

## ASSIGNMENT11.3

### Twitter Data Using Flume

Apache Flume is a Hadoop ecosystem component used to collect, aggregate and moves a large amount of log data from different sources to a centralized data store.

### PROCESS

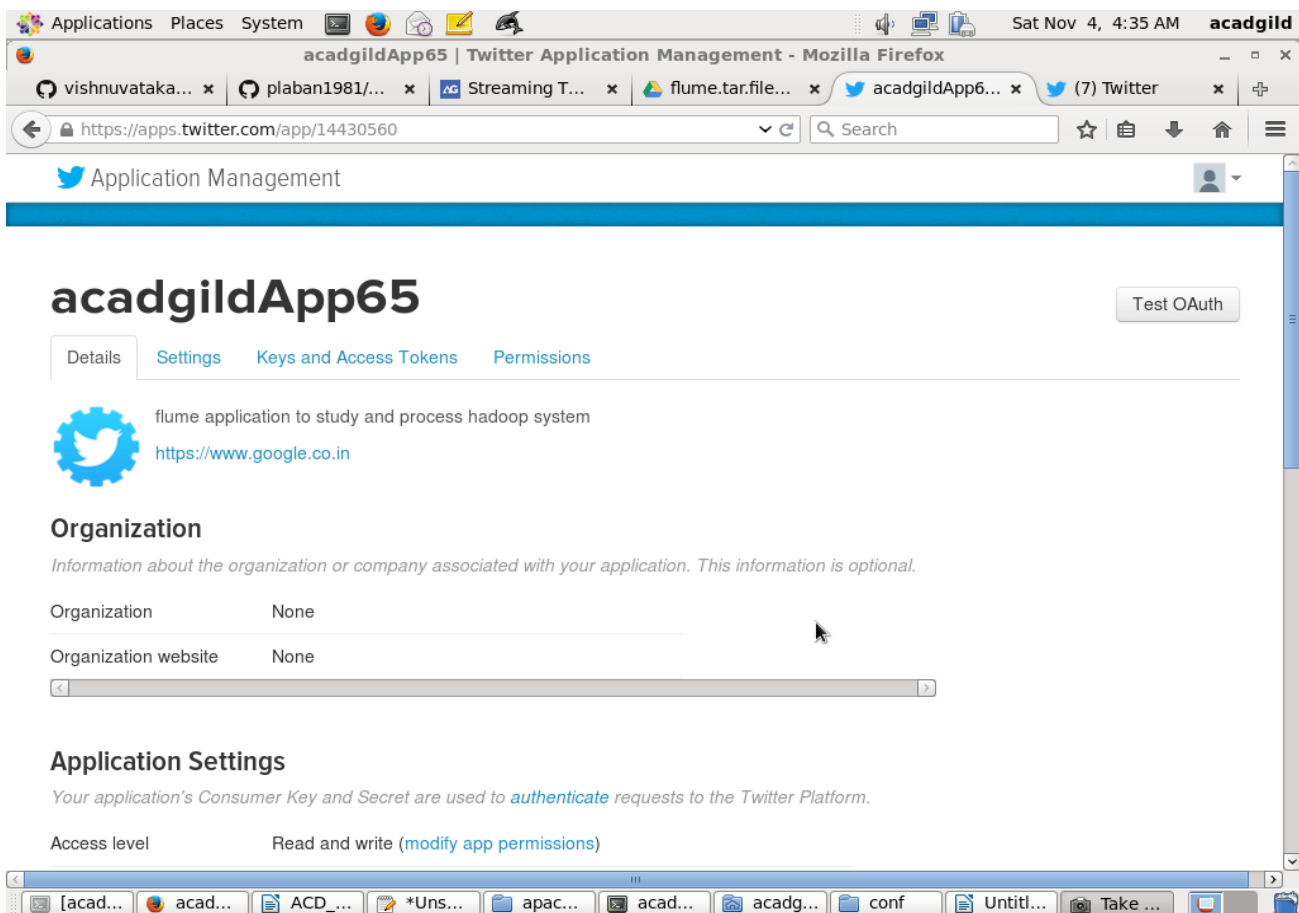
#### Fetching keys from twitter

Consumer Key (API Key) 1L2iVignj32qR6oiNVwJPagBw

Consumer Secret (API Secret) Qzt87kyoSg0JhlYb6SQD7Mis3iuUyjGE8I8EtPypRjDfVdbjwW

Access Token 926550535196459008-K4Z7ATvxvduW1C5JdO2gPlZ08csQBUA

Access Token Secret OJBCrroLNk0JzrzWoN9M5hmvCGNQkfoXtLSE2X6SIKyvz



#### Flume config file

```
