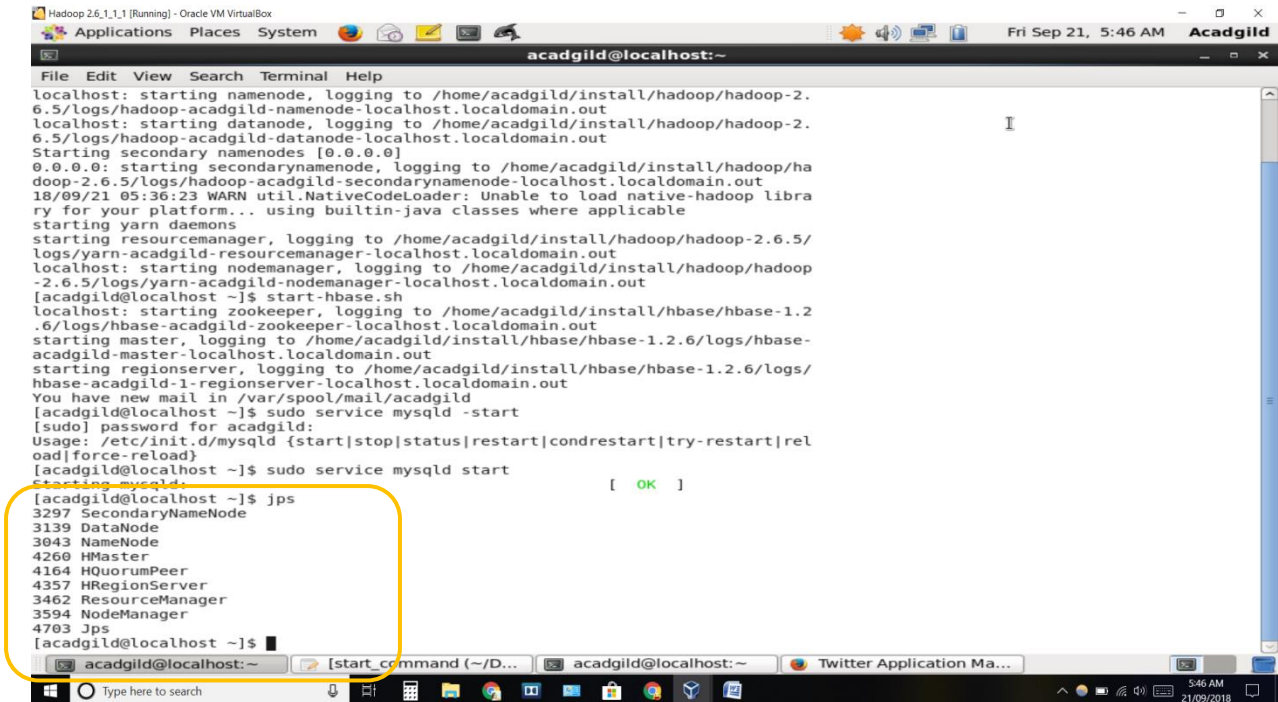


Big Data Hadoop 'Session 12: Oozie and Flume Assignment 1'

