

CS55*: Foundations and Frontiers of Large-Scale Computer Systems

I. Course Instructor

Instructor: Zhiling Lan

Web: <https://cs.uic.edu/profiles/lan-zhiling/>

Email: zlan@uic.edu

II. Course Information

Prerequisite(s)

CS 455; or consent of the instructor.

Course Goals

Large-scale computer systems play a pivotal role in both high-performance computing and cloud computing. This course is designed to help students stay current with advances in modern computing platforms, where hardware and software are co-designed to achieve high performance and efficiency. Its goal is to provide students with the foundational knowledge and exposure to frontier topics needed to diagnose, design, and deploy efficient, reliable, and sustainable large-scale systems. The lectures cover a broad array of topics including heterogeneous architecture, networking, storage, power management, availability and reliability, resource management, and emerging topics such as machine learning for systems. These concepts are reinforced with research paper readings and hands-on projects that involve computer system design and analysis. By the end of this course, students will:

- Students will be able to analyze the core principles and technologies that power large-scale systems for high-performance computing and cloud.
- Students will be able to create cross-layer solutions that leverage high-performance computing and cloud resources to solve real-world problems.
- Students will be able to evaluate trade-offs among performance, cost, energy, and availability in system design.
- Students will be able to critique research papers and distill insights to guide design decisions.

III. Course Materials and Topics

A recommended textbook is “Datacenter as a Computer: An Introduction to the Design of Warehouse-scale Machines” by L. Barroso, Jimmy Clidaras, U. Hoelzle (BCH), 3rd edition, Morgan & Claypool Publishers, 2019.

A number of research papers taken from premier cloud computing conferences (e.g., USENIX conferences) and high-performance computing conferences (e.g., SC, HPDC, IPDPS, Cluster) will be used for class reading. All the reading material will be provided to students as either PDF files or pointers to online resources.

Week	Topic
Week 1	Hardware architecture
Week 2	Software infrastructure
Week 3	Networking
Week 4	Storage and I/O
Week 5	Performance analysis
Week 6	Availability and reliability
Week 7-8	Resource management
Week 9-10	Power management
Week 11-12	AI for Systems
Week 13-14	Systems for AI
Week 15	Emerging technologies

IV. Course Assessment

We expect students to have a basic understanding of computer systems CS361, high-performance computing CS455, and machine learning knowledge. The course is a combination of lectures and paper reading, including reading, presenting, and a semester-long research project.

- (a) Exam and quiz assess ability to analyze core principles and technologies of large-scale systems; (b) In-class discussion discusses strengths, weaknesses, and applicability of various state-of-the-art technologies for large-scale systems.
- (a) Class project design cross-layer solution leveraging HPC/cloud for a specified application; (b) Project report justify design choices and results for application–hardware co-design optimizations on target platforms.
- Case studies evaluate architectural and operational trade-offs under realistic computing environments
- (a) Literature review critique research papers and synthesize design implications; (b) Presentation communicate distilled insights and implications for system design.

V. General Policies

This course focuses exclusively on the *emerging hardware and software trends in large-scale computer systems* deployed for cloud and high-performance computing. Upon completion, students will be well-prepared to tackle the challenges and opportunities presented by modern systems in these domains.

Generative AI Use Policy: AI usage is transparently disclosed; AI is used to aid learning and not as a replacement for understanding; AI is not permitted during quizzes, exams, or assessments.