CS525: Advanced Database Organization

Notes 0: Course Organization

Yousef M. Elmehdwi

Department of Computer Science

Illinois Institute of Technology

yelmehdwi@iit.edu

August 21$^{\text{st}}$, 2018

Welcome to CS525

Credits

- The slides are a modified version of the slides used by Hector Garcia-Molina for the CS 245 course at Stanford.
- Responsibility of any errors due to modifications belongs to *me*.

This Course

- Graduate-level introduction to the design and implementation of database management systems

Who we are?

- Course Webpage: `http://cs.iit.edu/~cs525/yousef/`
- Syllabus: Course Blackboard
- **Instructor:** Yousef Elmehdwi, Senior Lecturer, CS, 2nd year and 4th time teach this course!
  - Office Hours: T 1:00-2:00pm/F 4:10-5:10 pm or by appointment at Stuart Building, Room 237D
- **TA**: TBD.

Prerequisite(s)

- Courses: CS425
- Programming experience in C, C++
- Unix OS and file system knowledge is helpful
- Data structures (.e.g., CS401)

Lectures

- Lecture slides in PDF format will be posted shortly before or after the lecture (Blackboard/Website)
- Lecture slides cover essential material
- Try to cover same thing in many ways: Lecture, homework, exams

Workload and Grading

- Schedule and Important Dates
  - On blackboard & Website
- Programming Assignments 50% (10%10%15%15%)
  - 4 Assignments
  - Groups of 3 students (at most). Groups will be determined in the first day of class.
  - Plagiarism → 0 points and administrative action
- Homework assignments (5%)
- Quizzes (5%)
  - There will be two take home quizzes during the course
  - Quiz I: Handout: 09/25/2018, Due: 10/02/2018
  - Quiz II: Handout: 11/13/2018, Due: 11/19/2018
- Exams: Close book, close notes
  - Midterm Exam (20%): **10/18/2018**
  - Final (20%): TBA

Letter Grade Distribution

Points	Grade
90 - 100	A
80 - 89	B
70 - 79	C
60 - 69	D
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 and readings
- In the end there will be an optional assignment for extra credit.
- Each of the regular assignments will have optional parts that give extra credit.
- All assignments have to be implemented using C/C++.
- We will specify test cases for the assignments, but you are encouraged to add additional test cases.
- Source code managed in **git** repository on `Bitbucket.org`
- Handing in assignments = submit (push) to repository
- One repository per student
- You will get an invitation from the TA soon (wait one week, the if not, contact TA)

Course Policy

1. Make-up Exams: Only for officially proven health reasons.
2. Late Work:
   - -20% per day late

Attendance Policy

- I dislike mandatory attendance but attendance makes your life easier
- Students are expected to attend all classes and are responsible for all material covered in class, even when absent
- Students should understand that some material discussed in class is not covered in the textbook.

Course Overview

- File organization and access, buffer management, performance analysis, and storage management
- Database system architecture, query optimization, transaction management, recovery, concurrency control
- Reliability, protection, and integrity
- 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
- Elmasri and Navathe , Fundamentals of Database Systems , 6th Edition , Addison-Wesley , 2003
- Ramakrishnan and Gehrke , Database Management Systems , 3nd Edition , McGraw-Hill , 2002
- Silberschatz, Korth, and Sudarshan , Database System Concepts , 6th Edition , McGraw Hill , 2010

Important Dates

Week	Content
08/28	Assignment 1 handed out
09/18	Assignment 1 due
09/18	Assignment 2 handed out
10/11	Assignment 2 due
10/11	Assignment 3 handed out
11/06	Assignment 3 due
11/06	Assignment 4 handed out
11/20	Assignment 4 due
11/26	Optional Assignment due
09/25	Quiz 1 handout
10/02	Quiz 1 due
11/13	Quiz 2 handout
11/19	Quiz 2 due
10/18	Mid Term Exam, 5:00pm-6:15pm, Robert Pritzker Ctr 111
TBA	Final Exam, TBA

Notes 1: Introduction