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

1. Check Hadoop daemons running

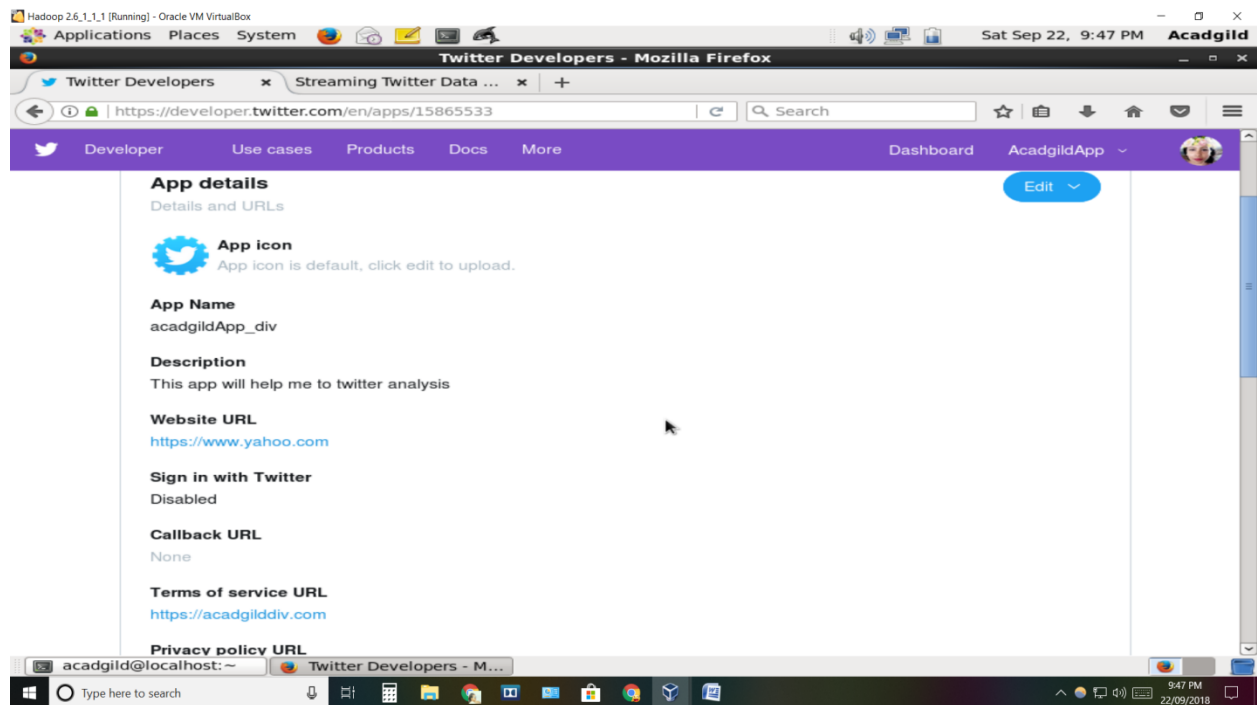


```
acadmild@localhost:~$ start-hbase.sh
localhost: starting namenode, logging to /home/acadmild/install/hadoop/hadoop-2.6.5/logs/hadoop-acadmild-namenode-localhost.localdomain.out
localhost: starting datanode, logging to /home/acadmild/install/hadoop/hadoop-2.6.5/logs/hadoop-acadmild-datanode-localhost.localdomain.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /home/acadmild/install/hadoop/hadoop-2.6.5/logs/hadoop-acadmild-secondarynamenode-localhost.localdomain.out
18/09/21 05:36:23 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
starting yarn daemons
starting resourcemanager, logging to /home/acadmild/install/hadoop/hadoop-2.6.5/logs/yarn-acadmild-resourcemanager-localhost.localdomain.out
localhost: starting nodemanager, logging to /home/acadmild/install/hadoop/hadoop-2.6.5/logs/yarn-acadmild-nodemanager-localhost.localdomain.out
[acadmild@localhost ~]$ start-hbase.sh
localhost: starting zookeeper, logging to /home/acadmild/install/hbase/hbase-1.2.6/logs/hbase-acadmild-zookeeper-localhost.localdomain.out
starting master, logging to /home/acadmild/install/hbase/hbase-1.2.6/logs/hbase-acadmild-master-localhost.localdomain.out
starting regionserver, logging to /home/acadmild/install/hbase/hbase-1.2.6/logs/hbase-acadmild-1-regionserver-localhost.localdomain.out
You have new mail in /var/spool/mail/acadmild
[acadmild@localhost ~]$ sudo service mysql -start
[sudo] password for admild:
Usage: /etc/init.d/mysql {start|stop|status|restart|condrestart|try-restart|reload|force-reload}
[acadmild@localhost ~]$ sudo service mysql start
[ OK ]
[acadmild@localhost ~]$ jps
3297 SecondaryNameNode
3139 DataNode
3043 NameNode
4260 HMaster
4164 HQuorumPeer
4357 HRegionServer
3462 ResourceManager
3594 NodeManager
4703 jps
[acadmild@localhost ~]$
```

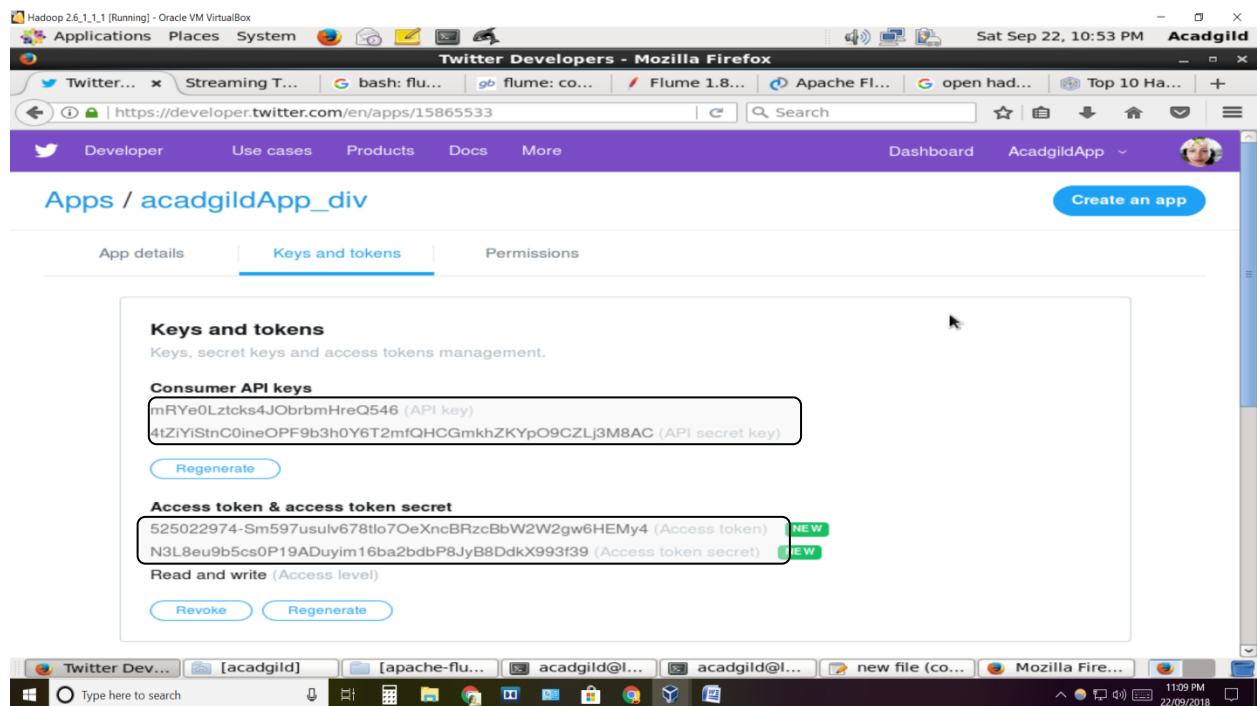
2. Go to the following link and click the 'create new app' button.

