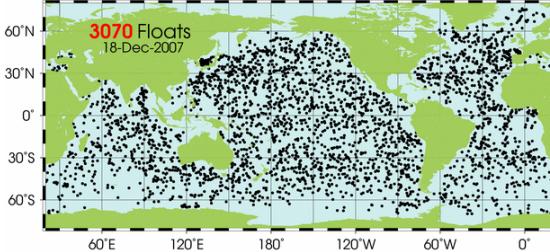


# Fundamentals of Database Systems

## CMPSC 174A

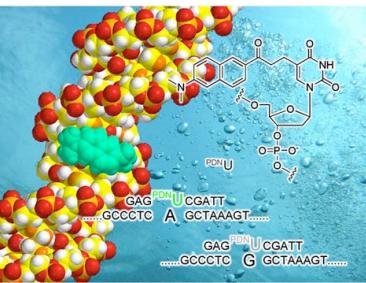
### Lecture 1: Introduction



# Class Goals



- The world is drowning in data!
- Need computer scientists to help manage this data
  - Help domain scientists achieve new discoveries
  - Help companies provide better services (e.g., Tiktok)
  - Help governments (and universities!) become more efficient
- Welcome to 174A: Fundamentals of Database Systems
  - Existing tools PLUS data management principles
  - This is not just a class on SQL!



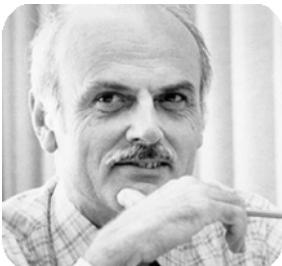
CMPSC 174A - 21 Wi



# Turing Awards in Data Management



Charles Bachman, 1973  
*IDS and CODASYL*



Ted Codd, 1981  
*Relational model*



Jim Gray, 1998  
*Transaction processing*



Michael Stonebraker, 2014  
*INGRES and Postgres*

## Most Enterprise AI Models Are Based on Relational Data\*



8,024 responses

- Retail: > 86% relational
- Insurance: > 83% relational
- Marketing: > 81 % relational
- Financial: > 77% relational

\* based on 2017 Kaggle survey of 16,000 ML practitioners

\* some of us helped bring the relational model to databases

# Staff

- Instructor: Shumo Chu
- TA's
  - Lianke Qin
  - Laboni Sarker
  - Chen Zhu

# Course Format

- Lectures: online
- Sections: online
- 6 homework assignments
- Midterm and final
- Class and section participation:  
Post and **answer** questions (in class, piazza, etc)

# Grading

- Homeworks 30%
- Midterm 20%
- Final 40%
- Class participation 10%
- Extra credit:
  - Some hw have extra credit questions
  - Large # of good answers on piazza

This is all subject to change

# Communications

- Web page:  
<https://gauchospace.ucsb.edu/courses/courseview.php?id=105054>
  - Everything is here
- Piazza:  
<https://piazza.com/class/kjjln4wpuj93g4?cid=5>
  - **THE** place to ask course-related questions
  - Log in today, enable notifications

# Textbook

Main textbook, available at the bookstore or pdf:

- *Database Systems: The Complete Book*,  
Hector Garcia-Molina,  
Jeffrey Ullman,  
Jennifer Widom  
**Second edition.**

REQUIRED READING !

# Six Homework Assignments

1. Sqlite intro (1 wk)
2. Sqlite basics (1 wk)
3. Datalog and Relational Algebra ( $1\frac{1}{3}$  weeks)
4. Spark (2 weeks)
5. Schema Design (1 wk)
6. Transactional Application ( $1\frac{2}{3}$  weeks)

# About the Assignments

- You will learn/practice the course material
- You will also learn lots of new technology
  - Time spent learning it is useful! Write on your CV!
- Note: we use anonymized assignments to improve the grading tool; to opt out, send me email after the final grade

# Deadlines and Late Days

- You have up to 4 late days
  - No more than 2 on any one assignment
  - Use in 24-hour chunks
- Late days = safety net, not convenience
  - You should not plan on using them
  - If you use all 4 you are doing it wrong

