CS525-04/05: Advanced Database Organization

Notes 0: Course Organization

Yousef M. Elmehdwi

Department of Computer Science

Illinois Institute of Technology

yelmehdwi@iit.edu

August 23rd 2023

Welcome to CS525

This Course

- Introduction to the design and implementation of disk-oriented database management systems.
- This is **not** a course on how to use a database to build applications.
  - See CS425
- This course will not teach you how to code in `C` but it is all about how you can build a DB system from scratch.

Who we are. . .

- **Instructor**
  - Yousef Elmehdwi
    - Associate Teaching Professor, 7^{th} year at IIT, not the 1^{st} time teaching CS525 ☺
    - Email: `yelmehdwi at iit dot edu`
    - Research: Data privacy and security
    - Office: Stuart Building, room `237D`
    - Office Hours: `Wednesdays 4:00-5:00 pm or by appointment`

Who we are. . .

- **TAs**: TBA

Prerequisite(s)

- Courses: CS425
- Programming experience in `C, C++`
  - I will not teach you how to write/debug `C,C++`
- Unix OS and file system knowledge is helpful

Course Info

- Time: `W 5:00 - 7:40 pm, RE 104`
- Lecture slides in PDF format will be posted before the lectures (Blackboard)
- Lecture slides cover essential material
- Lectures will be recorded and uploaded to course Blackboard right after each class.
- Students can access the recorded lectures whenever they need them.

What is expected from you

- Attend in-person lectures, if you can
- Be active and think critically
- Do programming Assignments
  - Start early and be honest
- Study for exams

Course syllabus

- You are expected to be familiar with the contents of the course syllabus
- Available on the course Blackboard
- If you haven't read it, read it after this lecture

Workload and Grading

- Schedule and Important Dates
  - On blackboard & Piazza
- Programming Assignments 50% (10%,10%,15%,15%)
  - 4 Assignments
  - Groups of 3 students (at most). Groups will be determined by Friday, September 8^{th}, midnight
  - Plagiarism $\rightarrow$ 0 points and administrative action (a lower grade, automatically failing a course)
- Quizzes (5%)
  - There will be two take home quizzes during the course
- Exams
  - Closed book, closed notes exams
    - Only **ONE** sheet of paper printed on front and back is allowed
  - Midterm Exam (20%): **11/08/2023**
  - Final (25%): During finals week **12/??/2023**

Letter Grade Distribution

Points	Grade
85 - 100	A
75 - 84	B
60 - 74	C
0 - 59	E

Programming Assignments

- 4 assignments one on-top of the other
- Starting from a storage manager you will be implementing your own tiny database-like system from scratch
- You will explore how to implement the concepts and data structures discussed in the lectures
- Each of the regular assignments will have optional parts that give extra credit. You can earn `2%` extra credit points per assignment.
- All assignments have to be implemented using `C/C++`.
- I will specify test cases for the assignments, but you are encouraged to add additional test cases.

Programming Assignments - Source Code Mange

- Code has to compile & run on server account
  - Email-ID@fourier.cs.iit.edu
  - Linux machine
  - SSH with X-forwarding
- Source code managed in **git** repository on Bitbucket.org
- Handing in assignments = submit (push) to repository
- Git tutorials:
  - http://www-cs-students.stanford.edu/~blynn/gitmagic/book.pdf
  - https://git-scm.com/documentation.

Programming Assignments: Details

- **Assignment 1 - Storage Manager:** Implement a storage manager that allows read/writing of blocks to/from a file on disk.
- **Assignment 2 - Buffer Manager:** Implement a buffer manager that manages a buffer of blocks in memory including reading/flushing to disk and block replacement (flushing blocks to disk to make space for reading new blocks from disk).
- **Assignment 3 - Record Manager:** Implement a simple record manager that allows navigation through records, and inserting and deleting records.
- **Assignment 4 - `B+-Tree` Index:** Implement a disk-based `B+-Tree` index structure.

Course Policy

1. Make-up Exams: Only for officially proven health reasons.

Course Overview

- File organization and access, buffer management, performance analysis, and storage management
- Database system architecture, query optimization, transaction management, recovery, concurrency control
- And more when time permits

Course Objectives

After attending the course students should be able to:

- Understand the design decisions behind textbook DBMS architectures
- Know the trade-offs of various storage organization techniques
- Be able to build parts of a small-sized data processing system from scratch
- Understand the basics of query optimization
- Know standard implementations of relational operators such as join, aggregation, and set operations
- Be able to estimate the cost of executing an operator/query based on DB statistics
- Know standard database indexing techniques
- Understand concurrency control and recovery mechanisms

Tentative Course Outline

The weekly coverage might change as it depends on the progress of the class.

Week	**Content**
Week 1	Introduction/ Hardware
Week 2	File and System Structure
Weeks 3-4	Indexing and Hashing
Weeks 5-8	Query Processing
Weeks 9-10	Crash Recovery
Weeks 11-12	Concurrency Control
Weeks 13-14	Transaction Processing
Week 15	Advanced topics

Suggested Texts, Readings & Materials

- Garcia-Molina, Ullman, and Widom, Database Systems: The Complete Book, $2^{nd}/3^{rd}$ Edition, Prentice Hall, 2008
- Silberschatz, Korth, and Sudarshan , Database System Concepts , 6th/7th Edition , McGraw Hill , 2010/2019
- Elmasri and Navathe , Fundamentals of Database Systems , 6th Edition , Addison-Wesley , 2003
- Ramakrishnan and Gehrke , Database Management Systems , 3nd Edition , McGraw-Hill , 2002

- I will also provide lecture notes that covers topics not found in textbooks.

Important Dates

Week	Content
08/24	Coding Assignment 1 handed out
09/18	Coding Assignment 1 due
09/19	Coding Assignment 2 handed out
10/11	Coding Assignment 2 due
10/12	Coding Assignment 3 handed out
11/05	Coding Assignment 3 due
11/06	Coding Assignment 4 handed out
12/03	Coding Assignment 4 due
10/19	Quiz 1 handout
10/22	Quiz 1 due
11/28	Quiz 2 handout
12/01	Quiz 2 due
11/08	MidTerm Exam
TBA	Final Exam, During Final week (12/??/2023)

Next

Notes 1: Introduction