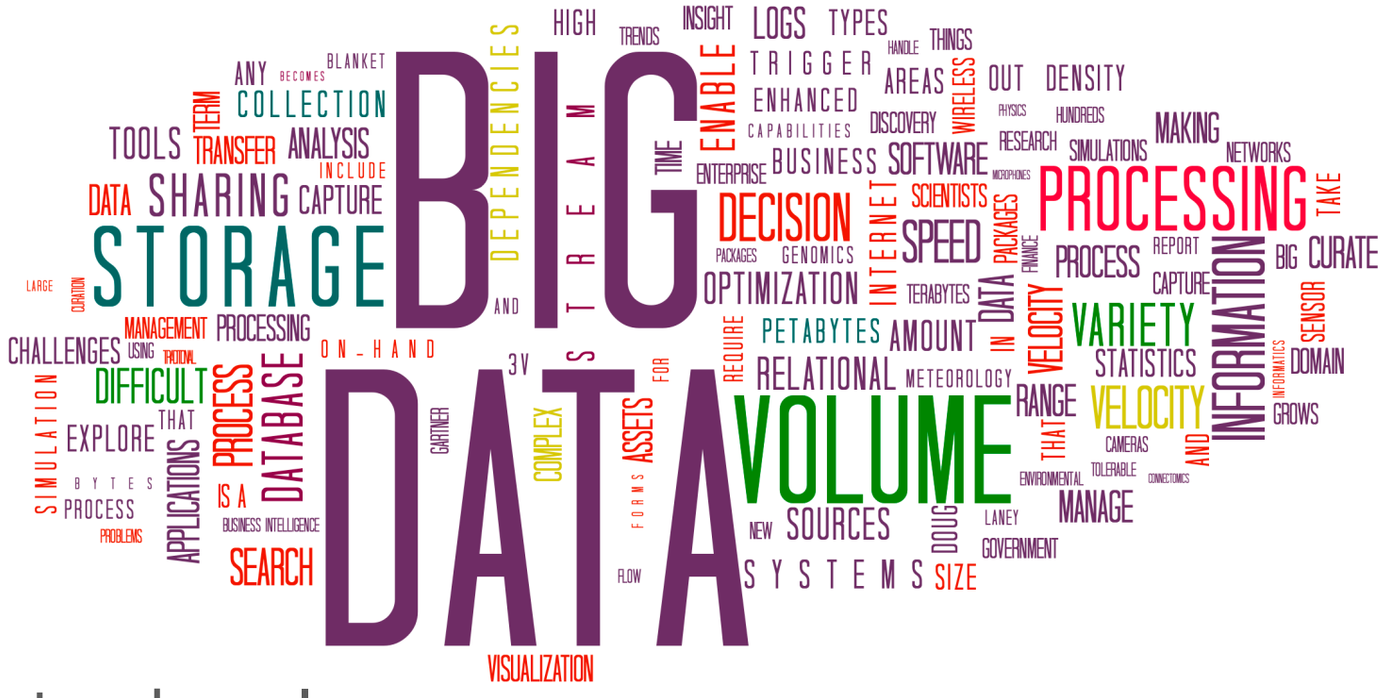
Introduction to Database Systems

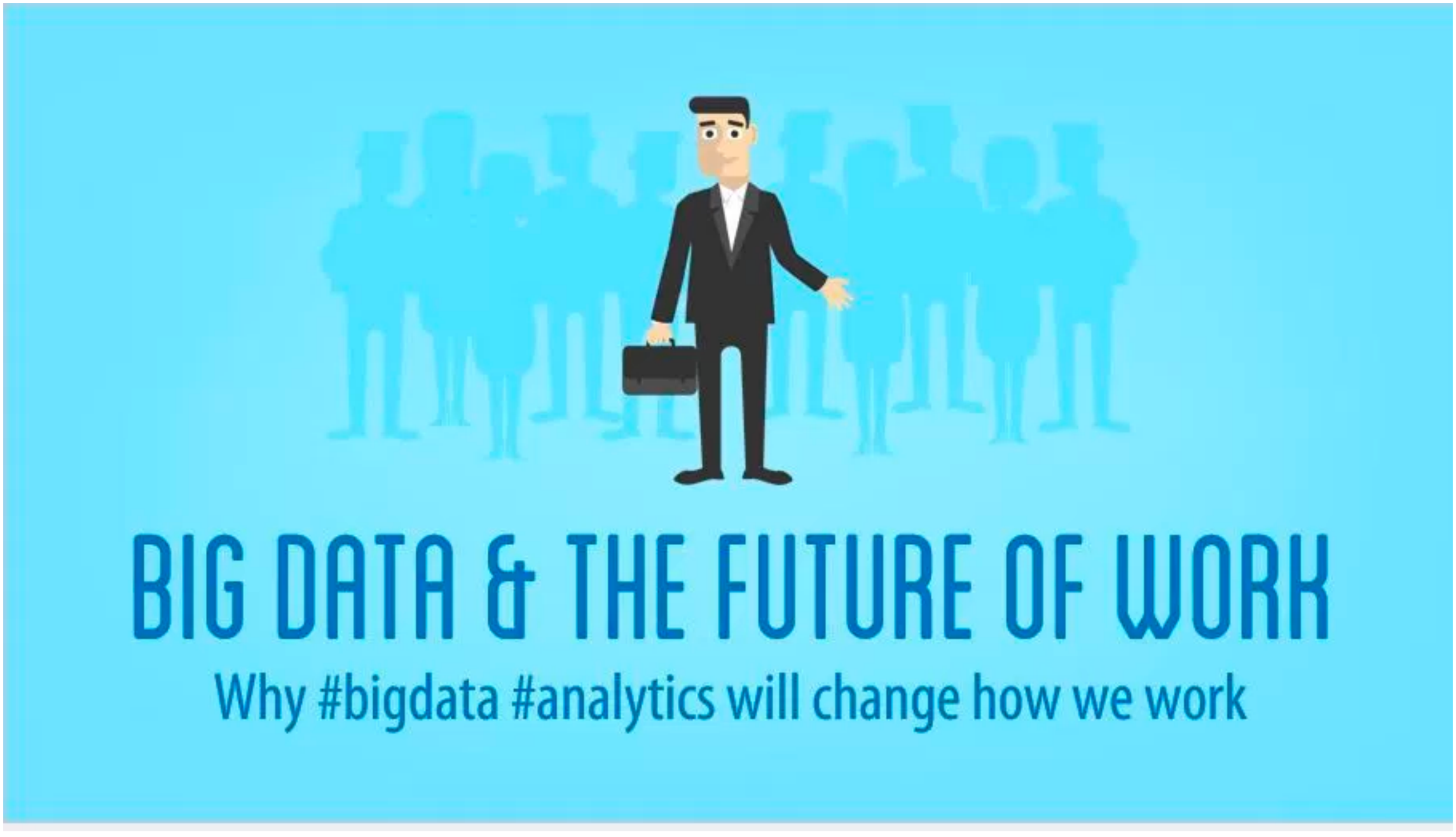
CS 377: Database Systems

Data is Everywhere

- Data management is necessary for:
  - Scientific discoveries
  - Service industries
  - Decision makers

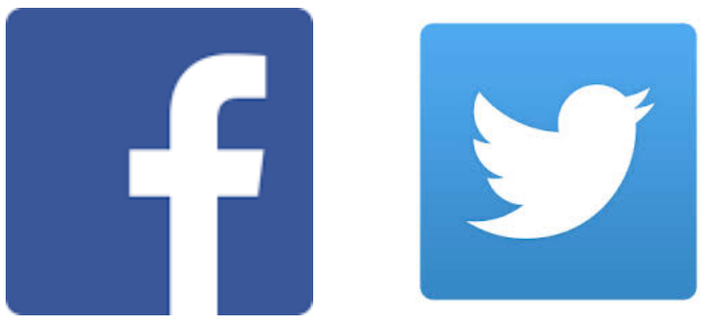


- Databases are the core technology



https://allthingsanalytics.com/2015/06/26/big-data-and-the-future-of-work/

World of “Data Driven” Companies



2014 A.M. Turing Award Winner

"Nobel Prize of Computing"

