CSCI 485 - File and Database Management
Units: 4
Fall 2023 - Mondays and Wednesdays 10-11:50 am

Instructor: Shahram Ghandeharizadeh
Office: SAL 208
Office Hours: Mon 12:30-1:30 pm, Wed 4-5:30 pm
Contact Info: shahram@usc.edu, 213-740-4781.

Course Description

This course introduces students to the basics of a file system, a relational database management system, and a graph data store. Topics covered include Entity-Relationship as a conceptual data modeling tool, relational data model and algebra, SQL, normal forms, fundamentals of a magnetic disk drive and solid state drives including their working details, transactions and their ACID properties, cache management techniques such as LRU-K, concurrency control protocols such as locking and time-stamp, and crash recovery protocols such as logging. We present and discuss several distributed and parallel file systems and key-value stores including Google File System (GFS) and FoundationDB (FDB), a peer-to-peer file system named CAN, and blockchain systems. This course includes a hands-on implementation of SQL and its execution engine using FDB.

Learning Objectives

By the end of the semester, students should have a grasp of fundamental concepts, design decisions, protocols and algorithms to build a file system (FS) and a relational database management system (RDBMS) using off-the-shelf software components. They should understand ACID semantics and their application to both FSs and RDBMSs. Finally, they should be able to provide an abstraction of today's hardware in support of database management systems.

Prerequisite(s): Principles of Software Development (CSCI 201)

Recommended Preparation: A good understanding of in-memory data structures such as record and their organizations in arrays, trees, hash tables, etc.

Course Notes:
- All lecture material will be posted on the USC blackboard system.
- The required textbook for this course is *Database System Concepts* by Silberschatz, Korth and Sudarshan, 7th edition, available in the bookstore.
- Grading breakdown
    - Exam 1: 30%
    - Exam 2: 30%
    - Assignments: 10%
    - Class Participation: 10%
    - Project: 20%
- This course has no comprehensive final exam. Exam 2 is held during the last lecture and covers material presented since Midterm 1.

Communication

This course uses the USC blackboard system to post lectures, make announcements, and disseminate assigned readings, assignments, and class project.

All assignments and projects must be submitted using the USC blackboard system. No e-mail submissions are accepted. All deadlines are firm. All late assignments receive a zero.

Technological Proficiency and Hardware/Software Required

Required Readings and Supplementary Materials

This course assumes students are Internet savvy and capable of using it as a digital library. This includes use of ACM/IEEE/Springer digital libraries. All USC students have access to these digital collections automatically.

Students must be proficient in use of the USC Blackboard system.

In addition, the course assumes students have used github (https://github.com/) for software management, Java programming language and an IDE such as Eclipse for software development, and a collaboration tool such as JIRA.

Description and Assessment of Assignments

Assignments are aligned with the learning objectives, meaning that each assignment serves to measure student performance on at least one learning objective. The first few assignments focus on ER, relational data model, implementation of relational data model using MySQL, and ability to write SQL queries. The project focuses on file system design, networked components, block-based devices, high data availability, and concurrency control protocols.

Grading Breakdown

Based on prior experience teaching this course, students who attend lectures do very well. In specific, students who skip lectures do poorly on Exam 1 and spend the second half of the semester concerned with their grade. To prevent this, classroom attendance is a requirement for this course. It is worth 10% of the final grade.

Assessment Tool (assignments)	Points	% of Grade
Class Participation	100	10%
Assignment	100	10%
Exam 1	100	30%
Exam 2	100	30%
Project	100	20%
TOTAL		**100%**

Class Participation: How it Works?

I will ask questions during the lectures and select a name from the roster randomly. The student present may answer the question. If the student is absent for 3 different lectures then their class participation grade will be zero.

Grading Scale

The final letter grade is based on a curve.

Course-specific Policies (Assignment Submission, Grading Timeline, Late work, and Technology)

All deadlines are firm. No late submissions are accepted. No e-mail submissions. Blackboard submissions are graded only.

Attendance

Attendance is mandatory. Based on prior experience, students who do not attend lectures perform poorly on exams. CSCI 485 accommodates student athletes with approved Travel Request Letters and students who give advance notice of religious observation.

