Georgia Institute of Technology
Course Syllabus: CS7210 Distributed Computing

OMSCS Spring 2021, College of Computing
Delivery: 100% Web-Based, Asynchronous on Canvas
Dates course will run: Jan. 14, 2021 – May. 8, 2021

Instructor Information

Dr. Ada Gavrilovska	Office: KACB 3348
	Email: ada@cc.gatech.edu
Weekly Office Hours via Blue Jeans	**Teaching Assistants:** William (Tony) Mason, fsgeek@gatech.edu Shangru Yi, syi73@gatech.edu Sreepraneeth Kotaguddam, sk304@gatech.edu

General Course Information

Description
This course will cover both fundamental concepts in distributed computing and discuss system designs enabling distributed applications. The objectives of the course include: In-depth understanding of core concepts of distributed computing, including study of both abstract concepts and practical techniques for building system support for distributed applications; construction of distributed system components by doing project work; and understanding of the current state of the art in several areas of distributed systems.

Pre- &/or Co-Requisites
No pre-requisites are enforced. However, you are expected to enter the course with graduate-level understanding of computer systems and/or computer networking, such as you would receive from CS6200, CS6210, or CS6250.

Course Goals and Learning Outcomes

Course Learning Outcomes
There are several learning outcomes of this course, centered around the following main topics:
1. *Distributed systems concepts, protocols and mechanisms*: Upon successful completion of this course, students will be able to understand the core concepts in distributed computing, such as logical clocks, consistent cuts, consensus, replication and fault tolerance, etc., as well as the common techniques to implement them, so as to build correct and performant distributed systems and applications.
2. *State-of-the-art developments in distributed systems*: Upon successful completion of this course, students will be familiar with multiple practical distributed systems which underpin popular real-world distributed applications and services and will be able to understand how such systems are realized by relying on core distributed computing concepts.
3. *Interplay of emerging technologies and paradigms and distributed systems*: Upon successful completion of this course, students will be able to understand how emerging technology trends concerning new hardware and applications, influence the designs and implementations of distributed systems.
4. *Practical experience with implementing distributed computing techniques*: Upon successful completion of this course, students will be able to solve distributed computing problems and to develop distributed services and applications, by gaining practical experience through a series of programming assignments.

Course Materials

Georgia Institute of Technology
Course Syllabus: CS7210 Distributed Computing

No textbook is required for this course. The course material includes required reading of research papers (often partial reading only) associated with each lecture. All research papers, required reading, and recommended reading will be provided as PDFs.

You can find free online textbook resources at these links:
- *Distributed Systems for Fun and Profit*, http://book.mixu.net/distsys/
- *Distributed Systems,* https://www.distributed-systems.net/index.php/books/ds3/

Classroom Management Tools
- Video Lectures: All video lectures are located on Canvas.
- Projects: are located on Canvas, submitted via Gradescope
- Reading Materials: are located on Canvas.
- Piazza Discussion: are located on Piazza and can be accessed via Canvas.
- Grades: are located on Canvas.
- Exams: are located on Canvas using HonorLock

Course Requirements, Assignments & Grading

Assignment Distribution and Grading Scale:

Assignments	Weight
Programming assignments	55%
Mid-Term Exam	20%
Final Exam	20%
Class Participation	5%
Total	**100%**

The course will include 5 programming assignments in Java, based on the Distributed Systems Labs (DSLabs): https://github.com/emichael/dslabs:
- **Project 0 DSLabs Intro**: students are guided to get familiar with the project framework and programming routines – *5% of course grade, 2 weeks estimated (due Jan. 31)*
- **Project 1 Client-Server**: an exactly-once RPC protocol is implemented and tested on top of an asynchronous network – *10% of course grade, 2 weeks estimated (due Feb. 14)*
- **Project 2 Primary-Backup**: a classical primary-backup protocol for fault-tolerance is implemented – *10% of course grade, 2 weeks estimated (due Mar. 7)*
- **Project 3 Paxos**: the *PAXOS* protocol for system consensus is implemented. Students will follow the lecture contents and some published papers – *15% of course grade, 3 weeks estimated (due Mar. 28)*
- **Project 4 Sharded KV Store**: a sharded key/value store out of multiple replica groups, with each of them using PAXOS for requests ordering internally, is built. A two-phase commit protocol is introduced for transactions – *15% of course grade, 4 weeks estimated (due Apr. 25)*

The exams will be based on the material covered in the course video lessons and on the assigned required reading. The exams will be open notes, proctored via HonorLock, and are not cumulative.