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=1L2iVignj32qR6oiNVwJPagBw
TwitterAgent.sources.Twitter.consumerSecret=0zt87kyoSg0Jhlyb6S0D7Mis3iuUyijGE8I8EtPypRjDfVdbjwW
TwitterAgent.sources.Twitter.accessToken=926550535196459008-K4Z7ATvxvduW1C5Jd02gPLZ08csQBUA
TwitterAgent.sources.Twitter.accessTokenSecret=0JBCrroLNk0JzrzWoN9M5hmvCGNQkfoXtLSE2X6SIKyvz
TwitterAgent.sources.Twitter.keywords=hadoop, bigdata, mapreduce, mahout, hbase, nosql
# Describing/Configuring the sink

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

TwitterAgent.sinks.HDFS.channel=MemChannel
TwitterAgent.sinks.HDFS.type=hdfs
TwitterAgent.sinks.HDFS.hdfs.path=hdfs://localhost:9000/user/flume/tweets
TwitterAgent.sinks.HDFS.hdfs.fileType=DataStream
TwitterAgent.sinks.HDFS.hdfs.writeformat=Text
TwitterAgent.sinks.HDFS.hdfs.batchSize=1000
TwitterAgent.sinks.HDFS.hdfs.rollSize=0
TwitterAgent.sinks.HDFS.hdfs.rollCount=10000
TwitterAgent.sinks.HDFS.hdfs.rollInterval=600

TwitterAgent.channels.MemChannel.type=memory
TwitterAgent.channels.MemChannel.capacity=10000
TwitterAgent.channels.MemChannel.transactionCapacity=1000

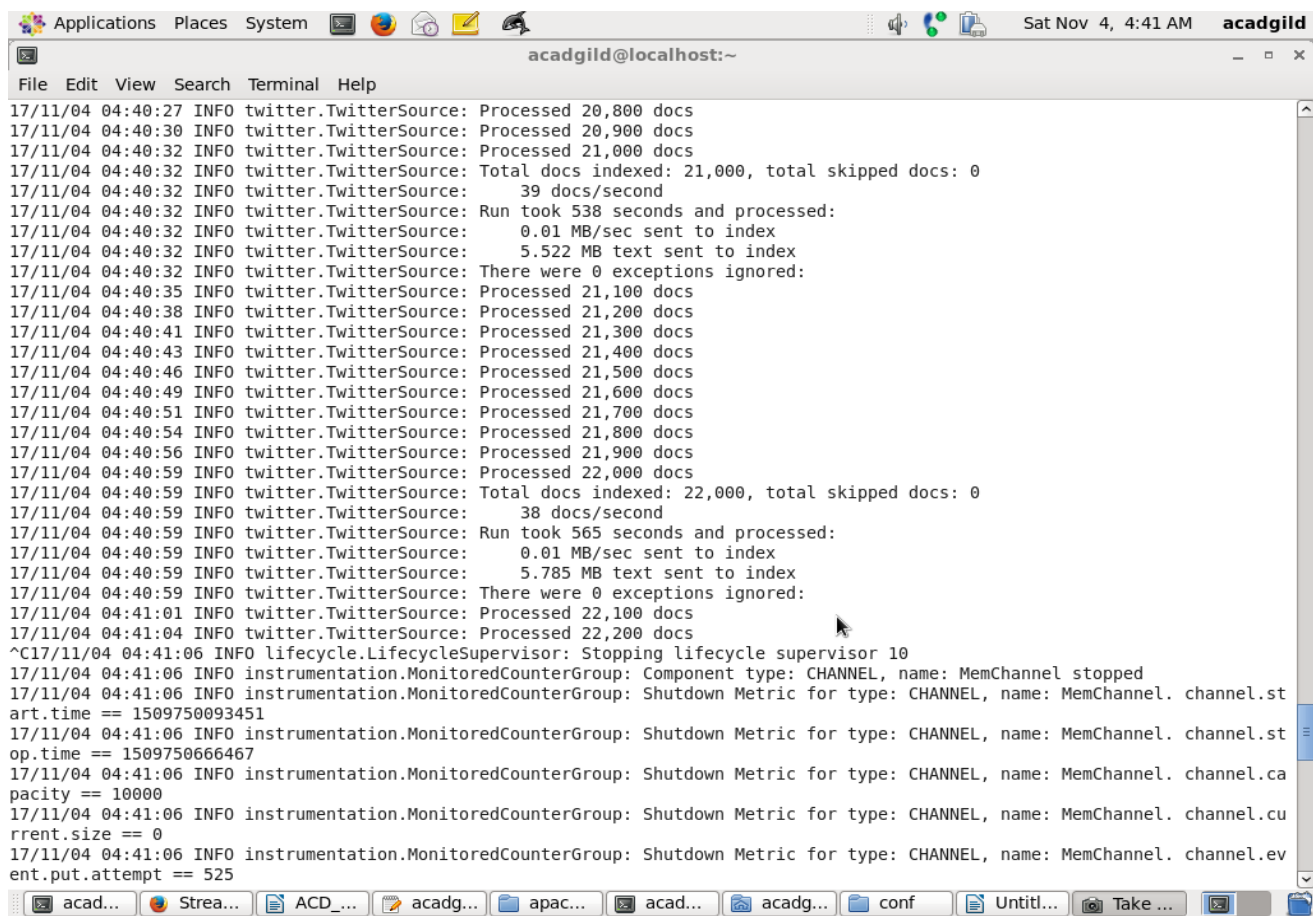
TwitterAgent.sources.Twitter.channels = MemChannel
TwitterAgent.sinks.HDFS.channel = MemChannel
```

