TASK

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

To stream data to our database from twitter we should have the following pre-requisites.

- Twitter account
- Hadoop cluster

Since, both prerequisites are available so moving to following steps:

Step 1: Created Twitter App by logging into my twitter account and from there I copied below keys:

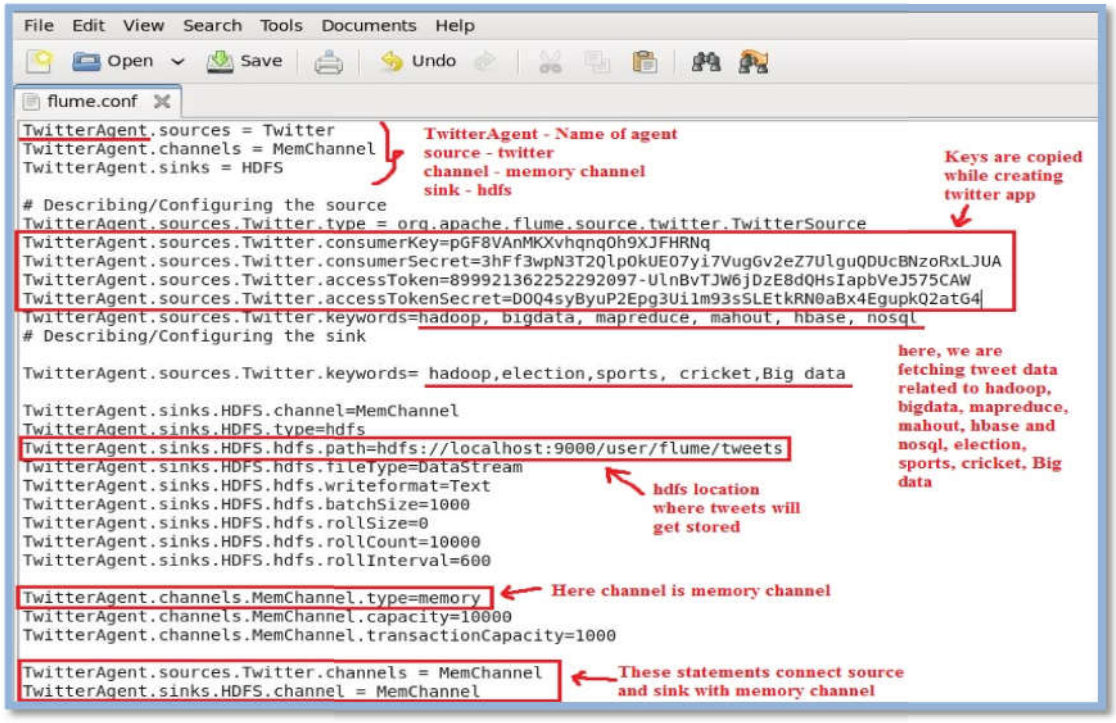
Cosumer Key (API Key) = pGF8VAnMKXvhqnqOh9XJFHRNq
Consumer Secret (API Secret) = 3hFf3wpN3T2QlpOkUEO7yi7VugGv2eZ7UlguQDUcBNzoRxLJUA
Access Token = 899921362252292097-UlnBvTJW6jDzE8dQHsIapbVeJ575CAW
Access Token Secret = DOQ4syByuP2Epg3Ui1m93sSLEtkRN0aBx4EgupkQ2atG4

Step 2: Since here in this task we have to stream and analyse twitter data, therefore below jar files are checked inside /$FLUME_HOME/lib [i.e. /usr/local/flume/lib] directory:

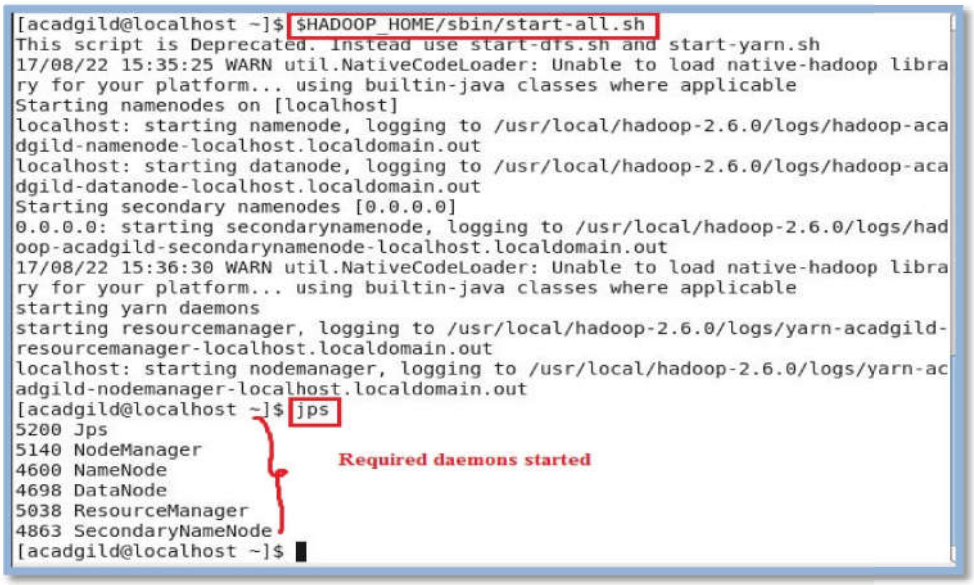
```
[acadgild@localhost lib]$ ls -lrt t*.jar
-rw-r--r--. 1 acadgild acadgild  56307 Aug 23  2014 twitter4j-stream-3.0.3.jar
-rw-r--r--. 1 acadgild acadgild 284077 Aug 23  2014 twitter4j-core-3.0.3.jar
-rw-r--r--. 1 acadgild acadgild  27698 Aug 26  2014 twitter4j-media-support-3.0.
3.jar
```

Above screenshot shows that all three required jars are present inside specified directory.

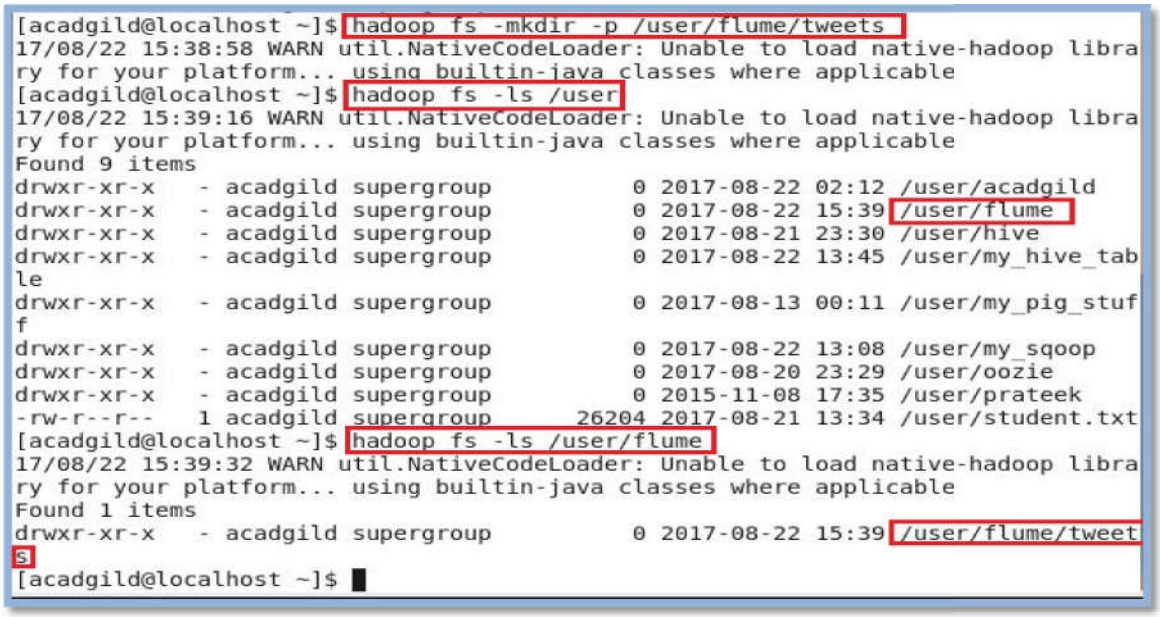
Step 3: Created flume.conf file [using gedit flume.conf command] inside /usr/local/flume/conf with above specified twitter app keys:



Step 4: In new terminal started all the Hadoop daemons before running the flume command to fetch the twitter data. Using 'jps' command checked whether all required daemons have started or not.