Class participation will consider your participation on the class Piazza site, or on the course informal Slack channel.

Grading Scale
Your final grade will be assigned as a letter grade according to the following scale:
A   90-100%
B   80-89%

Course Syllabus: CS7210 Distributed Computing

C	70-79%
D	60-69%
F	0-59%

Assignments Due Dates (Time Zone)
All assignment due dates are on Sundays at 11:59PM Anywhere on Earth time. We will not accept assignments submitted late due to time zone issues. We recommend changing your time zone in Canvas to show the due date in your local time. There are no exceptions.

Late Policy
No time extensions will be allowed, and late assignments or exams will not be accepted. If you are aware of a conflict with the scheduled time intervals for the exam, contact the instructor immediately.

In the event of an emergency (illness, family emergencies) please contact the Dean of Students office. The Dean of Students is equipped to verify emergencies and pass confirmation on to all your classes. For consistency, we ask all students to do this in the event of an emergency.

Office Hours
This class uses BlueJeans and the chat tool Piazza for its office hours. The weekly schedule for BlueJeans office hours will be announced as a pinned post in Piazza, and a recording link will be available after the session. The teaching team will also monitor Piazza and typically respond by end-of-day. Whenever possible, we encourage students to post Piazza questions as public questions, so as to allow for the entire class to benefit from the conversation, and to allow for faster responses from peers.

Feedback
Every semester, we make changes and tweaks to the course formula. As a result, every semester we try some new things, and some of these things may not work. We ask your patience and support as we figure things out, and in return, we promise that we, too, will be fair and understanding, especially with anything that might impact your grade or performance in the class. Second, we want to consistently get feedback on how we can improve and expand the course for future iterations. You can take advantage of the feedback box on Piazza (especially if you want to gather input from others in the class), give us feedback on the surveys, or contact us directly via private Piazza messages.

Technology Requirements and Skills

Computer Hardware and Software
- High-speed Internet connection
- Laptop or desktop computer with a minimum of a 2 GHz processor and 2 GB of RAM
- Windows for PC computers OR Mac iOS for Apple computers.
- Complete Microsoft Office Suite or comparable and ability to use Adobe PDF software (install, download, open and convert)
- Mozilla Firefox, Chrome and/or Safari browsers (*Note: HonorLock requires students to use Chrome*)

Canvas
This class will use Canvas to deliver course materials to online students. ALL course materials and assessments will take place on this platform.

Proctoring Information

Georgia Institute of Technology
Course Syllabus: CS7210 Distributed Computing

The midterm exam will be proctored with open notes, pen/paper, and calculator allowed. A proctored exam is similar to the one you would take in the classroom. These exams are delivered via a tool called HonorLock. HonorLock is an online proctoring service that allows you to take your exam from the comfort of your home. You DO NOT need to create an account, download software or schedule an appointment in advance. HonorLock is available 24/7 and all that is needed is a computer, a working webcam, and a stable Internet connection. To get started, you will need Google Chrome and to download the HonorLock Chrome Extension. You can download the extension at: www.honorlock.com/extension/install

When you are ready to take the exam, you would log into CANVAS, go to the course, and click on the exam link. Clicking Launch Proctoring will begin the HonorLock authentication process, where you will take a picture of yourself, show your ID, and complete a scan of your room. HonorLock will be recording your exam session by webcam as well as recording your screen. HonorLock also has an integrity algorithm that can detect search-engine use, so please do not attempt to search for answers, even if it's on a secondary device. If you encounter any issues, you may contact them by live chat, phone (844-243-2500), and/or email (support@honorlock.com).

