COS 480/580: Database Management Systems

Sudarshan S. Chawathe

University of Maine

Fall 2009

This course covers database systems from the perspective of database designers and database programmers (not to be confused with database system implementers). The emphasis is on fundamental topics that should be familiar to every computer scientist and good programmer. In addition to traditional topics such as Entity-Relationship modeling, relational database design theory, relational algebra and calculus, SQL, and Datalog, the course also covers object-oriented and object-relational databases, with topics such as ODL, OQL, and SQL3.

News and Reminders:

- Please read the newsgroup for timely announcements.
- Class newsgroup: Local group `umaine.cos480` on NNTP server `news.cs.umaine.edu`. Web interface to get started: `http://cs.umaine.edu/~chaw/news/`.
- The most recent version of this document may be found at `http://cs.umaine.edu/~chaw/cos480/`.
- Some sections below point to material in separate documents that are found on the class Web site, linked from the online version of this document.
- You may access a local copy of the PostgreSQL documentation (with a slightly improved formatting) at `pgsql/doc/html/`. In particular, the section describing *psql* is at `pgsql/doc/html/app-psql.html`.
- Please use the PDF version of this document for printing and reference: `cos480.pdf`

Contact Information

Class meetings:

Time: Tuesdays & Thursdays, 2:00–3:15 p.m.
Location: Neville Hall, Room 204.

Instructor: Sudarshan S. Chawathe

Office: Neville Hall, Room 224.
Office hours: (Please check for changes.)
Tuesdays & Thursdays: 8:00–8:30 a.m., 10:30–11:00 a.m., 1:45–2:00 p.m., 3:15–3:30 p.m.
Phone: +1-207-581-3930.
Please avoid calling except for truly urgent matters.
Email: `chaw@cs.umaine.edu`
Use email only for messages unsuitable for the newsgroup. (See below.) Please put the string *COS226* near the beginning of the Subject header of your messages to me.
Web: `http://cs.umaine.edu/~chaw/`.

Teaching Assistant: Mark Royer

Office: East Annex, Room 229.

Office hours: (Please check for changes.)
Mondays & Wednesdays: 1:00–4:00 p.m.
Phone: +1-207-581-2005.
Email: `mroyer@cs.umaine.edu`

Online Resources

Class Web site:
`http://cs.umaine.edu/~chaw/cos480/`
We will use the class Web site for posting homework assignments, hints, solutions, etc. Please monitor it.

Class Newsgroup: We will use the local USENET newsgroup `umaine.cos480` on the NNTP (net news) server `news.cs.umaine.edu` for electronic discussions. If you are unfamiliar with USENET, you may find the Web interface at `http://cs.umaine.edu/~chaw/news/` useful as a quick way to get started. You may find further information on USENET at `http://en.wikipedia.org/wiki/Usenet`. The newsgroup is the primary forum for electronic announcements and discussions, so please monitor it regularly, and post messages there as well. Unless there is a reason for not sharing your question or comment, please *use the newsgroup, not email*, for questions and comments related to this course.

Class mailing list: *Please make sure you are on the class mailing list.* A sign-up sheet is circulated at the first class meeting. If you miss it, please contact me to get on the list. We will use this mailing list only for urgent messages because all other messages will go on the class newsgroup. I anticipate fewer than a dozen messages on this list over the semester.

Grading Scheme

Grade components:

class participation	5 %
classroom exercises	5 %
homeworks	25 %
two quizzes (short exams)	10 %
two midterm exams	20 %
final exam	15 %
term project	20 %

Class Participation: Students are expected to contribute to learning by asking questions and making relevant comments in class and on the class newsgroup. Quality is more important than quantity. Disruptive activity contributes negatively. *Please make sure all disruptive devices are disabled while in class.* If you have a good reason for wanting to be disturbed in class, please contact me to make the appropriate arrangements.

Classroom Exercises: Our work in the classroom will include a number of short group exercises, meant to solidify understanding of the concepts being discussed. One or more such exercises are likely to be part of most class meetings. The exercises will be graded for correctness as well as effort and group work. Since attendance is not mandatory (see policies below), a significant number of low-scoring exercises will be dropped for each student. Please see me if you have concerns about the interaction of this component and the attendance policy.

Homeworks: Homeworks include programming and non-programming ones, often mixed. No collaboration is permitted. You are encouraged to discuss the problems and solution strategies *at a high level*, but the final solution and details must be your individual work. If you are unclear on the boundary between permissible and non-permissible interactions in this regard, please ask me.

Exams and Quizzes: All exams and quizzes are *open book, open notes*. You are free to bring with you any resources that you find useful. However, no communications are permitted other than between students and me. The use of computers during exams is strongly discouraged, but brief use is permitted *provided it does not cause a disturbance*. You may use the Internet, but only as a library to look up material you may find useful. As above, check with me if you are unclear on what is permitted. The exams are designed to require no equipment other than a pen and paper, along with the textbook and assigned readings.

