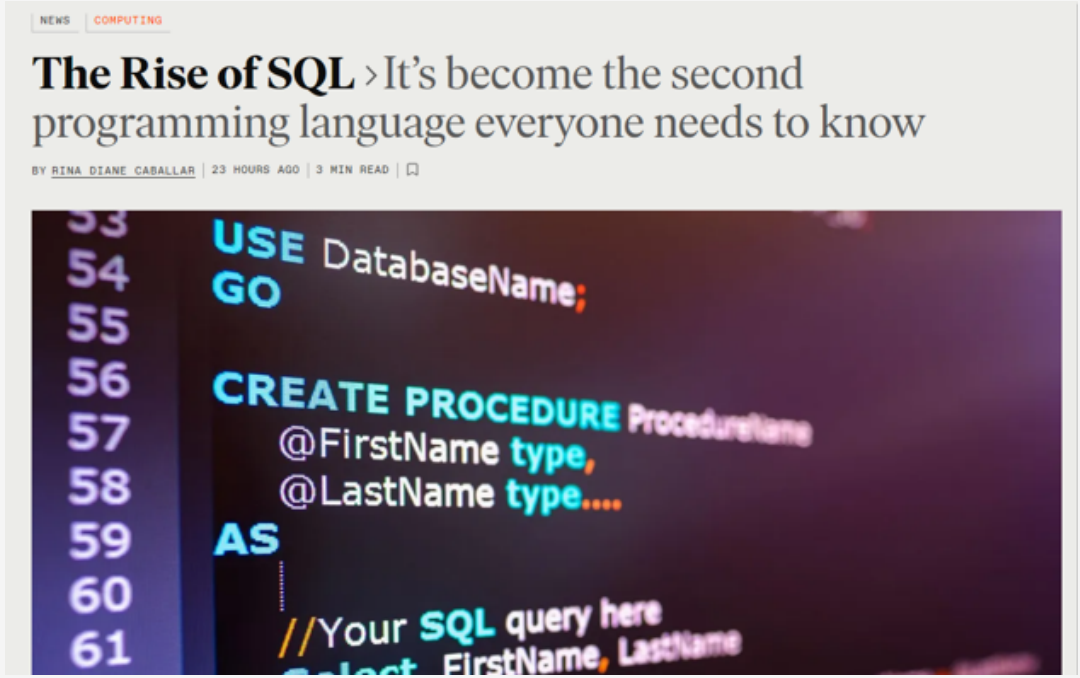
Welcome to CMSC508

Lecture 1

Wednesday - Aug 23, 2023

I’m glad you’re here!!!!



https://spectrum.ieee.org/the-rise-of-sql

CMSC 508 - Database Theory

- Semester course;
- 3 lecture hours. 3 credits.
- Prerequisite: CMSC 303 with a minimum grade of C.

Design and implementation of relational database systems. Emphasis is placed on entity-relationship diagrams, relational algebra, normal forms and normalization. Introduction to SQL. Discussion of physical level issues.

