*Project 1.3 – Twitter Sentiment Analysis*

*By: Krushna Chandra Debashram*

*02-Aug-2017*

**Introduction:**

Stream data from Twitter and you need to perform sentiment analysis on the data.

**Objective:**

To get sentiment from twitter data of key words like 'hadoop, bigdata, mapreduce, mahout, hbase , nosql'.

**Program used:**

Pig

Flume

**Prerequisites:**

Hadoop cluster installed in the system.

Flume

Twitter Account

**Twitter Account and tokens:**

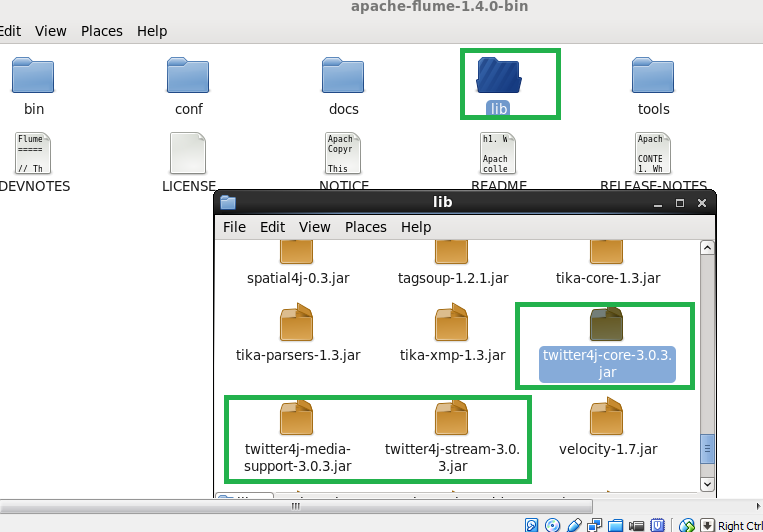
Create an account in twitter and application. Get the access tokens

|  |
| --- |
| Consumer Key (API Key) MIf7c7nIMAtZIDCBERDRq75Vp  Consumer Secret (API Secret) LbTLJW0ohv6Zx9Wk0HBfnfp3kzGW2JimSD4ZWHRK80RTqWiGXP  Access Token 880768243526270976-gDHpiQUb5EE1vw7rO4Dxaor9dg0d1TG  Access Token Secret ptPpGqjgJQhjNHu7xgKGHGq0CxSPn1O6lGA2GD11zFhnn |

