Syllabus

CSE 344 - Introduction to Data Management | Summer 2021

Administrivia

Course website: https://sites.google.com/cs.washington.edu/cse344-21su/home
Lectures: MWF 2:20 p.m., sections at various times on Thursdays, Zoom links on Canvas
Instructor: Ryan Maas, maas@cs.washington.edu
Office hours: TBA (Zoom links on Canvas)

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

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.

Remote Instruction

The course staff will make our best attempt to have live, interactive lectures and sections as usual. In the event that the staff's internet connections or other circumstances keep us from being able to provide stable, live instruction, we will provide recordings in advance.

We also recognize that some students may not be in Seattle during the quarter, lack stable internet connections, or have other situations that prevent them from joining synchronous instruction. To accommodate this, we will provide recordings of lecture (see the privacy statement below). All assessments will be done asynchronously, e.g. they will be due by a specific time but you will not have to be present during lecture to submit them. Make sure you participate in the class in other ways, such as asking and answering questions on Piazza.

We do encourage students to participate in the live instruction if possible and try to connect with other students as you would any other time! Consider using Ed or UW CSE Slack to set up study groups (just keep the academic misconduct rules in mind!)

Access the Zoom links on Canvas in order to join or view recordings.

Also, please bear with us if we need to make adjustments to any of this. This is new territory for all of us!

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 311 or CSE 321. 4 credits.

Here is a more specific list of potential topics that will be covered in this quarter's 344:

- Data models
- Query languages
- 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

- Homework assignments: 60%
- Quizzes: 30%
- Remaining 10% to be allocated between homeworks and quizzes depending on difficulty

Please note that these breakdowns are preliminary as this class is constantly evolving and we reserve the right to change them.
Estimated homework deadlines are on the website, details for each will be released after the previous homework is due. **All homeworks will be due at 11 pm PDT (Seattle time).**

More details about quiz logistics will be announced later. Quiz dates are posted on the course website. Make-up quizzes will only be given in case of a serious emergency. If you miss a quiz, even if you are sick or injured, you must contact Ryan BEFORE the quiz deadline (or arrange for someone to do so). You must show evidence that you are physically unable to take the quiz. No make-ups will be granted for personal reasons such as travel or conflicting schedules. Aim to take the quiz earlier during the day, no accommodations will be made for students who have trouble with last-minute submissions. The only accommodations for quizzes that will be made are those that correspond to the University's official accessibility guidelines, which must be reflected in your student account.

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

Regrades

If you think one of your homework or quiz questions was misgraded, make a private post on Piazza explaining which part you would like regraded and why you think you deserve more points. Do so within 2 weeks of the grade being posted.

Main Textbook

Required:

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. **Once you use-up your late days, no additional extensions are granted for any reason! 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 end up using all 6 late days, you are doing it wrong.

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
- Neo4j

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 CSE's Gitlab for submitting homework assignments.

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.

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 (**we check for this**.)

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!

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."

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.

Accomodations

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