Readme

Activity 1: WordCount on tweet:

Raw tweets: data\_2.csv

Cleaned data: data\_got.csv

Data without stop words: data\_stp.csv

Hadoop Commands:

Hdfs dfs –mkdir –p ~/input

Hdfs dfs –put ~/Data/ ~/input

hadoop com.sun.tools.javac.Main WordCount.java

jar cf wc.jar WordCount\*.class

hadoop jar wc.jar WordCount ~/input/Data ~/op\_wc

hdfs dfs -get ~/op\_wc

Output obtained in op\_wc was converted to a csv file: stp.csv

stp.csv was used to build the word cloud

From the data\_got the stop words were removed manually after word count and stored as a csv file: op\_got.csv