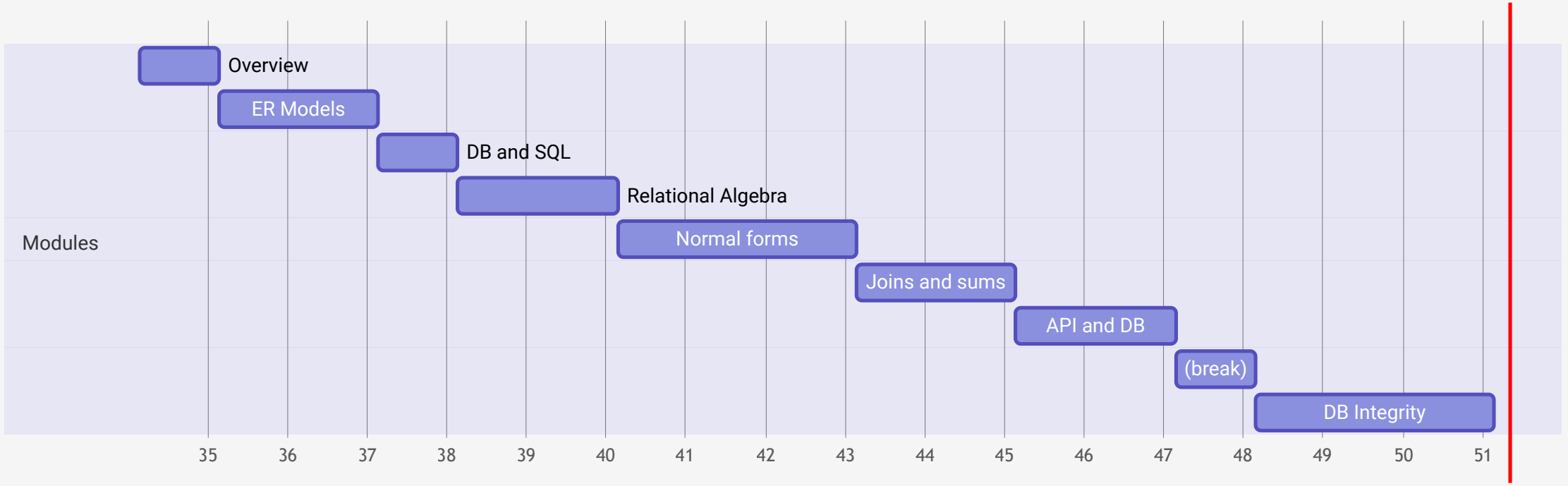
Students will be required to complete a design project and give an oral presentation of the project.

Learning objectives

- SLO1 - Analyze a complex computing problem and apply principles of computing and other relevant disciplines to identify solutions;
- SLO2 - Design, implement and evaluate a computing-based solution to meet a given set of computing requirements in the context of the program's discipline;
- SLO3 - Communicate effectively in a variety of professional contexts;
- SLO4 - Recognize professional responsibilities and make informed judgments in computing practice based on legal and ethical principles;
- SLO5 - Function effectively as a member or leader of a team engaged in activities appropriate to the program's discipline;
- SLO6 - Apply computer science theory and software development fundamentals to produce computing-based solutions;

Topics over time

The class organized in modules. The GANNT chart below highlights the topics and their order. Homework assignments reinforce topics discussed in class. Quizzes test understanding of the topics where appropriate. Material is presented in a way that complements the successful completion of the semester-long project.



Grading and deliverables

This will be a busy semester! There will be items due each week.

Overview

Below are tables presenting the grading scheme and deliverable list for the class.

Deliverables

Deliverable	Pct	Number of items
Homework	40%	11 submissions done individually
Quizzes	30%	4 quizzes taken on-line using respondus
Project	30%	7 submissions as part of a team

Grading

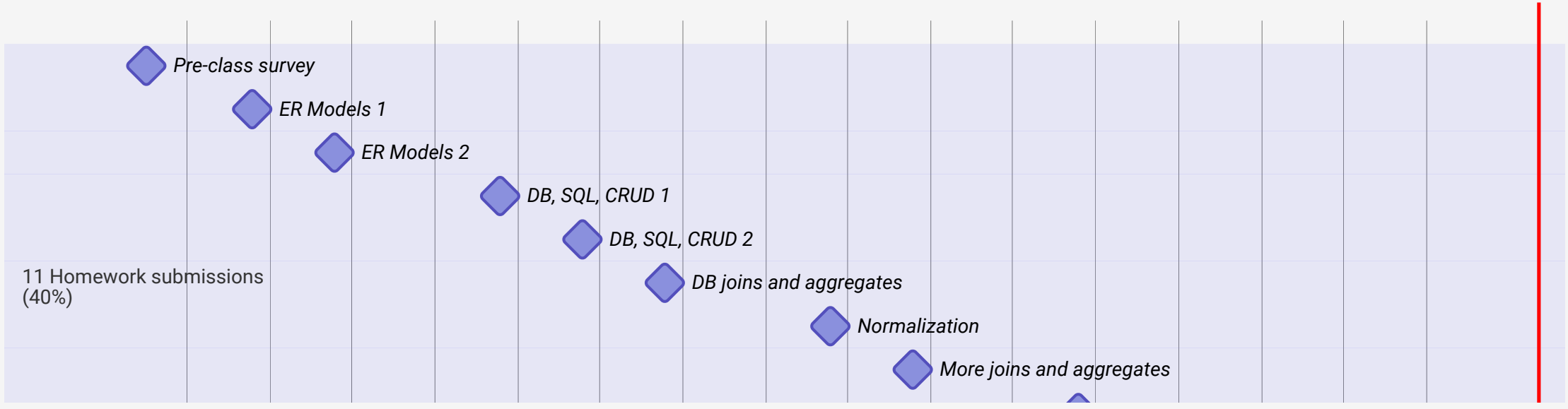
Score	Grade
90 <= X	A
80 <= X < 90	B
70 <= X < 80	C
60 <= X < 70	D
X < 60	F

