Syllabus

General Information

Parallel and Distributed System

Instructor: Yong Zhao, 523N

Contact: yong.zhao@scupi.cn

Teaching Assistant: Haowei Zhang

Contact: 1967159434@qq.com

Course Meeting Times

Lectures:

One session / week Thursday 8:15AM – 11:00AM

Office Hours:

Thursday 2:00pm -5:00 pm

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 Cluster Computing, Supercomputing, High Performance Computing, Grid Computing, Cloud Computing 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, programming models (data parallel, task parallel, process-centric, shared/distributed memory), scalability and performance studies, scheduling, synchronization, and tools (CUDA, Amazon AWS, OpenStack, Swift, MPICH, OpenMP, Hadoop, FUSE).

This course also includes a course project to help students get hands-on experience on the design, implementation, testing, and evaluation of a computer system, and to better understand the concepts involved in computer architecture.

Course Objectives

The goals of the course are to help the students understand the principles and applications of distributed and parallel systems, and look into details of the design and implementation of the key architectural components and programming models of distributed and parallel systems, and learn the key concepts and approaches to improve the performance and efficiency of these systems.

The course will prepare students for working in the Cloud and AI industry and can act as a footstep to more advanced material in graduate-level courses. This course can also provide a foundation for students interested in GPU programming, distributed systems, Cloud computing, and performance improvement; and it can provide system-level context for students interested in emerging technologies and artificial intelligence.

Prerequisite

Data Structure

Class Schedule

Lecture slides will be available for copying or posted on blackboard.

1. Introduction and Review
2. Parallel Architecture
3. Parallel Program Design
4. Message Passing Programming
5. Floyd's Algorithm
6. Introduction to CUDA
7. CUDA Parallelism Model
8. Parallel Computation Patterns
9. High Performance Computing
10. Cloud Computing
11. Memory and Data Locality
12. Performance Considerations
13. Joint CUDA-MPI Programming

Learning Outcome

At the end of the class, the students should be able to understand the key concepts related to parallel and distributed computing.

The students should also be able to:

- acquire knowledge and concepts and mechanisms related to distributed and parallel systems and explain how these concepts and mechanisms interact.
- program with MPI and CUDA programming models.
- evaluate and measure various distributed and parallel systems and improve the performance and efficiency.
- apply the corresponding technology and patterns to real world applications.

Grades

Grades will be based on homework, course project and final exam.

Homework and attendance: 30%

Course project (1 major project or 2 small projects): 40%

Final Exam: 30%

Collaboration and Academic Honesty Policy

Individual work on all homework and examinations is required, Cheating and copying other students' homework/exam are strictly prohibited. Any violation of this policy will be treated severely.

Collaboration amongst students to understand the course material and to work on course projects is strongly encouraged, however each student should take on different/distinguishable responsibilities in the course projects.

Course Reading Material

Textbook – Introduction to Parallel Programming Second Edition, China Machine Press. 2024

Reference Book: Designing and Building Parallel Programs, Ian Foster, 2003