Department of Computer Science and Engineering
University of Puerto Rico
Mayagüez Campus

CIIC 8995/5995 5016 – Big Data Analytics
Sprig 2017

Project 1: MapReduce for Twitter Analysis
Due Date: March 30, 2017, 11:55 PM

Objectives

1. Use Map Reduce to analyze trends contained in a collection of tweets.
2. Become familiar with HDFS and MapReduce

Overview

You will design, implement and test a series of MapReduce programs that will analyze a collection of tweets. The tweets will be provided to you. The analysis to be included are:

1. Count the number of occurrences for the words
    a. Trump
    b. MAGA
    c. Dictator
    d. Impeach
    e. Drain
    f. Swamp
    g. Change
2. Count the number of different keywords on in the message other than stop words: a, then, the, in , out, , which, etc.
3. Find the set of unique screen names that tweet.
4. Find all the retweets for each message.
5. Find all he replies for each message
6. Count and Find all messages posted by each user

Your solution will consist of a collection of MapReduce programs that perform tasks 1-6. It might be the case that some tasks require multiple-map reduce stages.

Visualization

Provide a means to visualize the results of the tasks 1-6, using the D3.js library . You are free to use the charts that you think best fits the visualization.

Deliverables

- **GitHub repo with all the code**

Grading

- **Project will be graded via demonstration of working code, forked from GitHub repo.**

PROJECT DUE DATE: **11:59 PM – March 30, 2017**.