Grading and deliverables

This will be a busy semester! There will be items due each week.

Homework Assignments

Homework assignments are done individually. Homeworks will be distributed using GITHUB classroom; a link will be provided in the Canvas assignment page. Homeworks will be submitted to Gradescope. Each assignment is worth 20 points. Rubrics will be provided. Homeworks will be written using Quarto. Homework 2 provides an opportunity to learn more about Quarto. Quarto will also be used for the semester project.

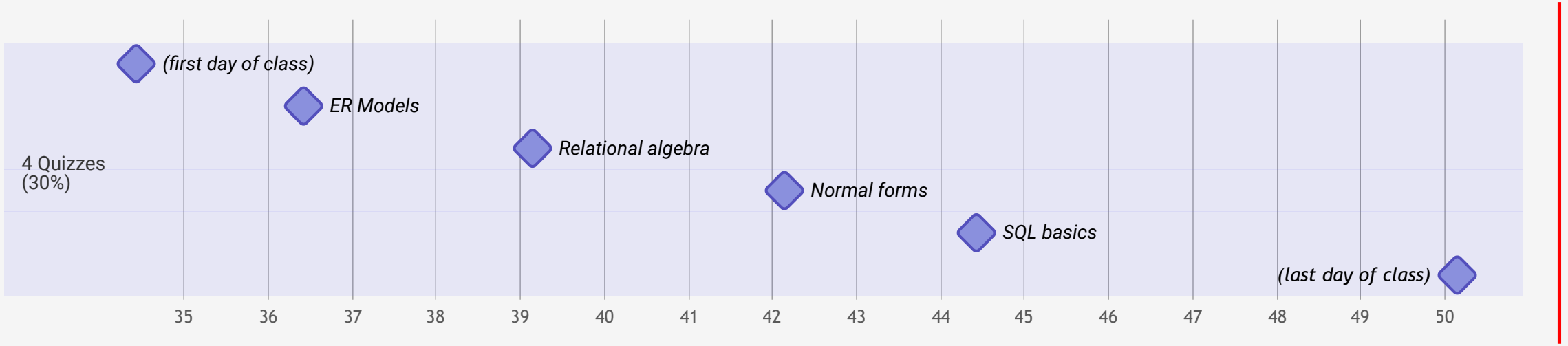


Grading and deliverables

This will be a busy semester! There will be items due each week.

Quizzes

There will be four quizzes throughout the semester. Each quiz will be worth 30 points. Quizzes will be administered on-line usint the Respondus lockdown browser. It is the responsibility of the student to ensure that the technology works. Homework assignment 1 offers practice to work out the bugs. Quizzes will be administered during class time.

