Syllabus

CSE 414 - Introduction to Data Management | Autumn 2021

Administrivia

Course website: https://sites.google.com/cs.washington.edu/cse414-22wi/
Lectures: A and B: MWF 12:30 - 1:20 p.m. SIG 134
Sections at various times on Thursdays. Online section 4:30
Recordings of lecture and the online section will be available after class on Panopto
Instructor: Ryan Maas, maas@cs.washington.edu
Office hours: TBA
Course message board: Ed https://edstem.org/us/courses/17047/discussion/

See the course website for more contact information, a full list of TA office hours, and an up-to-date calendar.

COVID - Related Protocols and Remote Accommodations

> **Our highest priority is the safety and well-being of our students and staff**.

This quarter 414 will be much like the class last quarter, but the first week of class will be held entirely online with the Zoom links posted in Canvas. A summary of the course format and options for in-person vs remote learning are listed below.

I will be lecturing in-person, but I want any students who are uncomfortable with being in the classroom to still have a good course experience. There will be **Panopto recordings of every lecture** posted to Canvas shortly after class, and there is no participation grade or mandatory requirement to attend the lecture.

Thursday sections will also be in-person, but instead of recording each one we will have an **alternative online section hosted on zoom** on Thursday evenings. The recordings of this will be posted, so even if you can't attend live on zoom, you can see view the recordings instead.

Office hours will be offered both online and in person, more details on this in the next week but you will have both options.
Our homework assignments are all turned in remotely on Gradescope. We will also have two exams in a **take-home format** (see more details below.)

The most important requirement at UW and the Allen School is that everyone **must wear a mask indoors at all times**. No food or drink is allowed in the classrooms, but there is one exception that if you need a drink of water you may briefly lift your mask to take a sip, and then return it to position. We will be strict about this rule, please contact us by email or use the anonymous feedback tool if you notice people breaking this rule and we don't.

In summary, all course participation has a remote option, and while I think the in-person option is more effective (and fun) you should feel no pressure to be in an environment where you feel unsafe. If you have any further questions or personal details to consider, please email me at maas@cs.washington.edu.

More links:

UW COVID Guidelines and Resources
Facemask Policy and FAQ

Course Goals

Databases are at the heart of modern commercial application development. Their use extends beyond this to many other environments and domains where large amounts of data must be stored for efficient update, retrieval, and analysis. The purpose of this course is to provide a comprehensive introduction to the use of management systems for applications. Some of the topics covered are the following: data models (relational and json), query languages (SQL, datalog, etc), transactions, parallel data processing, and database as a service. For the detailed list of topics and schedule, please see the course site.

Format

The class meets three times a week for lectures. We will follow parts of the textbook (see below). While the lectures are designed to be clear and self-contained, and to cover all the material used in class, you are strongly encouraged to read from the textbook, both in order to get a better understanding of the material covered and to learn about related topics, which are not covered in class. Please note that the lectures will be a combination of whiteboard, powerpoint, and live demos.

In addition to the lecture there will be sections taught by the TAs. The sections will discuss the material taught in class, will give detailed instructions on how to use some of the software needed for the homework assignments, and will give you an extra opportunity to ask questions.

There will be several homework assignments involving light programming and written questions. Additionally, there will be multiple quizzes.

Catalog Description

Introduction to Data Management: Introduces database management systems and writing applications that use such systems; data models (e.g., relational, semi-structured), query languages (e.g., SQL, datalog), language bindings, conceptual modeling, transactions, security, database tuning, data warehousing, parallelism, and Web-data management. Prerequisites: CSE 143 or CSE 163. 4 credits.

Here is a more specific list of potential topics that will be covered in this quarter's 414:

- Data models
- Query languages (e.g. SQL)
- Schema, logical, and physical design
- Database applications
- Transactions
- Data analytics
- NoSQL
- Cloud database systems

As this class is evolving, the list of topics might change over the quarter.

Grading and Exams

- Seven Homework assignments: 60%
- Two Exams: 30%
- Remaining 10% to be allocated between homeworks and exams depending on difficulty

Please note that these breakdowns are preliminary as this class is constantly evolving and we reserve the right to change them.
Each homework assignment will take approximately one or two weeks depending on the difficulty.. Details for each homework will be released after the previous homework is due. **All homeworks will be due at 11 pm PDT (Seattle time).**

We will have a midterm and final exam, but both are offered with a take-home option.
Both the midterm and the final will be take-home assignments with around 4 to 5 days to complete the test. You may use as much time as you want for the test.

