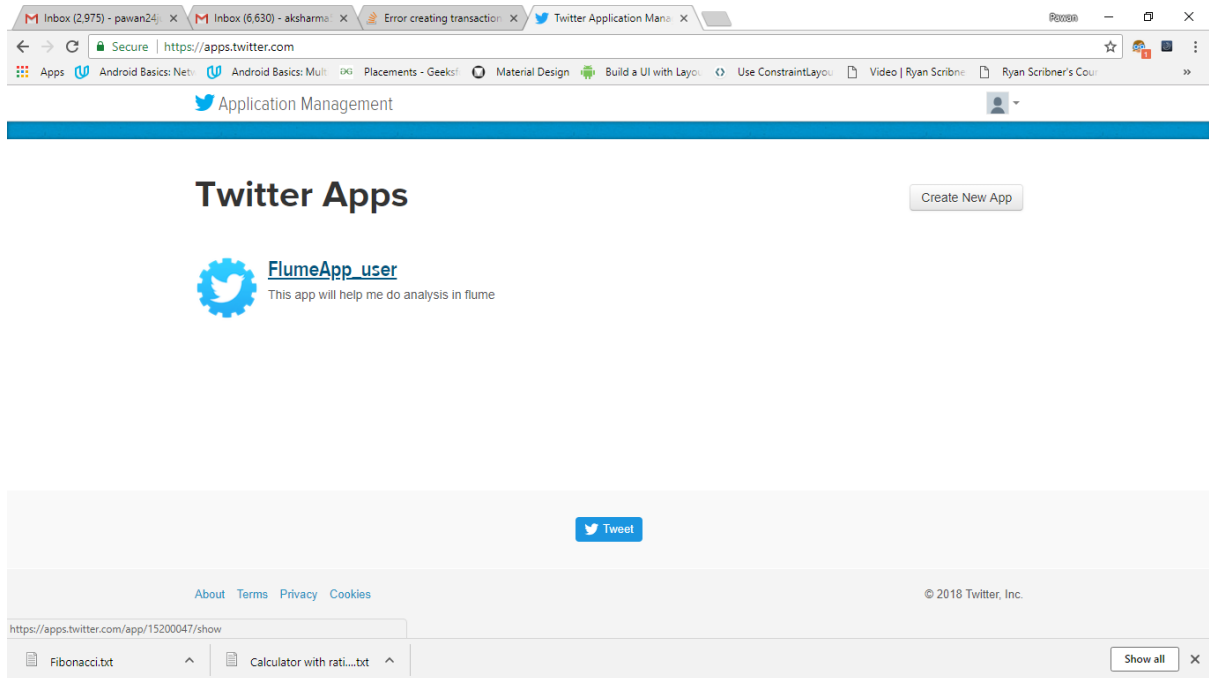


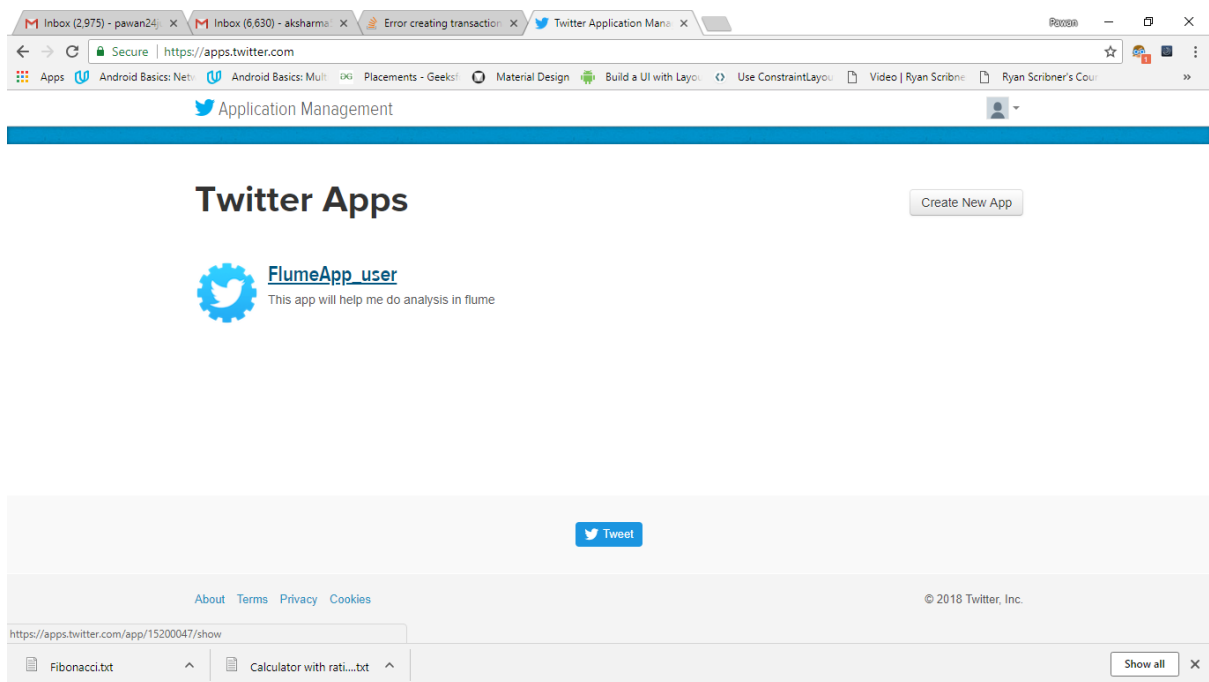
ASSIGNMENT

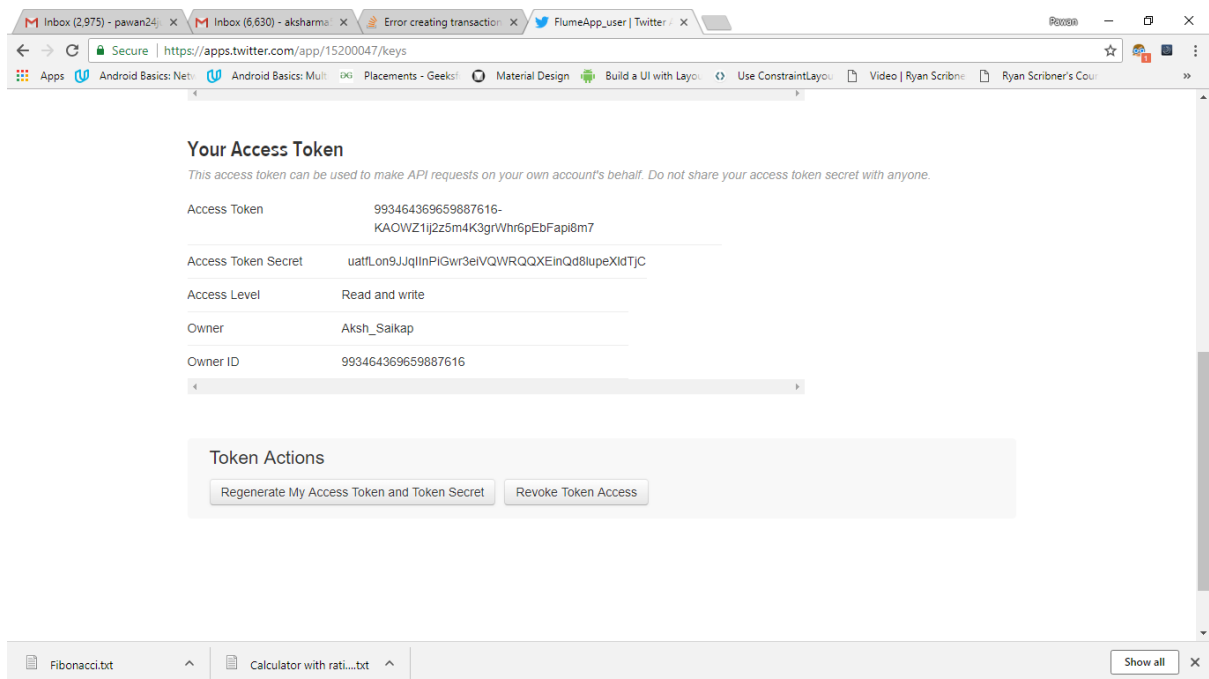
Task1: To create Flume Agent to stream data from twitter

Step 1: Create twitter application **FlumeApp_user**



Step2 : Take down consumer key, secret and Access token, Access secret





Step3: make a directory **flumeconf** , put **flumehdfs.conf** file into this directory.

