CSCI 599: Distributed Systems, a dance between complexity and performance

Units: 4
Fall 2024 TuTh 12:00-1:50PM

Location: TBD

Instructor: Seo Jin Park
Office: TBD
Office Hours: TBD
Contact Info: contact via Piazza or email via seojinpark.net if not enrolled yet.

Course Description

This is a first-year graduate class in distributed systems. Distributed systems is a powerful tool that enables scaling storage and computation power with relatively cheap cost. However, distribution brings in many complexities unseen in single computer systems. In this course, students will learn various such complexities and current solutions, both from industry and academia. This course requires reading 1 or 2 papers per week, four lab projects, and an (optional) final project. Each class will have a mix of lecture and discussion.

Learning Objectives and Outcomes

Students will learn

- How a large-scale systems structured
- Different replication techniques and their tradeoffs
- How to build a fault-tolerant systems
- Data consistency challenges and a few remedies
- How to speed up big data processing with data parallelism
- Emerging distributed programming model and framework
- How distributed systems may improve energy and cost efficiency

Recommended Preparation

Knowledge at the level of CSCI 201, CSCI 350, and CSCI 356 (or equivalent). Strong C++ experience.

Course Notes

Lectures slides and project descriptions will be posted online. Piazza will be used for announcements and discussions regarding lectures, projects, and exams.

Technological Proficiency and Hardware/Software Required

To work on projects, students must have a computer that can compile c++ code and run shell scripts.
For in-class participation, students must bring a smartphone with camera and internet access.

Required Readings and Supplementary Materials

No textbook. Required reading for each lecture will be posted online 1 week before each lecture.

Tentative reading lists:

[Illusion of a single image]

- [GFS] Sanjay Ghemawat, Howard Gobioff, and Shun-Tak Leung. 2003. The Google file system. In Proceedings of the nineteenth ACM symposium on Operating systems principles (SOSP '03). Association for Computing Machinery, New York, NY, USA, 29–43. https://doi.org/10.1145/945445.945450
- [RAMCloud] John Ousterhout, Arjun Gopalan, Ashish Gupta, Ankita Kejriwal, Collin Lee, Behnam Montazeri, Diego Ongaro, Seo Jin Park, Henry Qin, Mendel Rosenblum, Stephen Rumble, Ryan Stutsman, and Stephen Yang. 2015. The RAMCloud Storage System. ACM Trans. Comput. Syst. 33, 3, Article 7 (September 2015), 55 pages. https://doi.org/10.1145/2806887

[Fault-tolerance]

- [VM-FT] Scales, Daniel J., Mike Nelson, and Ganesh Venkitachalam. "The design of a practical system for fault-tolerant virtual machines." ACM SIGOPS Operating Systems Review 44, no. 4 (2010): 30-39.
- [Raft] Ongaro, Diego, and John Ousterhout. "In search of an understandable consensus algorithm." In 2014 USENIX Annual Technical Conference USENIX ATC 14, pp. 305-319. 2014.
- [ChainRep] Van Renesse, Robbert, and Fred B. Schneider. "Chain Replication for Supporting High Throughput and Availability." In *OSDI*, vol. 4, no. 91–104. 2004.
- [PBFT] Castro, Miguel, and Barbara Liskov. "Practical byzantine fault tolerance." In OSDI, vol. 99, no. 1999, pp. 173-186. 1999.

[Consistency]

- [2PC] Lampson, Butler, and David Lomet. "A new presumed commit optimization for two phase commit." In 19th International Conference on Very Large Data Bases (VLDB'93), pp. 630-640. 1993.

- [RIFL] Lee, Collin, Seo Jin Park, Ankita Kejriwal, Satoshi Matsushita, and John Ousterhout. "Implementing linearizability at large scale and low latency." In *Proceedings of the 25th Symposium on Operating Systems Principles*, pp. 71-86. 2015.
- [Spanner] Corbett, James C., Jeffrey Dean, Michael Epstein, Andrew Fikes, Christopher Frost, J. J. Furman, Sanjay Ghemawat et al. "Spanner: Google's Globally-Distributed Database." In *10th USENIX Symposium on Operating Systems Design and Implementation (OSDI 12)*, pp. 261-264. 2012.