# Exams

- Midterm (Feb 11) and Final (March 16)
- Location: in online

# Academic Integrity

- Anything you submit for credit is expected to be your own work
  - OK to exchange ideas, not detailed solutions
  - We all know difference between collaboration and cheating
- I trust you implicitly, but will come down hard on any violations of that trust

# Lectures and Sessions

- Lecture notes: Website
- Sessions:
  - 5 PM (Lianke)
  - 6 PM (Chen)
  - 7 PM (Laboni)

# Now onto the real stuff...

# Outline of Today's Lecture

- Overview of database management systems
- Course content

# Database

What is a database ?

Give examples of databases

# Database

What is a database ?

- A collection of files storing related data

Give examples of databases

# Database

What is a database ?

- A collection of files storing related data

Give examples of databases

- Accounts database; payroll database; UCSB's students' database; Amazon's products database; airline reservation database

# Database Management System

What is a DBMS ?

Give examples of DBMSs

# Database Management System

What is a DBMS ?

- *A big program written by someone else that allows us to manage efficiently a large database and allows it to persist over long periods of time*

Give examples of DBMSs

- Oracle, IBM DB2, Microsoft SQL Server, Amazon Redshift, Snowflake
- Open source: MySQL (Sun/Oracle), PostgreSQL
- Open source library: SQLite

We will focus on **relational** DBMSs most quarter

# An Example: Online Bookseller

- What data do we need?
  - 
  - 
  - 
  -
- What capabilities on the data do we need?
  - 
  - 
  - 
  -

# An Example: Online Bookseller

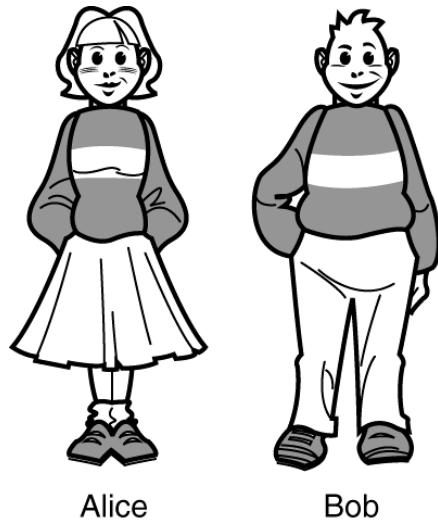
- What data do we need?
  - Data about books, customers, pending orders, order histories, trends, preferences, etc.
  - Data about sessions (clicks, pages, searches)
  - Note: data must be persistent! Outlive application
  - Also note that data is large... won't fit all in memory
- What capabilities on the data do we need?
  - 
  - 
  -

# An Example: Online Bookseller

- What data do we need?
  - Data about books, customers, pending orders, order histories, trends, preferences, etc.
  - Data about sessions (clicks, pages, searches)
  - Note: data must be persistent! Outlive application
  - Also note that data is large... won't fit all in memory
- What capabilities on the data do we need?
  - Insert/remove books, find books by author/title/etc., analyze past order history, recommend books, ...
  - Data must be accessed efficiently, by many users
  - Data must be safe from failures and malicious users

# Challenges for a DBMS

Alice and Bob receive a \$200 gift certificate as wedding gift

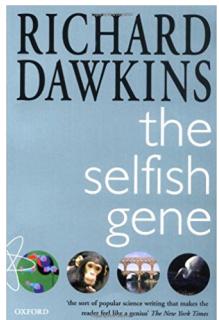


Alice

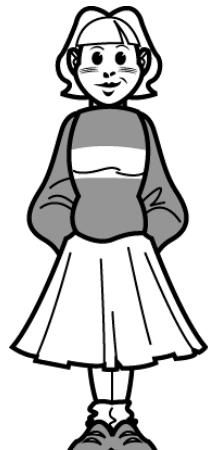
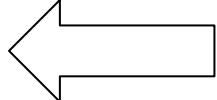
Bob

