University at Buffalo
CSE462/562 Datbase Systems (Spring 2022)

Instructor: Zhuoyue Zhao, zzhao35@buffalo.edu.
Lectures: T&H 11:00 am to 12:20 pm, Knox 110.
Teaching assistants: Meng Ding, mengding@buffalo.edu. Yunnan Yu, yunnanyu@buffalo.edu.
Office hours: TBD
Course website: `https://cse.buffalo.edu/~zzhao35/teaching/cse562_spring22/`
Annoucement and discussion board: We will be using Piazza for announcements, discussion and Q&A outside the lecture and office hours: `https://piazza.com/class/kxf79wjbzz52zw`
Course Description:
Database Management Systems (DBMS) are an important piece of software widely used in various data intensive applications. No matter what data model or query language it adopts, there are principles and methodologies commonly found in virtually all DBMS, in order to support efficient and fault-tolerant data storage, query and management. In this course, we will examine the internals of a traditional Relational DataBase Management System (RDBMS) and discuss the common principles and methodologies that may be useful in many other database and data processing systems beyond RDBMS. You will also gain practical experience of efficient data management through a semester-long project of implementing various components of a mini RDBMS.

The main topics of this course include database storage management, query processing, query optimization, transaction processing, concurrency control and recovery, in a relational database system. In addition to that, we will review a few select topics in the latest database research.

Please note that this is not a course on database design or database application development. These are covered in CSE460/560 Data Models and Query Languages. We will briefly review the necessary background on relational data model and the SQL language in the first a few lectures.

Learning Outcomes:
Upon completion of this course, a student is expected to

- know the common system architecture of RDBMS;
- understand how to organize data across storage hierarchy for efficient storage and retrieval;
- understand how to execute and optimize a relational query over a large database;
- understand how to ensure ACID properties in face of concurrent transactions and crashes;
- and obtain practical programming and collaboration skills in a complex software system project.

Course Prerequisites:
Required: undergraduate-level courses on the analysis of data structure and algorithms (e.g., CSE250, CSE331 or equivalent); Basic knowledge in C/C++ and Linux programming environment. Preferred: undergraduate-level courses on database design, relational model, SQL and database application development (e.g., CSE460/560 or equivalent).
No mandatory textbook: We will make all the lecture slides available on the course website.
Optional textbook and readings:

- Database System Concepts, Seventh Edition, by Abraham Silberschatz and Henry Korth and S. Sudarshan.

- The Red Book, Readings in Database Systems, 5th Edition (available online at `http://www.redbook.io/`), Peter Bailis, Joseph M. Hellerstein, Michael Stonebraker, editors.

Course Requirements:

- You are expected to understand and follow the academic integrity policy of UB, CSE and this course. We will *strictly* enfore the academic integrity policy (see below).
- In-person attendance is *required*. You should attend all the lectures and be familiar with all the course materials presented in class, which may appear in any course projects and exams. We will try to make the live streaming and recording of the lectures available, but we cannot guarantee the availablility of those in case of technical difficulties.
- There will be a mid-term exam and a final exam. They must be attended in person and there will be no make-up exam except for the exam/course conflicts made known to the instructor within three weeks after the add/drop deadline (2/28/2022 11:59 pm EST), or extremely unusual circumstances (e.g., quarantine due to COVID-19). The exams are open-book but you may only bring the paper-copy of lecture slides, the written assignments, the textbook or your lecture notes. You should not use any electronic device during the exame, except for a calculator that may be needed.
- There is a semester-long course project of building a mini RDBMS in C++, divided into five subprojects. You may work individually or collaborate in a team of two. The tasks in each project will include a coding component, which you may collaborate with your teammate, and a write-up, which you must complete independently without any collaboration or sharing with your teammate. The tasks of the projects remain the same regardless of the team size, and you will receive the same grade for the coding component of your team. The write-up will be graded individually. If you work in a team of two, you must also clearly describe the division of your implementation responsibilities in the write-up. We will use Autolab for code submission and UBLearns for write-up submission. Please refer to the course website for details.
- We will release a few homework assignments but they are not graded and you do not need to hand them in. They are meant to help you review the course materials and prepare for the exams. The problems in the exams will be similar to those in the homework assignments. We will release the solution to a homework assignments about one week after it is released.

Important Dates:

- Add/Drop deadline: Feburary 7, Monday, 2022.
- Midterm exam: March 17, Thursday, 2022, in the last lecture before spring recess.
- Final exam: May 17, Tuesday, 2022.
- Last day to resign from the course: April 22, Friday, 2022.
- Project due dates: please refer to the course website.

Grading Policy:
The course grade will be broken down as follows:

- Mid-term exam: 20%.
- Final exam: 20%.
- Projects: 60%.

And the following is the (tentative) assignment of letter grades:

- $[92, 100]$: A
- $[85, 92)$: A-
- $[77, 85)$: B+
- $[70, 77)$: B
- $[63, 70)$: B-
- $[55, 63)$: C+
- $[45, 55)$: C
- $[35, 45)$: C-
- $[15, 35)$: D

- $[0, 15)$: F

The grades will not be curved or differentiated between CSE460 and CSE560, but the instructor may slightly lower the minimum required grade for each letter grade at the end of the semester.

