

CSE6350 Syllabus, Fall 2016

No: CSE6350

Title: File and Storage System Infrastructure in Data Centers Supporting Internet-wide Services

Instructor: Dr. Song Jiang, Associate Professor of Computer Science and Engineering

Office Hours: M/W 2:00pm – 3:00pm

Office Location: ERB 559

Phone: (817) 272-3610

Email: song.jiang@uta.edu

Meeting Time: M/W 4:00-5:20pm

Course Meeting Location: GS 109

Web Page: <http://ranger.uta.edu/~sjiang/CSE6350-fall-16/index.htm>

Course Description

The course will allow students to know principles of Internet-wide distributed and cloud computing systems, to understand layers of system and application software in data centers for providing Internet-wide scalable and reliable services, to learn techniques for addressing critical issues such as data replication and consistency, failure management, system reliability, scalability, availability, and efficiency, to apply techniques of configuring a distributed computing platform matching expected workload characteristics, and to be prepared for better job opportunities in the area of distributed and cloud computing.

After completing this course, students should be able to do the following:

1. Explain various software layers in a data center supporting Internet-scale services.
2. Explain various critical issues that must be addressed in building a data-center's software infrastructure, and state-of-the-art techniques to address them.
3. Choose design alternatives based on workload characteristics.
4. Design resource management policies according to given optimization goals.

Textbook:

- 1) Operating System Concepts, 9th Edition, by Abraham Silberschatz, Peter B. Galvin, and Greg Gagne, Wiley Publishing, 2012.
- 2) Additional lecture notes and research papers provided by instructor

Prerequisites:

- 1) Operating System, Computer Architecture, and Parallel Computing
- 2) Knowledge of UNIX/Linux systems and high-level language C

Topics:

1. Revisiting components in a typical operating system
 - Separation of user- and kernel modes
 - Process management
 - Memory management
 - File system and distributed file system
2. Overview of distributed systems
 - Distributed system and cloud computing
 - The issues: reliability, scalability, availability, consistency, and performance.
3. Distributed Storage Systems
 - Google's GFS: the Google file system
 - Facebook's Haystack: Facebook's photo storage
 - Microsoft's Azure: a highly available cloud storage service with strong consistency
 - Ceph: a scalable, high-performance distributed file system
4. Key-value data management systems
 - LevelDB: a key-value store based on LSM tree
 - Google's Bigtable: a distributed storage system for structured data
 - SILT: a memory-efficient key-value store
 - Amazon's Dynamo: a highly available key-value store
5. Distributed Execution Frameworks
 - Google's MapReduce: simplified data processing on large clusters
 - Google's Pregel: a system for large-scale graph processing
 - Microsoft's Dryad: constructing distributed data-parallel programs from sequential building blocks
 - Berkeley's Spark and Shark

Grading Policy

Grading scale

- 1) A: [90, 100]
- 2) B: [80, 90)
- 3) C: [70, 80)
- 4) D: [60, 70)
- 5) F: below 60

Distribution of Points:

- 1) In-class discussion and attendance: 10%
- 2) Homework assignments: 15%
- 3) Projects: 20%
- 4) Presentation and Q&A report 15%
- 5) Midterm exam: 15%

6) Final exam: 25%

Makeup Exam and Makeup Assignment Policy:

No make-up exams will be given except for university sanctioned excused absences. If you miss an exam (for a good reason), it is your responsibility to contact instructor before the exam.

Late Policy

Each student has three late days. Clearly indicate this on your submission if late days are used. Beyond these days, homework assignments and projects must be handed in by due time and no late assignments will be accepted unless compelling reasons are provided and verified.

Course Policy

Americans with Disabilities Act

The University of Texas at Arlington is on record as being committed to both the spirit and letter of federal equal opportunity legislation; reference Public Law 93112 -- The Rehabilitation Act of 1973 as amended. With the passage of new federal legislation entitled Americans With Disabilities Act - (ADA), pursuant to section 504 of The Rehabilitation Act, there is renewed focus on providing this population with the same opportunities enjoyed by all citizens. As a faculty member, I am required by law to provide "reasonable accommodation" to students with disabilities, so as not to discriminate on the basis of that disability. Student responsibility primarily rests with informing faculty at the beginning of the semester and in providing authorized documentation through designated administrative channels.

Academic Integrity

It is the philosophy of The University of Texas at Arlington that academic dishonesty is a completely unacceptable mode of conduct and will not be tolerated in any form. All persons involved in academic dishonesty will be disciplined in accordance with University regulations and procedures. Discipline may include suspension or expulsion from the University. "Scholastic dishonesty includes but is not limited to cheating, plagiarism, collusion, the submission for credit of any work or materials that are attributable in whole or in part to another person, taking an examination for another person, any act designed to give unfair advantage to a student or the attempt to commit such acts." (Regents' Rules and Regulations, Part One, Chapter VI, Section 3, Subsection 3.2, Subdivision 3.22)

Please acknowledge the following Honor code in all submissions:

I pledge, on my honor, to uphold UT Arlington's tradition of academic integrity, a tradition that values hard work and honest effort in the pursuit of academic excellence. I promise that I will submit only work that I personally create or contribute to group collaborations, and I will appropriately reference any work from other sources. I will follow the highest standards of integrity and uphold the spirit of the Honor Code.

Drop Policy

Students may drop or swap (concurrently add and drop) classes through MyMav self-service throughout the registration period. After the late-registration period, students must see their academic advisor to drop a class or withdraw. Undeclared students must see an advisor in the University Advising Center. Drops can continue through a point two-thirds of the way through the term or session. It is the student's responsibility to officially withdraw, if he/she does not plan to attend after registering.

Emergency Exit Procedures

Should we experience an emergency event that requires us to vacate the building, students should exit the room and move toward the nearest exit. When exiting the building during an emergency, one should never take an elevator but should use the stairwells. Faculty members and instructional staff will assist students in selecting the safest route for evacuation and will make arrangements to assist individuals with disabilities. How best to exit the building will be explained on day one. (See also http://www.uta.edu/campus-ops/ehs/fire/Evac_Maps_Buildings.php and <http://www.uta.edu/police/EvacuationProcedures.pdf>.)

Student Support Services

UT Arlington provides a variety of resources and programs designed to help students develop academic skills, deal with personal situations, and better understand concepts and information related to their courses. Resources include tutoring, major-based learning centers, developmental education, advising and mentoring, personal counseling, and federally funded programs. For individualized referrals, students may visit the reception desk at University College (Ransom Hall), call the Maverick Resource Hotline at 817-272-6107, send a message to toresources@uta.edu, or view the information at <http://www.uta.edu/universitycollege/resources/index.php>.