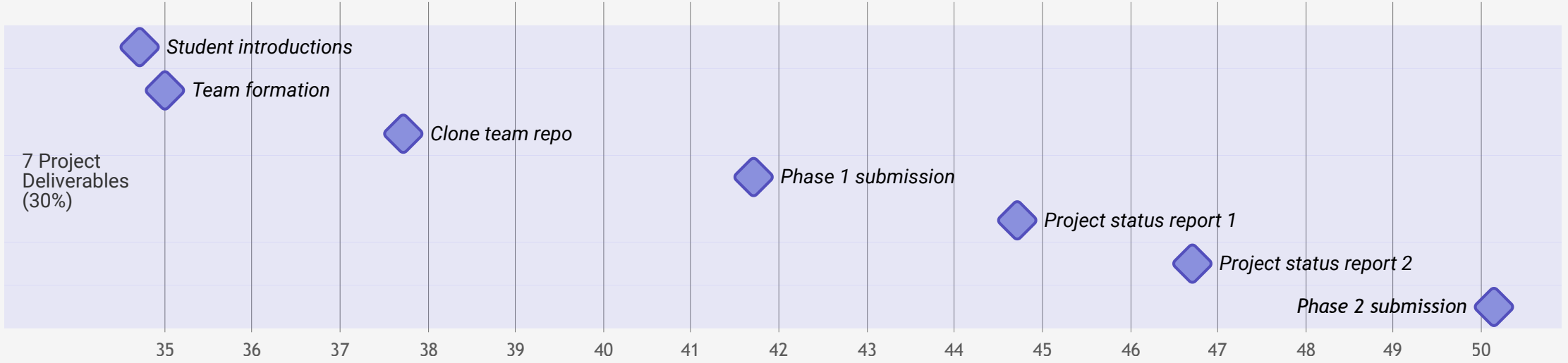


Grading and deliverables

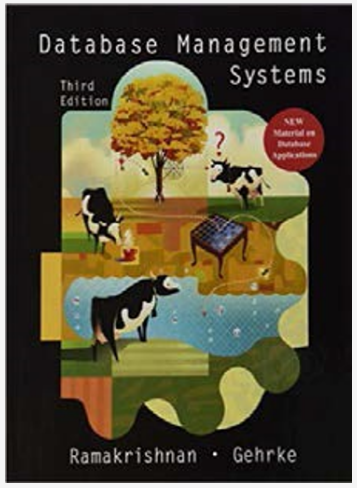
This will be a busy semester! There will be items due each week.

Semester-long project

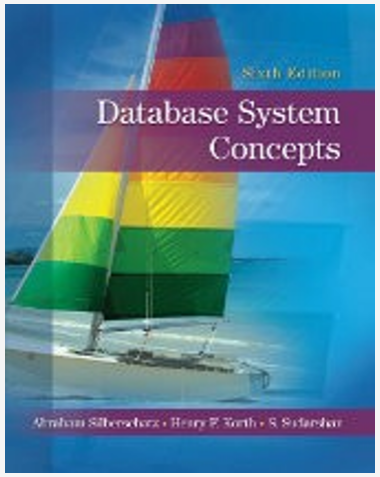
Students will work in teams of 2 on a semester-long project. The project will require the students design and build a database backend presenting an API. The domain of the project is chosen by the students. Tools include SQL, python, flask, MySQL, Quarto, and others.



Recommended Textbooks



Database Management Systems, Third Edition, R. Ramakrishnan, J. Gehrke McGraw-Hill ISBN 978-0072465631



Database System Concepts, Sixth Edition, A. Silberschatz, H. Korth, S. Sudarshan McGraw-Hill ISBN 978-0073523323

Do I need really need a text book?

No, you don't. These textbooks are *really good* and package lots of information into a single place.

Life is a database and from my perspective it's all about speed of retrieval. When I'm building databases, I can't EVER recall pulling out a text book to look something up!

Canvas as the OFFICIAL platform

YOU are responsible for your own LEARNING. My role is GUIDE and CURATOR.

WE are working together as a TEAM so that at completion of the course, you are comfortable with designing and building databases.

ANYTHING posted to the course canvas is FAIR GAME to be tested. If I think it's important enough to be shared as part of the course materials on Canvas, I believe that it's important enough for you to be familiar with it and possibly tested on it.

I will OFTEN not be able to cover all the material relevant to a specific topic in class. That does NOT mean that it isn't important. Even if I don't complete the lecture slides during a class session, you should be familiar with ALL the material in the slides. ANYTHING in the lecture slides is FAIR GAME for testing.

Statement on Generative AI

Grades for the course are assigned based on YOUR work, not the work of chatgpt, bard or some other generative AI tool. Grades help YOU understand where you are on the journey towards mastery of any pile of material.

Generative AI can be a real asset in computing. You don't need to memorize syntax, or struggle to hunt around to find some arcane code necessary to make your program work. But remember, it is supposed to be YOUR program.

