COSC 460 Databases
Fall 2018

Time	Lecture MW 1:20–2:35 pm
	Lab A Tu 1:20–3:10 pm
	Lab B W 3:20–5:10 pm
Location	314 McGregory (lecture), 328 McGregory (lab)
Instructor	Prof. Michael Hay (303 McGregory, mhay@colgate.edu)
Office hours	Tuesday 12:30-1:20
	Thursday 1:00-3:30
	and by appointment (see course website for details)
Open Lab Hours	McGregory 328, Sun-Thu, 7-10PM

1 Course Description

This course introduces the principles underlying modern database systems. These principles guide how information is represented as structured data, how computations on the data are expressed in query languages, and how systems are designed to enable efficient computation on large data sets. Topics include database design, data models, query languages, query processing and optimization, data storage and access, transaction management, and advanced topics as time permits. The required credit-bearing laboratory COSC 460L must be taken concurrently with COSC 460.

Prerequisites: COSC 290 and 301.

2 Materials & Resources

Course schedule/website: `http://cs.colgate.edu/~mhay/cosc460/` (link is also available through Moodle)

Textbook (required): *Database Management Systems (3rd edition)* by Raghu Ramakrishnan and Johannes Gehrke. `http://pages.cs.wisc.edu/~dbbook/` This is also known as the "Cow" book. It is available in the Colgate Bookstore.

Textbook (optional): *Database Systems Concepts (6th edition)* by Avi Silberschatz, Henry F. Korth, and S. Sudarshan. `http://codex.cs.yale.edu/avi/db-book/`. This is also known as the "Boat" book. This book covers the same topics as the Cow book and I have used this book in the past. If the Cow book isn't working for you, you are free to read the corresponding chapters from this book instead. I have a copy in my office which you are welcome to browse.

Textbook (recommended): The lab work will be conducted in Java and make heavy use of object-oriented programming and data structures. It is recommnded that you have available a textbook or some other resource that covers not only Java but also basic data structures (lists, maps, trees, hashtables, etc.). Your primary resources from COSC 102 should be sufficient.

Additional readings (required): Additional supplementary readings may be posted online or on reserve. Unless otherwise specified, you can assume these readings are *required*.

Software (optional): The lab computers have all of the software needed for this course installed. These computers are available during lab and open lab hours in the evenings (schedule TBD). If you prefer to work on a different machine, you are responsible for installing the necessary software and figuring out how to adapt the assignment instructions to match the particulars of your computer setup. This includes some kind of text editor (e.g., Atom, `https://atom.io/`), Java, sqlite, postgresql. Each assignment will include any additional software dependencies as well as instructions on how to get up and running.

Piazza (required): We will use Piazza for online discussion. It's accessible via Moodle.

3 Course & Lab Work

3.1 Course Work

Reading: Reading assignments for each lecture will be posted on the schedule. You are expected to complete the reading *before* class.

Problem sets: To help you assimilate the core concepts and prepare for the exams, there will be frequent "pencil-and-paper" problem sets.

Exams: There will be two midterms and a final exam. *The exam dates are on the schedule. If you have a conflict with the scheduled exam time, you are expected to notify me as soon as possible, at least one week prior to the exam.* All exams are closed book, closed notes, and collaboration is not permitted.

Participation in lecture: My goal is to make the classroom a fun and supportive learning environment. To achieve this goal, I need your help. Please come to class on time, mentally and physically ready to engage in the learning process (i.e., pay attention, ask questions, answer questions, give your best effort on in-class exercises, etc.).

Participation on piazza: You are expected to monitor piazza for course announcements and you are strongly encouraged to ask (and answer!) questions on Piazza.

"Beyond Relational Databases" Project The course focuses primarily on the relational model (and its associated programming language, SQL) but you will be expected to explore other models in the form of a research paper/project. More details about the project will be given later in the semester.

3.2 Lab Work

Lab: To complete this course, you must also sign up for the 2-hour weekly laboratory section. The lab is designed to reinforce your conceptual understanding of how databases work by building one yourself. In a sequence of labs that spans the entire semester, you will build ColgateDB, a lightweight database architecture that nevertheless supports SQL and transactions – meaning that it can handle multiple concurrent users querying and modifying the database and can even recovery gracefully from system crashes. You will implement a buffer manager, several relational algebra operators, a lock manager, and a recovery system. ColgateDB is written in Java.

Participation in lab: Since lab time will be used to work intensively on the lab assignments, it is important that you come prepared and attend every lab session. Please attend the lab section to which you are enrolled. Your participation grade will be determined by the following factors: on-time arrival, preparedness (completing relevant reading, reading the lab assignment closely, etc), productive use of lab time, appropriate and respectful behavior towards lab mates and support staff, on-time departure, etc.

4 Grading Guidelines

Lab work is a significant component of the course work. Therefore, you will receive a single grade for the course as a whole and that one grade will be submitted to the registrar for both the course (COSC460) and lab (COSC460L).

The composition of your final grade is as follows.