Course Policies, Expectations & Guidelines

Communication Policy
You are responsible for knowing the following information:
- Anything posted to this syllabus
- Anything emailed directly to you by the teaching team (including announcements via Piazza), 24 hours after receiving such an email.

Because Piazza announcements are emailed to you, you need only to check your Georgia Tech email once every 24 hours to remain up-to-date on new information. Georgia Tech recommends students to check their Georgia Tech email once every 24 hours. You will not be responsible for the contents of a time-sensitive message or announcement until 24 hours after it has been sent.

We generally prefer to handle communication via Piazza to help with collaboration among the teaching team, but we understand Piazza is not ideal for having information "pushed" to you. We may contact you via a private Piazza post instead of an email, but if we do so, we will choose to send email notifications immediately, bypassing your individual settings, in order to ensure you're alerted. As such, this type of communication will also spring under #2 above.

Note that this means you won't be responsible for knowing information communicated in several other methods we'll be using. You aren't responsible for knowing anything posted to Piazza that isn't linked from an official announcement. You don't need to worry about missing critical information so long as you keep up with your email and understand the documents on this web site. This also applies in reverse: we do **not** monitor our Canvas message boxes and we may not respond to direct emails. If you need to get in touch with the course staff, please post privately to Piazza (preferably to all Instructors, in order to ensure a timely response, or to an instructor individually, if you feel the content is sensitive and should not be shared) or tag the instructor in the relevant post.

Online Student Conduct and (N)etiquette
Communicating appropriately in the online classroom can be challenging. In order to minimize this challenge, it is important to remember several points of "internet etiquette"
that will smooth communication for both students and instructors:
- *Read first, Write later*. Read the ENTIRE set of posts/comments on a discussion board before posting your reply, in order to prevent repeating commentary or asking questions that have already been answered.

Course Syllabus: CS7210 Distributed Computing

- *Avoid language that may come across as strong or offensive.* Language can be easily misinterpreted in written electronic communication. Review email and discussion board posts BEFORE submitting. Humor and sarcasm may be easily misinterpreted by your reader(s). Try to be as matter-of-fact and professional as possible.
- *Follow the language rules of the Internet.* Do not write using all capital letters, because it will appear as shouting. Also, the use of emoticons can be helpful when used to convey nonverbal feelings. ☺
- *Consider the privacy of others.* Ask permission prior to giving out a classmate's email address or other information.
- *Keep attachments small.* If it is necessary to send pictures, change the size to an acceptable 250kb or less (one free, web-based tool to try is picresize.com).
- *No inappropriate material.* Do not forward virus warnings, chain letters, jokes, etc. to classmates or instructors. The sharing of pornographic material is forbidden.

NOTE: The instructor reserves the right to remove posts that are not collegial in nature and/or do not meet the Online Student Conduct and Etiquette guidelines listed above.

University Use of Electronic Email
A university-assigned student e-mail account is the official university means of communication with all students at Georgia Institute of Technology. Students are responsible for all information sent to them via their university-assigned e-mail account. If a student chooses to forward information in their university e-mail account, he or she is responsible for all information, including attachments, sent to any other e-mail account. To stay current with university information, students are expected to check their official university e-mail account and other electronic communications on a frequent and consistent basis. Recognizing that some communications may be time-critical, the university recommends that electronic communications be checked minimally twice a week.

Plagiarism & Academic Integrity
Students are expected to complete the assignments and exams independently and to submit their own original work. Students will be required to acknowledge they have read the course Collaboration vs. Cheating policy (details will be announced via Canvas).