**Twitter jar files:**

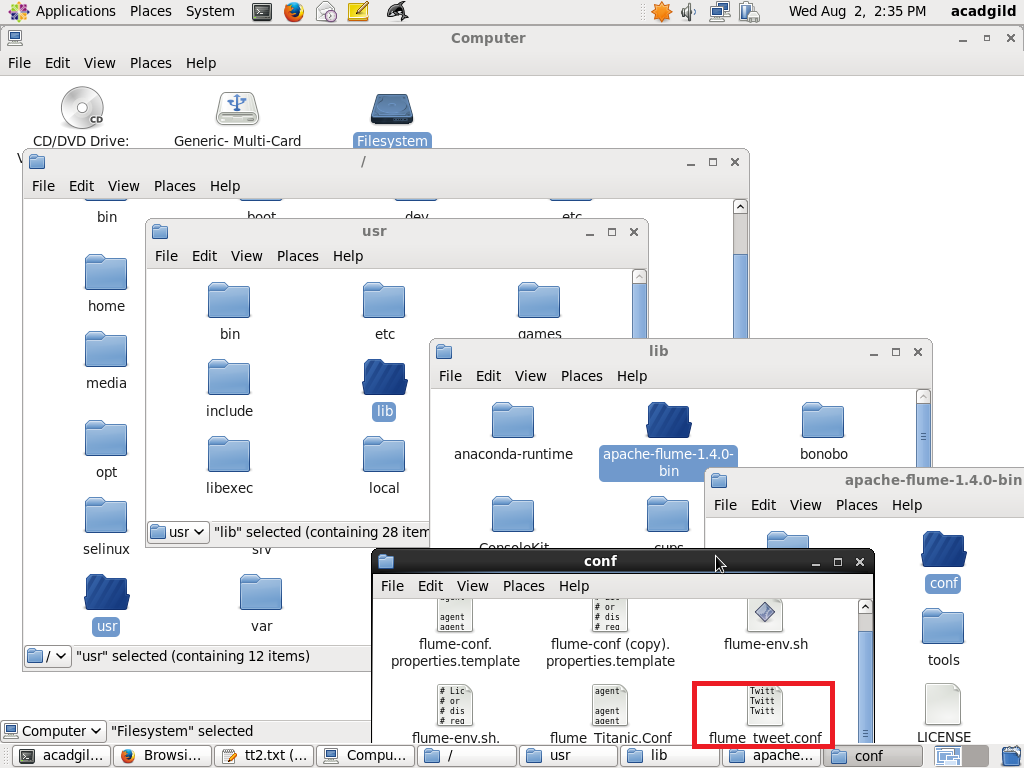
Put the twitter jar files in the folder

/usr/lib/apache-flume-1.4.0-bin/lib



**Twitter Flume configuration file:**

Create flume configuration file ‘flume\_tweet.conf’ in the folder /usr/lib/apache-flume-1.4.0-bin/lib



File:flume\_tweet.conf

|  |
| --- |
| 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=MIf7c7nIMAtZIDCBERDRq75Vp  TwitterAgent.sources.Twitter.consumerSecret=LbTLJW0ohv6Zx9Wk0HBfnfp3kzGW2JimSD4ZWHRK80RTqWiGXP  TwitterAgent.sources.Twitter.accessToken=880768243526270976-gDHpiQUb5EE1vw7rO4Dxaor9dg0d1TG  TwitterAgent.sources.Twitter.accessTokenSecret=ptPpGqjgJQhjNHu7xgKGHGq0CxSPn1O6lGA2GD11zFhnn  TwitterAgent.sources.Twitter.keywords=hadoop, bigdata, mapreduce, mahout, hbase, nosql  TwitterAgent.sources.Twitter.language=en  # Describing/Configuring the sink  TwitterAgent.sinks.HDFS.channel=MemChannel  TwitterAgent.sinks.HDFS.type=hdfs  TwitterAgent.sinks.HDFS.hdfs.path=/user/flume\_krushnadebashram/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 |

**Start Hadoop daemon:**

**[**acadgild@localhost ~]$ start-all.sh

[acadgild@localhost ~]$ mr-jobhistory-daemon.sh start historyserver

[acadgild@localhost ~]$ jps

3122 ResourceManager

3301 Jps

3669 JobHistoryServer

2966 SecondaryNameNode

3256 NodeManager

2733 NameNode

2830 DataNode

**Create a directory in Hadoop:**

Create a directory in hadoop ‘/user/flume\_krushnadebashram/tweets ‘ where the twitter data to store using flume.

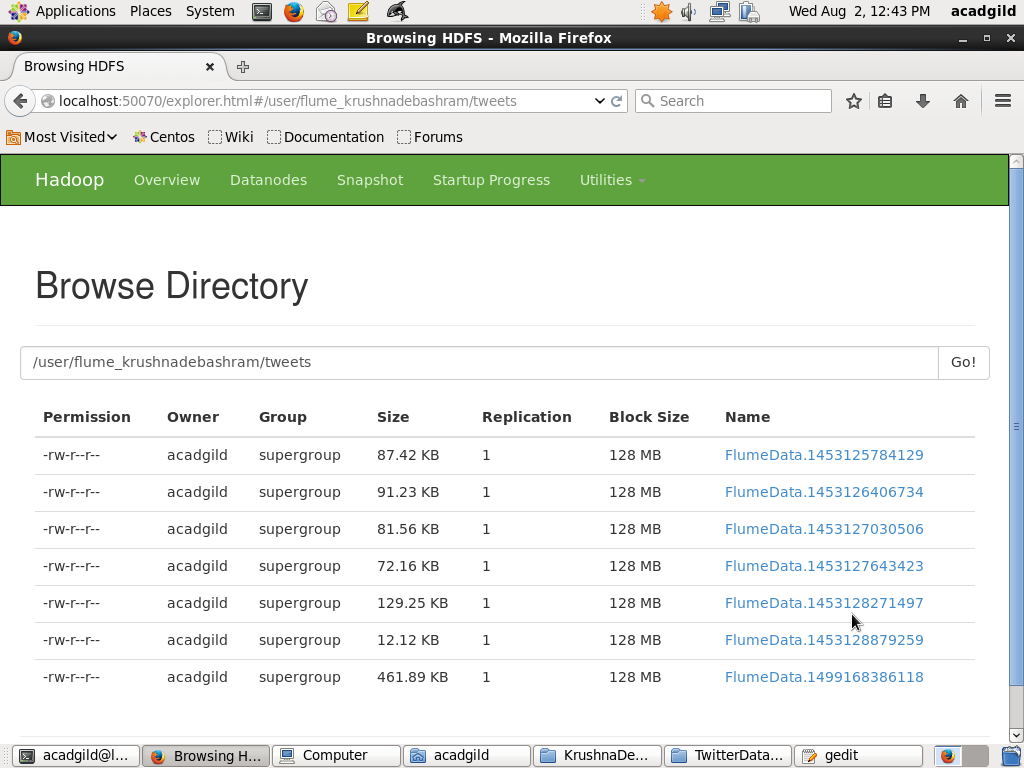
Hadoop dfs –mkdir /user/flume\_krushnadebashram/tweets

This path is in the flume\_twitter.conf

**Fetch twitter data:**

Create a directory in hadoop ‘/user/flume\_krushnadebashram/tweets ‘ where the twitter data to store using flume.

[acadgild@localhost ~]$ flume-ng agent -n TwitterAgent -f /usr/lib/apache-flume-1.4.0-bin/conf/flume\_tweet.conf;



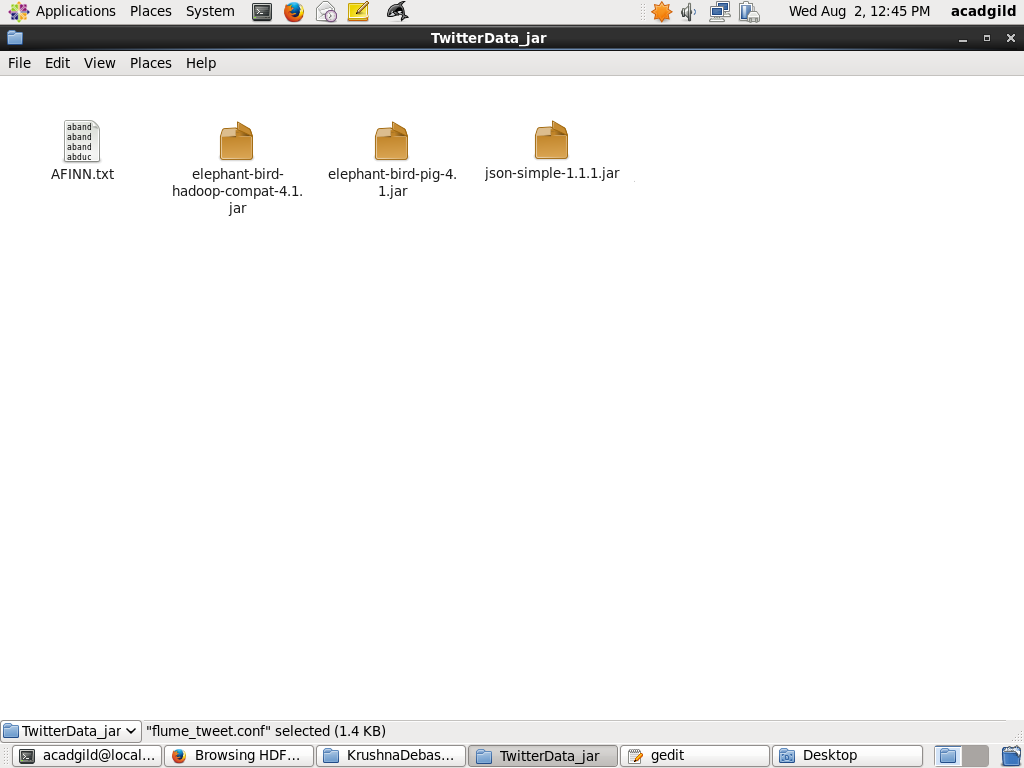
**Jars for JSON loader:**

Put the files

home/acadgild/KrushnaDebashram/TwitterData\_jar/elephant-bird-hadoop-compat-4.1.jar