[Data-parallel systems]

- [MapReduce] Dean, Jeffrey, and Sanjay Ghemawat. "MapReduce: Simplified data processing on large clusters." USENIX OSDI 2004.
- [Spark] Zaharia, Matei, Mosharaf Chowdhury, Tathagata Das, Ankur Dave, Justin Ma, Murphy McCauly, Michael J. Franklin, Scott Shenker, and Ion Stoica. "Resilient distributed datasets: A fault-tolerant abstraction for in-memory cluster computing." In *Presented as part of the 9th USENIX Symposium on Networked Systems Design and Implementation (NSDI 12)*, pp. 15-28. 2012.
- [ExCamera] Fouladi, Sadjad, Riad S. Wahby, Brennan Shacklett, Karthikeyan Balasubramaniam, William Zeng, Rahul Bhalerao, Anirudh Sivaraman, George Porter, and Keith Winstein. "Encoding, Fast and Slow: Low-Latency Video Processing Using Thousands of Tiny Threads." In *NSDI*, vol. 17, pp. 363-376. 2017.
- [MilliSort] Li, Yilong, Seo Jin Park, and John K. Ousterhout. "MilliSort and MilliQuery: Large-Scale Data-Intensive Computing in Milliseconds." In *NSDI*, pp. 593-611. 2021.

[Distributed programming model]

- [Ray] Moritz, Philipp, Robert Nishihara, Stephanie Wang, Alexey Tumanov, Richard Liaw, Eric Liang, Melih Elibol et al. "Ray: A distributed framework for emerging {AI} applications." In *13th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 18)*, pp. 561-577. 2018.
- [Nu] Ruan, Zhenyuan, Seo Jin Park, Marcos K. Aguilera, Adam Belay, and Malte Schwarzkopf. "Nu: Achieving Microsecond-Scale Resource Fungibility with Logical Processes.", 20th USENIX Symposium on Networked Systems Design and Implementation (NSDI 23)

[Energy-efficient distributed systems]

- [FAWN] Andersen, David G., Jason Franklin, Michael Kaminsky, Amar Phanishayee, Lawrence Tan, and Vijay Vasudevan. "FAWN: A fast array of wimpy nodes." In *Proceedings of the ACM SIGOPS 22nd symposium on Operating systems principles*, pp. 1-14. 2009.
- [E3] Liu, Ming, Simon Peter, Arvind Krishnamurthy, and Phitchaya Mangpo Phothilimthana. "E3: Energy-Efficient Microservices on SmartNIC-Accelerated Servers." In *USENIX annual technical conference*, pp. 363-378. 2019.

[Cluster scheduler]

- [Borg] Verma, Abhishek, Luis Pedrosa, Madhukar Korupolu, David Oppenheimer, Eric Tune, and John Wilkes. "Large-scale cluster management at Google with Borg." In Proceedings of the Tenth European Conference on Computer Systems, pp. 1-17. 2015.

Description and Assessment of Assignments

For each paper, students have to submit a short summary and answer questions through an online form before the due date. Inaccurate and purely guessed answers may negatively impact on grades.

There will be four lab (lab 0 - 3) assignments which will be in C++ programming language. Students may work in a group of two for all projects except the first one.

In class participation will be assessed based on a student's participation to discussions, in-class polls, and attendance. A generally active participants can get full credit even if not perfect on polls or attendance.

Students may opt in for final project. Each team should be with 2-3 students. The score of final project will replace the worst score of three in-class quizzes or lab 3.

