**Ali Shahzad**

**L1f18bscs0334**

**Section d3**

Analyzing Twitter Data using Appache Flume

FLUME:

Appache flume is highly reliable distributed service for

Collecting and transporting large ammount of

Streaming data from various sources to a centarlized data storage HDFS .

WE wil fetch Twitter Data using Twitter Api and Appache flume .

Requirements:

1: java (jdk)

2: Hadoop

3: Flume

4: Twitter Api

Step 1:

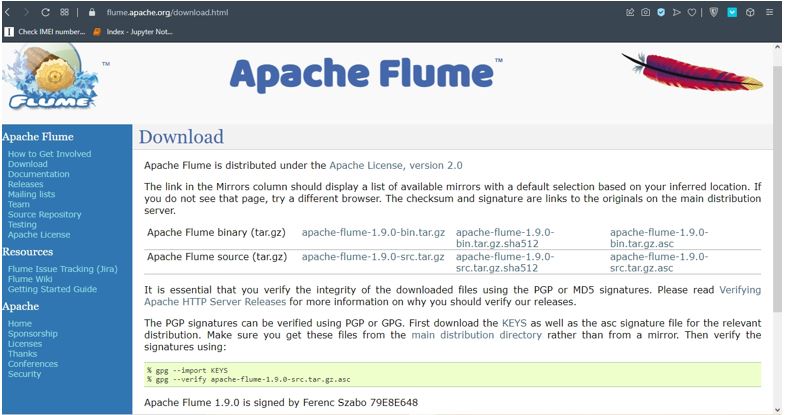
FLUME installation

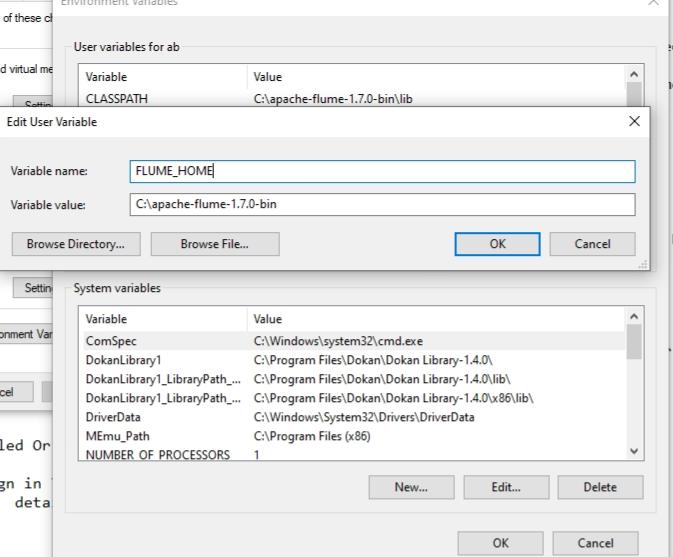
We need to install and configure Flume , Hadoop and JAVA (jdk) ∷

A: Download Flume From Appache and extract it to C Drive ∷

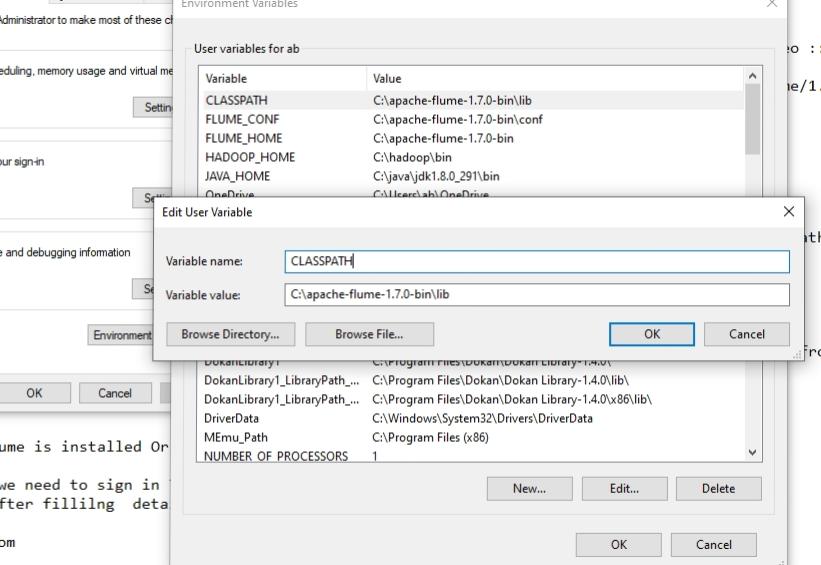
<http://www.apache.org/dyn/closer.lua/flume/1.9.0/apache-flume-1.9.0-bin.tar.gz>