Spring 2023 Course Schedule: A Weekly Breakdown

	Topics/Daily Activities	Readings/Preparation	Deliverables
Week 1 8/21 & 23	Introduction, ER	Chapter 7	Prepare for Assignments 1 and 2 on data modeling
Week 2 8/28 & 30	Relational and Graph Data Models	Sections 2.1-2.3	Logical data models
Week 3 Labor day, 9/6	Normal forms	A Simple Guide to Five Normal Forms in Relational Database Theory by W. Kent. Communications of the ACM, Volume 25, Issue 2, February 1983.	Duplicated data, data anomalies, and data loss
Week 4 9/11 & 13	SQL, Relational Algebra	Chapter 3, Section 6.2	Assignment 1 due
Week 5 9/18 & 20	Hard Disk Drive and file system design, index structures.	Chapter 10, ignore 10.3, Chapter 11	Assignment 2 due, Physical database design using B+-tree and hash indexes
Week 6 9/25 & 27	Query Processing & Optimization	Chapters 12 and 13	Cost driven heuristics
Week 7 10/2 & 4	Transactions & roll-out the project	Chapter 14	**ACID semantics of transactions.**
Week 8 10/9 & 11	**Review for Exam 1 Exam 1 is on 10/11**		
Week 9 10/16 & 18	Parallel file system design	The Google File System by S. Ghemawat, et. al., SOSP 2003. FoundationDB by Zhu, et. al., SIGMOD 2021.	Project Starts
Week 10 10/23 & 25	Spring Recess	Spring Recess	Spring Recess
Week 11 10/30 & 11/1	Crash Recovery Protocols	Chapter 16	Protocols to preserve atomicity.
Week 12 11/6 & 8	Concurrency Control	Chapter 15	Undesirable race conditions and data consistency
Week 13 11/13 & 15	Memory Mgmt + Performance modeling	The LRU-K Page Replacement Algorithm for Database Disk Buffering by E. J. O'Neil, P. E. O'Neil, and G. Weikum. SIGMOD 1993.	Algorithms to enhance cache hit

Week 14 11/20 & Thanksgiving	Peer-to-Peer file systems	The Scalable Content-Addressable Networks by S. Ratnasamy et. al., SIGCOMM 2001.	Decentralized file sharing
Week 15 11/27 & 29 Time permitting	Enterprise storage management, blockchain	A Hichhiker's Guide to the Blockchain Universe by J. Waldo. Communications of the ACM, March 2019, Vol. 62, No. 3.	Replicated ledgers
Exam 2 11/27 & 29	**Review for Exam 2** **Exam 2 is on 11/29**		

Statement on Academic Conduct and Support Systems

Academic Conduct:

Plagiarism – presenting someone else's ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in SCampus in Part B, Section 11, "Behavior Violating University Standards" policy.usc.edu/scampus-part-b. Other forms of academic dishonesty are equally unacceptable. See additional information in SCampus and university policies on Research and Scholarship Misconduct.

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
studenthealth.usc.edu/counseling
Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention.

National Suicide Prevention Lifeline - 1 (800) 273-8255 – 24/7 on call
suicidepreventionlifeline.org
Free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week.

Relationship and Sexual Violence Prevention Services (RSVP) - (213) 740-9355(WELL), press "0" after hours – 24/7 on call

studenthealth.usc.edu/sexual-assault
Free and confidential therapy services, workshops, and training for situations related to gender-based harm.

Office for Equity, Equal Opportunity, and Title IX (EEO-TIX) - (213) 740-5086
eeotix.usc.edu
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
usc-advocate.symplicity.com/care_report
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.usc.edu
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) 821-4710
campussupport.usc.edu
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
diversity.usc.edu
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
dps.usc.edu, emergency.usc.edu
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) 442-120 – 24/7 on call
dps.usc.edu
Non-emergency assistance or information.

Office of the Ombuds - (213) 821-9556 (UPC) / (323-442-0382 (HSC)
ombuds.usc.edu
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) 442-3340 or otfp@med.usc.edu
chan.usc.edu/otfp
Confidential Lifestyle Redesign services for USC students to support health promoting habits and routines that enhance quality of life and academic performance.