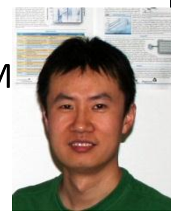
Syllabus

CS 495
Introduction to Distributed Computing

http://www.cs.iit.edu/~iraicu/teaching/CS495-F12/index.html

Semester: Fall 2012
Lecture Time: Tuesday/Thursday, 11:25AM - 12:40PM
Location: Stuart Building 239

Professor: Dr. Ioan Raicu (iraicu@cs.iit.edu, 1-312-567-5704)
Office Hours Time: Thursday 12:45PM-1:45PM
Office Hours Location: Stuart Building 237D



Teaching Assistant: Tony Li (tli13@iit.edu)
Office Hours Time: Monday/Tuesday 1PM-2PM
Office Hours Location: Stuart Building 006

Course Description

This course covers general introductory concepts in the design and implementation of distributed systems, covering all the major branches such as Cloud Computing, Grid Computing, Cluster Computing, Supercomputing, and Many-core Computing.

The specific topics that this course will cover are: scheduling in multiprocessors, memory hierarchies, synchronization, concurrency control, fault tolerance, data parallel programming models, scalability studies, distributed memory message passing systems, shared memory programming models, tasks, dependence graphs and program transformations, parallel I/O, applications, tools (Cuda, Swift, Globus, Condor, Amazon AWS, OpenStack, Cilk, gdb, threads, MPICH, OpenMP, Hadoop, FUSE), SIMD, MIMD, fundamental parallel algorithms, parallel programming exercises, parallel algorithm design techniques, interconnection topologies, heterogeneity, load balancing, memory consistency model, asynchronous computation, partitioning, determinacy, Amdahl's Law, scalability and performance studies, vectorization and parallelization, parallel programming languages, and power. Some of these topics are covered in more depth in the graduate courses focusing on specific sub-domains of distributed systems, such as Advanced Operating Systems (CS550), Parallel Computing (CS546), Cloud Computing (CS553), Data-Intensive Computing (CS554), Advanced Computer Architecture (CS570), and Fault Tolerant Computing (CS595).

While this CS495 course is not a pre-requisite to any of the graduate level courses in distributed systems, both undergraduate and graduate students who wish to be better prepared for these courses could take this CS495 course. Undergraduate students are highly encouraged to take CS495 prior to any of the graduate level courses in distributed systems. Graduate students who have already taken CS546, CS550, CS553, CS554, CS570, or CS595 should not take this CS495 class. Furthermore, this CS495 class should not be taken concurrently with CS546, CS550, CS553, CS554, CS570, or CS595.

Many of these graduate courses are part of the Master of Computer Science Specialization in Distributed and Cloud Computing. This CS495 course is part of the Undergraduate Specialization in Data Science and the Specialization in Distributed and Cloud computing.

Unique Opportunity -- Hands-on Practical Experience

Build a cluster: An important component of learning is to gain hands-on experience that a textbook just cannot teach. A portion of this course will cover practical aspects of distributed systems. For example, enrolled *students will participate in the design, assembling, configuring, and benchmarking of a real cluster*. The students would be exposed to practical issues in real cluster design, such as hardware tradeoffs, different operating systems, local and distributed storage, networking, virtualization, and grid/cloud middleware. The students will work in teams to build workstations/servers from scratch. The software stack will include Linux, XEN, Globus, Condor, OpenStack, NFS, PVFS, MPI, Swift, and Hadoop. This new cluster will then be used in subsequent assignments.

Use real Cloud Systems: Another set of assignments will deal with *real cloud systems*, such as Google App Engine, Amazon EC2/S3, or Hadoop (MapReduce framework).

Attend Conferences: Students will also get the opportunity to *attend local conferences in Distributed Systems*, specifically eScience 2012, UCC 2012, GlobusWorld 2013, and NSDI 2013.

SC13 Student Cluster Competition: I am also assembling *a team of undergraduate students to compete in the Supercomputing 2013 Student Cluster Competition*. The team has three slots (of the total 6 slots) available.

Paid Research Assistantship: Finally, I am also looking for an *undergraduate student to join my DataSys Laboratory for a paid assistantship*. If you are thinking about graduate school, or are excited about the opportunity to work at some of the largest technology companies (e.g. Microsoft, Google, Amazon, Facebook, Twitter, etc), then working in the DataSys Lab for several semesters will give you a significant advantage! Feel free to contact other students in my lab for feedback about the kinds of projects they are working on.

Textbooks

We will also use be using the textbook Distributed and Cloud Computing: Clusters, Grids, Clouds, and the Future Internet (DCC) by Kai Hwang, Jack Dongarra & Geoffrey C. Fox (Required). This is the most modern book about distributed systems I have found. Some of the fundamental topics in this book are not covered in enough detail, so for some topics, we will use another textbook, Andrew S. Tanenbaum and Maarten van Steen. "Distributed Systems: Principles and Paradigms" (DSPD), Prentice Hall, 2nd Edition, 2007 (Optional). I encourage you to buy both tetxbooks as they are both excellent, but if you have to choose just one, please buy the first (DCC), and the necessary optional reading material needed will be provided to the students in class.

Prerequisites

Systems Programming (CS351) or Operating Systems (CS450)

Mailing lists

There is a course mailing list; you can send mail to the list by sending email to cs495-f12@datasys.cs.iit.edu. Please see http://datasys.cs.iit.edu/mailman/listinfo/cs495-f12 for more information about the course mailing list.

Detailed Course Topics

Lecture topics:

- Distributed System Models
- System Architectures & Client-Server Models
- Programming Systems and Models
- Processes and threads
- Remote Procedure Call
- High-Performance Computing
- MapReduce
- Many-Task Computing
- Workflow Systems
- Grid Computing
- Cloud Computing
- Virtualization
- IaaS Clouds
- Filesystems
- Networked Filesystems
- Parallel Filesystems
- Distributed Filesystems
- Data-Intensive Computing
- Distributed Hash Tables
- Consistency Models
- Fault Tolerance
- Many-core Computing

Projects

There will be 6 projects throughout the semester, each worth 10% of the total grade. The projects will be completed in teams of 2 students. The first project will be hands-on, while the others will be primarily programming projects. Some projects will require knowledge of Java, while others will require knowledge of C and/or C++. It is expected that students know the basics of both of these languages.

Late Policy

Projects will be due at 11:59PM on the day of the due date, through BlackBoard. There will be a 15 minute grace period. Any late submissions beyond the grace period will be penalized 10% every day it is late.

Exams

There will be 2 exams, one covering the material from the first half of the class, and the second covering the material from the second half. The exams will be individual, but students will be allowed to use their textbooks and any notes they have (on paper). No electronic devices such as phones, eReaders, tables, or laptops will be allowed. Simple calculators can be used.

The exams are scheduled on:

- **10-09-2012 from 11:25AM - 1:25PM in SB239**
- **11-29-2012 from 11:25AM - 1:25PM in SB239**

Please note that they extend for 45 minutes after the usual end of class, but this should not interfere with anyone's other classes due to the lunch period.

There will be no makeup exams.

Grades

- **Projects (6): 60%**
- **Exam (2): 40%**

The following grading scale will be used. The scale will be adjusted downwards based on the overall performance of the entire class. Traditionally, in my classes, the class average score will typically fall in the B-grade range.

- **A: 90% ~ 100%**
- **B: 80% ~ 89%**
- **C: 70% ~ 79%**
- **D: 60% ~ 69%**
- **E: 0% ~ 59%**