Grading Breakdown

Assignment	% of Grade
Lab assignments	35%
Quiz	45%
Final project (optional)	Replaces the worst quiz or lab3
Homeworks	10%
Class Participation	10%
TOTAL	100%

Assignment Submission Policy

Students will submit their homework and labs through a web form. Projects will be due at 11:59:59 pm on the due date. Every student will have a total of 3 late days across all projects. In order to use a grace day, you must fill out this grace day request form before the lab's non-extended deadline. (You will need to be logged into your USC account to access this form).

Please note that grace days are in place of "excused late" submissions, not in addition to. If you request additional grace days from the instructor, you must have a documented reason for each grace day used to accompany your request. Once you have used your grace days, any late submission will not be accepted and graded as a 0.

Note: There is no grace period. Even if you submit a few minutes after the deadline, you will need to use a grace day (even if the wireless network in your dorm room is down or you have a github issue, etc.). It is your job to be on time and not cut it too close. Remember Murphy's Law and leave time for things to "go wrong."

Additional Policies

Missesd classes: lecture materials and assignments will be posted online, so there's no need to ask permission for missed classes. A single missed lectures won't prevent a student from receiving 100% of class participation scores.

Course Schedule: A Weekly Breakdown

	Topics/Daily Activities	Readings and Homework	Deliverable/ Due Dates
Week 1	Intro to distributed systems / RPC		Lab 0 out
Week 2	Illusion of a single image	GFS, RAMCloud	Lab 0 due
Week 3	Primary-backup replication and replicated state machine	VM-FT	
Week 4	Raft	Raft	Lab 1 out
Week 5	Quiz 1 and Chain replication	*ChainRep*	
Week 6	How to measure system performance and Byzantine consensus	PBFT	Lab 1 due
Week 7	Consistency	2PC, RIFL	Lab 2 out
Week 8	Consistency (2)	Spanner	
Week 9	Tail at scale and Quiz 2	TailAtScale	Lab 2 due, signup for final project.

Week 10	Data-parallel systems	MapReduce, Spark	Final project proposal due
Week 11	Distribution for interactivity	ExCamera, MilliSort	Lab 3 out
Week 12	Distributed programming model	Ray, Nu	
Week 13	Swarm of wimpy	FAWN	Lab 3 due
Week 14	Distributed job scheduling	Borg	
Week 15	Quiz3 and project demo		
FINAL	Final Project	Final Project	Due on the university-scheduled date of the final exam.

Statement on Academic Conduct and Support Systems

Academic Integrity:
The University of Southern California is a learning community committed to developing successful scholars and researchers dedicated to the pursuit of knowledge and the dissemination of ideas. Academic misconduct, which includes any act of dishonesty in the production or submission of academic work, comprises the integrity of the person who commits the act and can impugn the perceived integrity of the entire university community. It stands in opposition to the university's mission to research, educate, and contribute productively to our community and the world.

All students are expected to submit assignments that represent their own original work, and that have been prepared specifically for the course or section for which they have been submitted. You may not submit work written by others or "recycle" work prepared for other courses without obtaining written permission from the instructor(s).

Other violations of academic integrity include, but are not limited to, cheating, plagiarism, fabrication (e.g., falsifying data), collusion, knowingly assisting others in acts of academic dishonesty, and any act that gains or is intended to gain an unfair academic advantage.

The impact of academic dishonesty is far-reaching and is considered a serious offense against the university. All incidences of academic misconduct will be reported to the Office of Academic Integrity and could result in outcomes such as failure on the assignment, failure in the course, suspension, or even expulsion from the university.

For more information about academic integrity see the student handbook or the Office of Academic Integrity's website, and university policies on Research and Scholarship Misconduct.

Please ask your instructor if you are unsure what constitutes unauthorized assistance on an exam or assignment, or what information requires citation and/or attribution.

Course Content Distribution and Synchronous Session Recordings Policies
USC has policies that prohibit recording and distribution of any synchronous and asynchronous course content outside of the learning environment.

Recording a university class without the express permission of the instructor and announcement to the class, or unless conducted pursuant to an Office of Student Accessibility Services (OSAS) accommodation. Recording can inhibit free discussion in the future, and thus infringe on the academic freedom of other students as well as the instructor. (Living our Unifying Values: The USC Student Handbook, page 13).

