

COS 480/580: DATABASE MANAGEMENT SYSTEMS

Sudarshan S. Chawathe

University of Maine

Spring 2016

THIS COURSE COVERS DATABASE SYSTEMS FROM THE PERSPECTIVE OF DATABASE DESIGNERS AND PROGRAMMERS, different from database system implementers. The emphasis is on fundamental topics that should be familiar to every computer scientist and good programmer. The course covers topics such as Entity-Relationship modeling, relational database design theory, relational algebra and calculus, SQL, Datalog, object-oriented and object-relational databases, with OQL and SQL3, and semistructured databases, with XQuery.

Prerequisites: COS 350 (data structures and algorithms); programming maturity.

News and Reminders:

- Please read the newsgroup for timely announcements.
- Class newsgroup: Local group `umaine.cos480` on NNTP server `creak.um.maine.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.
- Please use the PDF version of this document for printing and reference: `cos480.pdf`

Goals and Outcomes

Goals

- Learn the fundamentals of database theory and practice.
- Acquire skills in forming declarative specifications in database query languages such as SQL, relational algebra, and XQuery.
- Learn techniques for mapping domain concepts to database constructs in a systematic manner with a good understanding of the consequences of each decision.
- Gain experience in reading the relevant research literature and other publications used to disseminate knowledge in the field.
- Practice the appropriate and ethical use of existing material of different kinds, such as source code, services, and documentation.
- Gain experience in contributing to the body of knowledge.
- Learn how to analyze the efficiency of data-intensive programs, with an emphasis on database queries.
- Gain experience in conducting and documenting experimental studies of database programs.
- Improve programming skills, with attention to data-intensive programs and systems.
- Improve communication skills, with particular emphasis on written communication and, further, well-written programs.
- Practice managing a self-directed project.

Student Learning Outcomes

Upon successful completion of this course, students should be able to

- Explain the key aspects of well-known data models.
- Interpret queries in languages such as SQL, OQL, XQuery, and Datalog.
- Precisely describe the semantics of queries.
- Determine and justify equivalence and non-equivalence of queries within and across query languages.
- Effectively demarcate the data-centric portions of an application or task.
- Map data processing tasks from abstract or informal specifications to concrete queries in various query languages.
- Justify the correctness of queries.
- Explain the primary determinants of performance in commonly encountered data management systems.
- Analyze the running times of queries and data-intensive programs using simple methods.
- Choose appropriate data management systems and programming interfaces for various tasks.
- Create applications that make effective use of database technology.
- Effectively read suitable publications related to the topic.
- Use resources such as others' code and writing in an ethical and professional manner.
- Contribute to the body of knowledge at an undergraduate level.
- Perform simple experimental studies of programs, with an emphasis on data-intensive portions.
- Program with attention to community standards and good practices.
- Communicate their programming work effectively.
- Meet Quantitative Literacy General Education requirements, such as being able to [following text is from U. Maine Gen. Ed. documents]:
 - Translate problems from everyday spoken and written language to appropriate quantitative questions.
 - Interpret quantitative information from formulas, graphs, tables, schematics, simulations, and visualizations, and draw inferences from that information.
 - Solve problems using arithmetical, algebraic, geometrical, statistical, or computational methods.
 - Analyze answers to quantitative problems in order to determine reasonableness. Suggest alternative approaches if necessary.
 - Represent quantitative information symbolically, visually, and numerically.
 - Present quantitative results in context using everyday spoken and written language as well as using formulas, graphs, tables, schematics, simulations, and visualizations.

Contact Information

Class meetings:

Time: Tuesdays & Thursdays, 11:00 a.m.–12:15 p.m.

Location: Little Hall, Room 203.

Instructor: Sudarshan S. Chawathe

Office: Neville Hall, Room 224.

Office hours: (Please check for changes.)

Tuesdays and Thursdays: 3:15–4:30 p.m.

Phone: +1-207-581-3930.

Please avoid calling except for truly urgent matters.

Email: sudarshan.chawathe@maine.edu

Use email only for messages unsuitable for the newsgroup. (See below.) Please use only this email address and put the string *COS350* near the beginning of the Subject header of the message. *All other messages may be ignored.*

Web: <http://cs.umaine.edu/~chaw/>.

