

Advanced Database Topics

Sentimental Analysis of Twitter Data using Hadoop

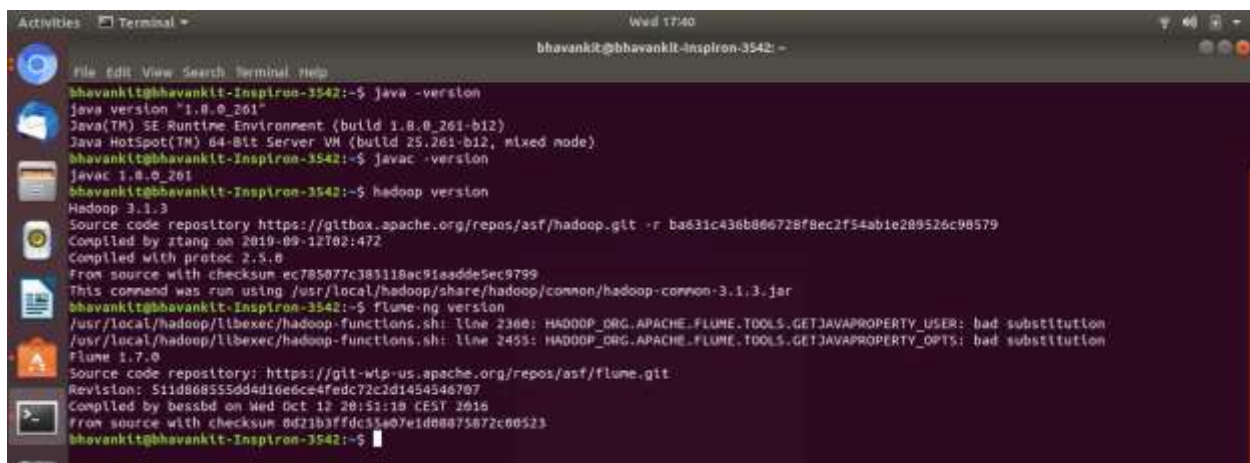
Bhavankit Navadiya (110008466) mansukh@uwindsor.ca Khushboo Ladha(110019963) ladhak@uwindsor.ca Shubham Singh(110019533) singh1zb@uwindsor.ca

Milestone II

Step 1:

During this step, we have configure Ubuntu 18.01 alongside Windows 10 to implement this project.

Step 2:



```

bhavankit@bhavankit-Inspiron-3542:~$ java -version
java version "1.8.0_261"
Java(TM) SE Runtime Environment (build 1.8.0_261-b12)
Java HotSpot(TM) 64-Bit Server VM (build 25.261-b12, mixed mode)
bhavankit@bhavankit-Inspiron-3542:~$ javac -version
javac 1.8.0_261
bhavankit@bhavankit-Inspiron-3542:~$ hadoop version
Hadoop 3.1.3
Source code repository https://gitbox.apache.org/repos/asf/hadoop.git -r ba631c436b866728f8ec2f54ab1e289526c98579
Compiled by ztang on 2019-09-12T02:47Z
Compiled with protoc 2.5.0
From source with checksum ec785077c383118ac91aadd5ec9799
This command was run using /usr/local/hadoop/share/hadoop/common/hadoop-common-3.1.3.jar
bhavankit@bhavankit-Inspiron-3542:~$ flume-ng version
/usr/local/hadoop/libexec/hadoop-functions.sh: line 2360: HADOOP_ORG.APACHE.FLUME.TOOLS.GETJAVAPROPERTY_USER: bad substitution
/usr/local/hadoop/libexec/hadoop-functions.sh: line 2455: HADOOP_ORG.APACHE.FLUME.TOOLS.GETJAVAPROPERTY_OPTS: bad substitution
Flume 1.7.0
Source code repository: https://git-wip-us.apache.org/repos/asf/flume.git
Revision: 51d868555dd4d1e0e4fedc72c2d1454546707
Compiled by bessbd on Wed Oct 12 20:53:10 CEST 2016
From source with checksum 0d21b3ffdc55a07e1d08075872c00523
bhavankit@bhavankit-Inspiron-3542:~$

```

Figure 1: Successful Installation of Java, Hadoop and Flume

In step 2, we have performed below operations:

- Firstly we configured the latest version of JDK, but after installation of Hadoop, we came to know that Hadoop is on working smoot with JDK versions <= 1.8, hence, you can find java –version as 1.8.0_261
- In this step we have installed latest version of Hadoop and run using ./start-all.sh but as mentioned in last point, nodemanager and resourcemanager not working properly so does localhost, hence changed the version of JDK.
- In final phase of step 2, we set up flume for live streaming where you can check two warning. That is not any problem but indicates that we did not create twitter.conf file as our Twitter API access application is under review.

