CSC 330 Spring 2024 Syllabus

Database Theory and Applications (CSC 330) Spring 2024

Marcus Birkenkrahe

January 4, 2024

1 General Course Information

- Meeting Times: Tuesday/Thursday, 2:30-3:45 hrs

- Meeting place: Lyon Building Room 104 (computer lab)

- Professor: Marcus Birkenkrahe

- Office: Derby Science Building 210

- Phone: (870) 307-7254 (office) / (501) 422-4725 (private)

- Office hours: Mon/Wed/Fri 10-10:50 am & 3-3.50 pm, Tue/Thu 4.15-4.45 pm & by appointment

- Textbook: we're going to use DataCamp instead of a textbook.

- Textbooks used to prepare this course - any of these would be worth having - they use SQLite, MySQL, and PostgreSQL.

  1. Lemahieu/Vanden Broucke/Baesens (2021). Principles of Database Management. Cambridge University Press. URL: pdbmbook.com.
  2. Teate (2021). SQL for Data Scientists: A Beginner's Guide for Building Datasets for Analysis. Wiley. URL: wiley.com.
  3. DeBarros (2022). Practical SQL - A Beginner's Guide to Storytelling with Data (2e). NoStarch. URL: nostarch.com.
  4. Beaulieu (2020). Learning SQL (3e). O'Reilly. URL: oreilly.com.
  5. Kreibich (2010). Using SQLite. O'Reilly. URL: oreilly.com.
  6. Allen/Owens (2010). The Definitive Guide to SQLite. Springer/Apress. URL: springer.com.

2 Standard and course policies

Standard Lyon College Policies are incorporated into this syllabus and can be found at: lyon.edu/standard-course-policies.

The **Assignments and Honor Code** and the **Attendance Policy** are incorporated into this syllabus also and can be found at: tinyurl.com/LyonPolicy.

3 Objectives

You will learn SQL, the (by far) dominant data science language in the real world (for relational databases), and one of the oldest languages in use. You will practice SQL and work with relational databases as part of your weekly assignments in DataCamp. We also work with SQLite, the world's most common database, and we model databases with UML and Entity-Relationship diagrams. We cover XML and NoSQL databases, Big Data, and aspects of physical database organization. You learn to access databases with your programming language of choice - C, C++, Java, Python, R, or whatever you like. For this iteration of the course (my 2nd since coming to Lyon), I want to experiment with using the texts that come with various DataCamp lessons as part of the "Data Engineer" career track (link).
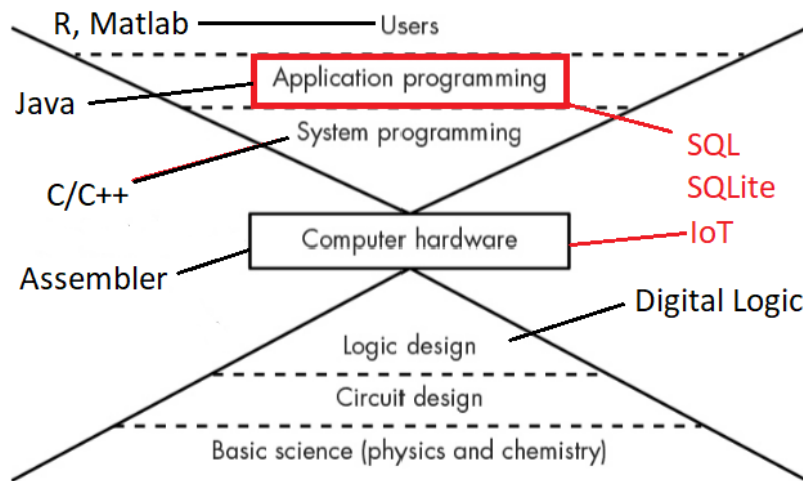

Figure 1: Computer landscape - modified after Steinhart, 2019

4   Student Learning Outcomes

Students who complete CSC 330 "Database Theory and Applications", will be able to:

- Understand the design principles of non/relational databases

- Know existing types of databases and their applications

- Master relational DB query languages like SQL and SQLite

- Design database applications using Entity Relationship Models

- Master computing infrastructure (editor, shell)

- Know how to effectively present assignment results

5   Course requirements

Some knowledge of, and experience with computers is useful but not critical. Curiosity is essential. You will gain data literacy skills by taking this course. The course will prepare you for further studies in computer and data science, or in other disciplines that use modern computing, i.e. every discipline, from accounting to zoology).

Formal prerequisites: MTH 115 (discrete math) or MTH 290 (Foundations of modern mathematics) and one of: CSC 100 (C/C++), CSC 109 (Python), or CSC 115 (Java).

6   Grading system

WHAT	WEIGHT
DataCamp assignments	25%
Final Project	25%
Tests	25%
Final exam	25%

Notes:

- To pass: 60% of all available points.