Course Schedule:
The following is the tentative course schedule and may be changed throughout the semetser. Please refer to the course website for the latest schedule.

Tentative Course Schedule

Week#	Topics	Textbook chapter	Project/Homework
1	Introduction to RDBMS; course logistics	Chapter 1	Project 1 assigned
	The relational model, and the SQL language	Chapters 2, 3	
2	The relational model and the SQL language (cont'd)		Course project sign-up due
	Physical storage systems	Chapter 12	HW 1 released
3	Data storage layout & buffer manager	Chapter 13	Project 1 due
	Overview of Taco-DB		Project 2 assigned
4	Indexing concepts	Chapters 14.1, 14.2	HW2 released
	Hash-based indexing	Chapters 14.5, 24.5	
5	Hash-Based Indexing (cont'd)		
	Tree-based Indexing	Chapters 14.3, 14.4	HW3 released
6	Tree-based Indexing (cont'd)		Project 2 due
	Single-table query processing	Chapter 15	Project 3 assigned
7	Single-table query processing (cont'd); mid-term recap and Q&A		
	Mid-term exam		
8	Spring recess, no lectures		
9	External sorting	Chapter 15.4	Project 3 due
	Join algorithm	Chapter 15.5	Project 4 assigned
10	Approximate query processing	Additional material	
	Query optimization	Chapter 16	
11	Query optimization (cont'd)		HW4 released
	Transaction processing	Chapter 17	
12	Pessimistic concurrency control	Chapters 18.1 - 18.4	
	Timestamp-ordering and optimistic concurrency control	Chapters 18.5, 18.6	
13	Multi-version concurrency control and snapshot isolation	Chapters 18.7, 18.8	HW5 released
	Crash recovery	Chapter 19	Project 4 due
			Project 5 assigned
14	Parallel and Distributed database	Chapter 20 and/or additional materials	
15	Hybrid transaction and analytical processing	Additional material	
	Final recap and Q&A		
16	Final exam, no lecture		Project 5 due

Academic Integrity Policy:
Academic integrity is critical to the learning process. It is your responsibility as a student to complete your work in an honest fashion, upholding the expectations your individual instructors have for you in this regard.

The ultimate goal is to ensure that you learn the content in your courses in accordance with UB's academic integrity principles, regardless of whether instruction is in-person or remote. Thank you for upholding your own personal integrity and ensuring UB's tradition of academic excellence. You should get familiar with the departmental and the university academic integrity policies and procedures for _graduate students_ and for _undergraudate students_

In this course, you may discuss with other students about your course projects and/or prepare for exams, but each team must submit their independent code implementation for the course project, and each individual must submit his/her independent write-up. We also require all students, whether enrolled, dropped or resigned from the class, to keep your course project repository inaccessible to public indefinitely, and never share it with any current or future students who may take the course. Examples that we consider as academic integrity violations include but are hot limited to: 1) copying any part of other team's code implementation in the course project, regardless of whether it is a verbatim copy or a copy with substantially similar structure; 2) collaborate with other teams in the course project in any manner except for discussing high-level concepts covered in the lectures (in particular, you should not exchange design and/or implementation details); 3) making your course project repository or any of its copy or fork publicly available or privately to any current or future students of the course; 4) submitting work that is not created by you or your team; 5) cheating or referring to any material not permitted in the exams.

Consistent with the CSE policy, we have **zero tolerance** towards academic integrity violations. Any academic integrity violation will result in an F grade for all students involved, unless the violation is purely accidental and does not provide any unfair advantage to any of the students involved .

Campus resources:

Accessibility Resources. If you have any disability which requires reasonable accommodations to enable you to participate in this course, please contact the Office of Accessibility Resources in 60 Capen Hall, 716-645-2608 and also the instructor of this course during the first week of class. The office will provide you with information and review appropriate arrangements for reasonable accommodations, which can be found on the web at: http://www.buffalo.edu/studentlife/who-we-are/departments/accessibility.html.

Counseling Services. As a student you may experience a range of issues that can cause barriers to learning or reduce your ability to participate in daily activities. These might include strained relationships, anxiety, high levels of stress, alcohol/drug problems, feeling down, health concerns, or unwanted sexual experiences. Counseling, Health Services, and Health Promotion are here to help with these or other concerns. You learn can more about these programs and services by contacting:

Counseling Services: 120 Richmond Quad (North Campus), phone 716-645-2720

202 Michael Hall (South Campus), phone: 716-829-5800

Health Services: Michael Hall (South Campus), phone: 716- 829-3316

Health Promotion: 114 Student Union (North Campus), phone: 716- 645-2837

Sexual Violence. UB is committed to providing a safe learning environment free of all forms of discrimination and sexual harassment, including sexual assault, domestic and dating violence and stalking. If you have experienced gender-based violence (intimate partner violence, attempted or completed sexual assault, harassment, coercion, stalking, etc.), UB has resources to help. This includes academic accommodations, health and counseling services, housing accommodations, helping with legal protective orders, and assistance with reporting the incident to police or other UB officials if you so choose. Please contact UB's Title IX Coordinator at 716-645-2266 for more information. For confidential assistance, you may also contact a Crisis Services Campus Advocate at 716-796-4399.