Step 3:

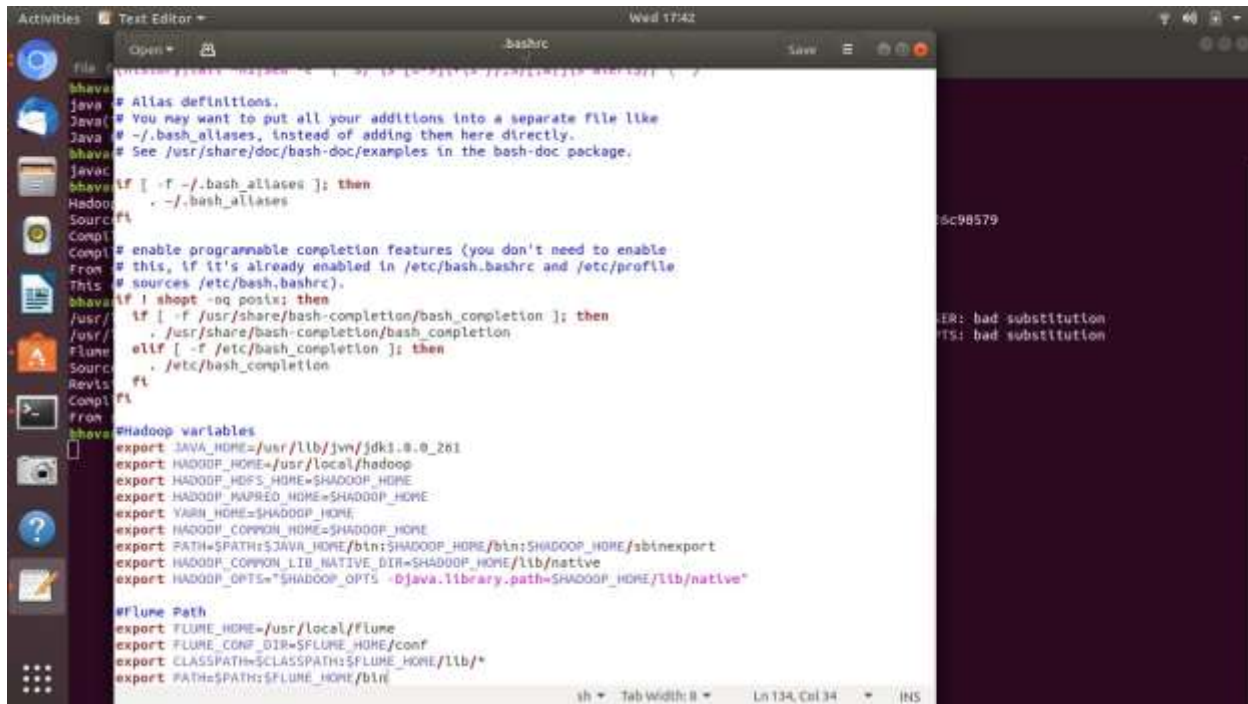


Figure 2: Environment variable setup for Java, Hadoop and Flume

In this step, we edited .bashrc file in order to set up environment for all the three components.

Step 4:

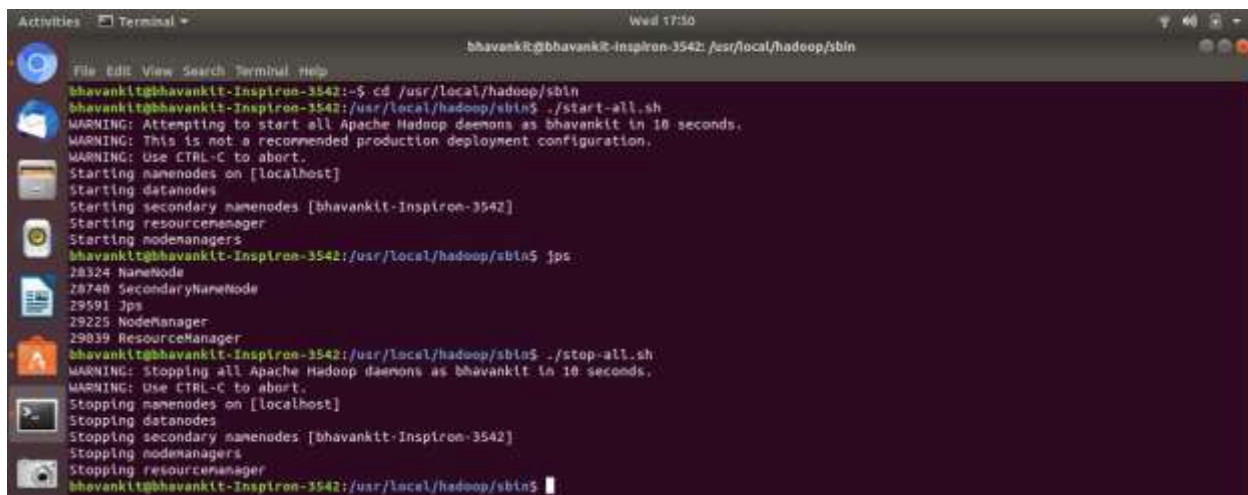


Figure 3: Hadoop start and stop

Here, we can see that all the Hadoop nodes are working properly using **jps** command. In this step, to run Hadoop properly we have configured core-site.xml, hdfs-site.xml, yarn-site.xml and maprd-site.xml.