More details to come as we get closer to the exams

We reserve the right to adjust assessment deadlines. We will provide due notice to the course email list and the class message board.

Grades will be curved, the typical median for the course is 3.5 - 3.6 but has been higher recently.

Regrades

If you think one of your homework or exam questions was misgraded, submit a regrade request on Gradescope explaining which part you would like regraded and why you think you deserve more points. The window for regrades will only be open for a few days after grades are released.

Main Textbook

Highly recommended:

Database Systems: the Complete Book, by Hector Garcia-Molina, Jennifer Widom, and Jeffrey Ullman. Available from the University Bookstore. Second edition.

Other texts:

The library has the following that you might find useful if you require another explanation of a topic.

- Fundamentals of database systems by Elsmasri and Navathe.
- Database management systems by Raghu Ramakrishnan and Johannes Gehrke.
- Foundations of database systems by Abiteboul, Hull and Vianu.

Late Policy

Homework assignments: you are allowed a total of 6 late-days with at most 2 late-days per assignment (unless otherwise noted) that you can use in 24-hour chunks at any time. There is no penalty to your grade for using late days. **Once you use-up your late days, no additional extensions are granted! Late submissions will not be graded.** Think of late days as a safety net in case of a true emergency, not as a convenience. Normally, you should use no late days during the entire quarter; if you do have an emergency, then you should use 1 or 2 late days. If you experience some difficult situation and need further accommodations, please contact Ryan at maas@cs.washington.edu

Grading Research Tool

We are using an automated verification tool for grading some of your homework assignments, and may use your class work to improve that tool. For example, we may use anonymized student assignments to design or improve the algorithms or build new tools to help database users. Any student who wishes to opt out can contact the instructor or TA to do so after final grades have been issued. This has no impact on your grade in any manner.

Programming

Some programming will be necessary in this course. One can only start to appreciate database systems by actually trying to use one. Databases only hold the data, the application logic needs to be written in some general purpose language. We will use the following tools in class:

- SQLite
- SQL Server on Windows Azure (Microsoft's Web Service)
- Java
- Amazon EC2
- Spark
- AsterixDB

Tools

The course website and mailing list will be used extensively to provide you with course information, such as the schedule mentioned above, homework assignments and solutions, class messages and many other things. There is a discussion board that everyone should use to keep in touch outside of class. Please see the main webpage of the course for details.

We often use Google Drive to share documents. Please use your CS GSuite account to access them.

We will use Gradescope for submitting homework assignments.

Collaboration policy

You are encouraged to discuss the content of this course with anyone you like. Unless otherwise stated, each homework and programming assignment is to be done individually. In part, this means: Do discuss concepts. Do help each other with practice problems. Don't show each other your code/answers. Don't look at past or present students' code/answers.

Consider the "Gilligan's Island Rule": after having a conceptual discussion with classmates, do something mindnumbing for half an hour (such as watching a Gilligan's Island rerun). If you still understand the concept after that, congratulations, you've probably learned instead of copying!

Academic integrity

Any attempt to misrepresent the work you submit will be dealt with via the appropriate University mechanisms, and your instructor will make every attempt to ensure the harshest allowable penalty. The guidelines for this course and more information about academic integrity are in a separate document (including, but not limited to, the Allen School's Academic Misconduct page and the College of Engineering's Academic Misconduct Process). You are responsible for knowing the information in these documents.

Privacy/FERPA statement

This course is scheduled to run synchronously at your scheduled class time via Zoom. These Zoom class sessions will be recorded. The recording will capture the presenter's audio, video and computer screen. Student audio and video will be recorded if they share their computer audio and video during the recorded session. The recordings will only be accessible to students enrolled in the course to review materials. These recordings will not be shared with or accessible to the public.

The University and Zoom have FERPA-compliant agreements in place to protect the security and privacy of UW Zoom accounts. Students who do not wish to be recorded should:

- Change their Zoom screen name to hide any personal identifying information such as their name or UW Net ID, and
- Not share their computer audio or video during their Zoom sessions.

Computer use policy

Some excerpts from the campus policies. Take them seriously: "You must use all UW [computing] resources in strict accordance with local, state, and federal laws. These laws cover such areas as illegal access to computer systems, networks, and files; copyright violations; and harassment issues... Software and information resources provided through the university for use by faculty, staff, and students may be used on computing equipment only as specified in the various software licenses. Unauthorized use of software, images, or files is regarded as a serious matter and any such use is without the consent of the University of Washington...If abuse of computer software, images, or files occurs, those responsible for such abuse will be held legally accountable."

Accomodations

Please refer to university policies regarding disability accommodations and religious accommodations.