CS525 - Advanced Database Organization - 2023 Summer

Class Period: Mondays/Wednesdays; Time: 13h:00 to 16h:10
Building: Stuart Building, Room 111

Syllabus

Instructor: Gerald Balekaki

Email: gbalekaki@iit.edu

Phone: 312 567 5176

Office: Stuart Building, room 206c

Office Hours: Mondays/Wednesdays, 11h:00 - 13h:00

Course Description

Database management systems are a crucial part of most large-scale industry and open-source systems. This course provides comprehensive coverage of issues associated with database system development and an in-depth examination of structures and techniques used in contemporary database management systems (DBMSs). Students will learn about the inner workings of these exciting systems: Which algorithms are used? What are typical architectures used to build a system as complex as a DBMS? What are implementation strategies? These questions and more will be answered during the course.

The course is highly applied, emphasizing practical skills and habits through a series of programming assignments during which students will develop their own tiny DBMS-like engine. We will cover the most important aspects/components of a DBMS: storage and buffer management, indexing, query optimization, query execution, and concurrency control and recovery.

Reading Material

The following text books will be helpful for following the course and studying the presented material. All four textbooks have their merits, but any one should be sufficient as reading material.

- 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, , *McGraw Hill*, 2010
- Garcia-Molina, Ullman, and Widom, Database Systems: The Complete Book, 2nd Edition, *Prentice Hall*,

The slides for the course will be made available on the course blackboard.

Prerequisites

- Courses: CS425
- Programming experience in C, C++ or other low level languages
- Unix OS and file system knowledge is helpful
- Data structures (e.g., CS331)

Workload

Programming Assignments

- 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.

There will be several programming assignments during the course. 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. The assignments will require the use of skills learned in this course as well as other skills you have developed throughout your program. Each assignment will build upon the code developed during the previous assignment. All assignments have to be implemented using C/C++. Furthermore, at the end of each assignment, each group is required to demonstrate solutions to their programming assignments.

The following criteria will be considered in the grading:

- Functionality:
  - Does the code do what it is supposed to do? Are all tests completed successfully?
- Documentation:
  - How well documented is the code?
- Code Organization:
  - Is the structure of the code clear and suits the goal?
- Innovation:
  - Does the solution use new and innovative ideas?
- We will check for plagiarism.
  - Plagiarism will result in zero points for the assignment, potentially academic sanctions, and may result in an E grade.

Late policies:

- -10% points per day

Midterm and Final Exam

There will be a mid term and a final exam covering the topics of the course.

- Midterm Exam
- Final Exams

Quizzes

There will be quizzes during the course. The main objective of the quizzes is for you and the instructor to evaluate how well you internalized the topics covered in the course.

- Quiz 1
- Quiz 2

Grading Policies

Weighting of Deliverable

- programming assignments: (10% + 10% + 15% + 15%)
- midterm exam: 20%
- final exam: 20%
- quizzes: (5% + 5%)

Grading Scheme

Your final course grade is determined based on your total score which is calculated as the weighted sum of the points for each of the deliverables. The weights are as shown above. For each deliverable you will receive between 0 and 100 points, and final grade will be computed and distributed as follows:
Points

Points	Grade
80 - 100	A
60 - 79	B
50 - 59	C
40 - 49	D
Below 40	E

Detailed Course Topics

- Introduction
  - Relational Algebra
  - DBMS Architecture
- Hardware Characteristics affecting DBMS Design
  - Read/Write Properties of Disks
  - RAID Storage
  - Memory Hierarchy
- Disk Storage and Buffer Management
  - Physical Tuple Layout
  - Page Layout
  - Tuple IDs
  - Buffer Replacement Strategies
- Indexing and Hashing
  - B-Tree-Family Indices
  - Hashing
- Query Optimization
  - Logical Optimization
  - Equivalence-preserving Transformations
  - Physical Optimization
  - Join Reordering
  - Cost Estimation
- Query Execution
  - Pipelining
  - Push vs. Pull based Execution

- ● Access Methods
- ● Join Methods
- ● Grouping and Aggregation
- ● Other Operator Implementations
- ● External Sorting
- ● Recovery
  - ● Write Ahead Log (WAL)
  - ● Algorithms for Recovery and Isolation Exploiting Semantics (ARIES)
- ● Concurrency Control
  - ● Serializability
  - ● Two-Phased Locking (2PL)
  - ● Implementing of Locks
- ● Bonus Topics (if times permits)
  - ● Distributed Database Systems
  - ● Data Warehousing
  - ● Parallel Query Execution
  - ● Techniques for Executing Nested Queries and Un-nesting
  - ● Additional Index Structures
  - ● Relation to Distributed Data Analytics (DISC)
  - ● Big data and data analytics

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

Academic Integrity

You are welcome to discuss assignments with classmates, but all final work must be your own. For details on what constitutes academic dishonesty, consult the university's Code of Academic Honesty at https://web.iit.edu/student-affairs/handbook/fine-

print/code-academic-honesty. Any confirmed cases of academic dishonesty will be reported to academichonesty@iit.edu, and any work involved will, at the very least, receive a reduction in grade deemed appropriate by the instructor.

Disability Resources

Reasonable accommodations will be made for students with documented disabilities. In order to receive accommodations, students must obtain a letter of accommodation from the Center for Disability Resources and make an appointment to speak with the instructor as soon as possible. The Center for Disability Resources is located in the *Life Sciences Building*, *room 218*, tel:312-567-5744 or disabilities@iit.edu.

IllinoisTech's Sexual Harassment

and Discrimination Information

- Sexual harassment, sexual misconduct, and gender discrimination by any member of the Illinois Tech community is prohibited. This includes harassment among students, staff, or faculty. Sexual harassment by a faculty member or teaching assistant of a student over whom they have authority or by a supervisor of a member of the faculty or staff is particularly serious. Such conduct may easily create an intimidating, hostile, or offensive environment.
- Illinois Tech encourages anyone experiencing sexual harassment or sexual misconduct to speak with the Title IX Office for information on the resolution process and support options.
- You can file a complaint electronically at http://iit.edu/incidentreport, which may be completed anonymously. You may also file a complaint in-person by contacting the Title IX Coordinator, Virginia Foster at tel:312-567-5725 / mailto:foster@iit.edu or the Deputy Title IX Coordinator tel:312-567-5726 / mailto:eespeland@iit.edu.
- If you are not ready to file a formal complaint but wish to learn about your rights and options, you may contact Illinois Tech's Confidential Advisor service at tel:773-907-1062. You can also contact a licensed practitioner in Illinois Tech's Student Health and Wellness Center at tel:312-567-7550
- For a comprehensive list of resources regarding counseling services, medical assistance, legal assistance and visa and immigration services, you can visit the Title IX Office's website at https://web.iit.edu/hea/resources