Homework assignments are designed to give you practice using the tools to create programs. The focus of the homework assignments is on the higher level cognitive skills - designing, comparing, judging, analyzing, assessing. **Use of Generative AI IS permitted on homework assignments and the semester project**, in so much as it can help the coding go faster. YOU are responsible for the design and should be able to explain it to anyone that asks.

Quizzes are designed to test your knowledge of the course material. **Use of Generative AI is NOT permitted on any of the quizzes.**

Housekeeping

Lots of stuff to do as we get this engine fired up:

- Visit the course canvas and get acclimated.
- Complete homework 1 (a student survey) using the Respondus lockdown browser.
- Get started on:
  - project deliverable 1 (due Friday), and
  - project deliverable 2 (due Sunday)!!!
- Join the course discord (Please change your server name!)

Module	Week	Date	Day	Lectures/Quizzes	Deliverables/Notes
Overview	1	8/23	Wed	MTG1: L1 (Welcome to CMSC508)	HW1 due (Preclass Survey); First day of CMSC508
Overview	1	8/25	Fri		PrjDel 1 due (Student introductions)
Overview	1	8/27	Sun		PrjDel 2 due (Team formation)
ER Models	2	8/28	Mon	MTG2: L2 (Entity-relation models 1)	Last day of add/drop
ER Models	2	8/30	Wed	MTG3: L3 (Entity-relation models 2)	
ER Models	2	9/1	Fri		HW2 due (ER Models 1)
ER Models	3	9/4	Mon		University closed: Labor day
ER Models	3	9/6	Wed	MTG4: Quiz 1 today (Entity-relation models)	

A long time ago … in a galaxy far, far away …

Early history

1950s and early 1960s:

- Data processing using magnetic tapes for storage
- Tapes provided only sequential access
- Punched cards for input

Early history

Late 1960s and 1970s:

- Hard disks allowed direct access to data
- Network and hierarchical data models in widespread use
- Ted Codd defines the Relational data model
- High-performance (for the era) transaction processing

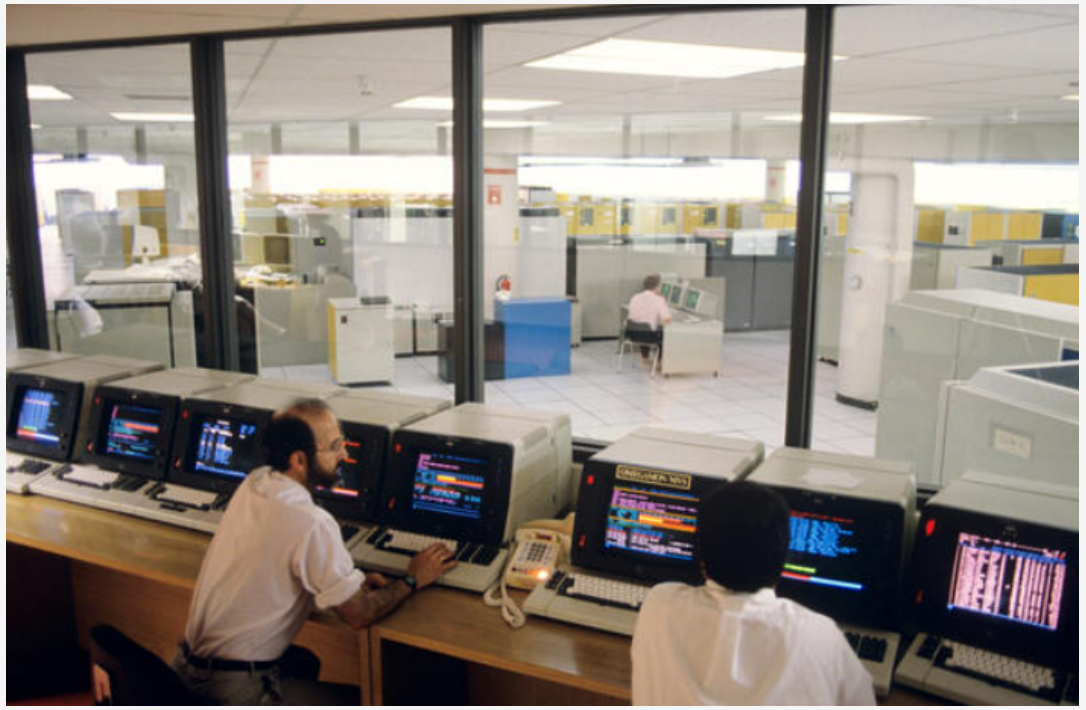
Modern systems

1980s:

- Research relational prototypes evolve into commercial systems
  - SQL becomes industrial standard
- Parallel and distributed database systems
- Object-oriented database systems

1990s:

- Large decision support and data-mining applications
- Large multi-terabyte data warehouses
- Emergence of Web commerce



Modern systems

Early 2000s:

- XML and XQuery standards
- Automated database administration

Later 2000s:

- Giant data storage systems
- Google BigTable, Yahoo PNuts, Amazon, …

2022 and Beyond:

https://cacm.acm.org/magazines/2022/8/262905-the-seattle-report-on-database-research/fulltext#body-4

DB vs DBMS

Important concept - DB vs DBMS

Database (DB)

- **Entities:** abstractions of data
- **Relationships:** semantic information that connects two or more entities to each other
- Logical architectures

Database Management System (DBMS)

- hosts multiple databases
- Set of programs to access and manipulate the data
- Controls access, manages users
- Environment that is convenient, efficient, secure, reliable, scalable
- Examples: Oracle, SQL Server, MySQL

Why databases?

Why not use spreadsheets?

Spreadsheets DO work and they ARE viable!

Here is view of a data base through a web site

https://vcu-ssg.github.io/ssg-quarto-cmsc-courses/

Here is the source data

https://docs.google.com/spreadsheets/d/1qrN3L7eRLsM-aVMHYaLQN-FMYtrTJf0_h6dLKlUdPkk/edit#gid=107368023

Why not use spreadsheets?

Drawbacks of using file systems to store data

- Data redundancy and inconsistency in multiple files and formats
- Need to write a new program to carry out each new task
- Difficulty in accessing data: where and how
- Data isolation
- Integrity problems and constraints
- Atomicity of updates
- Concurrent access by multiple users
- Security problems

Why use a DBMS?

Advantages of a DBMS

- Data independence (facilitates sharing of information)
- Efficient data access (indexes can be used for optimization)
- Data consistency and integrity
- Security and authorization
- Data administration
- Concurrent access and crash recovery
- Reduced application development time

Why use a DBMS?

Size and scale

Here is an example data warehouse build summary:

https://docs.google.com/spreadsheets/d/11BazRGXKIdAT4B5ln5O567-Ah0n68ulmjVqdXZgPyTE/edit#gid=1785128451

Why use a DBMS?

In summary: ACID

Atomicity: Either all operations of the transaction are properly reflected in the database or none are

Consistency: Execution of a transaction in isolation preserves the consistency of the database

Isolation: Although multiple transactions may execute concurrently, each transaction must be unaware of other concurrently executing transactions

Durability: After a transaction completes successfully, the changes it has made to the database persist, even if there are system failures.

Housekeeping

Lots of stuff to do as we get this engine fired up:

- Visit the course canvas and get acclimated.
- Complete homework 1 (a student survey) using the Respondus lockdown browser.
- Get started on:
  - project deliverable 1 (due Friday), and
  - project deliverable 2 (due Sunday)!!!
- Join the course discord (Please change your server name!)

Module	Week	Date	Day	Lectures/Quizzes	Deliverables/Notes
Overview	1	8/23	Wed	MTG1: L1 (Welcome to CMSC508)	HW1 due (Preclass Survey); First day of CMSC508
Overview	1	8/25	Fri		PrjDel 1 due (Student introductions)
Overview	1	8/27	Sun		PrjDel 2 due (Team formation)
ER Models	2	8/28	Mon	MTG2: L2 (Entity-relation models 1)	Last day of add/drop
ER Models	2	8/30	Wed	MTG3: L3 (Entity-relation models 2)	
ER Models	2	9/1	Fri		HW2 due (ER Models 1)
ER Models	3	9/4	Mon		University closed: Labor day
ER Models	3	9/6	Wed	MTG4: Quiz 1 today (Entity-relation models)	