- ➔ flumehdfs.conf contains the details of source=Twitter, sink=HDFS, channel=MemChannel, where the agent name is TwitterAgent
- ➔ File also contains the consumer key and access token details captured from twitter application.

```

acadgild@localhost:~/flumeconf
[acadgild@localhost ~]$
[acadgild@localhost ~]$ mkdir flumeconf
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ cd flumeconf
[acadgild@localhost flumeconf]$ cat > flumehdfs.conf
# mkdir flumeconf
# cd flumeconf
# copy flumehdfs.conf to flumeconf directory
# flume-ng agent -n TwitterAgent --conf /home/acadgild/flumeconf -f 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.TwitterSource
TwitterAgent.sources.Twitter.consumerKey = iL8WaOsnAvF4xj5h0FOKnvP0c
TwitterAgent.sources.Twitter.consumerSecret = Qn7YFMY1qb9RIQc7mbco6awhsDoQmK80E6
9rAKnO2jHYQClPgt
TwitterAgent.sources.Twitter.accessToken = 993464369659887616-KAOWZ1ij2z5m4K3grW
hr6pEbFapi8m7
TwitterAgent.sources.Twitter.accessTokenSecret = uatfLon9JjqlnPiGwr3eiVQWRQQXEInQd8lupeXldTjC
TwitterAgent.sources.Twitter.keywords = hadoop,bigdata,friends,MissUniverse,cric
ket,the
TwitterAgent.sources.Twitter.interceptors = il
TwitterAgent.sources.Twitter.interceptors.il.type = timestamp

# Describing/Configuring the sink
TwitterAgent.sinks.HDFS.type = hdfs

```

Step 4: Run the command to fetch data from twitter

➔ flume-ng agent -n TwitterAgent --conf /home/acadgild/flumeconf -f flumehdfs.conf

```
acadgild@localhost:~/flumeconf
[acadgild@localhost flumeconf]$
[acadgild@localhost flumeconf]$ flume-ng agent -n TwitterAgent --conf /home/acadgild/flumeconf -f flumehdfs.conf
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 -Xmx200m -cp '/home/acadgild/flumeconf:/home/acadgild/install/flume/apache-flume-1.8.0-bin/lib/*:/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:/usr/java/jdk1.8.0_151/lib/tools.jar:/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-i18n-2.0.0-M15.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/com
```

```
acadgild@localhost:~
-rw-r--r-- 1 acadgild supergroup 79540 2018-05-08 13:16 twitter/18-05-08/FlumeData.1525765595531
-rw-r--r-- 1 acadgild supergroup 68706 2018-05-08 13:16 twitter/18-05-08/FlumeData.1525765602641
-rw-r--r-- 1 acadgild supergroup 112149 2018-05-08 13:16 twitter/18-05-08/FlumeData.1525765609589
-rw-r--r-- 1 acadgild supergroup 83667 2018-05-08 13:17 twitter/18-05-08/FlumeData.1525765615580
-rw-r--r-- 1 acadgild supergroup 95483 2018-05-08 13:17 twitter/18-05-08/FlumeData.1525765620911
-rw-r--r-- 1 acadgild supergroup 61816 2018-05-08 13:17 twitter/18-05-08/FlumeData.1525765626290
-rw-r--r-- 1 acadgild supergroup 84087 2018-05-08 13:17 twitter/18-05-08/FlumeData.1525765631599
-rw-r--r-- 1 acadgild supergroup 74412 2018-05-08 13:17 twitter/18-05-08/FlumeData.1525765637870
-rw-r--r-- 1 acadgild supergroup 85044 2018-05-08 13:17 twitter/18-05-08/FlumeData.1525765644214
-rw-r--r-- 1 acadgild supergroup 109577 2018-05-08 13:17 twitter/18-05-08/FlumeData.1525765651635
-rw-r--r-- 1 acadgild supergroup 62142 2018-05-08 13:17 twitter/18-05-08/FlumeData.1525765657806
-rw-r--r-- 1 acadgild supergroup 69248 2018-05-08 13:17 twitter/18-05-08/FlumeData.1525765663122
-rw-r--r-- 1 acadgild supergroup 112348 2018-05-08 13:17 twitter/18-05-08/FlumeData.1525765668793
-rw-r--r-- 1 acadgild supergroup 59263 2018-05-08 13:17 twitter/18-05-08/FlumeData.1525765674515
-rw-r--r-- 1 acadgild supergroup 74425 2018-05-08 13:18 twitter/18-05-08/FlumeData.1525765679898
-rw-r--r-- 1 acadgild supergroup 121597 2018-05-08 13:18 twitter/18-05-08/FlumeData.1525765685917
-rw-r--r-- 1 acadgild supergroup 52851 2018-05-08 13:18 twitter/18-05-08/FlumeData.1525765692509
-rw-r--r-- 1 acadgild supergroup 110502 2018-05-08 13:18 twitter/18-05-08/
```