Teaching Assistant (COS 480): Zachary Hutchinson

Office: Boardman Hall, Room 249.
Office hours: (Please check for changes.)
Mondays and Wednesdays: 12:00–2:30 p.m.
Email: zachary.s.hutchinson@maine.edu

Online Resources

Class Web site:

<http://cs.umaine.edu/~chaw/cos480/>

We will use the class Web site for posting assignments, readings, notes, and other material. Please monitor it.

Class Newsgroup: We will use the local USENET newsgroup `umaine.cos480` on the NNTP (net news) server `creak.um.maine.edu` for electronic discussions. The Web interface at <http://cs.umaine.edu/~chaw/news/> provides convenient access. Some further, more general, information on USENET appears 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 a 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.* The mailing list will use the email address for each student as recorded in the official university records (*Maine Street* system). 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: *Students are expected to complete and submit all assigned coursework in good faith; those who fail to do so will earn a **failing grade, regardless of overall numerical score.***

component	% of grade
class participation	5
classroom exercises	5
homeworks	25
three quizzes (short exams)	9
two midterm exams	20
final exam	20
term project	16

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. See policies below.

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 primarily for effort, group work, and other contributions, and less so for simple correctness. Since attendance is not mandatory (cf. policies), some low-scoring exercises will be dropped for each student. Please ask for clarifications if there are 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. Everyone is encouraged to discuss the problems and solution strategies *at a high level*,

but the final solution and details must be individual work. If the boundary between permissible and non-permissible interactions is unclear, please ask for clarifications.

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 may be permitted provided it does not cause a disturbance, at the discretion of the proctor. You may use the Internet, but only as a library to look up material you may find useful. Ask for clarifications in case of any doubt. 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 meeting times, and often at a different location.

Term Project: In addition to the programming and other homeworks, this course features a term project. Group term projects are strongly encouraged, with groups of two to four students being typical, but individual term projects are also permitted. The details of the project are fairly flexible, and all students are encouraged to propose project that excites them. A few project ideas will also be provided in class for use as term projects, perhaps with some of modifications. The main requirement for the project is that it demonstrate the ability to work independently and apply the concepts studied in the course to an application. Projects will be graded based on a project submission that includes a project report, complete and well-documented source code and build instructions, a script for a demonstration, a demonstration, and a question-and-answer session following the demo. *These materials will be due two weeks before finals week.* Further details will be announced in class.

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 (or become) comfortable reading such papers. There will also often be additional or different questions on the exams and homeworks for COS 580. In general, COS 580 students are expected to perform to a higher standard on all submissions and tests. Although the class meetings and discussions are shared with COS 480, the assessment and grading is different and separate.

Policies

Due dates: All due dates and times, as announced in class, are strict, to the second. 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*, and fail the class if you don't turn it in at all (cf. Grade Components above).

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, on your own.* 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 tests in any case.

Classroom activities: This course is based on an active learning format, so effective classroom activities are critical to its success. Students are expected to contribute to their own learning and that of their classmates, and to devote 100% of their attention to these activities while in class. On a similar note, all electronic and other distractions (computers, phones, assorted gizmos, etc.) must be completely silenced and put away for the entire duration of the class. (Students who need any such devices

for disability accommodations should follow the guidelines outlined below. Others who need any accommodation in this regard due to special circumstances should make advance arrangements with the instructor.) No food or drink is allowed in class, other than water in a spill-proof container. Students who violate these rules or otherwise cause distractions in class will be asked to leave with *no warning*; habitual violators will face disciplinary action.

Office hours: All students are encouraged to make use of both the instructor's and TA's office hours to further their learning, obtain assistance on homework assignments, obtain feedback on their class performance, etc. However, office hours are not to be used as a substitute for attending and participating in class meetings (see above). Similarly, assistance with homework assignments will be limited to what is appropriate based on fairness to all; students are expected to demonstrate substantial effort on the assignment before seeking assistance.

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.

University of Maine administrative policy statements: [verbatim, standard wording]

Academic honesty: Academic honesty is very important. It is dishonest to cheat on exams, to copy term papers, to submit papers written by another person, to fake experimental results, or to copy or reword parts of books or articles into your own papers without appropriately citing the source. Students committing or aiding in any of these violations may be given failing grades for an assignment or for an entire course, at the discretion of the instructor. In addition to any academic action taken by an instructor, these violations are also subject to action under the University of Maine Student Conduct Code. The maximum possible sanction under the student conduct code is dismissal from the University.