Georgia Tech aims to cultivate a community based on trust, academic integrity, and honor. Students are expected to act according to the highest ethical standards. All students enrolled at Georgia Tech, and all its campuses, are to perform their academic work according to standards set by faculty members, departments, schools and colleges of the university; and cheating and plagiarism constitute fraudulent misrepresentation for which no credit can be given and for which appropriate sanctions are warranted and will be applied. For information on Georgia Tech's Academic Honor Code, please visit http://www.catalog.gatech.edu/policies/honor-code/ or http://www.catalog.gatech.edu/rules/18/.

Any student suspected of cheating or plagiarizing on an exam, exercise, or assignment will be reported to the Office of Student Integrity, who will investigate the incident and identify the appropriate penalty for violations.

Accommodations for Students with Disabilities
If you are a student with learning needs that require special accommodation, contact the Office of Disability Services at (404) 894-2563 or http://disabilityservices.gatech.edu/, as soon as possible, to make an appointment to discuss your special needs and to obtain an accommodations letter. Please also e-mail me as soon as possible in order to set up a time to discuss your learning needs.

Student-Faculty Expectations Agreement
At Georgia Tech we believe that it is important to strive for an atmosphere of mutual respect, acknowledgement, and responsibility between faculty members and the student body. See

Course Syllabus: CS7210 Distributed Computing

http://www.catalog.gatech.edu/rules/22/ for an articulation of some basic expectation that you can have of me and that I have of you. In the end, simple respect for knowledge, hard work, and cordial interactions will help build the environment we seek. Therefore, I encourage you to remain committed to the ideals of Georgia Tech while in this class.

Subject to Change Statement
The syllabus and course schedule may be subject to change. Changes will be communicated via the Canvas announcement tool. It is the responsibility of students to check Piazza, email messages, and course announcements to stay current in their online courses.

Georgia Institute of Technology

Course Syllabus: CS7210 Distributed Computing

Course Schedule

Tentative course schedule, any updates will be communicated via Canvas announcements.
Most of the required papers are only *partial readings*, based on the portion of the paper content covered in the lessons.

Week/Dates	Topics	Deliverables
1 **January 11/18**	**Lesson 1 Introduction to Distributed Systems** **Required Readings:** - Schneider, <u>What Good are Models and What Models are Good?</u> (partial) - <u>Fallacies Explained</u> (partial) **Optional Suggested Readings:** - <u>Google's Introduction to Distributed Systems</u> **Lesson 2 Primer of RPC** **Optional Suggested Readings:** - RPC in DS textbook, 4.1-4.2	**Acknowledge understanding of Collaboration and Cheating policy**
2 **January 25**	**Lesson 3 Time in Distributed Systems** **Required Readings:** - Raynal and M. Singhal. <u>Logical Time: A Way to Capture Causality in Distributed Systems (Links to an external site.)</u>. IRISA Technical Report. (up to Section 7) **Optional Suggested Readings:** - Lamport, <u>Time, Clocks and The Ordering of Events in Distributed Systems</u>	**Project 1: Intro to DS Labs (due 1/31)**
3 **February 1**	**Lesson 4 State in Distributed Systems** **Required Readings:** - Babaoglu, Marzullo, <u>Consistent Global States of Distributed Systems (Links to an external site.)</u>. (Sections 4, 6, and 13) **Optional Suggested Readings:** - Chandy, M. and Lamport, L., <u>Distributed Snapshots: Determining Global States of Distributed Systems</u>, ACM Trans. on Computer Systems, February 1985.	
4 **February 8**	**Lesson 5 Consensus** **Required Readings:** - J. Fischer, N. Lynch and M. S. Patterson, <u>Impossibility of distributed consensus with one faulty process</u>, JACM 32, 1985. **Optional Suggested Readings:** - Replication chapter in For Fun and Profit	**Project 2: Client-server (due 2/14)**
5 **February 15**	**Lesson 6 PAXOS and Friends** **Required Readings:** - Leslie Lamport, <u>Paxos Made Simple</u>, Microsoft Research (partial) - <u>A Brief Analysis of Consensus Protocols: From Logical Clock to Raft</u> (partial) **Optional Suggested Readings:** - Ongaro, Ousterhout, <u>In Search of an Understandable Consensus Algorithm</u>	
6 **February 22**	**Lesson 7 Replication** **Required Reading:**	

