

Databases and Data Management

Week 13-14

November 5-16

Teaching Team

Donald Grase, Executive Educational Instructor, (dgrasse@uchicago.edu)
Pavan Prathuru, Teaching Assistant, (pavanprathuru@uchicago.edu)

Recommended Texts

This course is based on two different textbooks, which would be a helpful reference.

Meier, Andreas and Kaufmann, Michael. “SQL & NoSQL Databases: Models, Languages, Consistency Options and Architectures for Big Data Management”

Silberschatz, Abraham, Korth, Henry, and Sudarshan, S. “Database System Concepts” (7th Edition)

Course Objectives

This course is designed to teach students the fundamentals of databases and database management. After completing the course, students will be able to:

- Design, implement, and manage relational databases effectively
- Write and optimize SQL queries for data retrieval and analysis
- Understand database administration, ensuring data integrity and optimal performance
- Apply principles in real-world scenarios

In more simple terms, after completing the course, I would want you to feel confident that you can hear about a database system and sketch a rough diagram of the schema, think about how it would be organized, and craft efficient queries from it to meet the needs of an organization, all while being mindful of how to be a steward of this information to ensure accuracy and security.

Course Schedule

Why the Relational Database (November 5)

- What is a database and why do we structure them the way we do?

- An introduction to the relational model, the most widely used model used to organize data within a database

Mathematical and Logical Foundations of DBMS (November 6)

- Effectively interacting with, and querying, databases is ultimately about formal logic, organization, and indexing.
- To understand the method to database management systems (DBMS), we will study set theory, a branch of mathematics that underpins nearly the rest of the entire course
- **Homework:** Set Theory Problem Set (Due End of Next Day)

Data Modeling (November 7)

- Conceptually, databases can be thought of in many ways, which has stakes for how they are organized
- We will study one of the most important tools for data modeling - the Entity Relationship model - and the rules for structuring a database to make sure it is efficient and easy to use
- **Homework:** Build Your Own E-R Diagram (Due End of Next Day)

Data Querying: Relational Algebra (November 8)

- The heart of database theory - and one of the tools that led to the development of relational database - is relational algebra, a defined algebraic structure based on set theory used for querying databases.
- Irrespective of how you will query databases in the future, relational algebra will be the foundation of your query
- **Homework:** Relational Algebra Problem Set (Due End of Next Day)

Synthesis and Assessment (November 9)

- How do set theory, databases, relational algebra, and data modeling fit together?
- This class will focus on synthesizing these ideas and placing the concepts in conversation with one another
- **Quiz 1: Due End of Day**

Data Querying in SQL (November 12)

- This class will bridge the theory of querying databases with practice
- We will learn the Structured Query Language (SQL) - the most important Database Language for DDL and DML
- **In Class Activity: Producing SQL Queries**

Data Security (November 13)

- Databases are incredibly valuable - sometimes so valuable that other actors who are not authorized access attempt to seize information from them
- This class will focus on key threats to data security, and how database administrators can help protect data

Data Governance (November 14)

- Data has to be managed appropriately during its whole life cycle - from the acquisition stage to the disposal stage
- This class will focus on data governance, an important and distinct concept from data management, including the best practices for administrators
- **Homework: Security and Governance Problem Set** (Due End of Next Day)

Beyond SQL (November 15)

- SQL is the most common - but not the only - tool for working with databases
- This class will focus on what is beyond SQL - including NoSQL and Postrelational databases

Synthesis and Assessment (November 16)

- Putting together the big picture
- **Quiz 2: Due End of Next Day**

Grading

The breakdown of grades will be as follows:

- Homework (Turn in on Canvas) (60%)
- Quizzes (Turned in on Canvas) (30%)

- Participation (Evaluated by attendance and by completion of in-class assignments) (10%)

Integrity

Each student is expected to turn in their own work. While collaboration is welcomed - and indeed expected - students should not directly copy their answers from either their peers or from online sources, including Artificial Intelligence (AI). Evidence of plagiarism is grounds for a failing grade on the assignment that was misrepresented or produced dishonestly.