Students with disabilities: If you have a disability for which you may be requesting an accommodation, please contact Disabilities Services, 121 East Annex, 581-2319, as early as possible in the term.

Course schedule disclaimer (disruption clause): In the event of an extended disruption of normal classroom activities, the format for this course may be modified to enable its completion within its programmed time frame. In that event, you will be provided an addendum to the syllabus that will supersede this version.

Sexual Discrimination Reporting: The University of Maine is committed to making campus a safe place for students. Because of this commitment, if you tell a teacher about an experience of sexual assault, sexual harassment, stalking, relationship abuse (dating violence and domestic violence), sexual misconduct or any form of gender discrimination involving members of the campus, your teacher is required to report this information to the campus Office of Sexual Assault & Violence Prevention or the Office of Equal Opportunity.

If you want to talk in confidence to someone about an experience of sexual discrimination, please contact these resources:

For confidential resources on campus: Counseling Center: 207-581-1392 or Cutler Health Center: at 207-581-4000.

For confidential resources off campus: Rape Response Services: 1-800-310-0000 or Spruce Run: 1-800-863-9909.

Other resources: The resources listed below can offer support but may have to report the incident to others who can help:

For support services on campus: Office of Sexual Assault & Violence Prevention: 207-581-1406, Office of Community Standards: 207-581-1409, University of Maine Police: 207-581-4040 or 911. Or see the OSASP website for a complete list of services at <http://www.umaine.edu/osavp/>

Programming

This course focuses on high-level concepts that are mostly oblivious to choices of programming languages and environments. However, in order to provide concrete realizations of these concepts, we will use PostgreSQL and SQLite as the primary database systems, Java as the primary programming languages and a POSIX environment as the primary operating system. Submissions will be in the form of packaged, well documented *source* files. *Proper documentation and packaging of source code and other material is a crucial component of assigned work and submissions failing in this regard will receive no credit.*

Programming Environment and Tools: You are free to choose details such as operating system, development environment, and editor based on your preferences. However, no matter what you use, the submission should be a *source-code* package that works on the host *aturing* (see below). Further details on the packaging, submission, and testing procedure will be provided in class and on the newsgroup.

Other Languages: If you prefer to use other programming languages or systems, please contact me by the second class meeting. I am quite open to the idea, and encourage interested students to explore it further. However, please check with me very early in the semester so that we can determine the specifics to make sure your submissions can be tested and graded fairly. You should avail of this option only if you are confident enough of your programming skills to not require any programming help, and are prepared to take on additional work. *This option is designed for students who are proficient in Java and wish to use this opportunity to master another language, not for students weak in Java and who wish to avoid them.* Anyone granted this option will still be responsible for all the material related to the default languages and systems used in the course.

Literate Programming: All submitted work must use a *literate programming style*: Your programs must be designed with *a human as the intended reader*, although they must also compile and run correctly. *Programs that do not meet this requirement are likely to receive a zero score with no further consideration.* Details will be discussed in class. The use of any specific literate-programming or documentation tool is neither necessary nor sufficient for this requirement.

Class accounts: Shell and PostgreSQL will be generated on the host `aturing.umcs.maine.edu` based on registration records. These accounts are required for successful completion of homeworks and other assignments. You should be able to access your accounts from anywhere on the Internet by using *ssh*. On most Unix-like hosts (GNU/Linux, Mac OS), the command `ssh -l username aturing.umcs.maine.edu` should suffice. For Windows hosts, the freely available *Putty* program works well.

Schedule

A rigid schedule is not conducive to effective learning, since it would limit our flexibility in exploring ideas as they arise in class. The actual schedule (both the timing and the selection of topics) will be determined by in-class interactions. Nevertheless, a partial and *approximate* schedule, to serve as a baseline, appears in Figure 1; it will be updated as we progress. 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.

At the beginning and end of each class, I typically announce the topics and textbook sections covered in that class and those due at the next class. It is important that students read the material *before* the class in which it is discussed and, in general, keep up with readings and studies.

Textbook and Readings

Textbook: *A first course in database systems.* Jeffrey D. Ullman and Jennifer Widom. Prentice-Hall, Upper Saddle River, New Jersey, third edition, 2008. The university bookstore carries this book, which is a required textbook for this course.

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 \star 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 $\star\star$ are extra credit for COS 580 students and double-extra credit for COS 480 students. Students who wish to receive credit for $\star\star$ items must discuss the specifics with me first. Everyone is encouraged to at least browse all the readings. This list will be updated to reflect class preferences this semester.