Midterm exams will be held during regular class meetings, and will be roughly an hour long. Each quiz is a short exam, roughly half an hour long, held during part of a class meeting. The final exam follows the usual university schedule, and is thus held outside of regular class meetings.

Project: In addition to the programming and other homeworks, the course features a semester-long group project. Students will work in groups of three or more to design and implement a substantial database application. Projects will be graded based on a written project report, the submitted source code, a demonstration, and a question-and-answer session following the demo. *These materials will be due the week before finals week*, but may be submitted earlier—there is no penalty for early submissions. Further details will follow.

COS 580: There will be additional readings assigned to COS 580 students. The readings will be a mix of some classic papers of the database field and more recent publications. COS 580 students are expected to be comfortable reading such papers. There will also be additional and/or different questions on the exams and homeworks. Similarly, COS 580 students will be held to a higher standard during the question-and-answer session following the project demo.

Policies

Due dates: All due dates (and times) are strict, as announced in class. If you believe your work was delayed by truly exceptional circumstances, let me know as soon as those circumstances are known to you and I will try to make a fair allowance. However, *the default is that you get a zero if you don't turn in the work on time.*

Attendance: Although I expect students to attend all class meetings, I will not be taking attendance. If you miss a class meeting, you are responsible for catching up on the lost material, including any important announcements made in class. If you have a valid reason for missing a class, let me know early and I will try to help you make up the class. There will be no make-up exams or quizzes. A missed test earns zero credit. If you have a valid reason for missing a test, let me know as early as that reason is known to you and I will make a fair allowance (but there will be no make up exam in any case).

Make-up classes: I may have to reschedule a few classes due to my other professional commitments. I will make every attempt to minimize the number of such occurrences and to reschedule for a time that works for most students. Further, I will make sure no student is penalized by such occurrences.

Academic honesty (standard university wording): Academic dishonesty includes cheating, plagiarism and all forms of misrepresentation in academic work, and is unacceptable at The University of Maine. As stated in the University of Maine's online undergraduate Student Handbook, plagiarism (the submission of another's work without appropriate attribution) and cheating are violations of The University of Maine Student Conduct Code. An instructor who has probable cause or reason to believe a student has cheated may act upon such evidence, and should report the case to the supervising faculty member or the Department Chair for appropriate action.

Disabilities (standard university wording): If you have a disability for which you may be requesting an accommodation, please contact Ann Smith, Director of Disabilities Services, 121 East Annex, 581-2319, as early as possible in the term.

Programming

Programming: We will use PostgreSQL as the database system for programming assignments. You are free to program in any programming language that has an open implementation but *if you plan to use a language other than C or Java, you should check with me very early in the course.* (As an initial check, your proposed environment should be easily testable on `gandalf`, our main Unix host.)

Class accounts: Class accounts for Unix and PostgreSQL will be generated based on the forms distributed at the first class meeting. If you missed them, please get in touch with me. You should be able to access your accounts from anywhere on the Internet (including the labs in Neville Hall and elsewhere on campus) by using *ssh* to connect to `cs.umaine.edu`. On most Unix hosts, the command `ssh -l` *`username`* `cs.umaine.edu` should suffice. For Windows hosts, the freely available *Putty* program works well: `http://www.chiark.greenend.org.uk/~sgtatham/putty/`. *Do not use unencrypted telnet sessions to connect to your account!*

Textbook and Readings

Textbook: *A first course in database systems.* Jeffrey D. Ullman and Jennifer Widom. Prentice-Hall, Upper Saddle River, New Jersey, second edition, 2001.

The textbook's Web site has many useful resources: `http://www-db.stanford.edu/~ullman/fcdb.html`. In particular, for a more detailed listing of course topics, please refer to the textbook's table of contents: `http://www-db.stanford.edu/~ullman/pub/fcdb-toc.txt`

Readings: Items marked with ⋆ are required for COS 580 students. COS 480 students may wish to read them if they plan to attempt the extra-credit questions on tests. Readings marked with ⋆⋆ are extra credit for COS 580 students and double-extra credit for COS 480 students. Students who wish to receive credit for ⋆⋆ items must discuss the specifics with me first. Everyone is encouraged to at least browse all the readings.

1. Edgar F. Codd. A relational model of data for large shared data banks. *Communications of the ACM*, 13(6):377–387, June 1970.
2. **Notes** on Codd's paper: `notes/rmodel.pdf`; `notes/rmodel/rmodel.html`.
3. [A recent paper for 480 and 580 will be added here.]
4. ⋆ Goetz Graefe. Query evaluation techniques for large databases. *ACM Computing Surveys*, 25(2):73–170, June 1993.
5. **Notes** on Graefe's paper: `notes/qeval.pdf`; `notes/qeval/qeval.html`.
6. [A recent paper for 580 will be added here.]
7. ⋆⋆ François Bancilhon and Raghu Ramakrishnan. An amateur's introduction to recursive query processing strategies. In *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD)*, pages 16–52, Washington, D.C., May 1986.

