CSCI 339: Distributed Systems

Syllabus for Spring 2023

General Info

Instructor:	Jeannie Albrecht
Email:	jeannie@cs.williams.edu
Office:	TCL 305
Class Meetings:	MTh 1:10 - 2:25
Webpage:	http://www.cs.williams.edu/~jeannie/cs339/index.html
Textbook:	*Distributed Systems*, by Van Steen and Tanenbaum, 4th ed.

Course Description

This course studies the key design principles of distributed systems, which are collections of independent networked computers that function as single coherent systems. Covered topics include communication protocols, processes and threads, naming, synchronization, consistency and replication, fault tolerance, and security. We also examine some specific real-world distributed systems case studies, ranging from the Internet to file systems. Class discussion is based on readings from the textbook and research papers. The goals of this course are to understand how large-scale, distributed computational systems are built, and to provide you with the tools necessary to evaluate new technologies after the course ends.

Grading Details

Grades will be computed as follows:

5%	Class Participation
10%	Written Homework
35%	Programming Projects
30%	Midterm Exam
20%	Final Project

Each of these items are explained in detail in the following sections. In general, beyond the 3 hours we spend together during our class meeting time, you should expect to spend (on average) at least 10 hours per week on work related to class.

Class Participation

Lectures are mandatory and you are expected to attend regularly. One goal of this course is to promote discussion of the issues in distributed systems among all class members. As such, you are encouraged and expected to ask questions, point out weaknesses, and make observations during class.

Written Homework

The homework portion of your grade will be determined by written evaluations of a number of technical papers that we will read during the course of the semester (roughly one per week). For each assigned paper, you will submit a 1–2 page evaluation that includes: (i) a high level summary that highlights the most important points addressed by the paper, (ii) the most glaring problem with the paper, and (iii) the conclusions you draw after reading the paper about building distributed systems. Your grade will largely be based on effort rather than correctness. **Evaluations must be submitted on GLOW before the beginning of class on the day the paper is assigned.** If you need to miss class, you must submit the evaluation before class ends in order to receive credit. Late homework will not receive credit for any reason.

Programming Projects

There will be 3 programming projects (in addition to the final project) during the semester. Details will be available on the course webpage. To complete the projects, you will need the ability to develop software programs using Java, C/C++, and Python. If you have not used C/C++ recently (or at all), you may want to refresh your knowledge using one of the many good books on the topic. In particular I recommend the classic, *The C Programming Language*, by Kernighan and Ritchie, because it is short and simple. I have a few reference books in my office that you are welcome to use. There are also many online tutorials. If you feel that need extra help, please come see me.

We will primarily use the department's Linux computers for projects. If you are not familiar with the Linux/UNIX computing environment, talk to me as soon as possible.

Programming assignments must be submitted by their due date. Late submissions will be penalized 25% of their value for each date late. Extensions will only be granted on a case-by-case basis for extenuating circumstances.

Midterm Exam

There will a written midterm examination in this course. The exam will be closed book, closed notes, and will stress conceptual understanding of the material. Details regarding the specific format of the exam will be discussed in class.

Final Project

In lieu of a final exam (and in addition to the three programming projects described above) you will complete a final project on a topic of your choosing. I will post some sample project ideas on the course webpage later in the semester. I encourage you to be creative and pick a topic related to distributed systems that will be fun and challenging to explore. You may work with a partner on your final project. In addition to completing the programming aspect of the project, you and your partner will present your results to your classmates and write a paper describing your work.

Collaboration

Written homework assignments and exams are to be completed individually. You are allowed to work with a partner on programming assignments. I encourage collaboration and assistance between groups in understanding material, but not in developing solutions. Please be sure to give explicit credit for any help received. If you have any doubts about this, ask me whether or not collaboration is appropriate. Uncredited collaborations will be considered a violation of the honor code. **The Computer Science honor code and computer usage policy applies to all material in this class. Please review this info (links available on course webpage) if necessary.**

Calendar

The following calendar is a (very) tentative schedule of topics that we will cover in class. This schedule will likely change, so you should check the course webpage frequently for updates.

Lecture	Topic
1	Course Overview (VST Ch 1–2, VST is an abbreviation for your textbook)
2	Intro to Networks and C/C++ (VST Ch 3–4), *Project 1 out*
3	Routing Basics
4	IP, TCP, UDP
5	Networks Wrapup
6	No class
7	Internet Services (VST Ch 6), *Project 1 due*
8	Resource Discovery and Search, *Project 2 out*
9	RPCs
10	Content Distribution Networks
11	Email and Porcupine
12	Time and Review (VST Ch 5), *Project 2 due*
13	*Midterm in class*
	Spring break
14	Coordination and Agreement, MapReduce, *Project 3 out*
15	Hadoop Cluster Setup
16	Storage Systems and RAID
17	BigTable and Distributed File Systems
18	Distributed File Systems, *Project 3 due*
19	Wide-Area Computing, *Project 4 out*
20	Security, (VST Ch 9)
21	Fault Tolerance and Replication (VST Ch 7–8)
22	P2P Systems Overview, *Project 4 checkpoint due*
23	Overlay Networks and DHTs
24	Project Presentations
25	Project Presentations, *Project 4 due during Reading Period/Finals Week*