Coursework	**Portion of grade**
Problem sets	5%
Lab assignments	40%
Midterm 1	10%
Midterm 2	10%
Final	25% (see note 1)
"Beyond Relational Databases" Project	10%

[Note 1] To pass the course, you must pass the final exam.

4.1 Lab Grading

The grading scheme is non-standard. Each lab has a "milestone," typically in the form of passing automated tests or turning in some worksheet exercises. Late submissions miss the milestone and will lose points according to the late policy.

In addition, I will periodically conduct code reviews (roughly 4 times over the semester). This is an opportunity to give you specific feedback on code quality. Code reviews may also include running additional tests for correctness.

Your final lab grade will be a weighted combination of the milestones (40%) and the code reviews (60%). The milestone portion is a simple average of the individual milestone grades; the code review portion is a simple average of the individual code review grades.

Be warned that the assignments are cumulative and thus errors in early labs must be resolved otherwise they can be impact the grade on later labs.

4.2 Problem Set Grading

Problem sets will be graded on a simple scale: satisfactory (1 point) or unsatisfactory (0 points). A student who gets satisfactory on (roughly) 80% or more of the collected problem sets will receive full credit. Late work is unsatisfactory.

4.3 Grading Scale

Grading is on an absolute scale (*i.e.*, no curve). Final grades are determined as follows. As a general rule, fractions are rounded down (e.g., an 89.9 is a B+, not an A-). A grade of A+ is awarded when the student demonstrates truly exceptional performance and is not simply determined by having a high final course grade.

F	D-	D	D+	C-	C	C+	B-	B	B+	A-	A	A+
< 60	60–62	63–66	67–69	70–72	73–76	77–79	80–82	83–86	87–89	90–92	≥ 93	*

5 Schedule & Topics

A schedule is outlined below. This schedule is *tentative*. Please refer to the online schedule for the most up to date schedule.

	Date	Title	Reading
0	2018-08-30	Half-day introduction	
1	2018-09-03	Relational model	1,3-3.2,3.4,3.6,skim 3.7
2	2018-09-04	Lab: Tuples	
3	2018-09-05	Storage 1: Memory hierarchy, buffer management	8-8.2.1, 9-9.1,skim 9.2,9.3,9.4
4	2018-09-10	Storage 2: Page and record formats	9.5-9.8
5	2018-09-11	Lab: Slotted Page	
6	2018-09-12	Relational algebra	4-4.2
7	2018-09-17	Indexing 1: Overview, trees	8.2.1-8.6,10-10.1
8	2018-09-18	Lab: Page Formatter	
9	2018-09-19	SQL 1: Introduction	5-5.3
10	2018-09-24	Indexing 2: B+-trees	10.2-10.9
11	2018-09-25	Lab: Buffer Manager	
12	2018-09-26	SQL 2: Aggregation (instructor travel)	5.4-5.5
13	2018-10-01	Indexing 3: Hash-based	11
14	2018-10-02	Lab: Heapfile	
15	2018-10-03	SQL 3: Nested queries	5.4-5.5
16	2018-10-03	Exam 1: 7-10pm, 314 McGregory	
17	2018-10-08	No class (fall break)	
18	2018-10-09	No lab (fall break)	
19	2018-10-10	Query Processing 1: Introduction, Sorting	12-12.4,13-13.2
20	2018-10-15	TBD (instructor travel)	
21	2018-10-16	Lab: Operators	
22	2018-10-17	QP 2: Sorting continued...	13
23	2018-10-22	QP 3: Joins	12.3.2-12.4,14.4
24	2018-10-23	Lab: TBD (instructor travel)	
25	2018-10-24	QP 4: Other operations	14
26	2018-10-29	Transactions 1: ACID	16-16.3
27	2018-10-30	Lab: Java concurrency tutorial	
28	2018-10-31	Transactions 2: Serializability, 2PL	16.4-16.5,17.1-17.3
29	2018-11-05	Transactions 3: Deadlocks	17.4-17.5
30	2018-11-05	Exam 2: 7-10pm, 314 McGregory	
31	2018-11-06	Lab: Lock Manager	
32	2018-11-07	Recovery 1: Undo only, Redo only	Boat Ch. 16
33	2018-11-12	Recovery 2: Undo+redo, checkpointing	review Boat Ch. 16
34	2018-11-13	Lab: Deadlock detection	
35	2018-11-14	Slack	
36	2018-11-19	No class (thanksigiving)	
37	2018-11-20	No lab (thanksgiving)	
38	2018-11-21	No class (thanksigiving)	
39	2018-11-26	TBD: beyond relational databases	
40	2018-11-27	Lab: Recovery	
41	2018-11-28	TBD: beyond relational databases	
42	2018-12-03	TBD: beyond relational databases	
43	2018-12-04	Lab: TBD	
44	2018-12-05	TBD: beyond relational databases	

Continued on next page

	Date	Title	Reading
45	2018-12-10	TBD: beyond relational databases	
46	2018-12-11	Lab: TBD	
47	2018-12-12	TBD: beyond relational databases	
48	2018-12-18	Final Exam (3-5pm)	

6 Policies

Attendance You are expected to regularly attend class, though formal attendance is not taken.