TUESDAY		THURSDAY	
January 19th Introduction; relational algebra. §§ 2.{1,2,4}.	C1	21st DBMS overview; queries; simple SQL. §§ 1.*, 2.3, 2.5.	C2
26th relational model and queries. §§ Readings 1 & 2.	C3	28th ★ Quiz 1 , regular class time & place.	C4
February 2nd SQL in detail. §§ 6.*	C5	4th SQL details, transactions. §§ 6.*	C6
9th Extended relational algebra. §§ 5.*	C7	11th Extended relational algebra. §§ 5.*	C8
16th ★ Midterm Exam 1 , regular class time & place.	C9	18th	C10
23rd	C11	25th	C12
March 1st	C13	3rd ★ Quiz 2 , regular class time & place.	C14
8th × <i>No class</i> . Spring break Mar. 7th–20th.		10th × <i>No class</i> . Spring break Mar. 7th–20th.	
15th × <i>No class</i> . Spring break Mar. 7th–20th.		17th × <i>No class</i> . Spring break Mar. 7th–20th.	
22nd	C15	24th	C16
29th	C17	31st ★ Quiz 2 , regular class time & place.	C18
April 5th	C19	7th	C20
12th	C21	14th	C22
19th ★ Midterm Exam 2 , regular class time & place.	C23	21st	C24
26th Special topic.	C25	28th Term project submissions due. Special topic.	C26
May 3rd ★ Term Projects Demonstrations.	C27	5th Synthesis and review.	C28
10th × <i>No class</i> . Finals week May 9th–13th.		12th × <i>No class</i> . Finals week May 9th–13th. ★ Check Univ. schedule for final exam.	

Figure 1: **Approximate** schedule, likely to change. Textbook items are in §§ chapter.section format.

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*. Pearson, 2nd edition, 2009. 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*. Pearson, 2004. A classic database-systems textbook.
4. Raghu Ramakrishnan and Johannes Gehrke. *Database Management Systems*. McGraw-Hill, 3rd edition, 2002. Another standard textbook with detailed coverage of some topics that we will cover only briefly.

Exercises, Homeworks, Tests, and Notes

Material will appear here as we move along the semester. It may be useful to refer to material from the previous session: <http://cs.umaine.edu/~chaw/201309/cos480/>.

- Class exercises:
 - Class Exercise 1: [hwq/ce01.pdf](#).
 - Class Exercise 2: [hwq/ce02.pdf](#).
 - Class Exercise 3: [hwq/ce03.pdf](#).
 - Class Exercise 4: [hwq/ce04.pdf](#).
 - Class Exercise 5: [hwq/ce05.pdf](#).
- Homework assignments:

- Homework 1: hwq/hw01.pdf.
- Quizzes and Exams:
 - Quiz 1: hwq/q01.pdf.

Homework and Project Submissions

Handwritten answers to non-programming problems should be submitted in class on the due date, at the beginning of class (within the first five minutes), unless prior alternate arrangements are made. If you prefer to type your answers, please make sure the result uses the proper symbolic notation for mathematical constructs. *Illegible, hard to read, or otherwise messy submissions, whether handwritten or typed, are likely to be returned without grading, for zero credit.* Answers to programming problems should be submitted electronically, using the packaging and submission procedure that will be described in class and on the class newsgroup.

All electronic submissions must be made using the upload interface at <http://cs.umaine.edu/~chaw/u/>. Electronic submissions in **all other forms**, such as email or physical media, will be **discarded and receive no credit**.

If your upload is successful, you will be presented with a confirmation Web page similar to the following sample. You should record the reported MD5 checksum and timestamp (important in case there are undetected problems).

```
SUCCESS: Please note the following for your records.
```

```
Successfully saved cos480-hw01-aardvark-alice-4233.tgz.  
MD5 checksum: 09ee098b83d94c7c046d6b55ebe84ae1  
Timestamp: 2016-01-23 14:37:42
```

If you do not see something very similar then your submission is unsuccessful.

If (and only if) there are unexpected problems and you are unable to submit your work as above, then you should save your file on your own computer (with some backups), compute its MD5 checksum using the md5sum utility on Unix-like systems (or other similar tools), and submit the file name, time stamp, and MD5 checksum (only, not the file itself) by email with a suitable Subject header.