Instructor Information: Professor Daniel Miranker
mailto:miranker@cs.utexas.edu

Lectures: Tuesdays and Thursdays 9:30AM-11:00AM, WAG 420

Classes will be held in person, unless there is a significant change in pandemic status. If Austin, Travis County returns to high-risk status, or campus health officials elevate risk concerns to similar conditions class will be held remotely by Zoom. All the rules about remote lectures will apply.

Lectures will be supplemented by on-line materials including video with and without associated SaaS interactive programming environments. Occasionally, scheduled class sessions will include students doing hands on work using software installed on their computer. If so, this will be announced in advance and details of all software to be installed on students computers prior to lecture will be also announced, in advance. A minimum personal computer configuration is not an issue. All class work that requires computer capability of any consequence will use cloud-based compute servers, (e.g. Amazon Web Services).

Office Hours:
Professor Miranker, Tuesdays, 1:00 – 2:00PM, Thursdays 11:00 – 12:00AM, GDC 5.436 (furthest office from the elevators), and by appointment set up by email. If you are not going to be at office hours at the beginning of the office hour, you must send an email as to when you will appear. (i.e. if no one appears at the beginning of office hours, and I haven’t been warned of a late arrival, the office hour will end).

TA: Zeyuan Hu, TBD

LMS and Communication Infrastructure: The University implements a broad range of learning management and communication platforms. In an ideal world we would use a single integrated computer program. Unfortunately the strengths and weaknesses of the suite of programs that the university has vetted are complementary and compel me to enlist 3 different programs. The class will use:

1. Canvas: Canvas will be used for you to submit assignments and provide you with your grades, and only those two functions. Canvas features will not be used to compute grades. Any question about class standing should be directed to Professor Miranker. Do not rely on any grade calculations implemented by Canvas. This is the only system authorized by the University to report homework and exam grades to students. It does not support a non-standard element of grading used in this class. (see grading, below)

2. Piazza: This is the class bulletin board system. Any questions you have about course organization or questions about the material should be posted to Piazza. Matters of a private nature should be sent to myself or the TA by email. Such emails should be rare. I assure you, if you have a question, no matter how certain you are that it’s a stupid question, at least two other students have the same question and the question deserves to be on Piazza.
3. **UT Box**: There will be a file structure hosted on UT Box. A link to its root will be posted on Piazza. The organization of the file structure will be obvious. All course materials will be pushed to Box. This includes homework assignments, homework solutions, non-textbook reading etc.
   - Unless you have a good reason to avoid this, you are **required** to use Box Drive or otherwise mount the Box folder on your local file system. i.e. do not use the Box web interface. The recent versions of Box have gotten very good at automatically determining, and enabling you to override, which files are cached locally on your computer, and which are not. Thus, available disk space on your personal computer should not be an issue and the most important files will be pushed to your machine, making them instantly available even if you are offline.
   - You are **required** to configure Box so that you are alerted, by notification and/or email, when a new file is pushed to the folder. Homework will be assigned by pushing a file to Box. There will be no other notice of assignments. **You are responsible** for borging yourself with the automated system and enabling it to send you a message that you will not miss. Thus, you are the single human responsible for you knowing when assignments have been posted and their due date. There will be multiple opportunities in the first two week of class to test and verify that you have configured your environment correctly.

**Course Description**

**Course Catalog Description**: Concepts of database design and database system implementation. Data models, query processing, database design theory, crash recovery, concurrent control, and distributed databases.

**Prerequisite**: Computer Science 429 or 429H with a grade of at least C-.

**Specific Description Per This Semester**: Primary skills developed in this course are data modeling, achieving performant SQL and enabling students to better configure relational database management systems (RDBMSs) as a result of learning how RDBMSs are organized, and the algorithms integral to each architectural component. A “from first-principles approach“ provides a foundation for understanding and addressing the same issues as they arise in other kinds of database management systems that are emerging as real alternatives to RDBMSs. i.e. Cloud native databases and databases using the graph and document (hierarchical) data models.