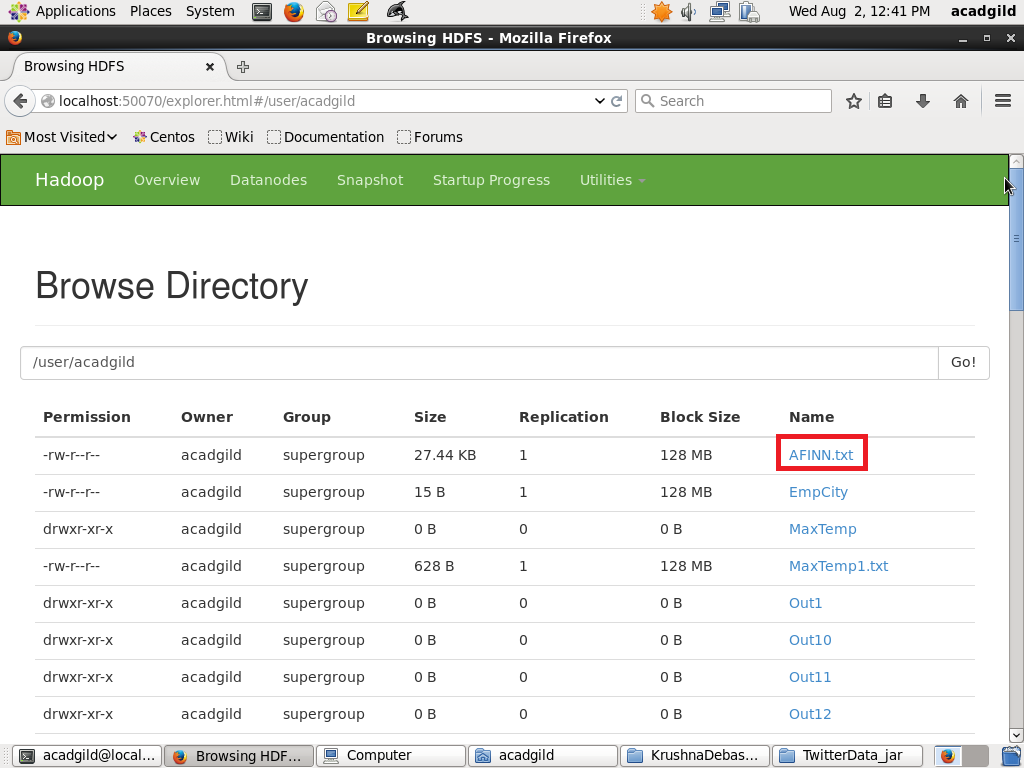
home/acadgild/KrushnaDebashram/TwitterData\_jar/elephant-bird-pig-4.1.jar

/home/acadgild/KrushnaDebashram/TwitterData\_jar/json-simple-1.1.1.jar



**Upload AFFIN.txt in Hadoop:**

hadoop dfs -put /home/acadgild/KrushnaDebashram/TwitterData\_jar/AFINN.txt



**Go to pig shell:**

Go to pig shell in mapreduce mode

[acadgild@localhost ~]$ pig -x mapreduce

grunt>

**Register the downloaded jars in pig:**

|  |
| --- |
| grunt>REGISTER '/home/acadgild/KrushnaDebashram/TwitterData\_jar/elephant-bird-hadoop-compat-4.1.jar';  grunt>REGISTER '/home/acadgild/KrushnaDebashram/TwitterData\_jar/elephant-bird-pig-4.1.jar';  grunt>REGISTER '/home/acadgild/KrushnaDebashram/TwitterData\_jar/json-simple-1.1.1.jar'; |

**Pig script to perform Sentiment Analysis:**

|  |
| --- |
| load\_tweets = LOAD '/user/flume\_krushnadebashram/tweets/F\*' USING com.twitter.elephantbird.pig.load.JsonLoader('-nestedLoad') AS myMap;  extract\_details = FOREACH load\_tweets GENERATE myMap#'id' as id,myMap#'text' as text;  tokens = foreach extract\_details generate id,text, FLATTEN(TOKENIZE(text)) As word;  dictionary = load '/user/acadgild/AFINN.txt' using PigStorage('\t') AS(word:chararray,rating:int);  dump dictionary;  word\_rating = join tokens by word left outer, dictionary by word using 'replicated';  describe word\_rating;  rating = foreach word\_rating generate tokens::id as id,tokens::text as text, dictionary::rating as rate;  word\_group = group rating by (id,text);  avg\_rate = foreach word\_group generate group, AVG(rating.rate) as tweet\_rating;  dump avg\_rate;  positive\_tweets = filter avg\_rate by tweet\_rating>=0;  dump positive\_tweets; |

