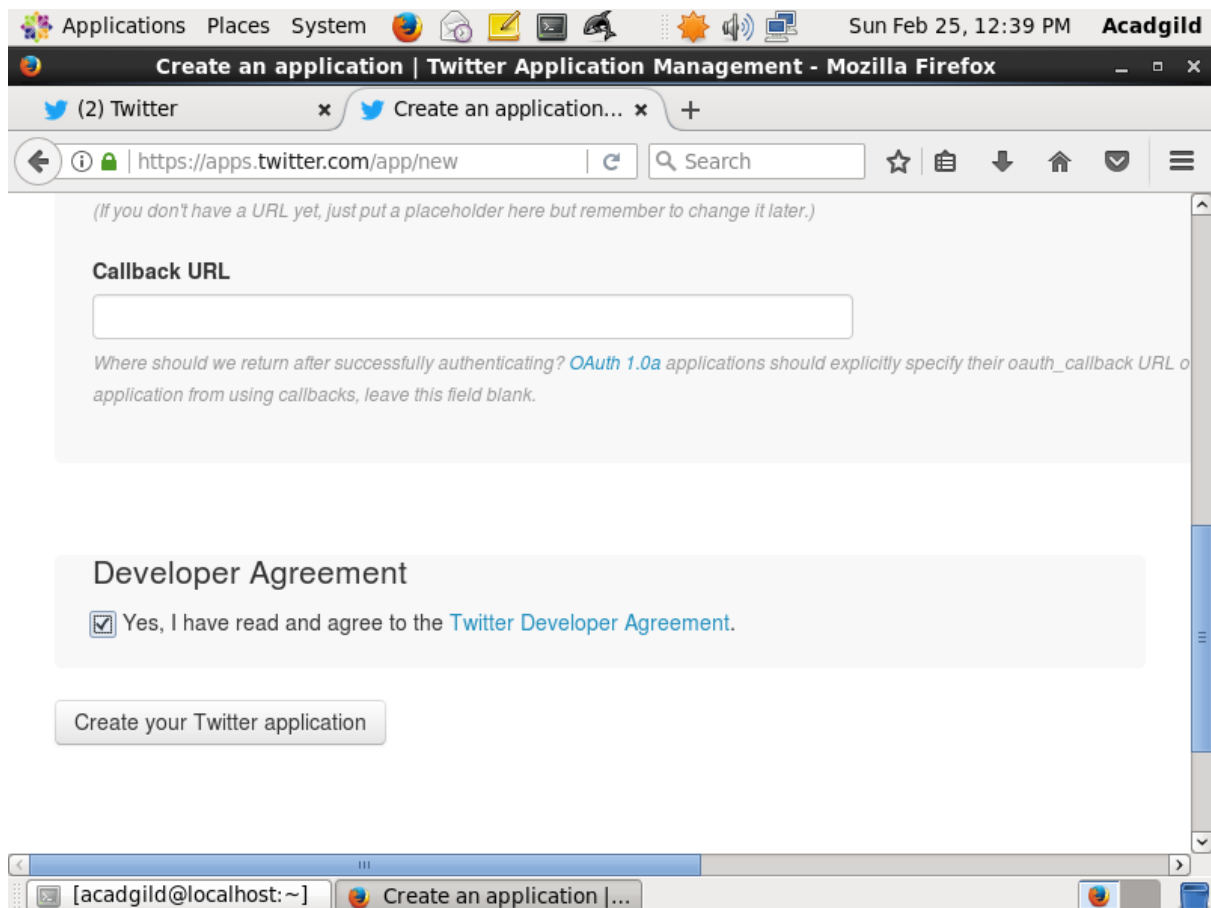Go to the following link and click the 'create new app' button.

https://apps.twitter.com/app

Step3:

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

Step 4:

Copy Consumer Key(API Key) and Consumer Secret(API Secret)

