CMSC 23500: Introduction to Databases

The University of Chicago, Spring 2015

Adam Shaw

`http://www.classes.cs.uchicago.edu/archive/2015/spring/23500-1`

Welcome! In CMSC 23500, we offer both a theoretical and practical introduction to the design and implementation of database systems. To enroll in the course, you must have taken the introductory computer science courses up to and including CMSC 15400, or have received explicit permission from the instructor.

A reasonable level of expertise in C programming is essential to participation in this course. You will be expected to implement a large project in C, yet C programming itself is not taught in this course.

This course has an old-fashioned focus on relational databases. It is grounded in relational database theory as originally put forth by E. F. Codd around 1970 and subsequently developed over the last four and a half decades. In our ten week time span, there is more than enough to discuss in this realm without treating more recent developments. It should be noted that traditional relational databases are not so much only common or widespread as they are *ubiquitous*, and your instructor, for one, believes they will always be.

The distinctive feature of this course is its unusually thorough treatment of the inner workings of a complex and full featured relational database system by way of the `chidb` project. *This is not merely a SQL course. You get to roll up your sleeves and hack.* `Chidb` is based closely on `sqlite` and was developed primarily by Borja Sotomayor, with front-end implementation by Allen Nelson and some input from yours truly. By working on `chidb`, you will gain direct experience with real, practical, complex data structures in great depth. Specifically, you will work with B-Tree data structures and SQL expression data structures that are substantially more advanced than the simple proof-of-concept data structures one so often works with in undergraduate classes.

The specific goals of the course are these:

- to learn to design relational databases, at the level of entity-relationship diagrams, schema diagrams, and SQL schemas
- to become acquainted with relational database theory, especially as it relates to modeling queries with relational algebra, and quantifying design properties by way of normal forms
- to develop a good working knowledge of SQL

- to build a substantial part of a relational database system, including creating files that match the `sqlite` file format, implementing B-trees for tables and indexes, and generating code for a variety of SQL queries, including natural joins

Beyond these main goals, we will also discuss transaction systems, database-backed web applications, and SQL injection attacks, among other topics.

Having taken this course, students will have a good working knowledge of design considerations for relational databases, both at an informal and a formal level, and have a notably deep, relative to our time frame, understanding of the inner workings of a relational database system.

Instructor

Adam Shaw, email: `ams@cs.uchicago.edu`, office: Ryerson 157.

Graduate Teaching Assistants

Your three TAs, all PhD students, are Haopeng Liu, Hannah Morgan, and Zhixuan Zhou.

Piazza: Online Support

If you have not yet done so, you will to need to register with *piazza*. Piazza is an online question-and-answer system that we use for that purpose as well as distribution of course materials on occasion.

Contacting Us

If you have questions about the course, and those questions are in a sense impersonal — that is, they are about course material or course logistics — we ask that you post those questions publicly on piazza, rather than contacting any of the staff members directly. This ensures you will receive the fastest, most consistent possible response from the staff. Since students usually have *common* questions, posting public questions is also very efficient for your classmates as well. As yet another advantage, it avoids duplication of work on the part of the staff.

In cases where you have a question that is about your own personal situation and not relevant to the class as a whole, you may ask a "private question" on piazza, which is invisible to your classmates, or send email to your instructor directly.

Please do not post anonymously to piazza. Piazza posts are better, more thoughtfully written, and more courteous when the author is identified. We reserve the right to delete anonymous posts when they appear.

Lectures

Lectures are in Ryerson 251, every Tuesday and Thursday from 1:30–2:50. The first lecture is on Tuesday, March 31; the last is on Tuesday, June 2.

The use of electronic devices during lecture is discouraged. In my 100-level classes, devices are simply banned. This quarter, in recognition of your more advanced standing, I will allow the use of laptops or tablets for note taking, only if the wireless capability of the device is disabled and the only open application is a text editor.

Office Hours

I will have office hours as follows:

- Mondays, 2pm–4pm in CSIL 4
- Thursdays, 3pm–4pm in Ryerson 157
- Fridays, 1pm–3pm in CSIL 4

Schedule of Topics by Week (subject to change)

Week	Topics
1	relational database concepts
2	database design, B-trees
3	relational algebra (σ, π, ρ, $\times$, $\bowtie$)
4	SQL
5	database design, ER diagrams
6	database design, web databases
7	**midterm exam**, transactions, conflict serializability
8	implementation of operations, query optimization
9	dependency theory, normal forms
10	synthesis and summary

Text

Database System Concepts, Sixth Edition, by Silberschatz, Korth and Sudarshan (McGraw-Hill).

Software

`sqlite3`, `git` and `gcc`. You will use `chisubmit` to submit your work (instructions will be available on piazza).

Coursework and Grading

There is one exam, to be held at the beginning of week 7 on Tuesday, May 12. There will be several problem sets to complete for homework throughout the quarter. Project 1 will be done in pairs, and Projects 2, 3 and 4 in a team of 4 students. Your course grades will be computed according to the following weights:

- Projects 1, 2, 3 and 4: 15%, 15%, 15% and 10%
- Homework: 20%
- Exam: 25%

What precisely constitutes an A, B, *etc.* will be determined by the collective performance of the class.

Late Work

Deadlines in this course are rigid. Since you submit your work electronically, deadlines are enforced to the minute. Late work will not earn credit.

(We will accept late work in the case of special circumstances, when those circumstances are extraordinary.)

Academic Honesty

In this course, as in all your courses, you must adhere to college-wide honesty guidelines as set forth at `http://college.uchicago.edu/policies-regulations/academic-integrity-student-conduct`. The college's rules have the final say in all cases. Our own paraphrase is as follows:

1. Never copy work from any other source and submit it as your own.
2. Never allow your work to be copied.
3. Never submit work identical to another student's.
4. Document all collaboration.
5. Cite your sources.