Further Reading: These books are *not* required reading and nothing in the course will depend directly on reading them. However, they are good sources for different explanations of some concepts, additional information on various topics, examples, and exercises.

1. Serge Abiteboul, Richard Hull, and Victor Vianu. *Foundations of Databases.* Addison-Wesley, 1995. This book is a textbook for COS 598, Advanced Topics in Databases, and focuses on Database Theory. The book is not light reading but it is much easier than reading the equivalent set of papers.

2. Hector Garcia-Molina, Jeffrey D. Ullman, and Jennifer Widom. *Database Systems: The Complete Book*. Prentice-Hall, 2002.
   The first half of this book is essentially identical to the main textbook. The second half covers topics in database system implementation, and is a good resource for learning more about how database systems are implemented. Since the terminology and style is consistent with the main textbook, it should be easy reading.

3. Christopher J. Date. *An Introduction to Database Systems*. Addison-Wesley, Reading, Massachusetts, 2000.
   A classic database-systems textbook.

4. Raghu Ramakrishnan and Johannes Gehrke. *Database Management Systems*. McGraw-Hill, third edition, 2002.
   Another standard textbook with detailed coverage of some topics that we will cover briefly.

5. Roderic Geoffrey Galton Cattell. *Object Data Management: Object-Oriented and Extended Relational Database Systems*. Addison-Wesley, Reading, Massachusetts, 1994.
   A good introduction to object and object-relational databases.

6. François Bancilhon, Claude Delobel, and Paris Kanellakis. *Building an Object-Oriented Database System: The Story of O2*. Morgan Kaufmann, 1992.
   Another good book on object databases.

7. Michael Stonebraker and Joseph M. Hellerstein, editors. *Readings in Database Systems*. Morgan Kaufmaann, San Francisco, California, third edition, 1998.
   This collection of papers, including some classics, provides a sampling of topics in database system implementation.

Homeworks and Tests

Homework assignments, exams, and solutions will appear here as we move along the semester.

It may be useful to refer to the homeworks and tests from the previous session: `http://cs.umaine.edu/~chaw/200809/cos480/`.

Schedule

At the beginning and end of each class, I will announce sections of the textbook covered in each class and those due at the next class. An approximate schedule appears in Figure 1. Please use it only as a rough guide to plan your studies. Do *not* use it to schedule travel or other events. If you need a definite answer on when something will or will not occur, you should check with me.

Tuesday	Thursday
September 1st C1 Introduction; simple Relational Algebra. §§ 3.0, 3.1, 5.0, 5.1, 5.2.	3rd C2 Simple SQL. §§ 6.1, 6.2.
8th C3 SQL, continued. §§ 6.3, 6.4.	10th C4 SQL, continued. §§ 6.5, 6.6.
15th C5 SQL, continued. §§ 6.7, 5.3, 5.4.	17th C6 ★ **Quiz 1**, regular class time & place.
22nd C7 §§ 8.1, 8.3, 8.4, 8.5	24th C8 §§ 2.1, 2.2. Readings 1 & 2.
29th C9 §§ 2.3, 2.4.	**October** 1st C10 ★ **Midterm Exam 1**, regular class time & place.
6th C11 Special tutorial.	8th C12 §§ 8.6, 8.7
13th ×*No class.* Fall break Oct. 10th–13th.	15th C13 §§ 3.2, 3.3.
20th C14 §§ 3.4, 3.5.	22nd C15 §§ 3.6, 3.7.
27th C16 §§ 7.1, 7.2.	29th C17 ★ **Quiz 2**, regular class time & place.
November 3rd C18 §§ 7.3, 7.4.	5th C19 §§ 4.1, 4.2, 4.3.
10th C20 4.4, 4.5, 4.6, 4.7.	12th C21 §§ 9.1, 9.2, 9.3.
17th C22 ★ **Midterm Exam 2**, regular class time & place.	19th C23 §§ 9.4, 9.5.
24th C24 §§ 10.1, 10.2; Reading 3.	26th ×*No class.* Thanksgiving break Nov. 25th–29th.
December 1st C25 §§ 10.3, 10.4.	3rd C26 §§ 8.2, special topic.
8th C27 Review.	10th C28 Review.
15th ×*No class.* Finals week Dec.14th–18th.	17th ×*No class.* Finals week Dec.14th–18th. ★ **Final exam will be as scheduled by the University.**

Figure 1: ***Approximate*** *schedule, likely to change.*