Project Description

CS 5540

(Rao)

Goal

- Develop a system to store, analyze, and visualize Twitter's tweets

- Each team should have 3 members only
  - Only one submission per team

- This project is divided into 3 phases

- Total weight: 40%

Phase 1

- Weight: 8%
- Due date: September 18, 2015
- Goal
  - Collect Tweets using Twitter's Streaming APIs (e.g., 100K Tweets)
    - https://dev.twitter.com/docs/streaming-apis
    - Search online for documentation
  - Run the WordCount example in Apache Spark on the collected tweets and collect the output and log files
- What to submit?
  - Your code and the output and log files

Phase 2

- Weight: 25%
- Due date: November 27, 2015
- Goal
  - Design and implement your ideas using Apache Spark
    - Figure out how to store the tweets
    - Write interesting analytic queries to explore and understand the data – at least 8 queries
    - Develop interesting visualizations (e.g., pie chart, heat map, bar graphs)
- What to submit?
  - A report describing the design and implementation of your tool/software along with details on testing
- You will run your code and show the output to the TA

Phase 3

- Weight: 7%
- Due date: on the day of the final exam
- Goal
  - To present your project in the form of an attractive poster and a demo of your software
  - You should report some evaluation (e.g., how much time did the analytic queries take to finish)
  - CSEE faculty and students will be invited to attend the poster/demo session