Assignment: Final Project - Build a Map-Reduce

Bucket: 6 - Final Project

Due Date: 2016-04-24

Grading Hidden? no

Teams? yes

Description:

In this assignment, you will build a Map-Reduce system like Apache Hadoop M-R.

Your system should run both in parallel on Amazon EC2 and sequentially (or threaded) on your local machine.

Sample interaction:

```
./start-cluster 8
./my-mapreduce WordCount.jar s3://foo/bar/input.txt s3://foo/bar/output
./stop-cluster
```

Your system should run the following programs, producing approximately the same output as Hadoop:

- Hadoop Sample: WordCount.java (use alice.txt (http://www.ccs.neu.edu/home/ntuck/courses/2015/09/cs6240/alice.txt) as the input)
- Hadoop Sample: WordMedian.java (use alice.txt as the input)
- Your Assignment 2.
- Your Assignment 5.
- Your Assignment 7.

You can modify these programs slightly, e.g. to change the package names of the Java import directives.

Include a README that describes how to set up and run your system.

Your report should include the following:

- What functionality you needed to implement.
- Your strategy for implementing each piece of functionality.
- What problems you ran into and how you solved them.
- How did each team member contribute?
- For each test program:
    - Did this test program require any new functionality to get to work?
    - How did performance compare to Hadoop?

    - Did you produce the expected output?
- Detailed conclusion.

The due date for this project is Sunday, April 24, at midnight. There will be no further extensions.

Assignment Download: ()

Your Submissions

Team Members: Yogiraj Awati, Sarita Joshi, Ashish Kalbhor, Sharmodeep Sarkar

New Team Submission (/assignments/81/submissions/new)

Date	Status	Automatic	Teacher	Score	Link

Course Page (http://www.ccs.neu.edu/home/ntuck/courses/2016/01/cs6240/index.html) | Piazza (https://piazza.com/class/ij4yepvnz8v3zf)

ajax-status: none