Consumer Key (API Key) CiCsmL1bXdsoCMV3MIfCXaC75
Consumer Secret (API
Secret) BvN0x6Z4WfTys4do2Y9QBJd8Fr5XalngYOyWHYD5sQj4G7EfEY
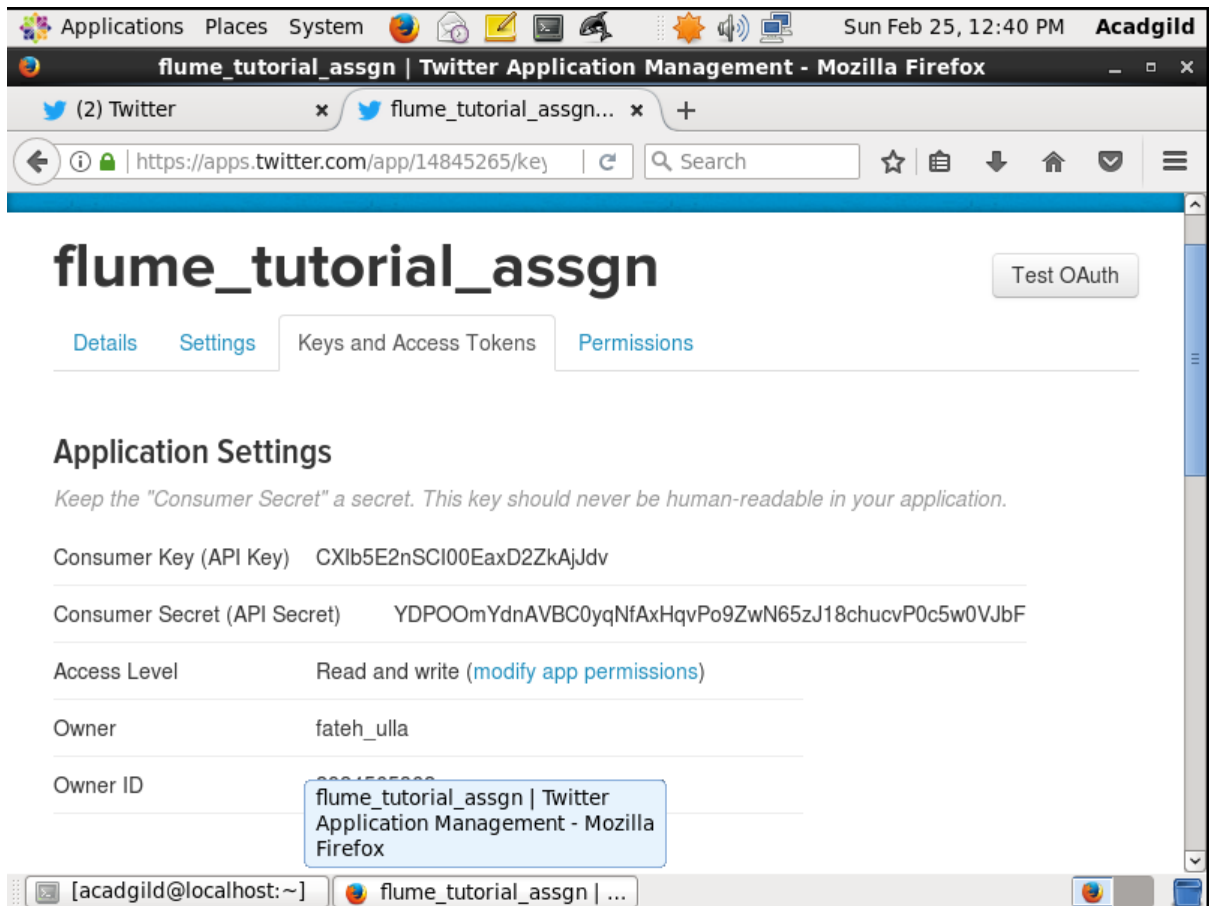
Step5:

Click on Create my access token button
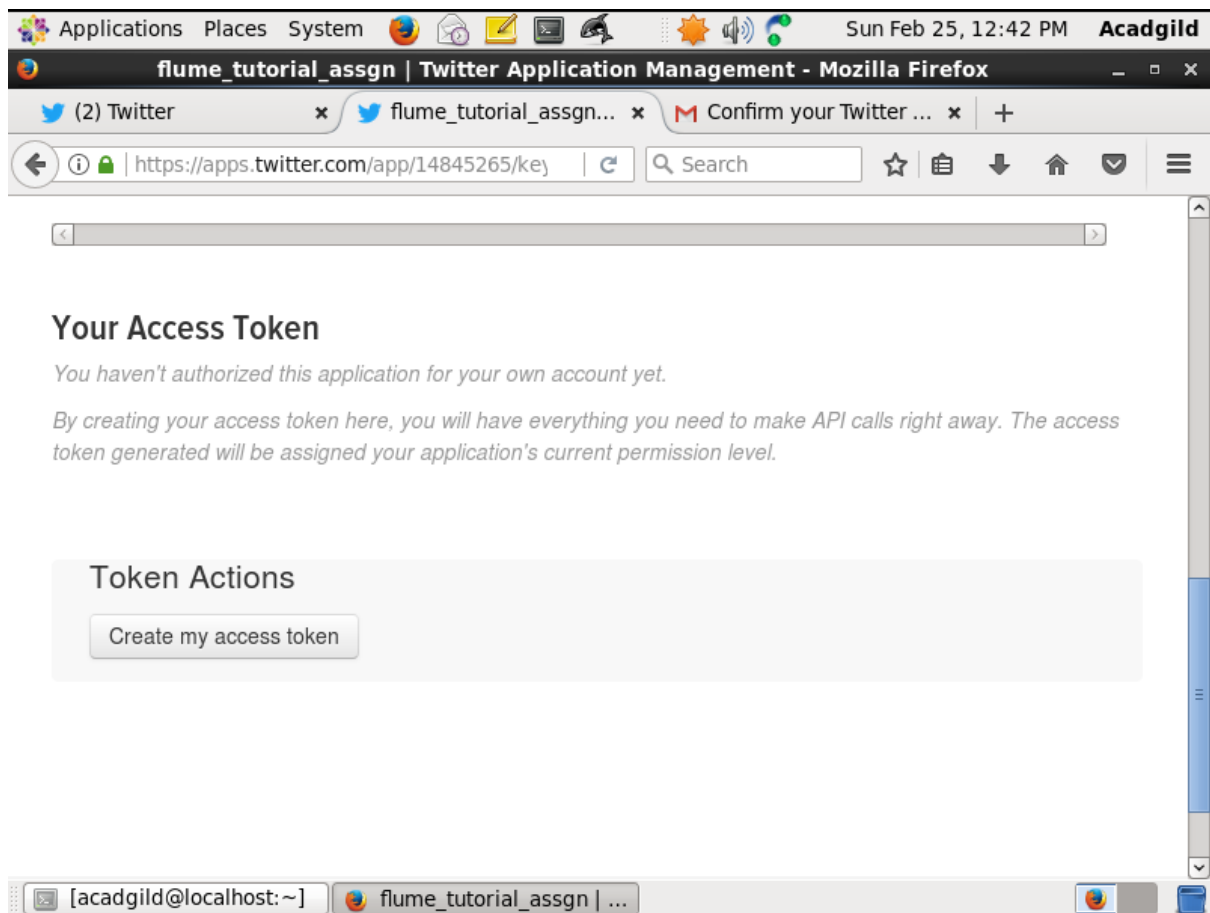
Step 6:

Copy Access Token and Access Token Secret

Access Token 2824565362-Asx4oCjtZpzxtkqpW3lvcw90KeEhcX9mNF2rb0i

Access Token Secret MiOpElNNy9D2NhGOUYlFyuKraPPFfuCDtcJDqmpX6mu31
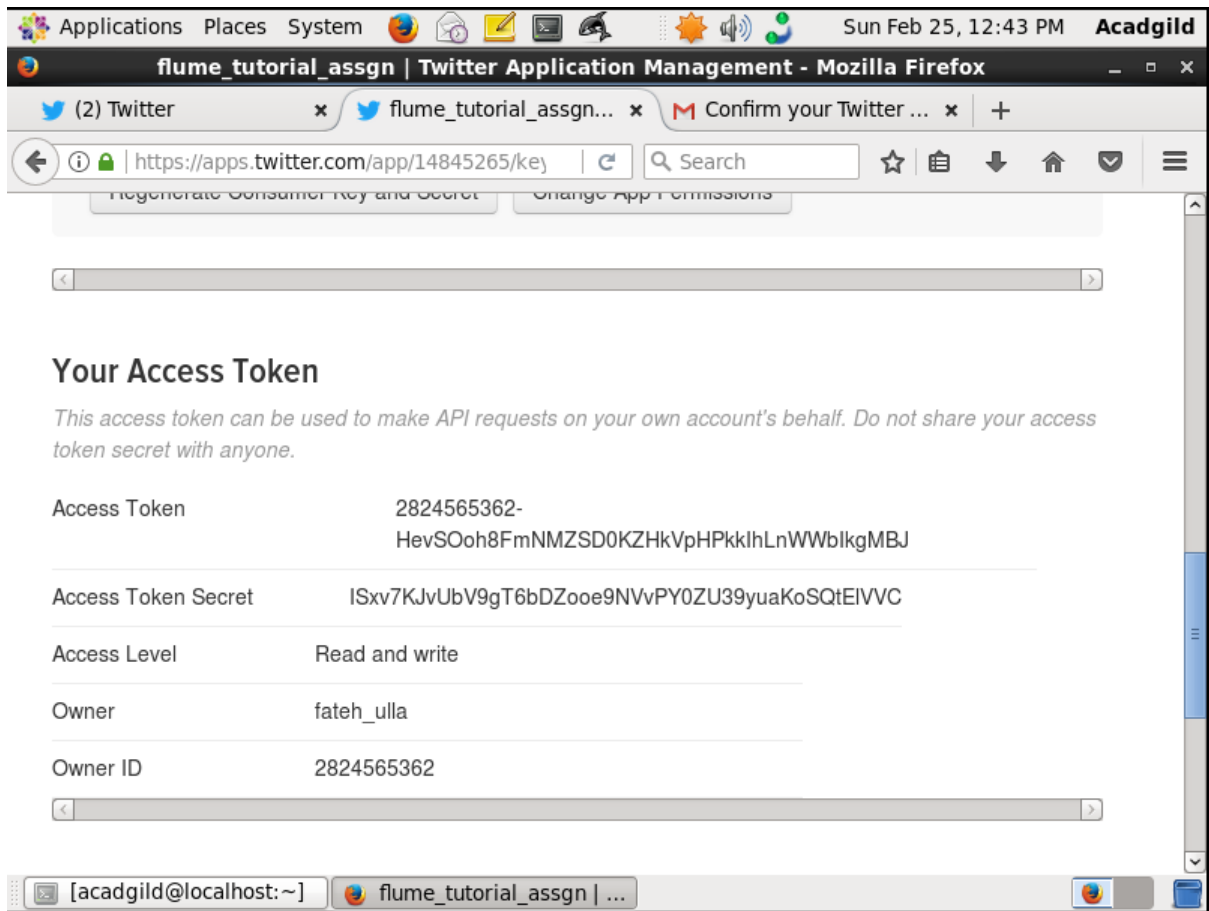
Step 7:
Make a directory of name flumeconf

Step 8:

Create a file with name flumehdfs.conf and Copy the Flume configuration code

```
acadgild@localhost:~/flumeconf

[acadgild@localhost flumeconf]$ cat>flumehdfs.conf

# Naming the components on the current agent.

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



# Describing/Configuring the source

TwitterAgent.sources.Twitter.type = org.apache.flume.source.twitter.TwitterSourc
e
TwitterAgent.sources.Twitter.consumerKey = dVKNTtz1e9BMZt7NcVduGFFdD
TwitterAgent.sources.Twitter.consumerSecret = fGOM63H1zDM4Ak9jI2KAB0fiyCpO9w2wCr
UwViZonFJ1cAYZOh
TwitterAgent.sources.Twitter.accessToken = 1441481497-jYAQmYA8DISqKIk8eRq3cCSpEP
D7dlpdQuuMaNk

TwitterAgent.sources.Twitter.accessTokenSecret 52wtOJsPyzxxW1W1Q80Zt2CsvmTRvkGPd
pmNUli8ZnnDJ
TwitterAgent.sources.Twitter.keywords = hadoop,bigdata,friends,MissUniverse,cric
ket,the
TwitterAgent.sources.Twitter.interceptors = i1
TwitterAgent.sources.Twitter.interceptors.i1.type = timestamp



# Describing/Configuring the sink


TwitterAgent.sinks.HDFS.type = hdfs
TwitterAgent.sinks.HDFS.hdfs.path = /user/acadgild/twitter/%y-%m-%d

TwitterAgent.sinks.HDFS.hdfs.fileType = DataStream

TwitterAgent.sinks.HDFS.hdfs.writeFormat = Text

TwitterAgent.sinks.HDFS.hdfs.batchSize = 100

TwitterAgent.sinks.HDFS.hdfs.rollSize = 0
```