Data modeling is the conceptual definition of the data structures (objects) stored in a database to support an application. Alternatively, a data model is a notation for describing data commonly comprising a way to describe the structure of the data and, operations and constraints on the data. These methods are well established for relational database management systems (RDBMS) as relational databases have been a monopolistic mainstay of courses, research and industry practice starting in the 1980’s. However, just in the last ten years that monopoly been breached. Graph and object-oriented databases (in the form of JSON) are gaining substantial traction. Thus, the class teaches both the use of conventional relational database modeling, E-R models, and a more abstract method, UML Class Diagrams, as the latter is able to model the contents of next generation databases a student may encounter in practice and E-R models may not apply at all. To better understand the value of the higher level of abstraction students will learn to use data modeling tools, compare the two forms of data model, and learn how to transform a more abstract model expressed in UML to the more detailed representation required when using an E-R model. As time permits, the course will explore the same process as it applies to non-relational, cloud-based databases.
The remainder of the course details the organization and algorithms constituent to each component in the architecture of a relational database management system. Concurrent with that focus are corresponding SQL programming assignments and laboratories that include database optimization, performance analysis and benchmarking. Thus, students experience, firsthand, how many of the algorithms behave in-situ.

There is no term project, per se, in this course. There is a sequence of text and lab homework problems that, if assigned all at one time, and with term project like instructions, would clearly be a term project. As a term project, the goal of the sequence is that you will have learned how to optimize a database management system for a particular workload. Note that in a job description this ability is typically the responsibility of the database administrator (DBA). Thus this course goes beyond the common core of an undergraduate introduction to database course.

Since many of the lab assignments, taken together, are equivalent to a term project, all labs must be completed to complete the course (i.e. to not complete all the labs is equivalent to not completing a term project).

Reading:

A copy of my personal library of online database textbooks is linked into the class UT Box file structure.

Required Texts:


Supplementary Texts:

In my library of on-line text books you will find

- The Ramakrishnan and Gehrke textbook.
- The solution manual for the Ramakrishnan and Gehrke textbook.

The Ramakrishnan and Gehrke text is very similar to this courses primary textbook. Thus, at your fingertips you have, an alternate explanation of the course material, homework questions, and their answers. It is not cheating to engage in extra problem solving and then check your answers in the solution manual. (But the resulting learning will be more effective if you don’t peek at an answer before you have given the problem a solid effort.)

Other Reading: In the class Box folder is a subfolder called Reading. Within that are folders titled by topic. E.g. Bloom Filter. Every file in a topic folder is required reading. In addition, each such folder may contain a folder called Recommended and/or Supplementary. If those additional folders are present, they will contain additional material that may help illuminate the topic, and, respectively, contain additional material selected by me as providing either more depth for those of you who are interested, or another approach to presenting the material for those of you who are having trouble grasping the ideas from the primary materials.
Grading: There will be 2 tests and a final. Homework is heavily weighted. Individual assignments will have different weights corresponding to the number of points for the homework. The grading components will be weighted as follows, test 1, 10%, test 2, 25%, final, 35%, homework, 30%. Labs are graded on a completion basis. Per above, all labs must be completed to complete the course, independent of late policy. In that regard, labs are graded on a three point scale, {0, 1, 2}. 2, completed, 0, no credit, you must redo the lab, and turn in a satisfactory result. (i.e. one that scores “2”.) A 1 means, you did not complete the lab satisfactorily, but you clearly tried. If you score “1”, on more than one lab you may be asked to redo the unsatisfactory labs. This ask is determined purely subjectively by the TA and the Instructor and is not subject to appeal.

There is no single formula that is considered the correct way to curve a grade. In this course the four grade components are normalized, weighted per the proportions above and combined. (This is not the same computation built into Canvas. So do not assign any meaning to a presentation of curved grades in Canvas.) Effectively the result of grading on a curve results in having class rank being a large determination of the final grade. I do assign, but minimally so, plus and minus grades. If you have any question about where you stand, contact Professor Miranker by email. The final is comprehensive and conducted at a time and location specified by the registrar.

Homework & Labs: There will be both homework and labs. Expect both a homework and a labs will be assigned weekly. You will be required to install PostgreSQL on a computer on which you have full administrative control. Some homework will require the use of Amazon Web Services(AWS). You will be furnished with an account per the generosity of Amazon. Unless you do something very strange or far beyond the scope of the assignment, and then pay no attention, (AWS will email you a warning), you will not incur any AWS charges.