- Michael Stonebraker recognized for contributions to modern database systems

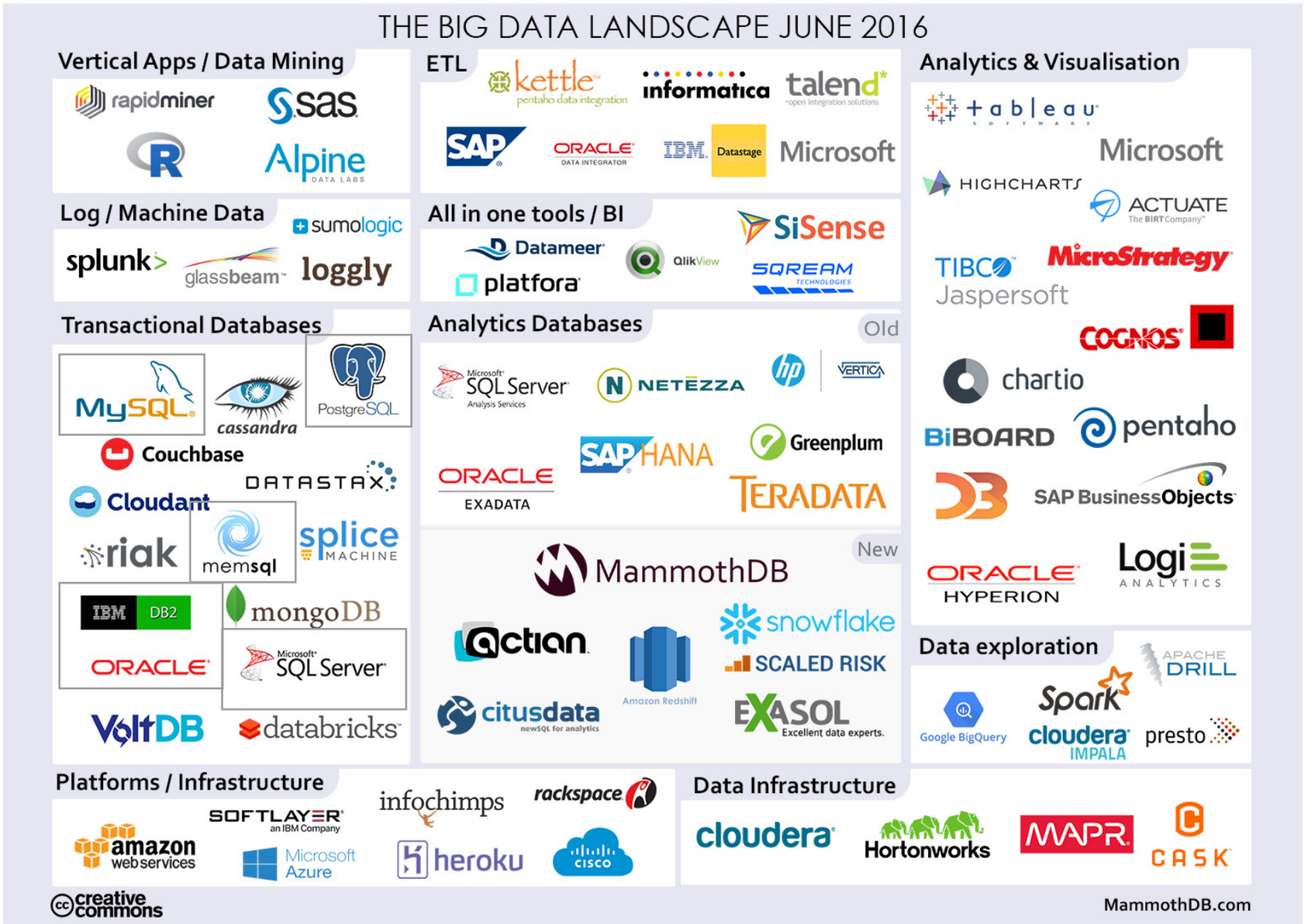
- Helped invent many relational database system (RDBMS) concepts

- Made / started many popular relational database systems implementations

http://amturing.acm.org/award_winners/stonebraker_1172121.cfm

Big Data Landscape: Infrastructure Change

Traditional RDBMS now surrounded by new tech — same principles though

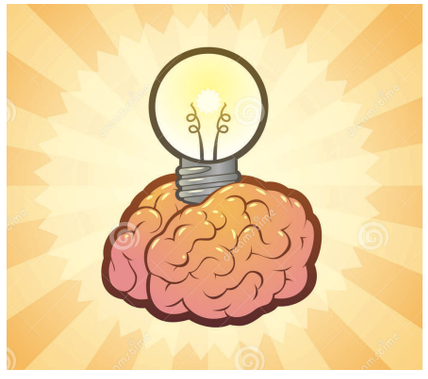


http://www.mammothdb.com/the-big-data-landscape-by-mammothdb/

Why Should I Care?

