

Distributed Big Data Analytics in the Cloud

Syllabus – Fall 2014

General Information

Supervisor: Prof. Stacy Patterson, Department of Computer Science

Office: Lally 301

Email: sep@cs.rpi.edu

Course Description

This course is an independent study, undergraduate research project in which students will (1) develop a cloud computing-based communication infrastructure for distributed big data analytics (2) implement a distributed compressed sensing algorithm on top of this infrastructure and (3) evaluate the performance of the infrastructure and algorithm in the Amazon cloud computing platform. A large part of this project involves the design and implementation of algorithms for “broadcast-convergecast” in a distributed computing environment. The students will develop novel solutions to this algorithmic design problem and evaluate their proposed solutions in a real-world cloud setting, potentially consisting of hundreds of nodes.

Pre-Requisites

Experience with software design, network programming, and algorithm development. Familiarity with linear algebra.

Learning Outcomes

At the end of this research project, the student will be able to

1. Create software design documentation using industry-standard practices and tools.
2. Design, develop and deploy software for cloud-computing platforms.
3. Design, implement, and test algorithms for multi-process communication in distributed computing environments.
4. Design and implement evaluation mechanisms for distributed computing applications.
5. Implement distributed data analytics algorithms.

Schedule

The students and professor will hold weekly meetings to review and improve algorithm and software design, discuss hurdles and next steps, and ensure that the project is on-schedule.

Grading

Grades will be assessed in two parts. 80% of the student grade will be based upon meeting the milestones below before the end of the semester. The professor will assess the remaining 20% at the end of the semester based upon the quality of the resulting work.

Milestones:

1. Software design plan created by students and approved by professor.
2. Successful demonstration of broadcast-convergecast algorithm in the cloud.
3. Successful demonstration of analytics algorithm in the cloud.
4. Delivery of short report, created by students, evaluating algorithm performance.

The professor may adjust these milestones as needed during the course of the project.

Academic Integrity

Participants in this program are required to comply with the requirements outlined in The Rensselaer Handbook of Student Rights and Responsibilities. Violations may result in penalties outlined in The Handbook.