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

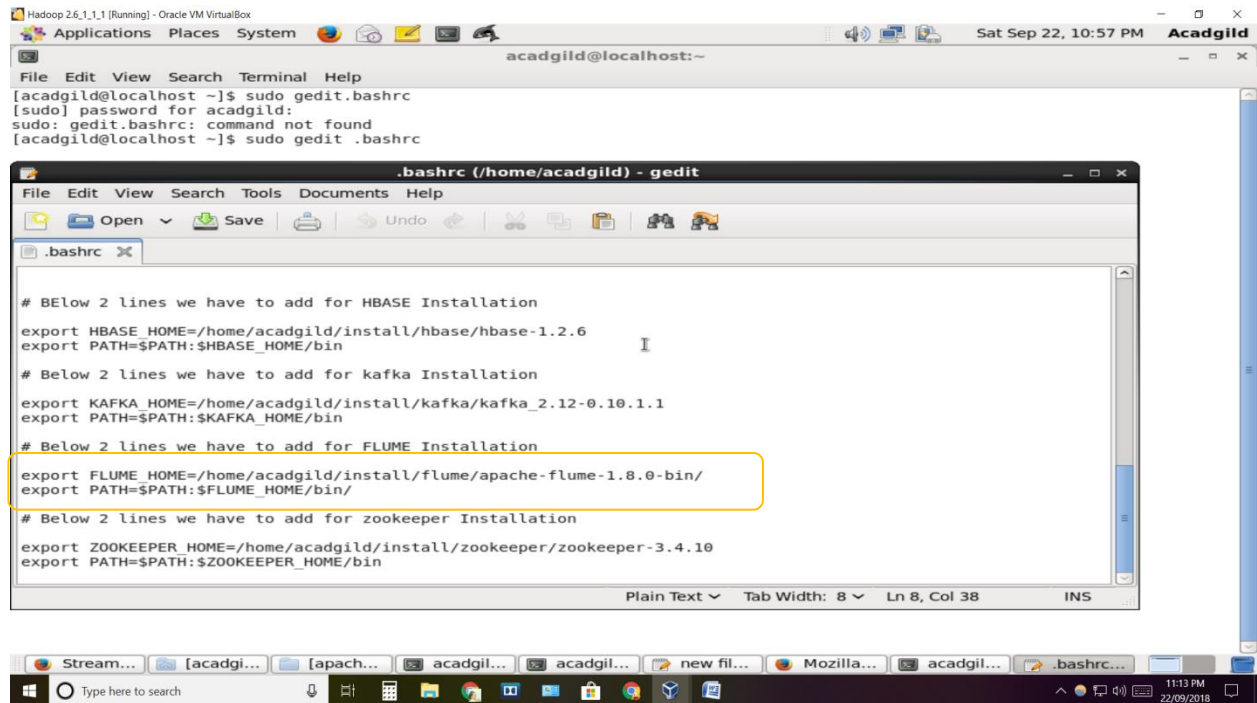
3. Created app named 'acadmildApp_div'