You must be careful to create an educational AWS account. Instructions will be given. It is easy to create a personal AWS account associated with your retail Amazon account. Such personal accounts are real AWS accounts and may incur charges.

Late Homework policy: Each student is permitted one late homework or one late lab, no questions asked, provided it is turned within 3 days of the due date. More than one homework or lab late, don’t ask. (Accomodation will be provided if a challenge arrises that is out of your control.) There will be a Canvas entry point for submitting late work. Just one entry point for a late homework and a late lab. Thus Canvas is configured to enforce the late policy. Recall all labs must be completed to complete the course. If you are late with more than one lab, contact the TA on how to turn it in.
Proforma Notifications

**Academic Integrity:** Study groups are encouraged, including working on homework and labs in groups. The line for cheating is very simple. No methodical, (a repeatable process e.g. copying), or mechanical, (use of cut and paste, or copying files) is allowed. A particular point of confusion on this policy is what it means to do labs in a workgroup. The answer is, each student must do the work on their own computer or account, even if the input typed is the same, on a per character basis, and consequently, the output may, similarly, be identical. Very simply, active engagement create more learning than passive engagement (i.e. watching).

Each student in the course is expected to abide by the University of Texas Honor Code: “As a student of The University of Texas at Austin, I shall abide by the core values of the University and uphold academic integrity.” Plagiarism is taken very seriously at UT. Therefore, if you use words or ideas that are not your own (or that you have used in a previous class), you must cite your sources. Otherwise you will be guilty of plagiarism and subject to academic disciplinary action, including failure of the course. You are responsible for understanding UT’s Academic Honesty and the University Honor Code which can be found at the following web address: [https://deanofstudents.utexas.edu/conduct/standardsofconduct.php](https://deanofstudents.utexas.edu/conduct/standardsofconduct.php)

Q Drop Policy: If you want to drop a class after the 12th class day, you’ll need to execute a Q drop before the Q-drop deadline, which typically occurs near the middle of the semester. Under Texas law, you are only allowed six Q drops while you are in college at any public Texas institution. For more information, see: [http://www.utexas.edu/ugs/csacc/academic/add drop/qdrop](http://www.utexas.edu/ugs/csacc/academic/adddrop/qdrop)

**Students with Disabilities:** Students with disabilities may request appropriate academic accommodations from the Division of Diversity and Community Engagement (DDCE), Services for Students with Disabilities (SSD) at [http://ddce.utexas.edu/disability](http://ddce.utexas.edu/disability)

**Sharing of Course Materials is Prohibited:** No materials used in this class, including, but not limited to, lecture hand-outs, videos, assessments (quizzes, exams, papers, projects, homework assignments), in-class materials, review sheets, and additional problem sets, may be shared online or with anyone outside of the class unless you have my explicit, written permission. Unauthorized sharing of materials promotes cheating. It is a violation of the University’s Student Honor Code and an act of academic dishonesty. I am well aware of the sites used for sharing materials, and any materials found online that are associated with you, or any suspected unauthorized sharing of materials, will be reported to Student Conduct and Academic Integrity in the Office of the Dean of Students. These reports can result in sanctions, including failure in the course.

The disclosure above is intended to prevent the unauthorized distribution of class materials through Chegg, CourseHero, and other online sharing sources that are often used for cheating.

**Class Recordings:** Class recordings are reserved only for students in this class for educational purposes and are protected under FERPA. The recordings should not be shared outside the class in any form. Violation of this restriction by a student could lead to Student Misconduct proceedings.

**Title IX Reporting:** Title IX is a federal law that protects against sex and gender-based discrimination, sexual harassment, sexual assault, unprofessional or inappropriate conduct of a sexual nature, dating/domestic violence and stalking at federally funded educational institutions. UT Austin is committed to fostering a learning and working environment free from discrimination in all its forms. When unprofessional or inappropriate conduct of a sexual nature occurs in our community, the university can: 1. Intervene to prevent harmful behavior from continuing or escalating. 2. Provide support and remedies to students and employees who have experienced harm or have become involved in a Title IX investigation. 3. Investigate and discipline violations of the university’s relevant policies.