# Challenges for a DBMS

Alice and Bob receive a \$200 gift certificate as wedding gift

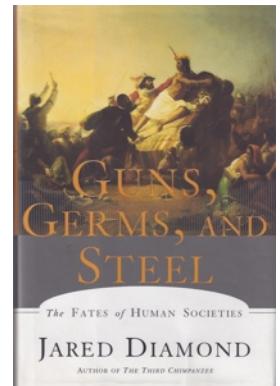
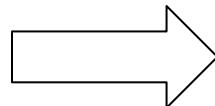


Alice @ her office orders  
"The Selfish Gene"



\$80

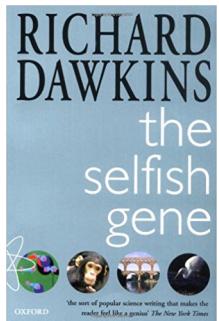
Bob @ home orders  
"Guns, germs, and steel"



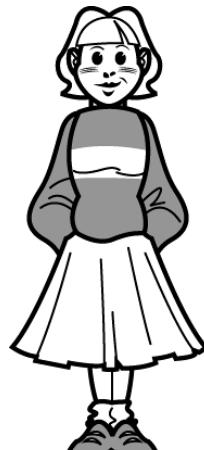
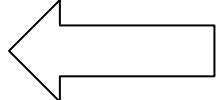
\$100

# Challenges for a DBMS

Alice and Bob receive a \$200 gift certificate as wedding gift

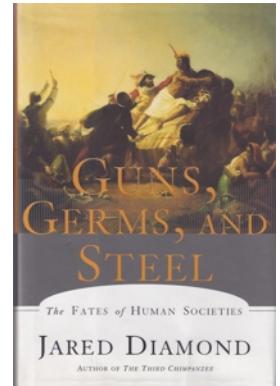
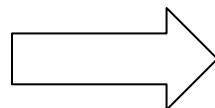


Alice @ her office orders  
"The Selfish Gene"



\$80

Bob @ home orders  
"Guns, germs, and steel"



\$100

Questions:

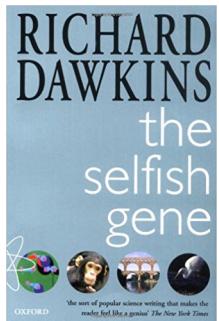
What is the ending credit?

What if second book costs \$130?

What if system crashes?

# Challenges for a DBMS

Alice and Bob receive a \$200 gift certificate as wedding gift



Alice @ her office orders  
"The Selfish Gene"



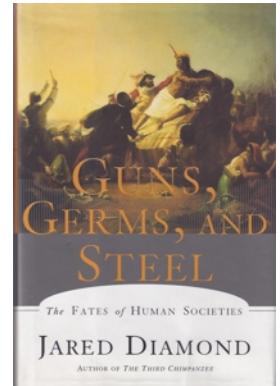
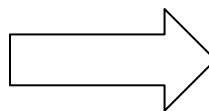
Alice

\$80

Bob @ home orders  
"Guns, germs, and steel"



Bob



\$100

Questions:

What is the ending credit?

What if second book costs \$130?

What if system crashes?

CMPSC 174A - 21 Wi

**Lesson:** a DBMS needs to handle various scenarios

# What a DBMS Does

- Describe real-world entities
- Store large datasets persistently
- Query & update efficiently
- Change structure (e.g., add attributes)
- Handle concurrent updates
- Crash recovery
- Security and integrity

# Key Players

- **DB application developers:** writes programs that query and modify data
- **DB designer:** establishes schema
- **DB administrator:** loads data, tunes system, keeps whole thing running
- **Data scientist:** learning insights from data
- **DBMS implementor:** builds the DBMS

# What is this class about?

- Unit 1: Intro (today)
- Unit 2: Relational Data Models and Query Languages
- Unit 3: Non-relational data
- Unit 4: RDMBS internals and query optimization
- Unit 5: Parallel query processing
- Unit 6: DBMS usability, conceptual design
- Unit 7: Transactions

# This Week

Homework 1 will be out.