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

**hadoop fs -mkdir -p /vihnu/tweet**

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

**[acadgild@localhost ~]\$ flume-ng agent -n TwitterAgent -f /home/acadgild/apache-flume-1.6.0-bin/conf/acadgild.conf**



```
acadgild@localhost:~  
File Edit View Search Terminal Help  
17/11/04 04:40:27 INFO twitter.TwitterSource: Processed 20,800 docs  
17/11/04 04:40:30 INFO twitter.TwitterSource: Processed 20,900 docs  
17/11/04 04:40:32 INFO twitter.TwitterSource: Processed 21,000 docs  
17/11/04 04:40:32 INFO twitter.TwitterSource: Total docs indexed: 21,000, total skipped docs: 0  
17/11/04 04:40:32 INFO twitter.TwitterSource: 39 docs/second  
17/11/04 04:40:32 INFO twitter.TwitterSource: Run took 538 seconds and processed:  
17/11/04 04:40:32 INFO twitter.TwitterSource: 0.01 MB/sec sent to index  
17/11/04 04:40:32 INFO twitter.TwitterSource: 5.522 MB text sent to index  
17/11/04 04:40:32 INFO twitter.TwitterSource: There were 0 exceptions ignored:  
17/11/04 04:40:35 INFO twitter.TwitterSource: Processed 21,100 docs  
17/11/04 04:40:38 INFO twitter.TwitterSource: Processed 21,200 docs  
17/11/04 04:40:41 INFO twitter.TwitterSource: Processed 21,300 docs  
17/11/04 04:40:43 INFO twitter.TwitterSource: Processed 21,400 docs  
17/11/04 04:40:46 INFO twitter.TwitterSource: Processed 21,500 docs  
17/11/04 04:40:49 INFO twitter.TwitterSource: Processed 21,600 docs  
17/11/04 04:40:51 INFO twitter.TwitterSource: Processed 21,700 docs  
17/11/04 04:40:54 INFO twitter.TwitterSource: Processed 21,800 docs  
17/11/04 04:40:56 INFO twitter.TwitterSource: Processed 21,900 docs  
17/11/04 04:40:59 INFO twitter.TwitterSource: Processed 22,000 docs  
17/11/04 04:40:59 INFO twitter.TwitterSource: Total docs indexed: 22,000, total skipped docs: 0  
17/11/04 04:40:59 INFO twitter.TwitterSource: 38 docs/second  
17/11/04 04:40:59 INFO twitter.TwitterSource: Run took 565 seconds and processed:  
17/11/04 04:40:59 INFO twitter.TwitterSource: 0.01 MB/sec sent to index  
17/11/04 04:40:59 INFO twitter.TwitterSource: 5.785 MB text sent to index  
17/11/04 04:40:59 INFO twitter.TwitterSource: There were 0 exceptions ignored:  
17/11/04 04:41:01 INFO twitter.TwitterSource: Processed 22,100 docs  
17/11/04 04:41:04 INFO twitter.TwitterSource: Processed 22,200 docs  
^C17/11/04 04:41:06 INFO lifecycle.LifecycleSupervisor: Stopping lifecycle supervisor 10  
17/11/04 04:41:06 INFO instrumentation.MonitoredCounterGroup: Component type: CHANNEL, name: MemChannel stopped  
17/11/04 04:41:06 INFO instrumentation.MonitoredCounterGroup: Shutdown Metric for type: CHANNEL, name: MemChannel. channel.st  
art.time == 1509750093451  
17/11/04 04:41:06 INFO instrumentation.MonitoredCounterGroup: Shutdown Metric for type: CHANNEL, name: MemChannel. channel.st  
op.time == 1509750666467  
17/11/04 04:41:06 INFO instrumentation.MonitoredCounterGroup: Shutdown Metric for type: CHANNEL, name: MemChannel. channel.ca  
pacity == 10000  
17/11/04 04:41:06 INFO instrumentation.MonitoredCounterGroup: Shutdown Metric for type: CHANNEL, name: MemChannel. channel.cu  
rrent.size == 0  
17/11/04 04:41:06 INFO instrumentation.MonitoredCounterGroup: Shutdown Metric for type: CHANNEL, name: MemChannel. channel.ev  
ent.put.attempt == 525
```

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

```
[acadgild@localhost ~]$ hadoop fs -ls /user/flume/tweets
```