- Money
  - Companies and startups need talent
  - Massive industry

- Intellectual merit
  - Fundamental to computer science
  - Understand how to handle data properly



Course Overview

- Fundamental concepts of database and database systems
  - Modeling and design
  - Database programming
- Preview of database implementation
- Preview of modern databases

GOAL: get a flavor of RDMBS and modern DB systems

What This Course is Not

- How to be a database administrator
- How to tune specific database systems
- How to build and implement databases

Course Logistics

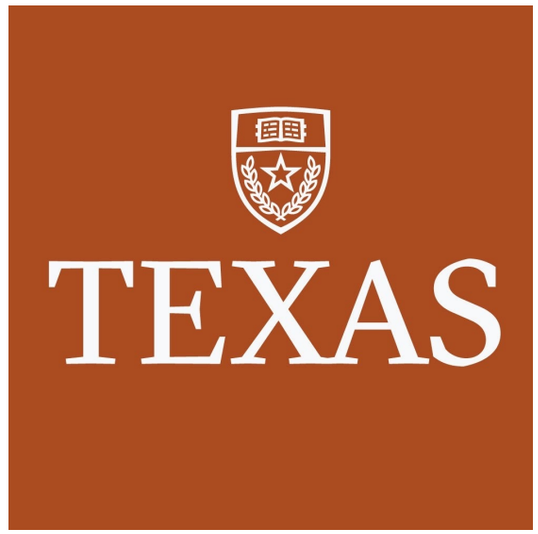
Course Website

http://joyceho.github.io/cs377-s17/index.html

- Lectures
- Assignments
- Example code (when applicable)

About Instructor (Me)

- Undergraduate / MEng from MIT
- PhD from University of Texas at Austin



- Research interests:
  - Data Mining / Machine Learning
  - Healthcare Informatics
- More information: http://joyceho.github.io

Teaching Staff

- Instructor: Joyce Ho
  - Email: joyce.c.ho@emory.edu
  - Office Hours @ MSC W414
    - M 1:00 pm-3:30 pm
    - W 9:30 am-12:00 pm
- TA: Camilo Valderrama
  - Email: cvalder@emory.edu
  - Office Hours: TBD

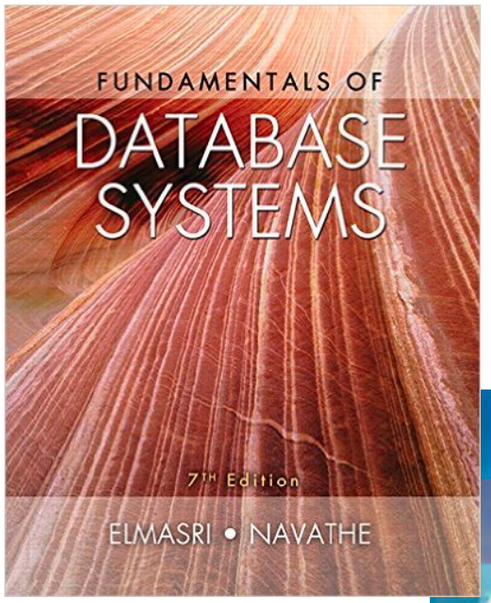
Communication

- Piazza: http://piazza.com/emory/spring2017/cs377
  - Announcements
  - Questions + Discussions
  - Assignment Clarifications + Slide Corrections
- Office Hours
- By Appointment

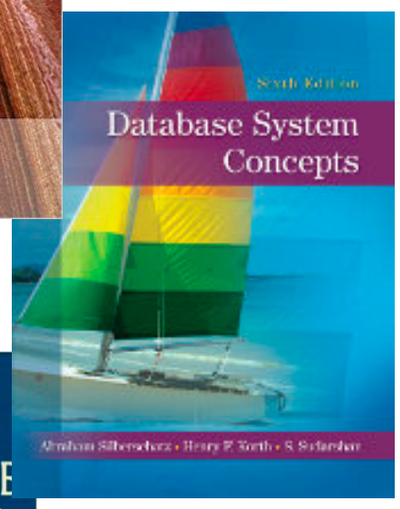
Important: Use an emory address or your OPUS name

