95-703 Database Management, Summer 2021

Carnegie Mellon University
Heinz College

Course Logistics

Class: Tuesday & Thursday, 9:00 am - 10:20 am ET (Zoom)
Instructor: Raja Sooriamurthi (raja@cmu.edu) TAs: Gloria Moon (gmoon) and Ruitao (Ray) Liu (ruitaol)

Course Description

We currently live in a data driven age. Data has emerged as the new oil that drives an organization: The successful operation of modern organizations relies on the effective use of their operational data. Database management systems (DBMS) are the engines of this data driven world.

Data collected and used by an organization is broadly divided into two types (i) line of business data and (ii) customer behavior data. Traditionally data management has focused on *line of business* data. For example, when a ride request is made to a ride sourcing company (Uber, Lyft, etc.), what data is needed to meet that request? When a purchase is made in a grocery store what is the flow of data during that transaction? Line of business data is used to support core business processes of the organization. Alternatively, based on the purchase patterns of a shopper or the volume or location of ride requests, how can a grocery store or a ride sourcing company make their operation more effective? The answer to this question is based on *customer behavior* data (who bought what, when etc.). Whatever type of data it may be, many fundamental questions are the same: How do you gather, organize, curate, and process data to help run an organization or extract actionable information to increase effectiveness?

The use of data involves three aspects (i) management (ii) security and (iii) analytics. In this course, we will study tools and techniques for managing data with database systems. At the highest level we will study two questions (a) how to *use* a database and (b) how to *build* a database. For more than three decades, the relational model has been the predominant model of data management. Most of this course will focus on the classic relational model. In the past several years, driven by evolving functional and non-functional (quality) needs of an organization, alternatives to the classic relational model have emerged. We will examine illustrative samples of these popular alternatives known as non-relational or NoSQL models.

The course is divided into the following four modules:

I. Querying the Relational Model
- Tables, rows, columns, keys
- SQL 1: SELECT, FROM, WHERE, ORDER BY, DISTINCT, LIKE, BETWEEN, IN
- SQL 2: JOIN, GROUP BY, HAVING, sub-queries, ANY, ALL, UNION
- Programming interfaces to a database (e.g., psycopg2)

II. Relational Database Design
- Integrity constraints: entity, referential, check
- Anomalies, functional dependencies, normalization
- Converting conceptual models to logical models to physical models

III. Additional Topics
- Functions and triggers
- Views, authorization, security, SQL injection
- Concurrency and recovery, ACID transactions

IV. Non-relational models of data
- Motivation
- Key-value stores

- Document databases

The exact themes we discuss and the depth to which we discuss them will depend on the pace of the course. As the course progresses the online schedule of the course on Canvas will be updated.

Readings via the ACM online library

There are a wide range of books on various aspects of database systems (from abstract theory to the details of a specific implementation). In addition to class discussions, we will use chapters and sections from various books available through the ACM online library. If you are not already an ACM student member you are highly encouraged to join. You will find the online library useful for many courses you will take in future semesters. Much beyond this course, for $19, you will find ACM student membership and online learning resources valuable for your advancement and career. All course content and the week by week sequence of themes will be available via Canvas and course communication will be via Piazza.

Learning Objectives

Upon successful completion of this course, students will have achieved the following learning objectives.

- Manipulate and extract information from a database using SQL
- Given an organizational need, identify the data needed to meet that need
- Build a user-story driven conceptual data model and implement the corresponding database with SQL
- Programmatically access a database from a front end application (written in PHP/Java/Python etc.)
- Understand transactions and security in data base systems
- Explore alternatives to the classic relational model

Component	Weight
Quizzes + Participation	15
Assignments	20
Project	20
Exam 1	15
Exam 2	15
Final Exam	15

Assessment

Performance in the course will be assessed along the adjacent components.
Final course grades will *not* necessarily be along raw score cut offs of 90 / 80/ 70. Rather, a number of factors are used: class participation, project peer evaluation, "performance clusters" (will be explained in class).

How to do well in the course

This is an analytical, problem solving course. The assignments provide opportunities to apply the knowledge and skills of this course. It is expected that the core knowledge and skills of this course be internalized. Reflecting this objective, the three exams have a combined weight of 45%. You may find the following guidelines helpful:

- *Participate*. Attend and ask questions in class. You will gain a lot more from the class by actively participating rather than passively observing. It will be more fun for all.
- *Prepare*. Peruse the slides before each class. Do the readings for each quiz.
- *Diligence*. Sincerely and independently do the assignments. Work submitted needs to be your own.
- *Rework*. The best way to perform well in the exams (and consequentially the course) is to redo class room exercises and quizzes.

- *Resources*. All people involved with delivering the course are eager to see you succeed. Avail of all available resources: Meet the instructor right after class with any immediate questions; stop by office hours of the instructor and TAs; effectively use Piazza. We are here to help.
- *Grades*. Focus on the above and don't sweat the raw score on Canvas. The final grades in the class are based not only on the raw absolute score on canvas but on other factors too.