Step 9:

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 --conf /home/acadgild/flumeconf -f flumehdfs.conf

```
[acadgild@localhost flumeconf]$ flume-ng agent -n TwitterAgent --conf /home/acadgild/flumeconf -f flumehdfs.conf
```
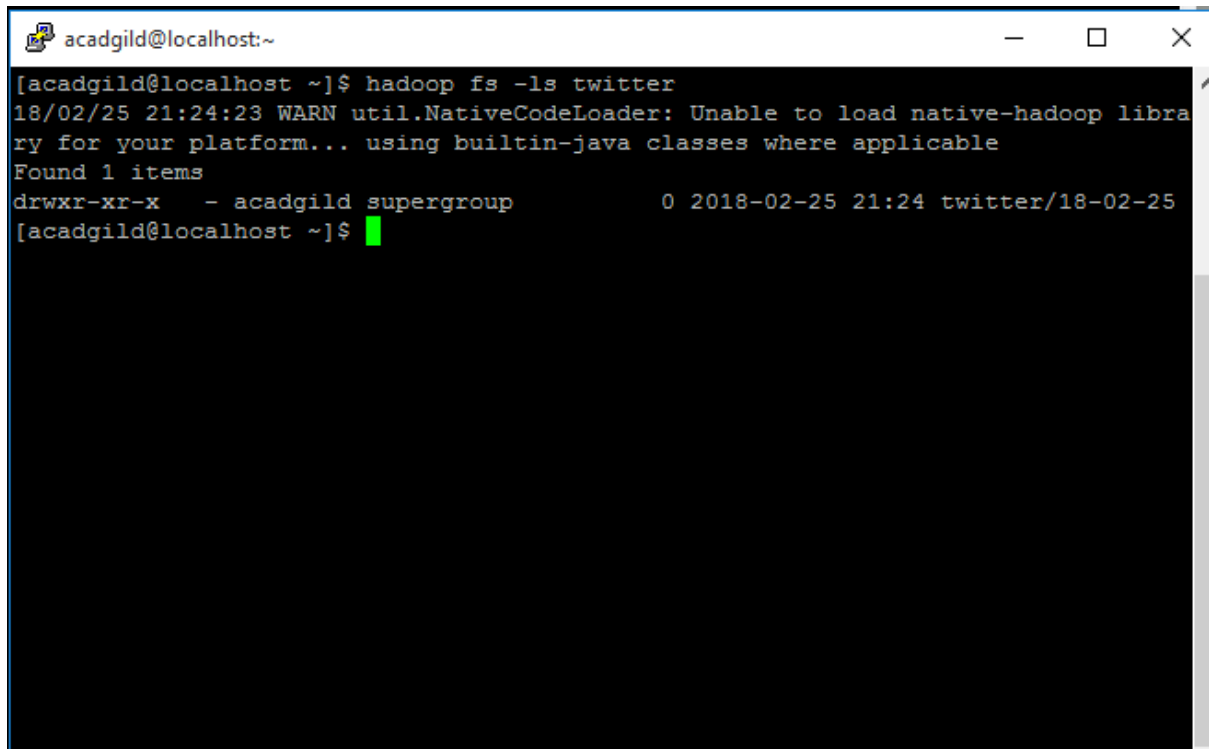
```
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.nextBlockOutputStream(DFSOutputStream.java:1345)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:587)
18/02/25 21:21:25 INFO twitter.TwitterSource: Processed 1,100 docs
18/02/25 21:21:26 INFO hdfs.HDFSDataStream: Serializer = TEXT, UseRawLocalFileSystem = false
18/02/25 21:21:27 INFO hdfs.BucketWriter: Creating /user/acadgild/twitter/18-02-25/FlumeData.1519573886949.tmp
18/02/25 21:21:27 INFO twitter.TwitterSource: Processed 1,200 docs
18/02/25 21:21:28 INFO twitter.TwitterSource: Processed 1,300 docs
18/02/25 21:21:30 INFO twitter.TwitterSource: Processed 1,400 docs
18/02/25 21:21:32 INFO hdfs.BucketWriter: Closing /user/acadgild/twitter/18-02-25/FlumeData.1519573886949.tmp
18/02/25 21:21:32 INFO twitter.TwitterSource: Processed 1,500 docs
18/02/25 21:21:32 INFO hdfs.BucketWriter: Renaming /user/acadgild/twitter/18-02-25/FlumeData.1519573886949.tmp to /user/acadgild/twitte
9
18/02/25 21:21:32 INFO hdfs.HDFSEventSink: Writer callback called.
18/02/25 21:21:32 INFO hdfs.HDFSEventSink: Bucket was closed while trying to append, reinitializing bucket and writing event.
18/02/25 21:21:32 INFO hdfs.HDFSDataStream: Serializer = TEXT, UseRawLocalFileSystem = false
18/02/25 21:21:33 INFO hdfs.BucketWriter: Creating /user/acadgild/twitter/18-02-25/FlumeData.1519573892880.tmp
18/02/25 21:21:34 INFO twitter.TwitterSource: Processed 1,600 docs
18/02/25 21:21:36 INFO twitter.TwitterSource: Processed 1,700 docs
18/02/25 21:21:37 INFO twitter.TwitterSource: Processed 1,800 docs
