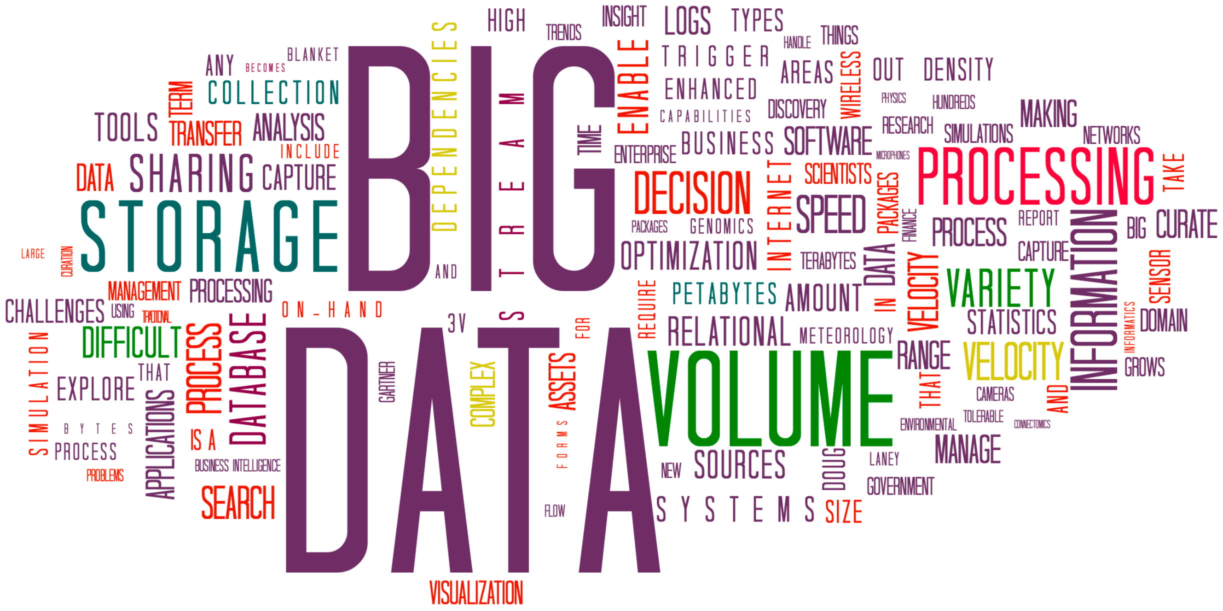
Introduction and Course Information

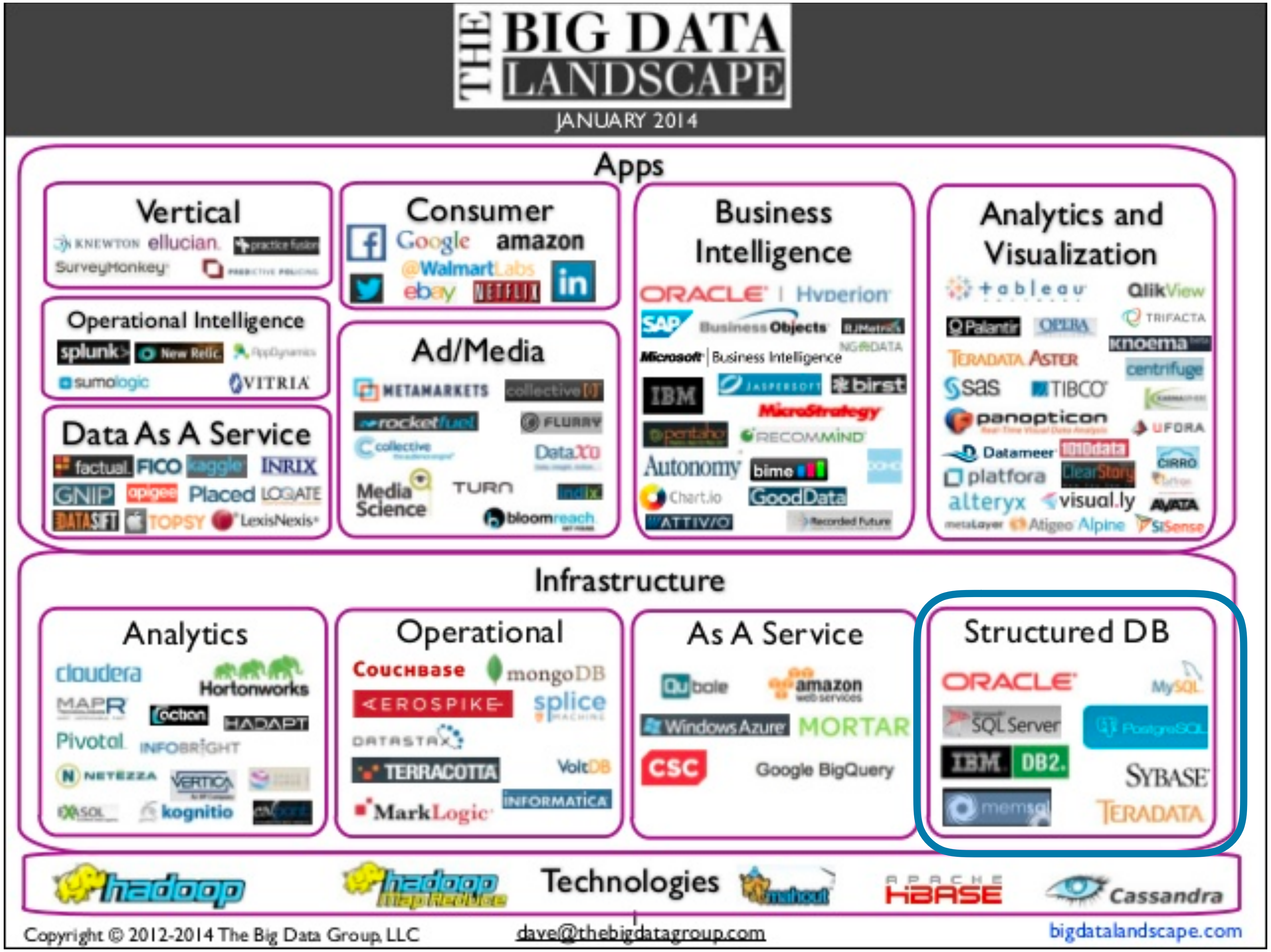
CS 377: Database Systems

Why Should I Care?

- Data can be found anywhere and everywhere
- Data management is necessary for:
  - Scientific discoveries
  - Service industries
  - Decision makers
- Databases are the core technology



Current Landscape



Course Overview

- Fundamental concepts of database and database systems
    - Modeling and design
    - Database programming
- Preview of database implementation
- Selected additional topics

Course Logistics

Teaching Staff

- Instructor: Joyce Ho
    - Email: joyce.c.ho@emory.edu
    - Office Hours: Tues/Thurs 1-4 pm @ MSC W414
- TA: Camilo Valderrama
    - Email: cvalder@emory.edu
    - Office Hours: Mon 2-4 pm @MSC N410

