

Homework 2

EECS 495

Hot Topics in Distributed Systems: Data-Intensive Computing

<http://www.eecs.northwestern.edu/~iraicu/teaching/EECS495-DIC/index.html>

Assigned: 01-14-2010

Due: 11:59PM 01-18-2010

Overview

A major component of your grade will be based on a project. The topic of the project will be chosen by the student(s), and will require possibly the implementation of a real/simulated system, a written report, and oral presentations.

Projects can fall in a number of different areas, which are loosely related to data-intensive distributed computing. Some of these areas are:

- Distributed file systems
- Data aware scheduling algorithms
- Distributed operating systems
- Distributed job management systems
- Parallel programming languages
- Distributed workflow systems
- Distributed monitoring systems
- Scientific computing with GPUs
- Scientific computing with MapReduce
- Distributed caching strategies
- Distributed cache eviction policies
- Distributed hash tables
- Applications porting/parallelization to clusters, grids, clouds, or supercomputers
- Virtualization impact for data-intensive computing

More project ideas from the Globus community can be found at http://dev.globus.org/wiki/Project_Ideas; not all these projects are suitable for this course, but they will help your brainstorming exercise in finding a suitable topic.

In the lecture on January 19th, I will take all your ideas, and use them to organize a brainstorming session for possible project ideas.

Assignment

You are to come up with 3 project ideas that you find interesting and you believe are relevant to the course. Please give each idea a title, and describe each one in at least 300 words each.