b: Now set The Environment variable for flume name As FLUME\_HOME

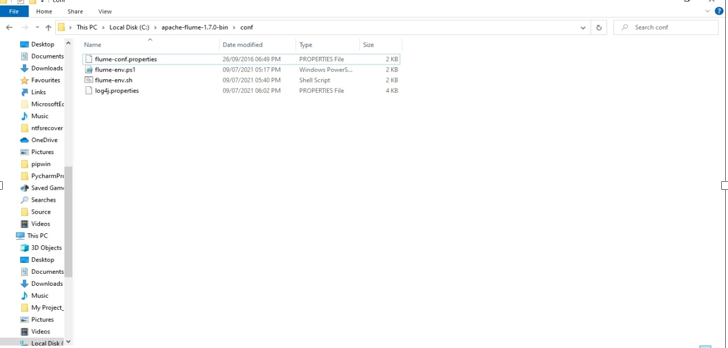




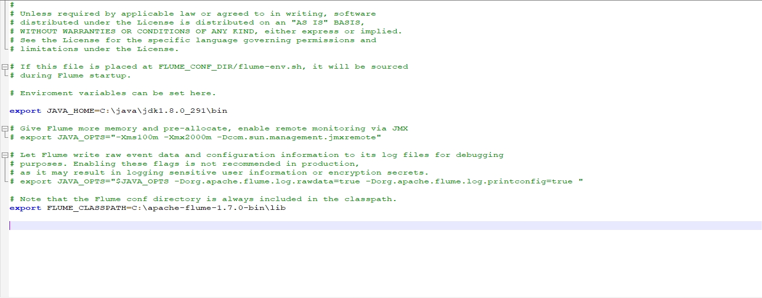
: and set env var for CLASSPATH which is the flume lib folder path .



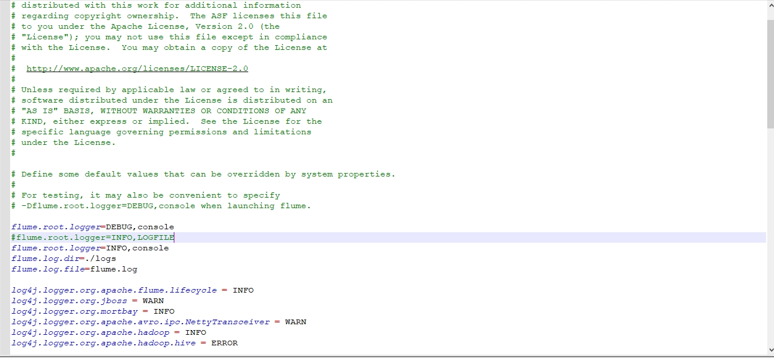
c: Open Conf folder Of and remove template from 3 files ( flume-conf.properties.template, flume-env.ps1.template ,flume-env.sh.template)



D: open flume-env.sh from flume conf folder and paste here java jdk path and classpath which we set earlier. You can copy path from environment var or from location in C drive



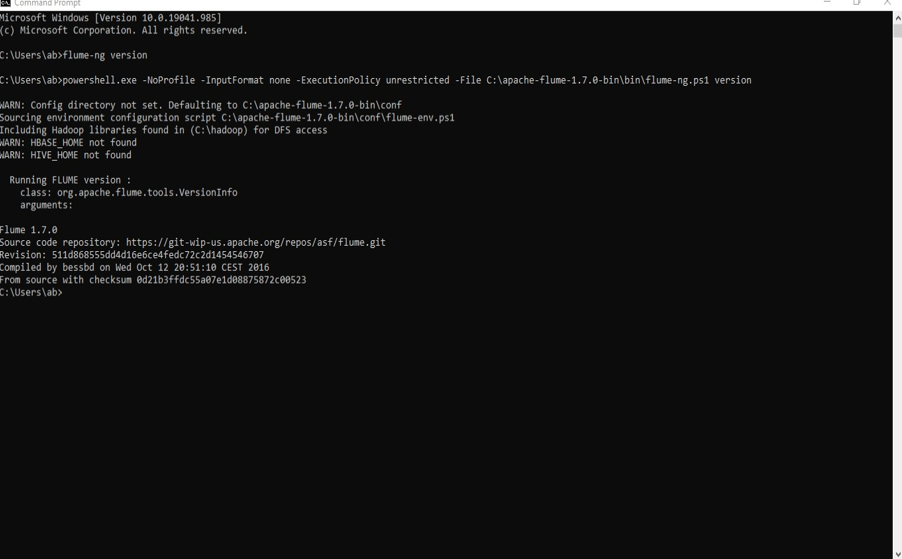
E: Open file log4j.properties in flume conf folder and uncomment the line flume.root.logger=DEBUG,console.. Then comment the next line which is #flume.root.logger=INFO,LOGFILE save it .