18/02/25 21:21:38 INFO hdfs.BucketWriter: Closing /user/acadgild/twitter/18-02-25/FlumeData.1519573892880.tmp
18/02/25 21:21:38 INFO hdfs.BucketWriter: Renaming /user/acadgild/twitter/18-02-25/FlumeData.1519573892880.tmp to /user/acadgild/twitte
0
18/02/25 21:21:38 INFO hdfs.HDFSEventSink: Writer callback called.
18/02/25 21:21:39 INFO hdfs.HDFSDataStream: Serializer = TEXT, UseRawLocalFileSystem = false
18/02/25 21:21:39 INFO hdfs.BucketWriter: Creating /user/acadgild/twitter/18-02-25/FlumeData.1519573899395.tmp
18/02/25 21:21:39 INFO twitter.TwitterSource: Processed 1,900 docs
18/02/25 21:21:41 INFO twitter.TwitterSource: Processed 2,000 docs
18/02/25 21:21:41 INFO twitter.TwitterSource: Total docs indexed: 2,000, total skipped docs: 0
18/02/25 21:21:41 INFO twitter.TwitterSource:      50 docs/second
18/02/25 21:21:41 INFO twitter.TwitterSource: Run took 40 seconds and processed:
18/02/25 21:21:41 INFO twitter.TwitterSource:     0.013 MB/sec sent to index
18/02/25 21:21:41 INFO twitter.TwitterSource:     0.534 MB text sent to index
18/02/25 21:21:41 INFO twitter.TwitterSource: There were 0 exceptions ignored:
18/02/25 21:21:43 INFO twitter.TwitterSource: Processed 2,100 docs
18/02/25 21:21:45 INFO hdfs.BucketWriter: Closing /user/acadgild/twitter/18-02-25/FlumeData.1519573899395.tmp
18/02/25 21:21:45 INFO hdfs.BucketWriter: Renaming /user/acadgild/twitter/18-02-25/FlumeData.1519573899395.tmp to /user/acadgild/twitte
5
18/02/25 21:21:45 INFO hdfs.HDFSEventSink: Writer callback called.
18/02/25 21:21:45 INFO hdfs.HDFSDataStream: Serializer = TEXT, UseRawLocalFileSystem = false
18/02/25 21:21:45 INFO twitter.TwitterSource: Processed 2,200 docs
18/02/25 21:21:46 INFO hdfs.BucketWriter: Creating /user/acadgild/twitter/18-02-25/FlumeData.1519573905713.tmp
18/02/25 21:21:47 INFO twitter.TwitterSource: Processed 2,300 docs
```

