User Guide

1. Initial Setup:

Following software and packages are used for set up of Public Opinion Analysis project.

1. Hadoop Pseudo Cluster: **Hadoop 2.6.1**
2. NoSQL Database to store the tweets in JSON format: **MongoDB 3.0.7**
3. Programming Language: **Python 2.7**
4. Web Server to host the User Interface web application: **Apache Tomcat 8.0.29**

Note*: Installable links for all the above requirements are mentioned in later sections following in this document.*

5. Integration of MongoDB with Hadoop: **mongo-hadoop-streaming-1.4.0.jar**

This Jar mongo-hadoop-streaming-1.4.0.jar is included when running the Mapper Reducer as mentioned in later sections of this document.

6. Integration of Python with Hadoop (Hadoop Streaming): **hadoop-streaming-2.6.1.jar**

This jar hadoop-streaming-2.6.1.jar is included when running the Mapper Reducer as mentioned in later sections of this document.

7. Type the following commands to install the libraries:

   - pip install mongopy
   - pip install tweepy
   - pip install jsonpickle
   - brew install mongodb

8. Start Hadoop node and MongoDB using the command **mongod** in the terminal
   Note: *All the commands and code have been tested on Mac OS.*

2. Creating data and Running Mapper Reducer using shell script (without hosting on Tomcat):

- After installing the above libraries, run the run_program.sh. The run_program.sh will first generate the data based on the keyword that is been passed and run the hadoop program automatically.
- It will give a message to open the html file that is used for User Interface.

  ***Note**: Open the file in **Mozilla Firefox** as the APIs used are compatible with Firefox.*
- The format of running shell script would be:

  sh run_program.sh KEYWORD HADOOP_PATH, KEYWORD is the name of the product. Our UI is configured to display 2 products currently that are **iphone** and **macbook**. We can run the script for any product but to view the analysis in frontend use only **iphone or mabook as the KEYWORD**.

  HADOOP_PATH is the home directory of the Hadoop installation.

Example command: sh run_program.sh iphone /Users/Downloads/hadoop-2.6.1/
run_program script does the following processing:

1. It checks whether both the arguments KEYWORD and HADOOP_PATH are present. It validates the HADOOP_PATH whether a valid path is been presented. The HADOOP_PATH should be the valid path.
2. It sets up the environment by copying the following scripts to the hadoop path
   1. publicmapper.py - the mapper code
   2. publicreducer.py - the reducer code
   3. hadoop_mongo - A set of library files that enable hadoop and mongo integration
   4. TwitterAnalysisUI.html - The final UI output file
3. Once setting up the environment the script runs data_load.py which fetches all the tweets that are related to the keyword specified and inserts the data to MongoDB
4. Once the data is ready, mapper reducer processes will run successfully. It will create output csv files such as iphone.csv and macbook.csv in <HADOOP HOME>.
5. The script will give the message to open the UI file from <HADOOP HOME> after completion.

Note: *For simplicity sake, we have given the static html file which displays the analysis of the data in graphical form. In the later section of the document, there are steps to create the WAR file and deploy it on Tomcat application server.*

3. Installation downloads

1. Install **MongoDB 3.0.7**

a. Use https://www.mongodb.org/downloads#production for downloading and installing MongoDB

2. Install **Python 2.7**

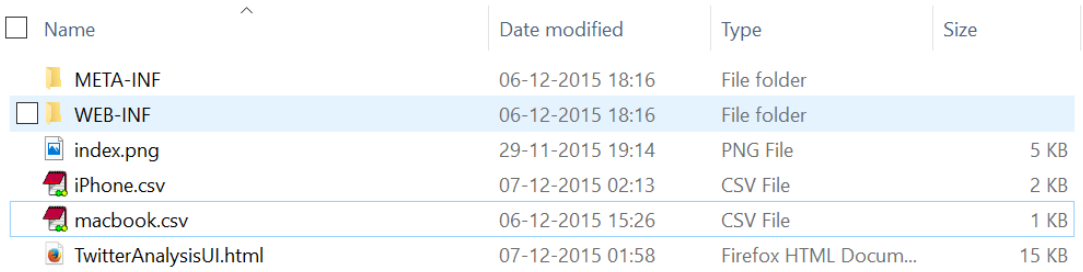
a. Use https://www.python.org/downloads/ for downloading and installing python

3. Install **Apache Tomcat 8.0.29**

a. Use https://tomcat.apache.org/download-80.cgi for downloading and installing Tomcat Server

4. Deploying and Running the application on Tomcat Server:

1. Copy the OpinionData folder from CD-ROM to local system.
2. Copy the iphone.csv and macbook.csv files (either both or atleast one) from <HADOOP_HOME> created in section 2 after running Mapper Reducer to OpinionData folder. The OpinionData folder would have following structure once files are copied:



Name	Date modified	Type	Size
META-INF	06-12-2015 18:16	File folder	
WEB-INF	06-12-2015 18:16	File folder	
index.png	29-11-2015 19:14	PNG File	5 KB
iPhone.csv	07-12-2015 02:13	CSV File	2 KB
macbook.csv	06-12-2015 15:26	CSV File	1 KB
TwitterAnalysisUI.html	07-12-2015 01:58	Firefox HTML Docum...	15 KB

3. Now run following commands in cmd prompt:

   cd <local_folder>/OpinionData

   jar -cvf OpinionData.war *

4. Start Tomcat Server. Go to <Tomcat Installation directory>. Go to bin folder and run startup.bat file.

5. Go to http://localhost:8080/manager/html/

   Deploy the WAR file (OpinionData.war created in step 2) by uploading the WAR file from the path <local_folder>/OpinionData as shown in below screenshot.

6. For avoiding steps 1-5, Instead of following steps 1-5, deploy the sample OpinionData.war file provided in CD-ROM.
7. In http://localhost:8080/manager/html/ link, verify whether the WAR is running or not.
8. Run the application using Mozilla or Internet Explorer 11 using the URL:

   http://localhost:8080/OpinionData/

Following web page will be shown to the user: