CSC 460 (Database Design)

Spring 2024

GS906, Tuesday & Thursday 8:00 – 9:15AM

Course Description

Emphasis on DBMS architecture and implementation issues such as storage structures, multidimensional index structures, query optimization, concurrency control and recovery, and distributed database systems. This course will cover database design or SQL programming as well

Course Prerequisites

CSC 335 and CSC 345

Instructor and Contact Information

Lei Cao
712 Gould-Simpson
caolei@arizona.edu
520-621-4632

Office hours: Wednesday 2:00PM – 3:30PM at my office or by email appointment
Instructor home page: https://www.cs.arizona.edu/~caolei/
Piazza link: https://piazza.com/arizona/fall2024/csc460001, access code: csc460
D2L: https://d2l.arizona.edu/d2l/home/1491466
Grade Scope: https://www.gradescope.com/courses/821300, access code: 7EENB7
Course home page: https://www.cs.arizona.edu/~caolei/teach.html

Course Format and Teaching Methods

Classes will consist of lectures, discussions, and student presentation based on readings from the database literature. There will be two exams and 8 assignments – 5 programming "Labs" and 3 problem sets.

Course Objectives

This course is designed to introduce senior undergraduate students to the foundations of database systems, focusing on basics such as the relational algebra and data model, query optimization, query processing, and transactions.

Students will:

- Learn how information is stored in a database.
- Learn the foundational algorithms used to implement core DBMS operations.
- Learn how DBMS operations are policed to ensure correct executions.
- Learn the core ideas of recovery from incomplete data manipulation actions.

Expected Learning Outcomes

Course learning outcomes include all of the following:

- The ability to implement a basic data I/O subsystem.

- Understand the key principles of database query optimizer and implement a simple prototype.
- Explain in detail multiple implementation options for relational algebra operators.
- Demonstrate the ability to implement core DBMS algorithms.
- Discuss the pros and cons of transaction concurrency control techniques.
- Demonstrate a clear understanding of the key principles of recovery systems based on write-ahead logging.

Makeup Policy for Students Who Register Late

Students who register after the first class meeting may make up missed assignments/projects at a deadline set in consultation with the instructor.

Course Communications

We will use official UA email and Piazza as the primary mode of contact. D2L will be used to provide grading and feedback

Required Texts or Readings

The course readings will primarily be drawn from the 5th Edition of ''Readings in Database Systems'', edited by Stonebraker and Hellerstein. It is available online at this website. Note that PDFs of all the papers in the book are not necessarily linked from the website; we will include PDFs in reading assignments.

Other textbooks:

Database System Concepts (Silberschatz/Korth/Sudarshan), U.S. 7th ed., McGraw-Hill, 2019. It is available through D2L ebook.
Database Management Systems (Raghu Ramakrishnan, Johannes Gehrke), 3rd Edition.

Supplemental Readings:
There will be several other readings that will be posted on the course web site.

Assignments and Examinations: Schedule/Due Dates

8 assignments including 5 programing labs and 3 problem sets. See scheduled topics and activities for due dates.
2 exams: mid-term (10/3) and final exam (12/19).

Final Examination

Friday, December 19, 2024, 8:00 a.m. - 10:00 a.m. The final is required, is comprehensive, and will be given on this date at this time. Make your end-of-semester travel plans accordingly.

Link to the Final Exam Regulations and Final Exam Schedule: https://registrar.arizona.edu/faculty-staff-resources/room-class-scheduling/schedule-classes/final-exams

Grading Scale and Policies

Grades are assigned based on labs, home works, 2 exams, and class participation. The grading breakdown is as follows:

- Assignment (Labs and Problem Sets): 60% total

- Exams: 15% each
- Class Participation: 10%

The final grade in the course is determined by the better of a per-class grading curve and overall performance:

- 90% or better: A;
- 80% or better: B;
- 70% or better: C;
- 60% or better: D;
- below 60%: E.

University policy regarding grades and grading systems is available at https://catalog.arizona.edu/policy/courses-credit/grading/grading-system

Incomplete (I) or Withdrawal (W):
Requests for incomplete (I) or withdrawal (W) must be made in accordance with University policies, which are available at https://catalog.arizona.edu/policy/courses-credit/grading/grading-system.

Dispute of Grade Policy:

If you wish to dispute your grade for an assignment, you have two weeks after the grade has been turned in. In addition, even if you only dispute one portion of the grading for that unit, I reserve the right to revisit the entire unit (assignment or project).

Scheduled Topic and Activities (Tentative)

The schedule may be adjusted based on instructor discretion.

Week	Date	Description
1	8/27	Lecture 1: course introduction, logistics, introduction to database. Assignments: Assigned: Lab 0
1	8/29	Lecture 2: SQL Readings: • Section 1 of "A Practical Introduction to Databases", by Christopher Painter-Wakefield. • Michael Stonebraker and Joseph Hellerstein. What Goes Around Comes Around. In "Readings in Database Systems" (aka the Red Book), or online here (link to PDF). • Section 3.1 of "A Practical Introduction to Databases", by Christopher Painter-Wakefield. Assignments: **Assigned: PS 1**

2	9/3	Lecture 3: Schema design Readings: Section 3.2 and 3.3 of "A Practical Introduction to Databases", by Christopher Painter-Wakefield and Section 2 of of "A Practical Introduction to Databases", by Christopher Painter-Wakefield. Assignments: **Assigned: Lab 1** Due: Lab 0
2	9/5	Lecture 4: Introduction to Database Internals Readings: We will start discussing database system internals, based on the content of the paper: Joseph Hellerstein and Michael Stonebraker. "Architecture of a Database System". PDF You only need to read sections: 1, 2, and 4 (up to subsection 4.5 included).
3	9/10	Lecture 5: Database Operators and Query Processing Readings: You will read the same paper as last lecture: Joseph Hellerstein, Michael Stonebraker and James Hamilton. "Architecture of a Database System". PDF This lecture will focus on different physical database operators, their implementation, and how query plans are actually executed. Assignments: Due PS 1
3	9/12	Lecture 6: Indexing and Access Methods Readings: This lecture will cover various issues related to the physical storage of relations on disk, as well as index data structures we might use to efficient access those stored relations. • Read Chapter 2 on B+Tree basics in Database Internals by Alex Petrov. • "The R*-Tree: An Efficient and Robust Access Method for Points and Rectangles." Beckmann et al, in The Red Book. Assignments:

		Assigned: Lab 2
4	9/17	Lecture 7: Join Algorithms Readings: This lecture will cover join algortihms, focusing in particular on Hash Join and Sort-Merge join and the relative tradeoffs of the two approaches. Read the following paper: L.D. Shapiro. Join Processing in Database Systems with Large Main Memories. Red Book. Assignments: **Assigned: PS 2**
4	9/19	Lecture 8: Query Optimization Readings: In this lecture, we will discuss query optimization, focusing on the algorithms proposed in the classic "Selinger" paper. Read the following paper: • Patricia Selinger, M. Astrahan, D. Chamberlin, Raymond Lorie, and T. Price. Access Path Selection in a Relational Database Management System. Proceedings of ACM SIGMOD, 1979. Pages 22-34. Red Book. [PDF] • *Optionally*, you may also wish to look at: Michael Mannino, Paichen Chu, and Thomas Sager. Statistical Profile Estimation in Database Systems. ACM Computing Surveys 20(3), 1988. Pages 191-221. [PDF]. This paper discusses many of the techniques that used to make query optimization and cost estimation practical in modern database systems. Assignments: Due: Lab 1
5	9/24	Lecture 9: Query Optimization (CONT.)
5	9/26	Lecture 10: Database Layout for Analytic Databases Readings: In this lecture, we will discuss column-oriented databases, which represent a different way of building a relational database that is optimized for large-scale read-intensive. • Daniel J. Abadi, Samuel R. Madden, Nabil Hachem. "Column-Stores vs. Row-Stores: How Different Are They Really?" SIGMOD, 2008. [PDF]

		Optional Reading: The original C-Store paper gives a bit more background about the features of C-Store, focusing less on the storage system and executor. Mike Stonebraker, Daniel Abadi, Adam Batkin, Xuedong Chen, Mitch Cherniack, Miguel Ferreira, Edmond Lau, Amerson Lin, Sam Madden, Elizabeth O'Neil, Pat O'Neil, Alex Rasin, Nga Tran and Stan Zdonik. "C-Store: A Column Oriented DBMS". VLDB, pages 553-564, 2005. [PDF]
6	10/1	Mid-term Review Assignments: • Due: PS 2
6	10/3	Mid-term
7	10/8	Lecture 11: Transactions And Locking Readings: In this class, we will begin our discussion of concurrency control and recovery. Please read: Michael J. Franklin. Concurrency Control and Recovery. The Computer Science and Engineering Handbook, 1997. [PDF] Assignments: **Assigned: Lab 3**
7	10/10	Lecture 12: Optimistic Concurrency Control and Snapshot Isolation Readings: H.T. Kung and John T. Robinson. "On Optimistic Methods for Concurrency Control." TODS, June, 1981. In Red Book. Assignments: Due: Lab 2
8	10/15	Lecture 13: Optimistic Concurrency Control and Snapshot Isolation (CONT.)
8	10/17	Lecture 14: Recovery Readings: "ARIES: A Transaction Recovery Method Supporting Fine-Granularity Locking and Partial Rollbacks Using Write-Ahead Logging". C. Mohan et al. In the Red Book. Read Sections 1-7, and skim Sections 12 and 13.

9	10/22	Lecture 15: Recovery (cont.)
9	10/24	Lecture 16: Advanced Cardinality Estimation Other: Project discussion Assignments: **Assigned: Lab 4** Due: Lab3
10	10/29	Lecture 17: Distributed Databases Assignments: **Assigned: PS 3**
10	10/31	Lecture 18: Distributed Transactions Readings: • C.Mohan, Bruce Lindsay, and R. Obermarck. Transaction Management in the R* Distributed Database Management Systems. ACM Transactions On Database Systems 11(4), 1986. In Red Book. This paper discusses distributed transactions, addressing the problem of providing ACID-style semantics in a shared nothing environment.
11	11/5	Lecture 19: Distributed Transactions (CONT.)
11	11/7	Lecture 20: NoSQL/Eventual Consistency Readings: • DeCandia et al. Dynamo: Amazon's Highly Available Key-value Store. In SOSP, 2007. [PDF] • Werner Vogels. Eventually Consistent - Revisited. All Things Distributed (Blog), December 2008. [Link] In this class, we will survey a range of "NoSQL" systems, which offer different consistency properties, or different query languages (or both) than those offered by relational database systems.
12	11/12	Lecture 21: High Performance Transaction Readings:

		• M. Stonebraker, S. Madden, D. J. Abadi, S. Harizopoulos, N. Hachem, and P. Helland. VLDB 2007. The End of an Architectural Era (It's Time for a Complete Rewrite) • [Optional] Alexander Thomson, Thaddeus Diamond, Shu-Chun Weng, Kun Ren, Philip Shao, and Daniel J. Abadi. SIGMOD 2012.Calvin: fast distributed transactions for partitioned database systems. • [Optional] Yi Lu, Xiangyao Yu, Lei Cao, Samuel Madden. VLDB 2021. Epoch-based Commit and Replication in Distributed OLTP Databases
12	11/14	Lecture 22: Cluster Computing (Spark) Readings: • Matei Zaharia, Mosharaf Chowdhury, Tathagata Das, Ankur Dave, Justin Ma, Murphy McCauley, Michael J. Franklin, Scott Shenker, Ion Stoica. Resilient Distributed Datasets: A Fault-Tolerant Abstraction for In-Memory Cluster Computing.[PDF]
13	11/19	Lecture 23: SnowFlake Readings: SnowFlake is an analytical database designed for the cloud. • Benoit Dageville, Thierry Cruanes, Marcin Zukowski, Vadim Antonov, Artin Avanes, Jon Bock, Jonathan Claybaugh, Daniel Engovatov, Martin Hentschel, Jiansheng Huang, Allison W. Lee, Ashish Motivala, Abdul Q. Munir, Steven Pelley, Peter Povinec, Greg Rahn, Spyridon Triantafyllis, Philipp Unterbrunner. The Snowflake Elastic Data Warehouse. Proceedings of ACM SIGMOD 2016. [PDF] Assignments: Due: Lab4
13	11/21	Lecture 24: Learned Data Systems Readings: Some recent work on using machine learning to optimize database systems. • Tim Kraska, Mohammad Alizadeh, Alex Beutel, Ed H. Chi, Jialin Ding, Ani Kristo, Guillaume Leclerc, Samuel Madden, Hongzi Mao, Vikram Nathan. SageDB: A Learned Database System. [PDF]

		Assignments: **Assigned: Lab 5** Due: PS 3
14	11/26	Lecture 25: Learned indexes Readings: • Tim Kraska, Alex Beutel, Ed H Chi, Jeffrey Dean, Neoklis Polyzotis. The case for learned index structures. [PDF] Optionally, you can read the following paper on steering a query optimizer: • Ryan Marcus, Parimarjan Negi, Hongzi Mao, Nesime Tatbul, Mohammad Alizadeh, Tim Kraska. Bao: Making Learned Query Optimization Practical. [PDF]
14	11/28	Thanksgiving
15	12/3	Lecture 26: Additional Topic (Buffer 1)
15	12/5	Lecture 27: Additional Topic (Buffer 2)
16	12/10	Final exam review Assignments: Due: Lab5

Classroom Behavior Policy

To foster a positive learning environment, students and instructors have a shared responsibility. We want a safe, welcoming, and inclusive environment where all of us feel comfortable with each other and where we can challenge ourselves to succeed. To that end, our focus is on the tasks at hand and not on extraneous activities (e.g., texting, chatting, reading a newspaper, making phone calls, web surfing, etc.).

Students are asked to refrain from disruptive conversations with people sitting around them during lecture. Students observed engaging in disruptive activity will be asked to cease this behavior. Those who continue to disrupt the class will be asked to leave lecture or discussion and may be reported to the Dean of Students.

Some learning styles are best served by using personal electronics, such as laptops and iPads. These devices can be distracting to other learners. Therefore, students who prefer to use electronic devices for note-taking during lecture should use one side of the classroom.

Safety on Campus and in the Classroom
For a list of emergency procedures for all types of incidents, please visit the website of the Critical Incident Response Team (CIRT): https://cirt.arizona.edu/case-emergency/overview

Also watch the video available at https://arizona.sabacloud.com/Saba/Web_spf/NA7P1PRD161/app/me/ledetail;spf-url=common%2Flearningeventdetail%2Fcrtfy000000000003841

University-wide Policies link
Links to the following UA policies are provided here, http://catalog.arizona.edu/syllabus-policies:

- Absence and Class Participation Policies
- Threatening Behavior Policy
- Accessibility and Accommodations Policy
- Code of Academic Integrity
- Nondiscrimination and Anti-Harassment Policy

Department-wide Syllabus Policies and Resources link
Links to the following departmental syllabus policies and resources are provided here, https://www.cs.arizona.edu/cs-course-syllabus-policies :

- Department Code of Conduct
- Class Recordings
- Illnesses and Emergencies
- Obtaining Help
- Preferred Names and Pronouns
- Confidentiality of Student Records
- Additional Resources
- Land Acknowledgement Statement

Subject to Change Statement
Information contained in the course syllabus, other than the grade and absence policy, may be subject to change with advance notice, as deemed appropriate by the instructor.