Step 5: Created new directory inside HDFS path, where the Twitter's tweet data would be stored:



Step 6: Used below command to fetch the twitter's tweet data into the HDFS cluster path:

The above command will start fetching data from Twitter and steams it into the HDFS given path.

```
17/08/22 15:54:00 INFO twitter4j.TwitterStreamImpl: Waiting for 250 milliseconds
17/08/22 15:54:01 INFO twitter4j.TwitterStreamImpl: Establishing connection.
17/08/22 15:54:03 INFO twitter4j.TwitterStreamImpl: Connection established.
17/08/22 15:54:03 INFO twitter4j.TwitterStreamImpl: Receiving status stream.
17/08/22 15:54:07 INFO twitter.TwitterSource: Processed 800 docs
17/08/22 15:54:12 INFO twitter.TwitterSource: Processed 900 docs
17/08/22 15:54:16 INFO twitter.TwitterSource: Processed 1,000 docs
17/08/22 15:54:16 INFO twitter.TwitterSource: Total docs indexed: 1,000, total s
kipped docs: 0
17/08/22 15:54:16 INFO twitter.TwitterSource:     19 docs/second
17/08/22 15:54:16 INFO twitter.TwitterSource: Run took 52 seconds and processed:
17/08/22 15:54:16 INFO twitter.TwitterSource:     0.005 MB/sec sent to index
17/08/22 15:54:16 INFO twitter.TwitterSource:     0.262 MB text sent to index
17/08/22 15:54:16 INFO twitter.TwitterSource: There were 0 exceptions ignored:
17/08/22 15:54:23 INFO twitter.TwitterSource: Processed 1,100 docs
```

Once, the tweet data started streaming into the given HDFS path, used 'Ctrl+c' command to stop the streaming process.

Step 7: To check whether file containing tweet data got created successfully inside hdfs or not, used the following command:

```
[acadgild@localhost ~]$ hadoop fs -ls /user/flume/tweets
17/08/22 15:46:37 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
Found 1 items
-rw-r--r--   1 acadgild supergroup    1815770 2017-08-22 15:46 /user/flume/tweet
s/FlumeData.1503396832990
[acadgild@localhost ~]$
0
```

Twitter data got stored inside this file

Step 8: Below cat command displays the tweet data inside the file created in above step:

```
[acadgild@localhost ~]$ hadoop fs -cat /user/flume/tweets/FlumeData.150339683299
0
```

Tweet Data

```
しかし直後、１人の女性が全員を殺害！

⇒https://t.co/q3BE2shsVQ

地獄へ突き落とした理由がヤバイ…
https://t.co/vsPSX2Z641a href="https://twitter.com/" rel="nofollow">最新の話題
 Tweet</a>https://pbs.twimg.com/media/DHzKyvXVoAEAgdd.jpg https://twitter.com
/livenewstweet/status/899817008417562624/photo/1$899938208720531457 3
                                                                FranceB#Des
perateHousewives #PrisonBreak 7 Gueule d'ange Katris_Prideen(2017-08-22T15:46:03
Z RT @proportionnelle: Sérieusement @AUCHAN_France ? https://t.co/9ubHUtA2wk
a href="http://twitter.com/download/android" rel="nofollow">Twitter for Androi
d</a>https://pbs.twimg.com/media/DHxJsqMXUAEfVAB.jpg https://twitter.com/prop
ortionnelle/status/899675187737710593/photo/1 9'+J]
[acadgild@localhost ~]$
```