Course Textbook(s)

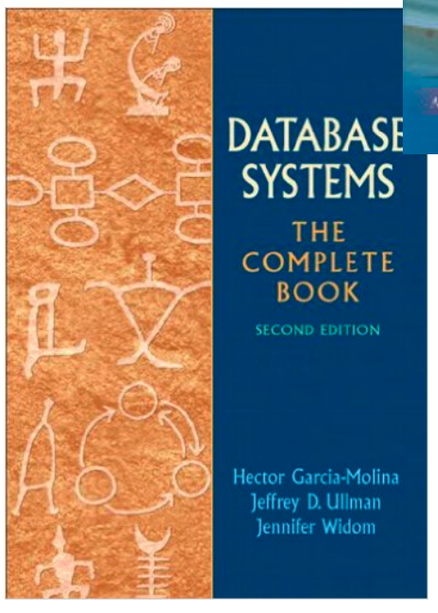
- Fundamentals of Database System (Recommended)
  Ramez Elmasri and Shamkant Navathe



- Database System Concepts (Supplemental)
  Abraham Silberschatz, Henry Korth, and S. Sudarshan



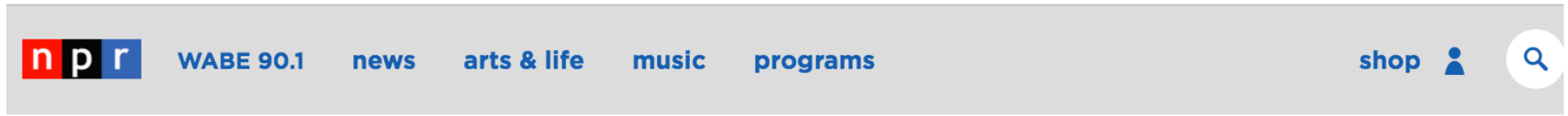
- Database Systems: The Complete Book (Supplemental)
  Hector Garcia-Molina, Jeffrey D. Ullman, and Jennifer Widom



Lectures

- Lecture slides cover essential material
  - Best reference for the course
  - Book may provide more details and the appropriate chapters are noted in the book
- Regular class attendance will make your life easier
  - Not all solutions to exercises will be posted in the slides

Laptops & Tablets



EDUCATION

Attention, Students: Put Your Laptops Away

ATTENTION / STUDENT LEARNING / TECHNOLOGY

3:19

+ Queue

April 17, 2016 · 6:00 AM ET

Heard on Weekend Edition Sunday

Download

Laptop Use in Class: Effects on Learning and Attention

August 22, 2015 | Beth Fisher

Research on Teaching and Learning

Your choice to use it or not!

Wonkblog

Why smart kids shouldn't use laptops in class

Graded Elements

- 4 Homeworks (20%)
  - Concepts practice — written assignments
- 4 Projects (30%)
  - SQL practical experience
  - Longer than homework
- Midterm (20%)
- Final (25%)
- Participation (5%)

Grades will likely be curved (up) so the class mean falls at least in a B/B+ range

Assignments: Honor Code

- College Honor Code and Departmental Policy
- Acceptable and encouraged to discuss assignments with other students but **ANY WRITEUP AND CODE MUST BE YOUR OWN**
- All program assignments must include a README file with the following comment at the top of the file:

```
/*
THIS CODE IS MY OWN WORK, IT WAS WRITTEN WITHOUT
CONSULTING CODE WRITTEN BY OTHER STUDENTS.
_Your_Name_Here_
*/
```

Late Assignment Policy

- 6 “flexible” late days to be distributed across 8 assignments (4 projects, 4 homeworks)
- 10% off per day
- Maximum number of late days per assignment is **3** days

Exam Policies

- Midterm rescheduling can be possible if request is made **at least a week** prior to the date
    - A different exam (of similar difficulty) than what will be administered
- Otherwise must be taken at the required time
    - Exceptions must be made through the Office of Undergraduate Education (OUE)

