**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:

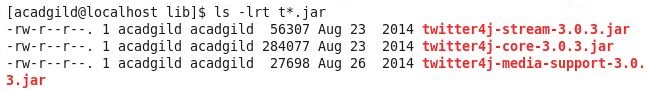
**Step 1: Created Twitter App by logging into my twitter account and from there I copied below keys:**

**Cosumer Key (API Key) = pGF8VAnMKXvhqnqOh9XJFHRNq**

**Consumer Secret (API Secret) = 3hFf3wpN3T2QlpOkUEO7yi7VugGv2eZ7UlguQDUcBNzoRxLJUA**

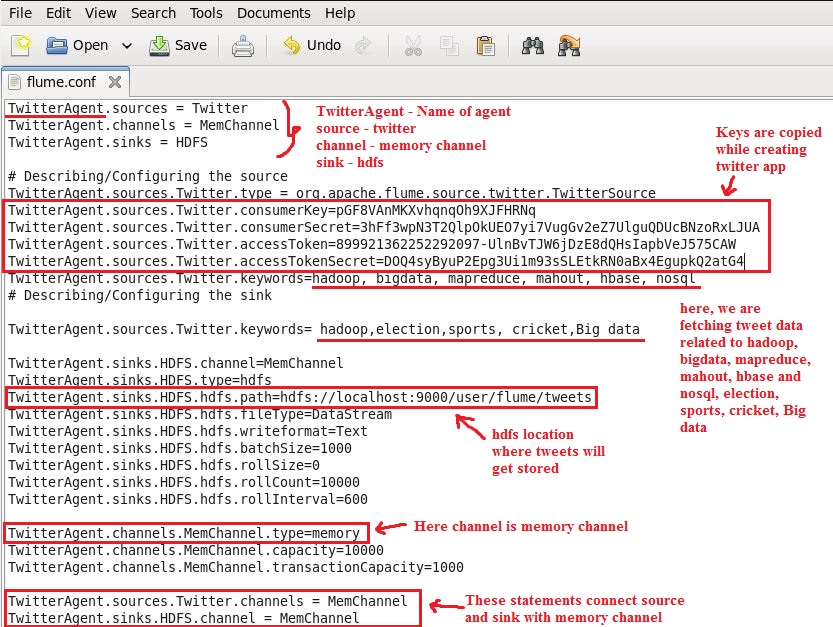
**Access Token = 899921362252292097-UlnBvTJW6jDzE8dQHsIapbVeJ575CAW**

**Access Token Secret = DOQ4syByuP2Epg3Ui1m93sSLEtkRN0aBx4EgupkQ2atG4**

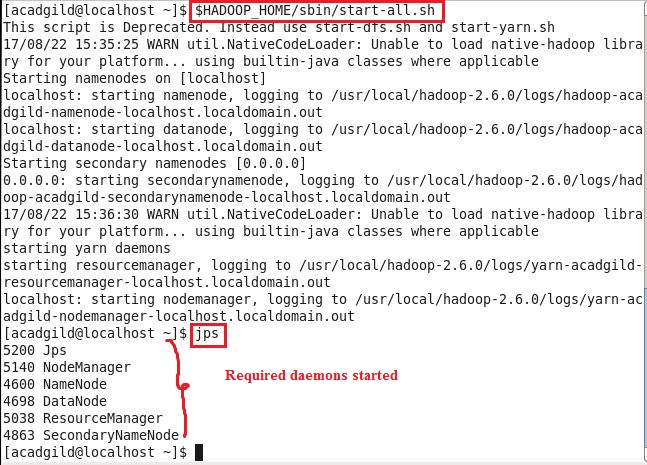
**Step 2: Since here in this task we have to stream and analyse twitter data, therefore below jar files are checked inside /$FLUME\_HOME/lib [i.e. /usr/local/flume/lib] directory:**

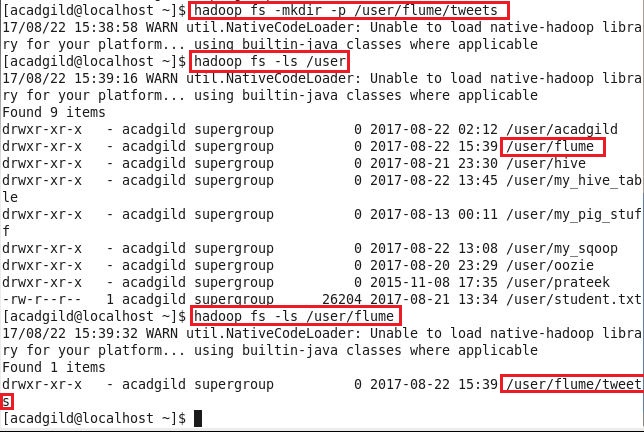
**Above screenshot shows that all three required jars are present inside specified directory.**

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

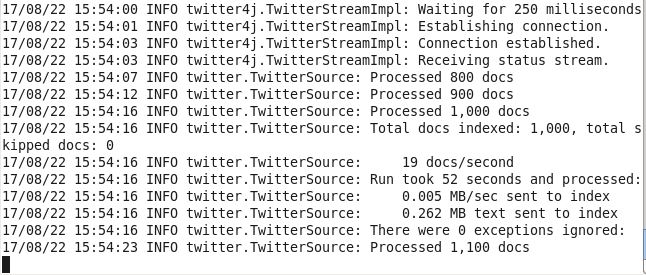
****

**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.**

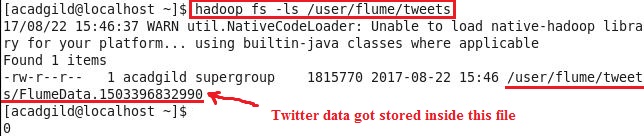
****

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

**G:\ACADGILD\course material\Hadoop\Sessions\Session 11\Assignments\Assignment3\Screenshots\4.1.JPGStep 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.**

**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:**

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

**G:\ACADGILD\course material\Hadoop\Sessions\Session 11\Assignments\Assignment3\Screenshots\d.JPG**

****