**Results:**

**Average**: in the attached file AverageTweets.txt.

Sample result as below

|  |
| --- |
| ((689088744444563456,Urgent Need: Hadoop Developer\_Sunnyvale, CA\_6+ Months https://t.co/3Lii4uE0EV Need: Hadoop Developer\_Sunnyvale, CA\_6+ Months),)  ((689088825897914368,Urgent Need: Hadoop Developer\_Sunnyvale, CA\_6+ Months https://t.co/wY7I6SbDh4 Need: Hadoop Developer\_Sunnyvale, CA\_6+ Months),)  ((689088827890208770,Which do you choose for a #Data lake? #Hadoop or an in-memory database? https://t.co/TSei1vcDoz),)  ((689088999214960640,Urgent Need: Hadoop Developer\_Sunnyvale, CA\_6+ Months https://t.co/tql29fgfOJ Need: Hadoop Developer\_Sunnyvale, CA\_6+ Months),)  ((689089406528040960,Blend, munge, and prep your #Hadoop #data faster w/ #spark and #impala https://t.co/bC5grOUci7 https://t.co/EXhBBCN45k),)  ((689089513793163265,https://t.co/puba3rJPht - toucharcade-forums - #ndclondon #arcade #gamesjobs #aapp #hadoop #opensource #animation https://t.co/oXq8cSK4oO),)  ((689089634723319811,Urgent Need: Hadoop Developer\_Sunnyvale, CA\_6+ Months https://t.co/6mO2IzA30B Need: Hadoop Developer\_Sunnyvale, CA\_6+ Months),)  ((689089657972363264,#MO Java Developer - Hadoop Spark Python R Java: London-London, Java Developer - Hadoop Spark Python R Locatio... https://t.co/khfi8SWIok),)  ((689089663269777410,Data Scientist - Python, Hadoop, R, Hive, Belfast - https://t.co/bh9Gic7uYU),)  ((689089837513707525,Apply Today: Senior Manager-Hadoop Development Needed in Silicon Valley CA in San Jose, CA https://t.co/8MM4vayerq #job),)  ((689089909542531072,RT @rick\_vanderlans: If you want to have a simple explanation of #HDFS works and you like cartoons, this is the place: https://t.co/WQqkx7Q…),1.5)  ((689089951296827392,RT @rick\_vanderlans: If you want to have a simple explanation of #HDFS works and you like cartoons, this is the place: https://t.co/WQqkx7Q…),1.5)  ((689089952232173568,RT @kristinevick: Got #bigdata? Manage it on your own terms with #Hadoop and @SASDataMGMT https://t.co/n4xFS7UaEM https://t.co/OM9f1z2z5c),)  ((689089980459790336,We are hiring: AWS Kumo Software Developer - Knowledge Pla https://t.co/piws8wSsxv #job @awscloud #java #aws #machine\_learning #hadoop),)  ((689090036877361152,https://t.co/X86NbDAkrJ - toucharcade-forums - #ndclondon #arcade #gamesjobs #aapp #hadoop #opensource #animation https://t.co/tS2i8uUjbk…),)  ((689090160164782080,RT @AvanteTIC: Ethical hacking, Hadoop, business intelligence...consulta nuestras convocatorias iClass, online en directo https://t.co/rTPn…),)  ((689090221691006981,@FXS\_Finance\_HU - Look forward to staying in touch! Do you use #Hadoop for your #data needs?),) |

**PositiveTweet**: in the attached file PositiveTweets.txt.

Sample result as below

|  |
| --- |
| ((689085590822891521,@filmfan hey its time for you guys follow @acadgild To #AchieveMore and participate in contest Win Rs.500 worth vouchers),2.0)  ((689085611639205888,@sujitjohn Follow @acadgild To #AchieveMore &amp; participate in contest Hurry up! &amp; win Flipkart vouchers),4.0)  ((689085636456923138,@sujitlalwani Hey tweeps want to win Flipkart vouchers so follow @acadgild To #AchieveMore contest Prizes of Rs.500),2.5)  ((689085663334035456,@im\_bicky Hey Friends &amp; Tweeples Please Follow @acadgild To #AchieveMore and participate in contest Be lucky get Rs.500 vouchers),3.0)  ((689085686390063104,@ShreyVithalani You don't wanna miss this Go follow @acadgild To #AchieveMore and play in contest &amp; win Flipkart vouchers),1.0)  ((689085710079537154,@SANGEETAAGRAWA Tweethearts! Follow @acadgild To #AchieveMore and participate in contest and win Flipkart vouchers! Aye!),4.0)  ((689085776731213824,@AryanSarath Contestest freaks! Follow @acadgild To #AchieveMore and play in contest &amp; win shopping vouchers Wohooo!),4.0)  ((689085820674945025,@RaviThakkar Its time to follow @acadgild To #AchieveMore and participate in contest You can win shopping vouchers),4.0)  ((689085845538762752,@iwant\_us Follow @acadgild To #AchieveMore and participate in contest &amp; Win Flipkart vouchers! Win Rs.500 worth vouchers),2.0) |