Step 10:

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

hadoop fs -ls twitter


```
[acadgild@localhost ~]$ hadoop fs -ls twitter
18/02/25 21:24:23 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
Found 1 items
drwxr-xr-x   - acadgild supergroup          0 2018-02-25 21:24 twitter/18-02-25
[acadgild@localhost ~]$
```

hadoop fs –ls twitter/18-02-25


```
acadgild@localhost:~
[acadgild@localhost ~]$
[acadgild@localhost ~]$ hadoop fs -ls twitter
18/02/25 21:24:23 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
Found 1 items
drwxr-xr-x   - acadgild supergroup          0 2018-02-25 21:24 twitter/18-02-25
[acadgild@localhost ~]$ hadoop fs -ls twitter/18-02-25
18/02/25 21:26:08 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
Found 51 items
-rw-r--r--   1 acadgild supergroup          0 2018-02-25 21:21 twitter/18-02-25/
FlumeData.1519573866825
-rw-r--r--   1 acadgild supergroup     642615 2018-02-25 21:21 twitter/18-02-25/
FlumeData.1519573886949
-rw-r--r--   1 acadgild supergroup     162673 2018-02-25 21:21 twitter/18-02-25/
FlumeData.1519573892880
-rw-r--r--   1 acadgild supergroup     155142 2018-02-25 21:21 twitter/18-02-25/
FlumeData.1519573899395
-rw-r--r--   1 acadgild supergroup     157044 2018-02-25 21:21 twitter/18-02-25/
FlumeData.1519573905713
-rw-r--r--   1 acadgild supergroup     168914 2018-02-25 21:21 twitter/18-02-25/
FlumeData.1519573912376
-rw-r--r--   1 acadgild supergroup     120224 2018-02-25 21:22 twitter/18-02-25/
FlumeData.1519573918620
-rw-r--r--   1 acadgild supergroup     167547 2018-02-25 21:22 twitter/18-02-25/
FlumeData.1519573924108
-rw-r--r--   1 acadgild supergroup     145111 2018-02-25 21:22 twitter/18-02-25/
FlumeData.1519573930395
-rw-r--r--   1 acadgild supergroup     135162 2018-02-25 21:22 twitter/18-02-25/
FlumeData.1519573936137
-rw-r--r--   1 acadgild supergroup     140931 2018-02-25 21:22 twitter/18-02-25/
FlumeData.1519573941843
-rw-r--r--   1 acadgild supergroup     155814 2018-02-25 21:22 twitter/18-02-25/
FlumeData.1519573947425
-rw-r--r--   1 acadgild supergroup     149193 2018-02-25 21:22 twitter/18-02-25/
FlumeData.1519573953428
-rw-r--r--   1 acadgild supergroup     108025 2018-02-25 21:22 twitter/18-02-25/
FlumeData.1519573959400
-rw-r--r--   1 acadgild supergroup      68263 2018-02-25 21:22 twitter/18-02-25/
FlumeData.1519573965037
-rw-r--r--   1 acadgild supergroup     219306 2018-02-25 21:22 twitter/18-02-25/
FlumeData.1519573970548
-rw-r--r--   1 acadgild supergroup     129103 2018-02-25 21:23 twitter/18-02-25/
FlumeData.1519573976360
```

hadoop fs -cat twitter/18-02-25/FlumeData.1519574124574

You have new mail in /var/spool/mail/acadgild

[acadgild@localhost ~]$ hadoop fs -cat twitter/18-02-25/FlumeData.1519574124574

18/02/25 21:28:35 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":"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"]}]}C₀蒸!?███6N█$967790164981424134f█초보 연류독╇ / 정유진 배우님 앞이러❤ / 애정배우 오조오억령 / 덕인일치가 꿈!🙆█.♪달올❁m00n_halo(2018-02-25T21:25:30Z█RT @300HTERA_NEW: 어느 분께서 위튜블 의미하는 쯤 꽃을 선물하셨는데

꽃 잘 모시겠다며 배우님으로 생각하겠다며
"꽃 배우님... 아 그 분 아닙니다!"