Step 5:

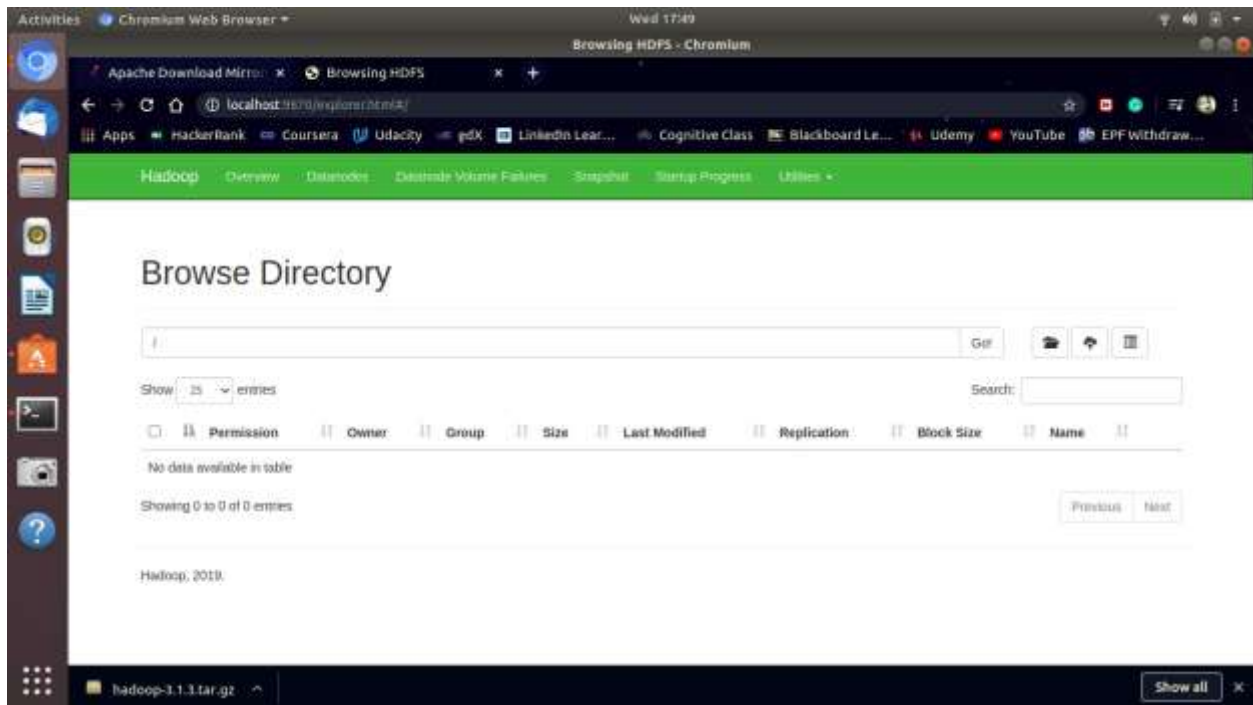
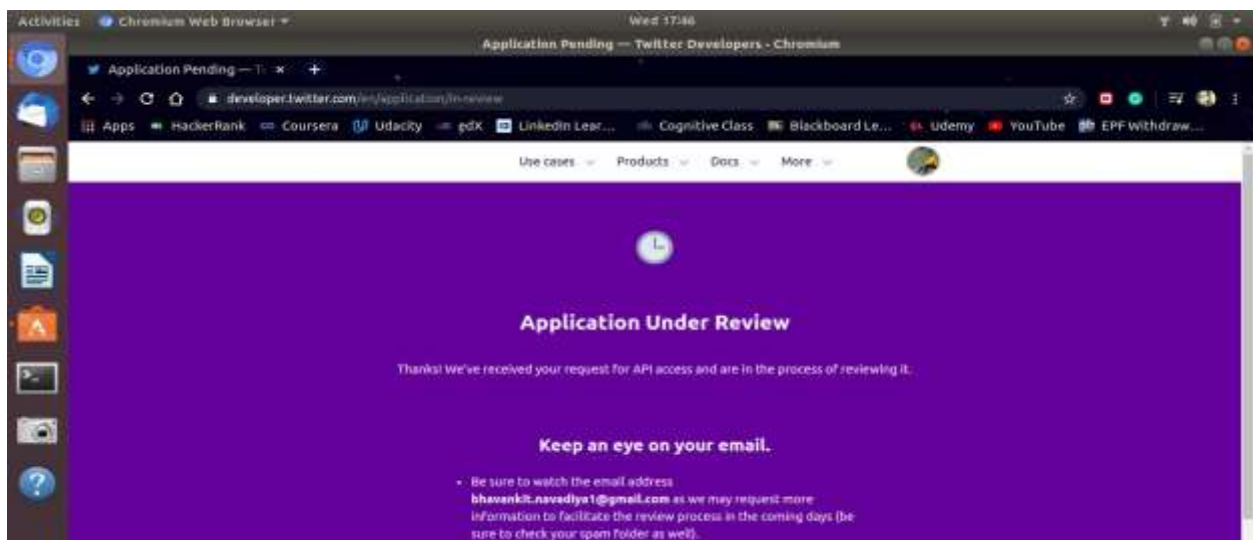