Distribution or use of notes, recordings, exams, or other intellectual property, based on university classes or lectures without the express permission of the instructor for purposes other than individual or group study. This includes but is not limited to providing materials for distribution by services publishing course materials. This restriction on unauthorized use also applies to all information, which had been distributed to students or in any way had been displayed for use in relationship to the class, whether obtained in class, via email, on the internet, or via any other media. (Living our Unifying Values: The USC Student Handbook, page 13).

Students and Disability Accommodations:
USC welcomes students with disabilities into all of the University's educational programs. The Office of Student Accessibility Services (OSAS) is responsible for the determination of appropriate accommodations for students who encounter disability-related barriers. Once a student has completed the OSAS process (registration, initial appointment, and submitted documentation) and accommodations are determined to be reasonable and appropriate, a Letter of Accommodation (LOA) will be available to generate for each course. The LOA must be given to each course instructor by the student and followed up with a discussion. This should be done as early in the semester as possible as accommodations are not retroactive. More information can be found at osas.usc.edu. You may contact OSAS at (213) 740-0776 or via email at osasfrontdesk@usc.edu.

Support Systems:
Counseling and Mental Health - (213) 740-9355 – 24/7 on call
Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention.

988 Suicide and Crisis Lifeline - 988 for both calls and text messages – 24/7 on call
The 988 Suicide and Crisis Lifeline (formerly known as the National Suicide Prevention Lifeline) provides free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week, across the United States. The Lifeline is comprised of a national network of over 200 local crisis centers, combining custom local care and resources with national standards and best practices. The new, shorter phone number makes it easier for people to remember and access mental health crisis services (though the previous 1 (800) 273-8255 number will continue to function indefinitely) and represents a continued commitment to those in crisis.

Relationship and Sexual Violence Prevention Services (RSVP) - (213) 740-9355(WELL) – 24/7 on call
Free and confidential therapy services, workshops, and training for situations related to gender- and power-based harm (including sexual assault, intimate partner violence, and stalking).

Office for Equity, Equal Opportunity, and Title IX (EEO-TIX) - (213) 740-5086
Information about how to get help or help someone affected by harassment or discrimination, rights of protected classes, reporting options, and additional resources for students, faculty, staff, visitors, and applicants.

Reporting Incidents of Bias or Harassment - (213) 740-5086 or (213) 821-8298
Avenue to report incidents of bias, hate crimes, and microaggressions to the Office for Equity, Equal Opportunity, and Title for appropriate investigation, supportive measures, and response.

The Office of Student Accessibility Services (OSAS) - (213) 740-0776
OSAS ensures equal access for students with disabilities through providing academic accommodations and auxiliary aids in accordance with federal laws and university policy.

USC Campus Support and Intervention - (213) 740-0411
Assists students and families in resolving complex personal, financial, and academic issues adversely affecting their success as a student.

Diversity, Equity and Inclusion - (213) 740-2101
Information on events, programs and training, the Provost's Diversity and Inclusion Council, Diversity Liaisons for each academic school, chronology, participation, and various resources for students.

USC Emergency - UPC: (213) 740-4321, HSC: (323) 442-1000 – 24/7 on call
Emergency assistance and avenue to report a crime. Latest updates regarding safety, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible.

USC Department of Public Safety - UPC: (213) 740-6000, HSC: (323) 442-1200 – 24/7 on call
Non-emergency assistance or information.

Office of the Ombuds - (213) 821-9556 (UPC) / (323-442-0382 (HSC)
A safe and confidential place to share your USC-related issues with a University Ombuds who will work with you to explore options or paths to manage your concern.

Occupational Therapy Faculty Practice - (323) 442-2850 or otfp@med.usc.edu
Confidential Lifestyle Redesign services for USC students to support health promoting habits and routines that enhance quality of life and academic performance.