Sharing work (anonymously) At times throughout the semester, I may want to share your work with other students in the course and on occassion, share your work with the entire class. Any time student work is shared, it will be done anonymously and only short selections of student work will be shared at any one time (e.g., one short code snippet). *If you are uncomfortable with me sharing your work anonymously, please let me know.*

Late assignment submissions A late submission is one that is not submitted before the deadline. Therefore, an assignment submitted one second after the deadline is considered late. Late submissions can earn partial credit. The default penalties for late assignments are as follows: If an assignment is submitted within the first 24 hours after the deadline can receive a maximum score of 90%; within 24-48 hours, a maximum score of 80%; and so on. There may be some assignments where the policy on late submission differs from the above; in those cases, the policy will be stated on the assignment.

Academic honesty You are expected to abide by Colgate's academic honor code: `http://www.colgate.edu/student-handbook2/academic-dishonesty-and-the-academic-honor-code`.

The guidelines for permissible collaboration and use of external resources are as follows:

- For exams and quizzes, you are not permitted to collaborate or use outside resources of any kind unless explicitly stated on the exam/quiz.
- Any artifacts (problem set answers, lab write up, code, etc.) that you submit must be *entirely your own work*. Copying from another student or resource, in whole or in part, is strictly forbidden.
- Appropriate citation of collaborators/external resources. If an assignment permits you to discuss high-level ideas with other students and you do so, you are expected to clearly acknowledge any collaboration (e.g., a brief note at the top of the submitted work saying something such as, "Talked over problem 5 with Joe Smith" would suffice). If you use any outside resources (beyond assigned readings, course notes, and other instructor-provided materials), you must provide appropriate citation. Again, a simple note indicating the resource and how it was used (e.g., "For problem 4, I used `www.example.com`, which describes an algorithm for [X]."). *Failing to acknowledge your collaborators or outside resources can be considered a violation of the honor code.*
- Group work. If an assignment permits you to work in groups, you may of course collaborate with your group mates. However, generally speaking collaboration between groups is forbidden.

Academic Support and Disabilities Services If you feel you may need an accommodation based on the impact of a disability, please contact Lynn Waldman, Director of Academic Support and Disability Services at 315-228-7375 in the Center for Learning, Teaching, and Research. `http://www.colgate.edu/cltr/academic-support-and-disability-services`

7 Student Resources

Open Labs The department organizes open lab hours (schedule TBA, but generally 7-10 most evenings) where you can use department labs to work *collaboratively* with your class mates and seek help from the available tutors (many of whom have taken this course).

One-on-one tutoring It is possible to arrange one-on-one tutoring support through CLTR. This is arranged on a per student basis. Please come see me if you are interested in exploring this option. You may also wish to visit the CLTR website: `http://www.colgate.edu/cltr`.

Borrowing computing equipment The department has a limited number of computers available for temporary loan. You *must* request and obtain permission before borrowing equipment – simply taking a laptop from the classroom is *not* permitted! If you are interested in borrowing a computer, contact me.

NASC Liaison Group NASC liaisons are a group of natural science and mathematics faculty members dedicated to providing science-interested students from underrepresented groups with mentorship, motivation, and individualized support as they navigate their paths in the sciences at Colgate. NASC liaisons do not replace the role of an academic advisor or offer formal academic advising. Rather a NASC liaison may meet one-on-one with a student to give another perspective on their academic plan; give tips on effective studying; or introduce a student to upper-class peers, alumni, or other faculty members that might be able to help them. The roles of NASC liaisons will depend on students' needs, and we encourage students to reach out for mentorship and moral support.

The NASC Liaison Group includes professors Gerry Gogel (Chemistry), Engda Hagos (Biology), Silvia Jiménez Bolaños (Mathematics), Patricia Jue (Chemistry), Spencer Kelly (Psychology & Neuroscience), Amy Leventer (Geology), Rebecca Metzler (Physics & Astronomy), Jason Meyers (Biology), and Elodie Fourquet / Joel Sommers (Computer Science). See the website for the most up-to-date information. (`http://www.colgate.edu/academics/departments-and-programs/division-of-natural-sciences-and-mathematics`)

Case Library/Informational Literacy and Reference: Use of the stellar library offerings, including the services of the outstanding reference and informational literacy librarians, is something to be made the most of during your time at Colgate. I suggest you get to know the librarians and to use their exceptional and imaginative expertise for assistance in ways that will enrich and enliven your intellectual studies and academic work.

Counseling Center: Dawn LaFrance, Director. `http://www.colgate.edu/offices/support/counseling`. Life at college can sometimes get bumpy; if you are experiencing emotional and personal difficulties (related to college or not), the Counseling Center offers completely confidential and highly professional services, both for individuals and groups.

ITS: IT Service Desk. Support and expertise related to computer and technology questions and problems, such as Moodle, email, Internet and public access computers on campus. Phone: (228-7111) Location: Third Floor of Case Geyer Library `http://www.colgate.edu/offices-and-services/information-technology`