17/11/04 05:13:12 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

Found 4 items

```
-rw-r--r--  1 acadgild supergroup  9618092 2017-11-04 04:41  
/user/flume/tweets/FlumeData.1509750097545
```

```
-rw-r--r--  1 acadgild supergroup  5004481 2017-11-04 04:56  
/user/flume/tweets/FlumeData.1509751295968
```

```
-rw-r--r--  1 acadgild supergroup  9419766 2017-11-04 05:07  
/user/flume/tweets/FlumeData.1509751661597
```

```
-rw-r--r--  1 acadgild supergroup  3230141 2017-11-04 05:11  
/user/flume/tweets/FlumeData.1509752272196
```

```

Applications Places System [Icons] Sat Nov 4, 5:21 AM acadgild
[Terminal Icon] acadgild@localhost:~
File Edit View Search Terminal Help

@larryelder: So, Hillary colluded with the DNC to steal the nomination and then colluded with Russia to try to win--and it's TRUMP who g...http://twitter.com/download/android rel="nofollow">Twitter for Android</a>92658758741377843360 Fresno, CA Casey Campagna Caseycampagnaa\(2017-11-04T04:41:10Z\)@SarahMackAttack: Some squid are very brave during their checkups and others...pretend to be rocks. https://t.co/JGhq4SPwXWa href="http://twitter.com/download/iphone">a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>https://pbs.twimg.com/media/DNqNa9tUEAEvYQ.jpg0 https://twitter.com/SarahMackAttack/status/926208176160169984/photo/1\$926587587409489920\$1 Denver, Colorado0 Denver's #1 Source for Breaking News and Intelligent Talk - Live-Local Talk with Peter Boyles, Dan Caplis, Chuck & Julie, and Craig Silverman0 0z710 KNUS Denver710KNUS (2017-11-04T04:41:10Z)Crestani Talks With Us Before His Speech in Boulder on Tuesday - Nov 3, 2017 - Hr 3 https://t.co/nMrCu sf3fE #KNUS #GreatRadioa href="https://dlvr.it/com" rel="nofollow">dvlr.it</a>9265875873970831377The Come UpNYOUNG Dum b Nigga Sticking To The Code.000Uncle S.B.LoveShreu\(2017-11-04T04:41:10Z\)NMy dick be turning these bitchies into stalkers a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>9265875874348605470Eggward, USA0Sam MotleySammotley1\(2017-11-04T04:41:10Z\)@StephenKing: Only 11 minutes without Trump's divisive poison, but wasn't it sweet?a href="http://twitter.com/download/android" rel="nofollow">Twitter for Android</a>9265875874097274880(Perth shire, Scotland0Contributor to https://t.co/X5cvZWz6cf Sometime guest @robsaulshow. Also at https://t.co/OtpYAqvYni. https://t.co/Xb5ESFjIch Trying to wake people up! Born in a href="https://twitter.com/download/android" rel="nofollow">Twitter for Android</a>9265875874348974080One beatles lyric every five minutes.00 &Every Beatles Lyricevbeatleslyric\(2017-11-04T04:41:10Z\)I thinking of youa href="http://www.apeuro.xyz z" rel="nofollow">everybeatleslyrics</a>9265875874012200960Αττική, Ελλάδα0 love music, hanging out, the sea and videogames especially Aqw.0 DXelsearRogers Rayne\(2017-11-04T04:41:10Z\)x@VerlyrusAE Can we get this emo vampire dudes jacket in Aqw?0 a href="http://twitter.com" rel="nofollow">Twitter Web Client</a>92658758742207283340retoria0ather of 3.00lanhla Knosi0lanhlaJ Knosi(2017-11-04T04:41:10Z)Today's stats : 26 non-followers. found via https://t.co/LQ8K3CsYyg0a href="http://whounfollowedme.org/" rel="nofollow">WhoUnfollowedMe Org</a>9265875873968988160Monterrey México0carrera s, Maestrias en; Administración Pública, Derecho Privado, en Derecho Fiscal, Estudios de Maestría en Psicología. DOCTOR EN AD MINISTRACIÓN ESTRATEGICA.000Pedro Cantu Elizondo0pedrocantue(2017-11-04T04:41:10Z)RT @elnorte: SMNews joven baja Loma Larga en patinata por Av. Gómez Morín. #LibreAcceso https://t.co/UccBmAaC7 https://t.co/X4L08hZvhva href="http://twitter.r.com" rel="nofollow">Twitter Web Client</a>https://pbs.twimg.com/media/DNVfZSAUIAA8-N.jpgzhttps://twitter.com/elnordestatus/926579776244117504/photo/1\$92658758739714457700i think you'll know – @fixfosters0@alizonaparrish\(2017-11-04T04:41:10Z\)ZRT @teamgreys4ever: Sobre Photograph: https://t.co/xawgTYxB1Ba href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>https://pbs.twimg.com/tweet\_video\_thumb/DNvICnWwAAA\_mcy.jpg0https://twitter.com/teamgreys4ever/status/926585992886341632/photo/10#v609265875874348769280Brugge, België0Beleidsmedewerker internationale sa menwerking - Departement Mobiliteit & Openbare Werken0Katrien Criel0KatrienCriel(2017-11-04T04:41:10Z)RT @EEN EU: Got a project for the sustainable #BlueEconomy? Come to @EU\_MARE @EU\_EASME InfoDay to discover new funding opps... a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>9265875874348769280#v60[acadgild@localhost ~]$ 62;9;c^c [acadgild@localhost ~]$ █

```