Week/Dates	Topics	Deliverables
	- Chain Replication for Supporting High Throughput and Availability (partial) **Lesson 8 Fault-tolerance** **Required Reading:** - A Survey of Rollback-Recovery Protocols in Message-Passing Systems (partial)	
7 **March 1**	**Lesson 9 Distributed Transactions** **Required Readings:** - Spanner: Google's Globally-Distributed Database, OSDI 2012. **Optional Suggested Readings:** - Amazon Aurora: On Avoiding Distributed Consensus on I/O, Commits and Membership Changes	**Project 3: Primary-backup (due 3/7)**
8 **March 8**	**Midterm review and exam**	**Midterm exam window: 3/11-3/15**
9 **March 15**	**Lesson 10 Consistency and Geo-Distributed Data Stores** **Required Readings:** - Scaling Memcache at Facebook - Wyatt Lloyd, et. al., Don't Settle for Eventual: Scalable Causal Consistency for Wide Area Storage with COPS, SOSP 2011. (partial)	**GT "drop-date" 3/17 at 4pm ET**
10 **March 22**	**Lesson 11 Peer-to-peer, Mobility** **Required Readings (partial):** - Stoica, I., Morris, R., Karger, D., Kaashoek, M. F. and Balakrishnan, H., Chord: A Scalable Peer-to-Peer Lookup Service for Internet Applications, TON. - Badrinath et. al., Designing Distributed Algorithms for Mobile Computing Networks. (partial)	**Project 4: Paxos (due 3/28)**
11 **March 29**	**Lesson 12 Distributed Data Analytics** **Required Readings:** - Zaharia et al., Resilient Distributed Datasets: A Fault Tolerant Abstraction for In-Memory Cluster Computing, NSDI'12. **Optional Suggested Readings:** - MapReduce: Simplified Data Processing on Large Clusters, OSDI'04	
12 **April 5**	**Lesson 13 Distributed Machine Learning** **Required Readings:** - Hsieh et al., Gaia: Geo-Distributed Machine Learning Approaching LAN Speeds, NSDI'17. **Optional Suggested Readings:** - Federated Learning - Cartel: Collaborative Transfer Learning for the Edge	
13 **April 12**	**Lesson 14 Support for Datacenter-based Distributed Computing**	

Week/Dates	Topics	Deliverables
	Required Readings (partial): - FaSST: Fast, Scalable and Simple Distributed Transactions with Two-Sided (RDMA) Datagram RPCs - Verma, et al., Large-scale Cluster Management at Google with Borg, Eurosys'15. - LegoOS: A Disseminated, Distributed OS for Hardware Resource Disaggregation, OSDI'18 **Optional Suggested Readings:** - Tumanov et al., TetriSched: Global Rescheduling with Adaptive Plan-ahead in Dynamic Heterogeneous Clusters (Links to an external site.), Eurosys'17 - Serverless in the Wild: Characterizing and Optimizing the Serverless Workload at a Large Cloud Provider, ATC'20	
14 **April 19**	**Lesson 15 Byzantine Fault Tolerance, Blockchain** **Required Readings:** - Castro, B. Liskov, Practical Byzantine Fault Tolerance, OSDI, Feb. 1999.	**Project 5: FT KV store (due 4/25)**
15 **April 26**	**Lesson 16 Edge Computing, IoT** **Required Readings:** - Satya et al. The Computing Landscape of the 21st Century, HotMobile'19 **Optional Suggested Readings:** - A. Sengupta, et al., Transactuations: Where Transactions Meet the Physical World, USENIX ATC, 2019.	
16 **May 3**		**Final Exam window: 4/29 – 5/4**
17 **May 6**		**End-of-Course Survey, CIOS Survey, Peer Feedback**