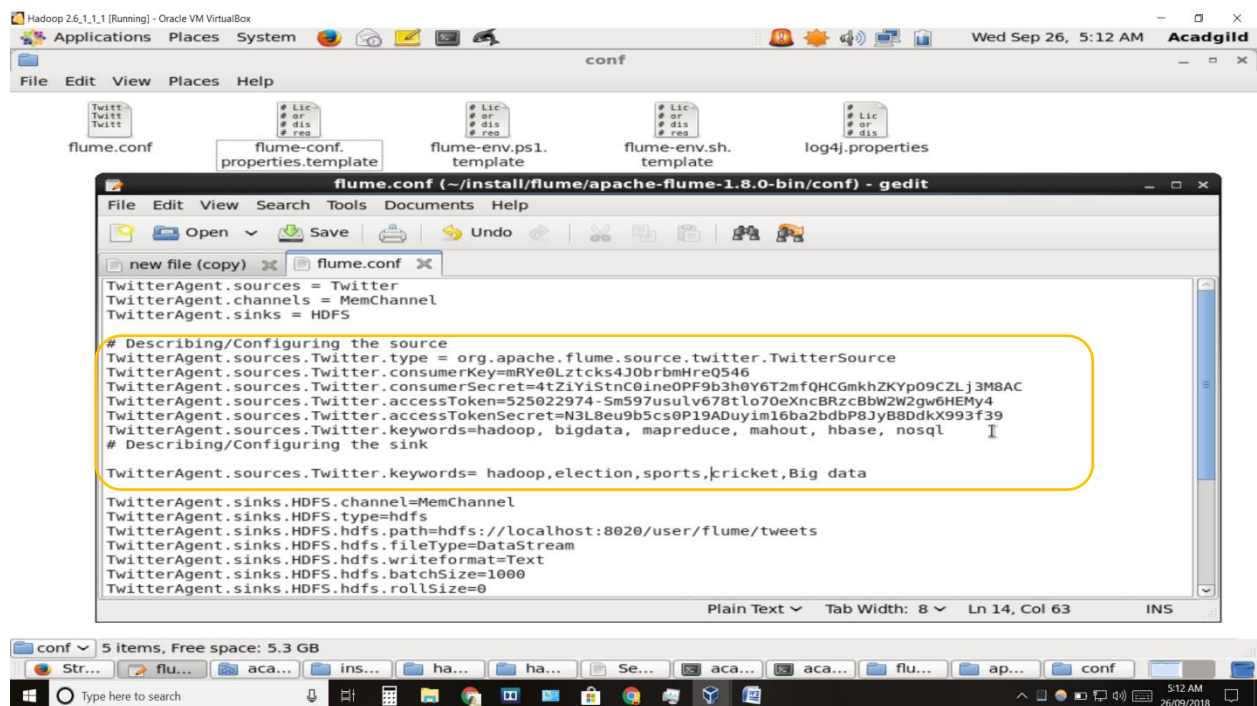
4. Generated Access Token and Access token Secret code.



