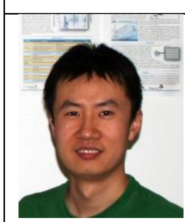
Syllabus

CS 451
Introduction to Parallel and Distributed Computing

http://www.cs.iit.edu/~iraicu/teaching/CS451-S14/index.html

Semester: Spring 2014
Lecture Time: Tuesday/Thursday, 11:25AM-12:40PM
Location: Stuart Building (SB) 104

	**Professor:** • **Dr. Ioan Raicu** (iraicu@cs.iit.edu, 1-312-567-5704) ○ Office Hours Time: Thursday 2PM-3PM (SB237D)
	**Teaching Assistant:** • **Iman Sadooghi** (isadoogh@iit.edu) ○ *Office Hours Time:* Monday/Wednesday 12:45PM-1:45PM (SB003b)
	**Teaching Assistant:** • **Tonglin Li** (tli13@iit.edu) ○ *Office Hours Time:* Thursday 10AM-11AM, Friday 12:45PM-1:45PM (SB002)

Office Hours Summary:

- **Monday:** 12:45PM-1:45PM (Sadooghi SB003b)
- **Tuesday:** 2PM-3PM (Raicu SB237D)
- **Wednesday:** 12:45PM-1:45PM (Sadooghi SB003b)
- **Thursday:** 10AM-11AM (Li SB002)
- **Friday:** 12:45PM-1:45PM (Li SB002)

Course Description

This course covers general introductory concepts in the design and implementation of parallel and 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: asynchronous/synchronous computation/communication, concurrency control, fault tolerance, GPU architecture and programming, heterogeneity, interconnection topologies, load balancing, memory consistency model, memory hierarchies, Message passing interface (MPI), MIMD/SIMD, multithreaded programming, parallel algorithms & architectures, parallel I/O, performance analysis and tuning, power, programming models (data parallel, task parallel, process-centric, shared/distributed memory), scalability and performance studies, scheduling, storage systems, synchronization, and tools (Cuda, Swift, Globus, Condor, Amazon AWS, OpenStack, Cilk, gdb, threads, MPICH, OpenMP, Hadoop, FUSE).

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 CS451 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 CS451 course. This course involves lectures, programming assignments, and exams. 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 CS451 course is part of the Undergraduate Specialization in Data Science and the Specialization in Distributed and Cloud computing, and can even be used for the Master of Computer Science Specialization in Distributed and Cloud Computing. This course was offered as CS495 in 2012 (http://www.cs.iit.edu/~iraicu/teaching/CS495-F12/).

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

This course will use Piazza to facilitate discussions for assignments, at http://piazza.com/iit/spring2014/cs451/home (it has not been setup yet, more instructions will follow).

Course Topics

Lecture topics:

- Distributed System Models
- High-Performance Computing
- Grid Computing
- Cloud Computing
- Many-core Computing
- Many-Task Computing
- Programming Systems and Models
- Processes and threads
- MapReduce
- Workflow Systems
- Virtualization
- Distributed Storage & Filesystems
- Data-Intensive Computing
- Distributed Hash Tables
- Consistency Models
- Fault Tolerance
- Performance analysis and tuning
- Parallel architectures
- Multithreaded programming
- GPU architecture and programming
- Message passing interface

Projects

There will be 5 projects throughout the semester, each worth 10% of the total grade. The projects will be completed in teams of 2 students, and will be programming intensive. 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 C/C++ and Java. For some projects, you will be allowed to choose the language in which to implement your work.

Computer Usage

Computer systems that can be used for development of projects (more information about access to these will be passed in the first several lectures):

- **Sirius:** 12-node (100-core) private cloud running OpenStack using virtualization and Linux
- **Jarvis:** 10-node (80-cores) Linux Cluster running SGE and NVIDIA GPUs
- **Amazon AWS:** $100 credit per student

Schedule

Tentative schedule:

- 01/14/2014: Syllabus
- 01/16/2014: Introduction to Distributed Systems; Project #1
- 01/21/2014: Introduction to Distributed Systems; Quiz #1
- 01/23/2014: TBA
- 01/28/2014: TBA; Last day to add/drop class; Quiz #2
- 01/30/2014: TBA
- 02/04/2014: TBA; Quiz #3; Project #2
- 02/06/2014: TBA
- 02/11/2014: TBA; Quiz #4
- 02/13/2014: TBA
- 02/18/2014: TBA; Quiz #5
- 02/20/2014: TBA; Project #3
- 02/25/2014: TBA; Quiz #6
- 02/27/2014: TBA
- 03/04/2014: Midterm Exam (11:25AM-1:25PM, SB104)
- 03/06/2014: TBA
- 03/11/2014: TBA; Quiz #7
- 03/13/2014: TBA; Project #4
- 03/18/2014: **Spring Break (NO CLASS)**
- 03/20/2014: **Spring Break (NO CLASS)**
- 03/25/2014: TBA; Quiz #8
- 03/27/2014: TBA
- 04/01/2014: TBA; Quiz #9
- 04/03/2014: TBA
- 04/08/2014: TBA; Quiz #10; Project #5
- 04/10/2014: TBA
- 04/15/2014: TBA; Quiz #11
- 04/17/2014: TBA
- 04/22/2014: TBA; Quiz #12
- 04/24/2014: TBA; Project #5 due
- 04/29/2014: TBA
- 05/01/2014: Final Exam (11:25AM-1:25PM, SB104)

Late Policy

Assignments will be due at 11:59PM on the day of the due date, through BlackBoard. There will be a 15 minute grace period. There will also be a 7-day late pass, where students can submit late assignments without penalty; the late pass can be used in 1-day increments spread out over multiple assignments. Any late submissions beyond the grace period and beyond the 7-day late pass, will be penalized 10% every day it is late. Assignments will not be accepted after the solutions have been posted.

Exams & Quizzes

There will be weekly short 5~10 min quizzes covering material from the prior week. If you are an in-class student, you must take the quiz in class; for remote students, you do not have to go to a testing center to take the quiz, I'll send you an online quiz (at a pre-arranged time) which you will have 5~10 minutes to complete and submit. The quizzes 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. There will also 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), just like for quizzes. Please note that the exams 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. The dates of the exams have been listed in the schedule (03/04/2014 and 05/01/2014), please mark them in your calendar to make sure you will not have any scheduling conflicts.

There will be no makeup quizzes or exams if you miss them.

Of the 12 quizzes, I'll drop your lowest 2 scores when computing your final grade.

Grades

- **Projects (5): 50% -- can use late day passes**
- **Quizzes (12): 20% -- can drop the lowest 2 grades**
- **Exam (2): 30%**

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%**

Course Outcomes

When a student completes this course, s/he should be able to:

1. Explain the range of requirements that modern parallel/distributed systems have to address.
2. Define the functionality that parallel/distributed systems must deliver to meet some need.
3. Articulate design tradeoffs inherent in large-scale parallel and distributed system design.
4. Describe how the resources in a parallel and distributed system are managed by software.
5. Justify the presence of concurrency within the framework of a parallel and distributed system.
6. Demonstrate the potential run-time problems arising from the concurrent operation of many (possibly a dynamic number of) tasks in a parallel and distributed system.
7. Summarize the range of mechanisms (in a distributed system) that can be employed to realize concurrent systems and be able to describe the benefits of each.
8. Understand the memory hierarchy and cost-performance tradeoffs.
9. Explain what virtualization is and how it is realized in hardware and software.
10. Examine the wider applicability and relevance of caching.

11. Summarize the features of a parallel and distributed system.
12. Understand the difference between a local, shared, parallel, and distributed filesystem.
13. Summarize the full range of considerations that support parallel and distributed file systems.
14. Understand the difference between various parallel/distributed systems paradigms: Cluster Computing, Grid Computing, Supercomputing, Cloud Computing, and Peer-to-Peer Computing.
15. Understand the different programming paradigms of parallel and distributed systems, such as high-performance computing, high-throughput computing, and many-task computing.
16. Understand GPU architectures and programming
17. Understanding the difference between SIMD and MIMD architectures, and their implications on programming models.
18. Have the ability to performance evaluation and tuning of parallel and distributed applications.