Step5: To check the data captured, run the command

➔ `hadoop fs -cat twitter/18-05-08/FlumeData.1525765766131`

```
acacgild@localhost:~$  
FlumeData.1525765748712  
-rw-r--r-- 1 acacgild supergroup 48307 2018-05-08 13:19 twitter/18-05-08/  
FlumeData.1525765754213  
-rw-r--r-- 1 acacgild supergroup 58179 2018-05-08 13:19 twitter/18-05-08/  
FlumeData.1525765759419  
-rw-r--r-- 1 acacgild supergroup 160650 2018-05-08 13:19 twitter/18-05-08/  
FlumeData.1525765766131  
-rw-r--r-- 1 acacgild supergroup 54759 2018-05-08 13:19 twitter/18-05-08/  
FlumeData.1525765771919  
-rw-r--r-- 1 acacgild supergroup 105116 2018-05-08 13:19 twitter/18-05-08/  
FlumeData.1525765778213  
-rw-r--r-- 1 acacgild supergroup 49245 2018-05-08 13:19 twitter/18-05-08/  
FlumeData.1525765783544  
You have new mail in /var/spool/mail/acacgild  
[acacgild@localhost ~]$ hadoop fs -cat twitter/18-05-08/FlumeData.1525765766131  
18/05/08 13:20:11 WARN util.NativeCodeLoader: Unable to load native-hadoop libra  
ry 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"]}  
  ]  
}  
{"id": "993759726704918529", "user_friends_count": 0, "user_location": "womennawomenna (2018-05-08T13:19:16Z)", "user_description": "Евгений Плющенко - биография, фото, личная жизнь фигуриста https://t.co/UFjn26yBuK", "user_statuses_count": 2320, "user_followers_count": 0, "user_name": "a href=\"http://twitter.com\" rel=\"nofollow\">Twitter Web Client", "user_screen_name": "a href=\"http://twitter.com\" rel=\"nofollow\">Twitter Web Client", "created_at": "2018-05-08T13:19:16Z", "text": "a href=\"http://twitter.com\" rel=\"nofollow\">Twitter Web Client", "retweet_count": 0, "retweeted": false, "in_reply_to_user_id": null, "source": "a href=\"http://covellin", "in_reply_to_status_id": null, "media_url_https": "a href=\"http://covellin", "expanded_url": "a href=\"http://covellin", "type": "photo"}  
DcQLM_gVwAAkoLn.jpgzhttps://twitter.com/luicexx/status/993759726683832320/photo/
```