F: Now open cmd and type flume-ng version if there is no error Then we successfully installed the Appache Flume

( Note: “ if you get an error msg on line 189 then Open flume-env.ps1 from bin folder in Flume go to line169 .. Here remove the javahome path and add the java bin folder path you can copy it from C drive java bin folder

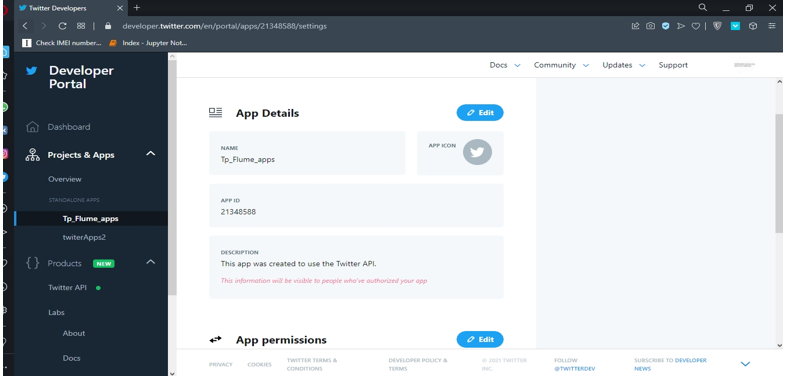
After edit line169: \_\_return “$env:C:\java\bin\java.exe”\_\_ “ … )

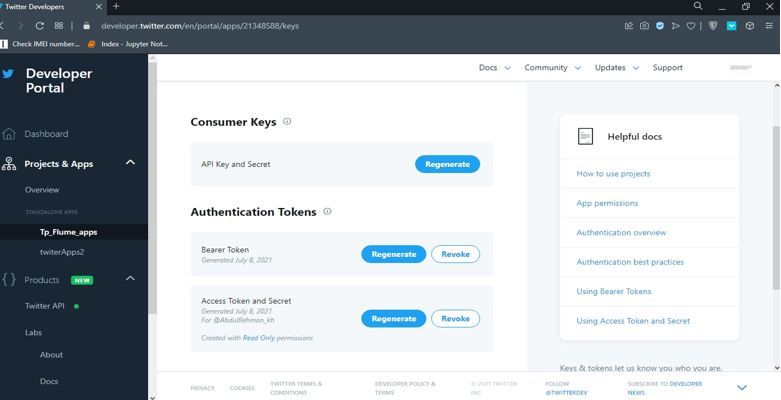


Step 2:

Create Twitter Developer Account <https://apps.twitter.com> and then click On create app Fill the Details

After that copy the keys,Secreatkeys,token, and secreatoken .

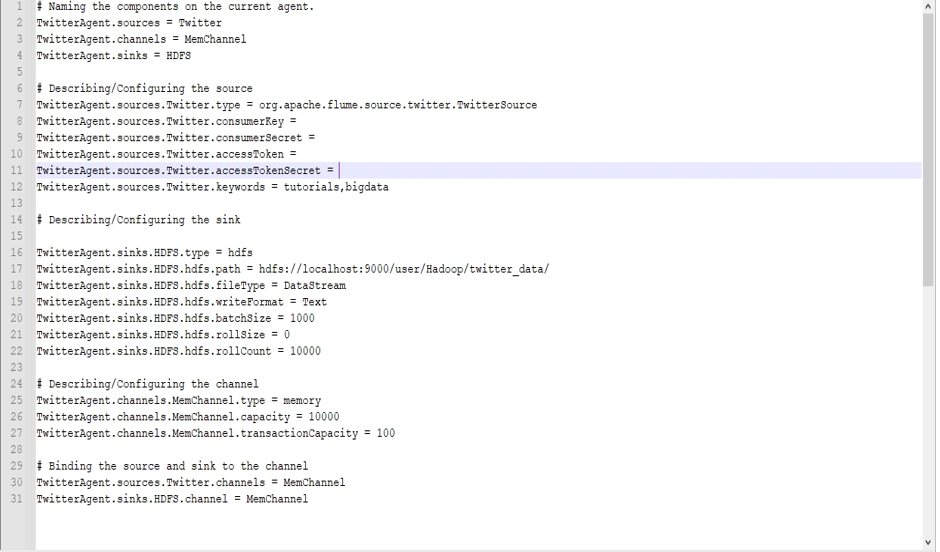




Step 3:

Now We have to add file in conf folder of Flume name as twitter.conf… You can type these lines in notepad and save as .conf in flume conf folder or can copy from internet

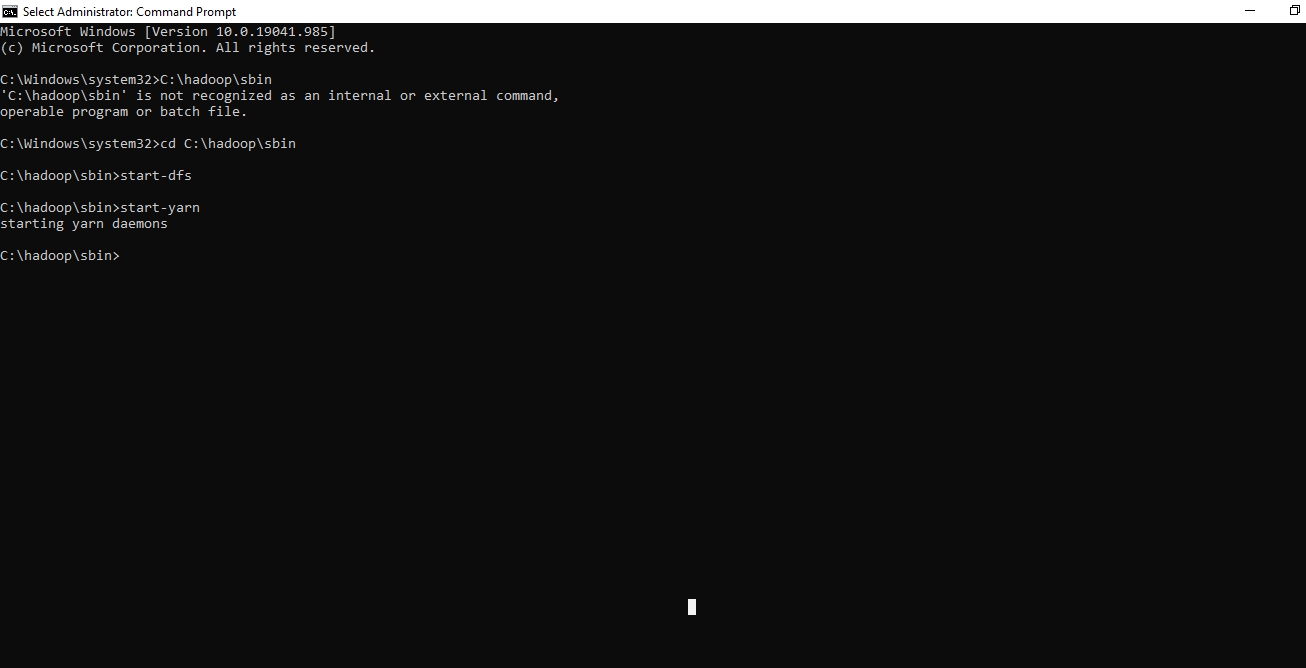
Note : “In line 8, 9, 10, 11, Here you have to add your api key ,secreatkey, token ,secreattoken Which you generated from twitter app created earlier”



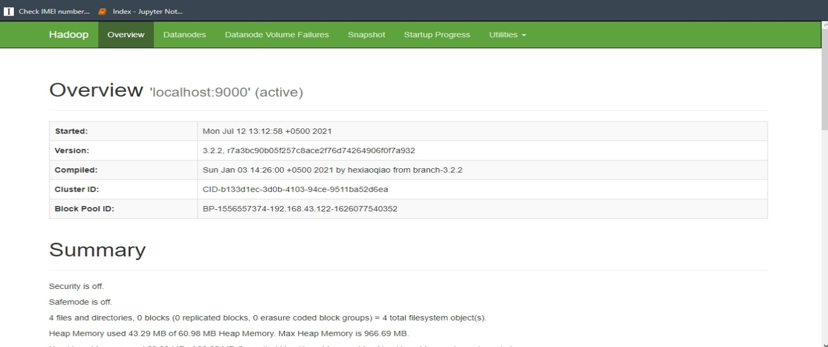
Step 4:

Start Hadoop

A: Open cmd and go to Hadoop sbin folder and type start-dfs enter then type and start-yarn