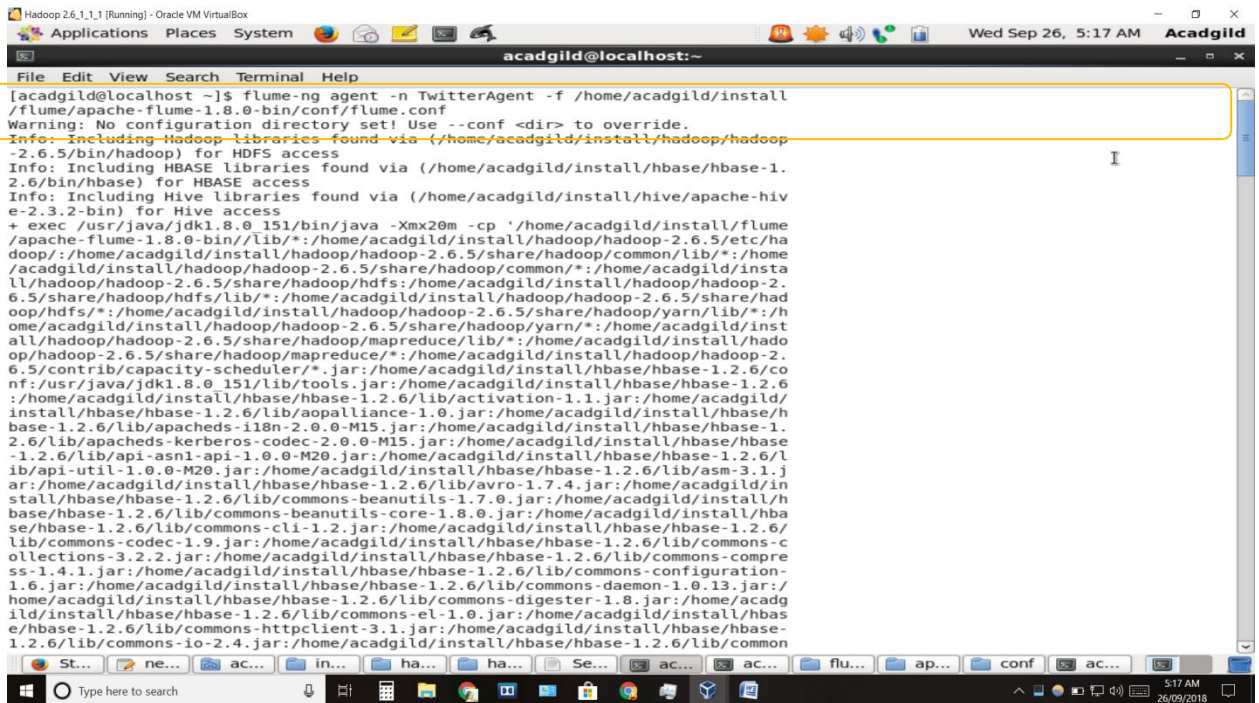
5. Edit bash file



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

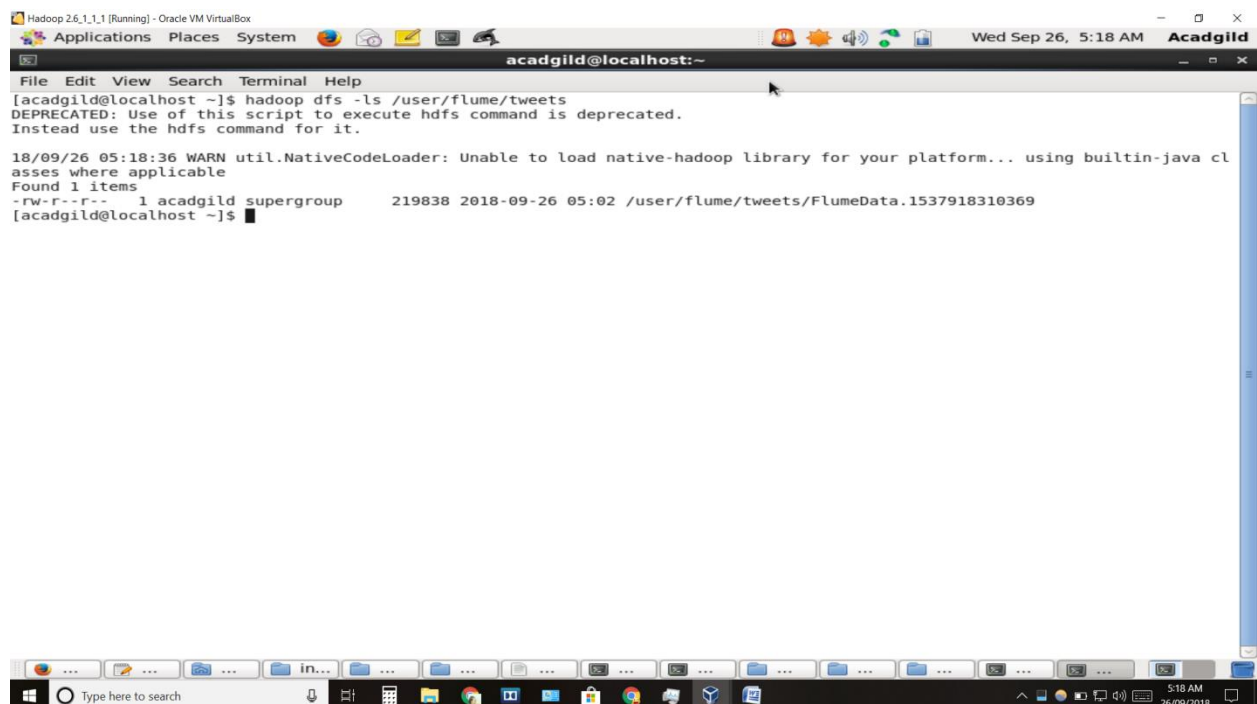


7. Fetch the twitter tweet data into the HDFS cluster path by flume



```
acadgild@localhost:~$ flume-ng agent -n TwitterAgent -f /home/acadgild/install/flume/apache-flume-1.8.0-bin/conf/flume.conf
Warning: No configuration directory set! Use --conf <dir> to override.
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 -Xmx20m -cp '/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/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-1.2.0-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/commons-daemon-1.0.13.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-digester-1.8.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-el-1.0.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-httpclient-3.1.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/commons-io-2.4.jar:/home/acadgild/install/hbase/hbase-1.2.6/lib/common
```