Participation Details

- Commitment to help you learn
- Bi-Weekly “quizzes” to assess class comprehension
  - Format will be 1 question, 10 minutes
  - Not meant to require much (if any) studying
- Bi-Weekly “polls” to tailor class for the students

In theory, if you attend class regularly and pay attention this should be automatic

Student Expectations

- Attend lectures
  - Miss them at your own peril
- Be active and think critically
  - Ask questions in class
- Do projects and homework
  - Start early and utilize the teaching staff
- Study for exams

Course Content: 1st Half

- Foundations (3 lectures)
  - Basic concepts
  - Data modeling and design (Entity-Relationship Model, relational data model)
- Query languages (7-8 lectures)
  - Relational algebra & calculus
  - Structured query language (SQL)
  - MySQL
- Database design (2 lectures)
  - Designing relational schema to prevent data corruption

Course Content: 2nd Half

- Applications of SQL (2 lectures)
  - JDBC & PHP
- Introduction to database systems (6 lectures)
  - Indexing
  - Basics of query optimization
  - Transactions & concurrency
- Specialized and New Data Systems (2 lectures)
  - Distributed RDMBS + MapReduce
  - NoSQL

Databases: High-Level Introduction

What does it mean?

- **Data**: A set of known facts that can be recorded and have an implicit meaning

- **Database**: A collection (files) of related data

- **Database Management System (DBMS)**: A software package or system that facilitates the creation and maintenance of a computerized database, allowing data to persist over long periods of time

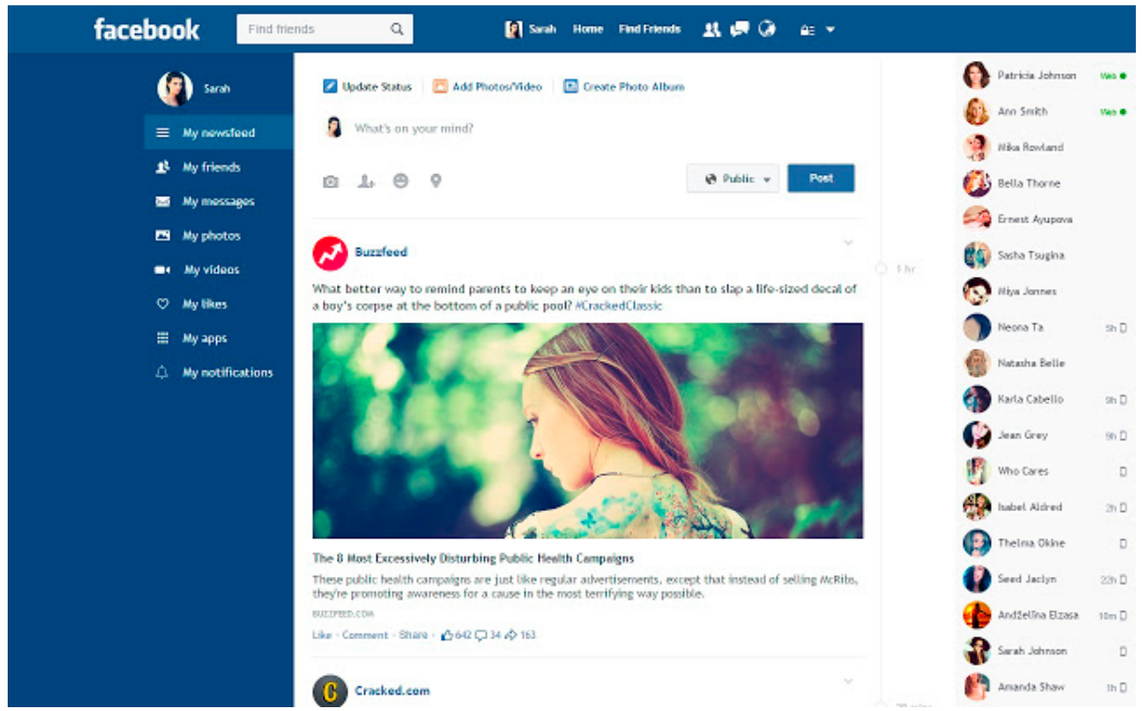
Examples of DBMS

- SQL Server, Microsoft Access (Microsoft)
- DB2 (IBM)
- Oracle
- MySQL, PostgreSQL, SQLite (Open Source)