Class Policies

Attendance and Preparation for Class: To fully engage in classroom discussions, you are required to attend all class sessions and come prepared for each class (e.g., charged laptop for in-class exercises, functioning software). Attendance will be taken at the beginning of class. Arriving after attendance has been taken will count as an absence. In the event of a situation requiring you to be absent (e.g., job interview) please contact the professor and head TA in advance. Your final course grade may be lowered by 1/3 of a letter grade for each unexcused or undocumented absence.

In class Quizzes / Exercises: There will be periodic short quizzes at the beginning of some classes based on the material of the previous classes. Explicit readings will be specified beforehand. As the learning objective arises, there will also be in-class exercises. Students who have an unexcused absence or are tardy will not be able to make up these quizzes.

Assignments and Flex days: Part of professional behavior is submitting deliverables on time. Due dates of all deliverables (assignments, projects etc.) will be specified when issued and it is expected that assignments will be submitted on time. Equipment failure, last minute network load issues etc. are not valid reasons for missing a deadline. At the same time 'life happens' — you may have to travel for an interview, may fall sick, it may be an extremely busy week etc. To accommodate such situations, each student has 3 flex days. Unless explicitly specified otherwise, you may apply 1 flex day (24 hours) for submitting an assignment beyond the due date. After that, submissions will be deducted 20% of the assignment grade for each day late. Please email Raja when you avail of a flex day.

Academic Integrity: Unless explicitly stated otherwise, *all work needs to be individually done*. While it is fine to discuss general ideas, all submitted work must be your own. Sharing of work with another student or using the work of another's when completing your own will result in a grade of zero. *In short: do not show your work (code etc.) to a classmate nor look at the work of a classmate*. We will be using tools (such as MOSS[1]) to detect program similarity. Any case of suspected cheating will be brought to the Dean's attention. If you referred to external sources or consulted with others be sure to clearly indicate so. Be sure to familiarize yourself with the University policies on academic integrity http://www.cmu.edu/policies/student-and-student-life/academic-integrity.html .

Reassessment: If you would like a component of the course (assignment, exam etc.) to be reevaluated, submit your request in writing (email Raja will suffice) explaining in detail why you feel your response needs to be re-assessed. Any reassessment requests need to be submitted within two weeks of the assignment or exam being returned.

For Students with Learning Disabilities: If you wish to request an accommodation due to a documented disability, please inform your instructor and contact the Office of Disability Resources http://www.cmu.edu/disability-resources

[1] https://theory.stanford.edu/~aiken/moss

Take care of yourself

Do your best to maintain a healthy lifestyle this semester by eating well, exercising, avoiding drugs and alcohol, getting enough sleep, and taking time to relax. Despite what you might hear, using your time to take care of yourself will actually help you achieve your academic goals more than spending too much time studying.

All of us benefit from support and guidance during times of struggle. There are many helpful resources available on campus. An important part of the college experience is learning how to ask for help. Take the time to learn about all that's available and take advantage of it. Ask for support sooner rather than later – this always helps.

If you or anyone you know experiences any academic stress, difficult life events, or difficult feelings like anxiety or depression, we strongly encourage you to seek support. Consider reaching out to a friend, faculty or family member you trust for assistance connecting to the support that can help. Counseling and Psychological Services (CaPS) is here for you: call 412-268-2922 and visit their website at http://www.cmu.edu/counseling/. Over 25% of students reach out to CaPS some time during their time at CMU. *http://www.cmu.edu/teaching/designteach/design/syllabus/syllabussupport.html*

Every individual must be treated with respect.

We are diverse in many ways, and this diversity is fundamental to building and maintaining an equitable and inclusive campus community. Diversity can refer to multiple ways that we identify ourselves, including but not limited to race, color, national origin, language, sex, disability, age, sexual orientation, gender identity, religion, creed, ancestry, belief, veteran status, or genetic information. Each of these diverse identities, along with many others not mentioned here, shape the perspectives our students, faculty, and staff bring to our campus. We, at CMU, will work to promote diversity, equity and inclusion not only because diversity fuels excellence and innovation, but because we want to pursue justice. We acknowledge our imperfections while we also fully commit to the work, inside and outside of our classrooms, of building and sustaining a campus community that increasingly embraces these core values. Each of us is responsible for creating a safer, more inclusive environment.

Unfortunately, incidents of bias or discrimination do occur, whether intentional or unintentional. They contribute to creating an unwelcoming environment for individuals and groups at the university. Therefore, the university encourages anyone who experiences or observes unfair or hostile treatment on the basis of identity to speak out for justice and support, within the moment of the incident or after the incident has passed. Anyone can share these experiences using the following resources:

- Center for Student Diversity and Inclusion: csdi@andrew.cmu.edu, (412) 268-2150
- Report-It online anonymous reporting platform: www.reportit.net username: tartans password: plaid

All reports will be documented and deliberated to determine if there should be any following actions. Regardless of incident type, the university will use all shared experiences to transform our campus climate to be more equitable and just.

Let's have a fun and productive course!

May 25, 2021