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



```
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 classes 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


```
Hadoop 2.6.1_1.1 (Running) - Oracle VM VirtualBox
Applications Places System
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

```
Hadoop 2.6.1_1.1 (Running) - Oracle VM VirtualBox
Applications Places System
acadgild@localhost:~
File Edit View Search Terminal Help
i'm funny
i'm ugly

you'll always be laughing at something<[REDACTED] href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>[REDACTED]
{"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":"use
r_statuses_count","type":["int","null"]},{"name":"user_followers_count","type":["int","null"]},{"name":"user_name","type":["s
tring","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 statu
s_id","type":["long","null"]},{"name":"media url https","type":["string","null"]},{"name":"expanded_url","type":["string","nu
ll"]}]}}[REDACTED]
056666e176[REDACTED]10447312702851850246[REDACTED]lgerierالجزائر لوسبر دم حم لالا انا هلا[REDACTED]نريوما غريما[REDACTED]n8H2NCwebbKhNA(2018-09-
26T05:01:59Z[REDACTED]RT @elmoder1980: يشغلها بةتيغ وي جيتو غم شة انا
يُقَالُ اِي فِي قُرْدَا لَ تَقُولُوا اَنَا
يُقَالُ اِي فِي قُرْدَا لَ تَقُولُوا اَنَا
<[REDACTED] href="http://twitter.com/download/android" rel="nofollow">Twitter for Android</a>[REDACTED]10447312702726
184966[REDACTED]
MARS[REDACTED]India013(2018-09-26T05:01:59Z[REDACTED]RT @SenBillNelson: Los venezolanos siguen sufriendo a manos del régimen corrup
to de Maduro. Insto a mis colegas a que aprueben nuestro proy[REDACTED] href="https://mobile.twitter.com" rel="nofollow">Twitter
Lite</a>[REDACTED]104473127028922776[REDACTED]A6Republic of the PhilippinesJtoo much wondering makes me depressed[REDACTED]psukhe[REDACTED]sychedelic3a
m(2018-09-26T05:01:59ZVRT @TheDailyWords : https://t.co/PlwtJlf3n5[REDACTED] href="https://about.twitter.com/products/tweetdeck"
rel="nofollow">TweetDeck</a>[REDACTED]https://pbs.twimg.com/media/Dnm7x3hU4AAuo_p.jpg[REDACTED]https://twitter.com/TheDailyWords /status/1043
071267861975940/photo/161044731270272634880[REDACTED]West Virginia, USA I fly a starship[REDACTED]anner Staten[REDACTED]teaksauccyyy(2018-09-26T05
:01:59Z-I'll tell you something you may want to hear, but I'll lie.[REDACTED] href="http://twitter.com/download/iphone" rel="nofol
low">Twitter for iPhone</a>[REDACTED]10447312702767472646[REDACTED]albania_villalbaalbav31(2018-09-26T05:01:59Z[REDACTED]RT @GOPChairwoman:
Beto O'Rourke wants to:
* abolish ICE & decriminalize unauthorized border crossings
* have govt take over health care
*-[REDACTED] href="http://twitter.com/download/android" rel="nofollow">Twitter for Android</a>[REDACTED]1044731270276816896[REDACTED]
únior[REDACTED]unior0lin(2018-09-26T05:01:59ZRT @PabloNews #MelhorClipeTVZIndestrutivel[REDACTED] href="http://twitter.com/download/
android" rel="nofollow">Twitter for Android</a>[REDACTED]10447312702976081936[REDACTED]
東京[REDACTED]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」@Utah[REDACTED]Kentaro HAYASHI[REDACTED]Kent Hayashi
(2018-09-26T05:01:59ZRT @ICRRPr: 東京宇宙線研究所は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
```