All are "relational" (or "object-relational") database systems

Example: Facebook

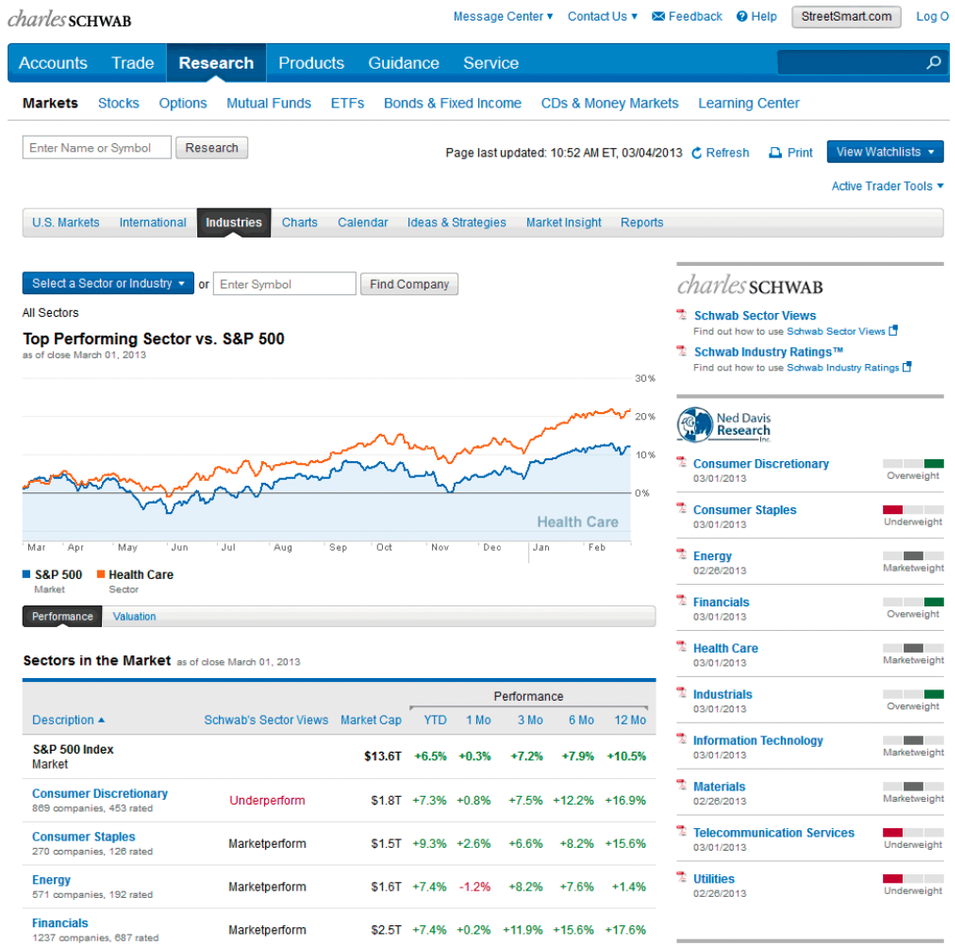
- What data needs to be stored?
  - Status Updates
  - Requests
  - Alerts
  - …
- How is data stored used?



Facebook uses MySQL

Example: Charles Schwab

- What data needs to be stored?
  - Customers
  - Transaction histories
  - Balances
  - …
- How is data stored used?