그 분 아닙니다 ㅋㅋㅋㅋㅋㅋㅋㅋㅋㅋㅋㅋㅋㅋ█Twitter for Android$967790164972994560█ガスト 公式アカウント
✈フェアやキャンペーン情報をお届け！一部店舗では実施していない商品もあります。※店舗やメニューに関するご意見やお問合わせは当アカウントでの返信が出来かねますので、恐れ入りますが下記窓口へお願い致します。→http://twd.ac/t5laLY TEL:0120-125-807(平日9〜18時)█❖ガスト 公式アカウント gusto_official(2018-02-25T21:25:30Z█@marimooo88 ごめんなさい ハズレ …☺
3/7までは毎日チャレンジ♪

さらにWチャンス！
ステーキガストでは抽選で3,000円当たるキャンペーンを実施中！
詳しくは ▶ @steakgusto029 https://t.co/BxBtUaPxRZ█████.█BelugaCampaignSKY^https://pbs.twimg.com/media/DWYAVH-VMAAnXMT.jpg█https://twitter.com/gusto_official/status/965460824805466112/photo/1$967790164998172672(룔 피 스 .█(
다 음 DOWWWWNECK(2018-02-25T21:25:30Z█복 복 슐하욜 유아하개때 일단 내가 나이드 제일 많고 하넌거 절조찼어█Twitter for Android$967790164985565184█+62█□AyuvaniaY(2018-02-25T21:25:30Zh"What hurts you today, makes you stronger tomorrow."|twittbot.net$967790165006548995█北海道█ウェルスアンバサダー・協会認定 認定資格合格番号：C1115179599 以下の３条件、全てを満たしている方以外、僕の【超極秘案内】█車 直樹ADC72264(2018-02-25T21:25:30Z█去年（２０１６年）の紅白、審査員席が槙っちょに有り、審査員の人ステージを見るとき横を見なければならず、首が痛くならないのかな？と思った！|twittbot.net$967790164977188864█ジュレット Rそしゃげ 鍵は身バレ防止用！███
れ 点 repura321(2018-02-25T21:25:30Z█玄武の拳抜きだと一番奏義まわしたいしウーノの配置は一番後ろでいいんだろうか 適当に並べてたけどアルタイルカトルウーノにするか█Twitter Web Client$967790164977176576█Fdon't judge me if you don't know me█Afเคี เคีย เคีย เคียkor_kengk(2018-02-25T21:25:30Z█RT @cuttolipta: บางทีการให้เกิดคนเลย
อาจช่วยให้เราทำใจได้ง่ายขึ้น█Twitter for Android$967790164998148096█olsh junior█jennifer rosejenniferrossee(2018-02-25T21:25:30Z█I just added this to my closet on Poshmark: Coral Halter Bralette. https://t.co/ykJI0u8neU via @poshmarkapp #shopmyclosettPoshmark$967790164977229824█> Yugendra LilhareYugendraL(2018-02-25T21:25:30Z█RT @SohanRajawat5: #संविदा _विरोधी _बी जेपी
#हमारी _भूल कमरे का मूल
बी जेपी के रा ष्ट्रीय नेता पुरभात झा ते भुंडन का मणा उठ्ठ्ते हुए कहा की बा …█Twitter for Android$967790164981440512█*6`แฟ ต้อยจ้ก.ลาๆ.บบDewduckk(2018-02-25T21:25:30Z█RT @poupaep: Are you ready for the B.E.S.T!!!! 🙆⬇□⬇□⬇□⬇□⬇□⬇□ แต่ก็ต้องแย่ะะไล่ยแ.คียงคนท้องตามเต้าดังสนามยมนลดกไปหมดแล้วาาา😊⬇□⬇□⬇□♦♦♦ #CLa…█Twitter Web Client$967790164981399553█B헤 더 수덜먼 인정 수덜먼█싯타나,부미P [귤위푸] [봬█코드Ricocomet#3304] [훌찐쵸드 '개배소빈'] [뿔뿔뿔셩 스님 '리코헤셩']4█
티 크 Rico_Commu(2018-02-25T21:25:30ZT허 헝 .....페에야 널수있오려나 …█TweetDeck$967790164968990265█في غبدي اللي طوافي السهم ماجورمنا �😊 ❤بام ♪
3_likl(2018-02-25T21:25:30Z<@aaaa_hii في ليبيه في 🙆❤█████Twitter for Android█████967790164994031616█
a href="http://twitter>█.ب.مت█4Wajdan Hamad²⁴sᴀ.Wjadan24(2018-02-25T21:25:30Z█هً█.فهها █أمتلً فيه أجمل من اللي دايم منبهر فيك رقم إنك أصبحت بامت ٢٠* ديسمبرt█طهّن أيفرك أيسروً █بام ♪

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