CPSC 4770/6770
Distributed and Cluster Computing

Class Location/Meeting Time:
Zoom on Canvas / Tuesday and Thursday 11:00am-12:15pm

Instructor: Dr. Shuangshuang Jin
Email: jin6@clemson.edu Phone: (206) 601-7654
Office Hours: Mon 11:30am-12:30pm

Teaching Assistant/Grader: Liwei Wang
Email: liweiw@clemson.edu Phone: (864) 633-7758
Office Hours: Wed 1:00pm-2:00pm

Course Description

This course will investigate issues in modern distributed platforms by examining a number of important technologies in the areas of parallel and distributed computing in computational and data-intensive problems.

By the end of the course, each student should understand and be able to apply several specific tradeoffs for parallel application and algorithms development, performance, and management on a number of distributed platforms.

Prerequisites: Working knowledge of C/C++ and Linux system. Working knowledge of data structures and algorithms. Familiarity with Java and Python.

Course Objectives

At the completion of the course, students should be able to

- Apply mathematical foundations, algorithmic principles, and computer science theory in the modeling and design of computer-based systems in a way that demonstrates comprehension of the tradeoffs involved in design choices.
- Analyze a problem and identify and define the computing requirements appropriate to its solution.
- Apply design and development principles in the construction of large-scale computing systems.
- Function effectively on teams to accomplish a common goal.

Course Materials

- Lecture slides and videos will be available on canvas.
- Example codes will be available in slides and on Palmetto (/scratch1/jin6/cpsc4770)
- Links to papers on subjects we will be discussing in class will also be listed and/or embedded in the slides.

Laptop Requirements

- Having access to a laptop during class time is critical
  - Working with supercomputing resources in class
  - Working on in-class electronic quizzes on Canvas

Topical Outline

- Parallel Computing
- Palmetto Cluster
- Distributed File Systems

- Message Passing Interface (MPI)
- Pleasant Parallelism
- Partitioning and Divide and Conquer Strategies
- Synchronous Computing

- Big Data and Data Intensive Computing
- Hadoop Distributed File System (HDFS)
- Hadoop MapReduce
- Memory Centric Data Intensive Computing
- PySpark

- (TBD)
- Shared Memory Programming – OpenMP
- Shared Memory Programming -- Pthreads

Grading

- **Grade Scale:**
  - **Undergraduate Students:**
    - A – 90 - 100;
    - B – 80 to < 90;
    - C – 70 to < 80;
    - D – 60 to < 70;
    - F – < 60
  - **Graduate Students:**
    - A – 92 - 100;
    - B – 82 to < 92;
    - C – 70 to < 82;
    - F – < 70

- **Grade Components:**
  - Assignment: 60%
  - Exam: 30%
    - Midterm: 20%
    - Final: 10%
  - Quiz: 10%

Additional Policies

- **Attendance**: Attendance is critical to the success of students in the class. Substantial project information will be provided in class lectures. Random quizzes or roll calls will be made in class. No individual lectures will be given.
- **Late-work**: An assignment submitted within 0 to 24 hours after the due time will only be eligible for 80% of the maximum number of points allotted; An assignment submitted within 24 to 48 hours after the due time will only be eligible for 50% of the maximum number of points allotted; Assignments submitted more than 48 hours after the due time will not be accepted.
- **Re-grade**: All requests for re-grades must be submitted within one week of the graded assignments being returned. Mistakes occasionally happen during the grading process. If you think a mistake has been made regarding your grades, you should send me an email with detailed justification within one week of the date the grades are available. No changes on grades will be made after one week from the date the grades are posted.
- **Academic Integrity:** "As members of the Clemson University community, we have inherited Thomas Green Clemson's vision of this institution as a "high seminary of learning." Fundamental to this vision is a mutual commitment to truthfulness, honor, and responsibility, without which we cannot earn the trust and respect of others. Furthermore, we recognize that academic dishonesty detracts from the value of a Clemson degree. Therefore, we shall not tolerate lying, cheating, or stealing in any form."
- **Collaboration Policy:** Collaboration between students on homework assignments in this class is permitted. Plagiarism is not allowed. Taking assignments from other classmates or downloading completed assignments from websites is not allowed.
- **Disability Accommodations**: If you have a documented disability that requires an accommodation, please contact me so we can set up an appointment to discuss your needs. Or contact: Student Disability Services, G20 Redfern, 864-656-6848.
- **Mobile Devices**: Please refrain from using mobile devices during our class sessions.
- **Web Camera**: Please turn on your web camera while taking the class for better face-to-face communications and in-class interactions.

Important dates:

- First Class: Wednesday August 19, 2020
- Last Day to Add a Class: Tuesday August 25, 2020
- Last Day to Drop a Class without a "W" Grade: Tuesday September 1, 2020
- Last Day to Drop a Class without a Final Grade: Friday October 23, 2020
- Fall Break: Monday November 2 – Tuesday November 3, 2020
- Final Exams: Monday December 7 – Friday December 11, 2020
- Tuesday September 8, 2020: Last day to apply for December graduation