Figure 4: Working Localhost in hdfs

This is the UI where we will get tweets using flume and Hadoop (Hadoop distributed file system).

Step 6:



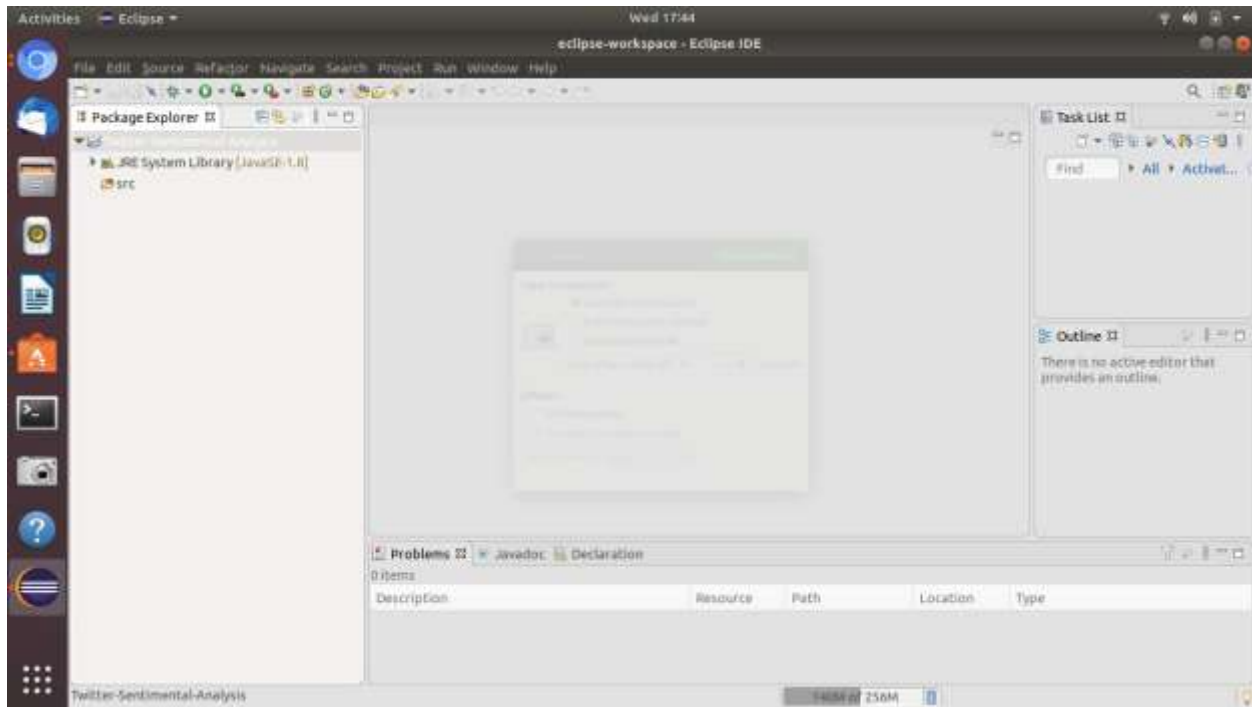


Figure 6: Eclipse Environment Setup

In this final step, we have installed eclipse IDE to write java or python code for sentimental analysis where we are probably going to use AFFIN library to rate the tweets in range of -5 to 5 where zero perform as neutral.

Note: Here we can consider **Twitter API access** and **AFFIN compatibility** as key components for this project, it means, if these two works properly, we are able to complete this sentimental analysis successfully.

Required Resource:

[1] "Apache Hadoop 3.1.3," Hadoop. [Online].

Available: <https://hadoop.apache.org/docs/r3.1.3/> [Accessed: 05-Aug-2020].

[2] "Apache Flume - Configuration," Tutorialspoint. [Online].

Available: https://www.tutorialspoint.com/apache_flume/apache_flume_configuration.htm. [Accessed: 05-Aug-2020].