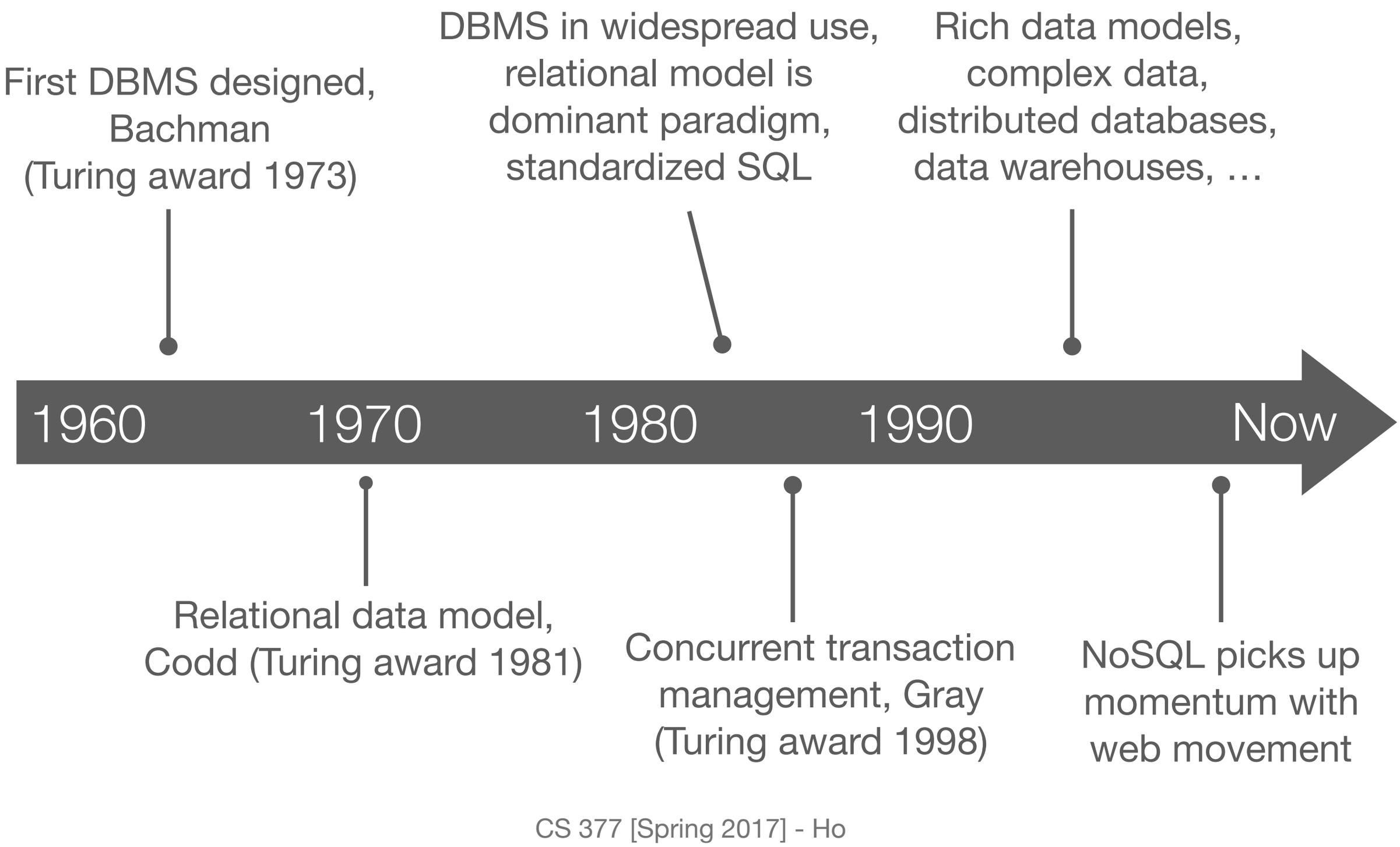


Charles Schwab uses multiple DBMS including Oracle and SQL Server

Standard DBMS Functionality

- Database manipulations: insertions, deletions, and modifications
- Efficient querying
- Concurrent processing and sharing by multiple users
- Consistent and valid data
- Recovery after crashes
- Security and user authorization

Evolution of DBMS



People

- **Database administrator**: sets up software/hardware, authorizes access, monitors its use, and tunes system

- **Database designer**: define and set up schema, sometimes loads the data

- **Database user**: modifies and queries the data

- **Database application developer**: builds applications to query/modify data

- **DBMS implementor**: builds the DBMS system


I THINK WE SHOULD BUILD AN SQL DATABASE.
UH-OH
E-mail: SCOTTADAMS@AOL.COM
S. Adams
DOES HE UNDERSTAND WHAT HE SAID OR IS IT SOMETHING HE SAW IN A TRADE MAGAZINE AD?
© 1995 United Feature Syndicate, Inc.(NYC)
WHAT COLOR DO YOU WANT THAT DATABASE?
I THINK MAUVE HAS THE MOST RAM.