- Tests: weekly online quizzes, which are previewed and reviewed in class.

- Final exam: selection of the most challenging weekly quiz questions.

- You only have to write the final exam if you want to improve your grade at the end of the course. If the final exam result is below your final grade average up to this point, it will be ignored.

- There will be numerous (optional) bonus assignments to improve your midterm and final grades.

7   Schedule and session content

For **important dates**, see the 2023-2024 Academic Calendar at: catalog.lyon.edu/202324-academic-calendar.

No class on: Tuesday/Thursday 19/21 March (Spring break) - Friday 29 March (Easter break). Last day of classes: Wednesday, 8 May. Final exams (non-graduating students): 10-15 May.

Schedule:

Week	Sessions
1	Introduction to SQL: RDBMS/Querying
2	Introduction to SQL (continued)
3	Querying with SQL
4	SQL Variants: PostgreSQL and SQL Server
5	Intermediate SQL: SELECT, WHERE, GROUP BY
6	Intermediate SQL (continued)
7	Introduction to RDB: Constraints & Keys
8	Introduction to RDB: (continued)
9	Database Design
10	Database Design (continued)
11	Exploratory Data Analysis in SQL I
12	Exploratory Data Analysis in SQL II
13	Tutorial: SQLalchemy (Python) I
14	Tutorial: SQLalchemy (Python) II
15	Tutorial: Running SQL queries in R/Python
16	Course Review and Project Work

DataCamp:

- Introduction to SQL: relational db and querying (2h) - Much of the world's raw data—from electronic medical records to customer transaction histories—lives in organized collections of tables called relational databases. Being able to wrangle and extract data from these

databases using SQL is an essential skill within the data industry and in increasing demand.

In this two-hour introduction to SQL, you'll get to know the theory and the practice through bite-sized videos and interactive exercises where you can put your new-found skills to the test.

SQL is an essential language for building and maintaining relational databases, which opens the door to a range of careers in the data industry and beyond. You'll start this course by covering data organization, tables, and best practices for database construction.

The second half of this course looks at creating SQL queries for selecting data that you need from your database. You'll have the chance to practice your querying skills before moving on to customizing and saving your results.

PostgreSQL and SQL Server are two of the most popular SQL flavors. You'll finish off this course by looking at the differences, benefits, and applications of each. By the end of the course you'll have some hands-on experience in learning SQL and the grounding to start applying it on projects or continue your learning in a more specialized direction.

- Intermediate SQL: SELECT, WHERE, GROUP BY (4h)

    - How to filter and compare data
    - How to use aggregate functions to summarize data
    - How to sort and group your data
    - How to present your data cleanly using tools such as rounding and aliasing

- Intro to relational DB: constraints and keys (4h) - You've already used SQL to query data from databases. But did you know that there's a lot more you can do with databases? You can model different phenomena in your data, as well as the relationships between them. This gives your data structure and consistency, which results in better data quality. In this course, you'll experience this firsthand by working with a real-life dataset that was used to investigate questionable university affiliations. Column by column, table by table, you'll get to unlock and admire the full potential of databases. You'll learn how to create tables and specify their relationships, as well as how to enforce data integrity. You'll also discover other unique features of database systems, such as constraints.

- Database design (4h): A good database design is crucial for a high-performance application. Just like you wouldn't start building a house without the benefit of a blueprint, you need to think about how your data will be stored beforehand. Taking the time to design a database saves time and frustration later on, and a well-designed database ensures ease of access and retrieval of information. While choosing a design, a lot of considerations have to be accounted for. In this course, you'll learn how to process, store, and organize data in an efficient way. You'll see how to structure data through normalization and present your data with views. Finally, you'll learn how to manage your database and all of this will be done on a variety of datasets from book sales, car rentals, to music reviews.

- Exploratory data analysis in SQL (4h): You have access to a database. Now what do you do? Building on your existing skills joining tables, using basic functions, grouping data, and using subqueries, the next step in your SQL journey is learning how to explore a database and the data in it. Using data from Stack Overflow, Fortune 500 companies, and 311 help requests from Evanston, IL, you'll get familiar with numeric, character, and date/time data types. You'll use functions to aggregate, summarize, and analyze data without leaving the database. Errors and inconsistencies in the data won't stop you! You'll learn common problems to look for and strategies to clean up messy data. By the end of this course, you'll be ready to start exploring your own PostgreSQL databases and analyzing the data in them.

- Tutorial: SQLalchemy (Python)

- Tutorial: Run SQL queries in R and Python

Workload: approx. 4-5 hours per week.

1. Class time = 16 * 3 * 50/60 = 40 hours

2. Tests (home) = 16 * .5 = 8 hours

3. DataCamp assignments: 15 * 2 = 30 hrs (approx)