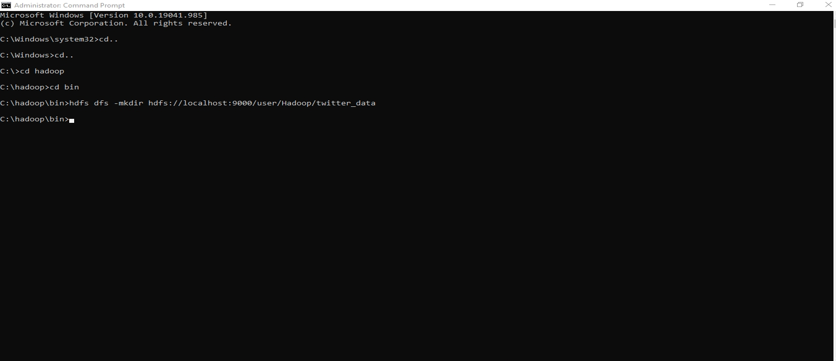
You can check if the hadoop service is started open Browser and type https:// localhost:9870



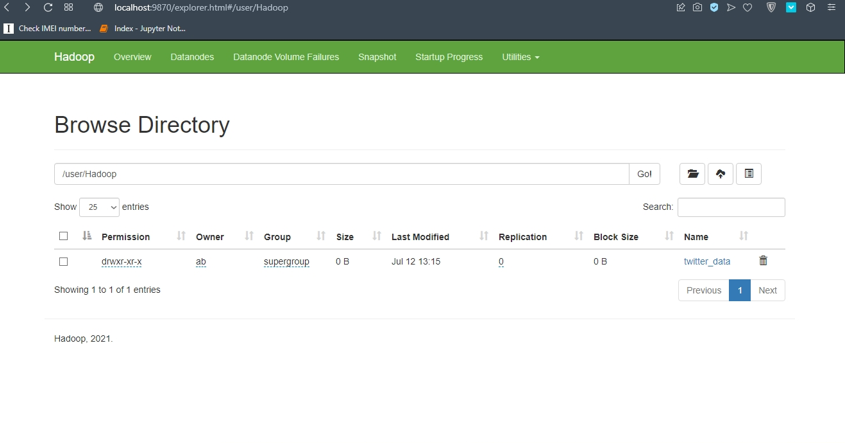
B: In Hadoop dfs we can create Directory using mkdir command

Open new cmd and Go to hadoop bin folder then type

Hdfs dfs -mkdir hdfs://localhost:9000/user/Hadoop/twitter\_data If there is no error then we have created dir in hadoop dfs



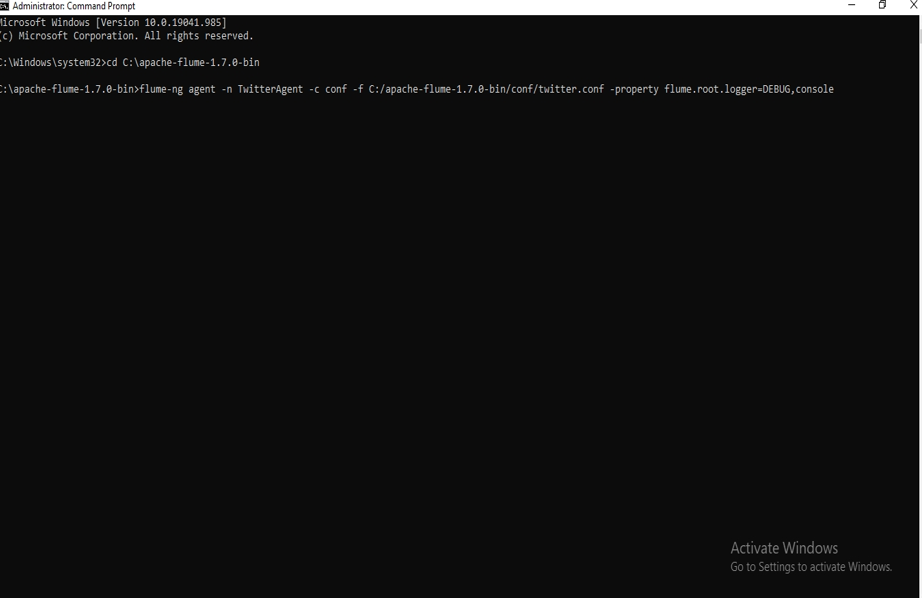
Step 5:



: open cmd go to flume folder and then type

Flume-ng agent -n TwitterAgent -c conf -f C:/apache-flume-1.7.0-bin/conf/twitter.conf -property flume.root.logger=DEBUG,consol

You have to change the path in this line C:/’ADD your PATH OF CONF FOLDER ‘/twitter.conf



If the process of fetching twittes into hdfs started then work is done.

Wait for few seconds and then terminate the process and check the data is saved in dir you created in HDFS.