About 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

Course Information

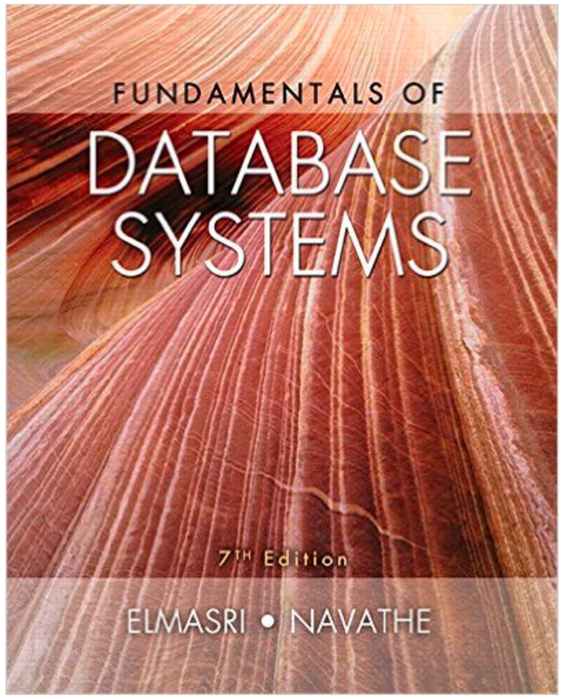
- Webpage: http://joyceho.github.io/cs377-s16/index.html
  - Lectures
  - Assignments
- Piazza: http://piazza.com/emory/spring2016/cs377
  - Announcements
  - Questions + Discussions
  - Note: Please use an emory address or your OPUS name

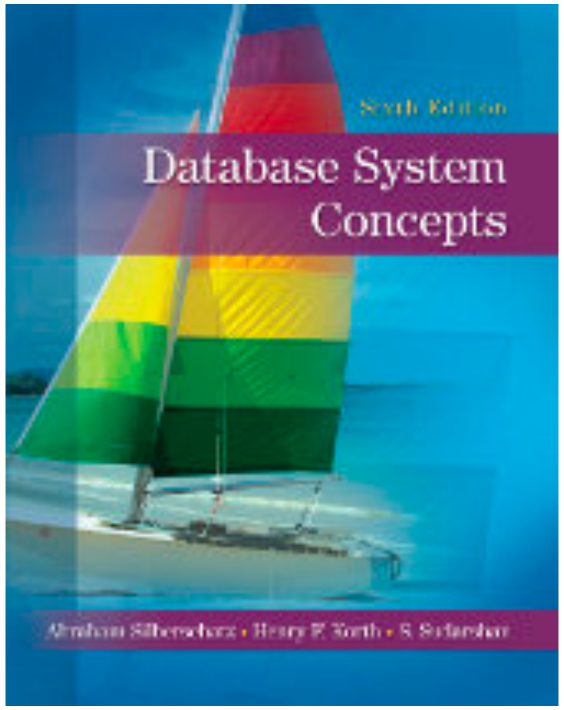
Course Textbook

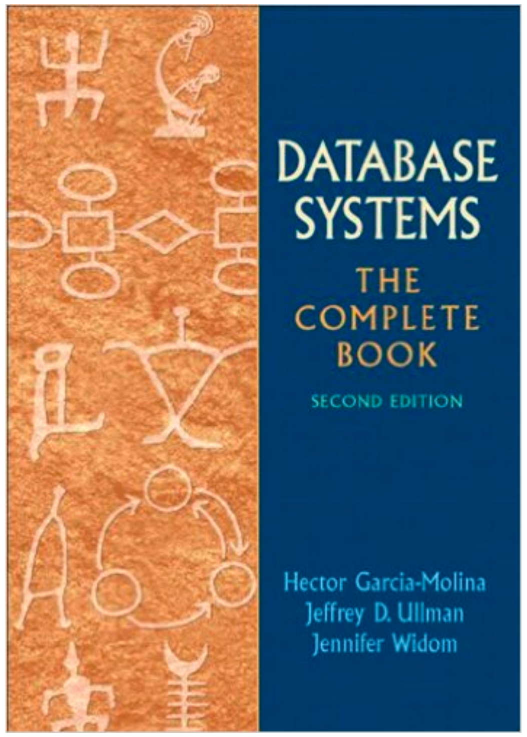
- Fundamentals of Database System (Recommended)
  Ramez Elmasri and Shamkant Navathe

- 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







Grading

Assignments / Projects	50%
Midterm	20%
Final	25%
Participation	5%

Grades may be curved (up) so the class mean falls at least in a B range.

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

Policies

- Exams
  - Must be taken at the required time
  - Midterm rescheduling can be possible if request is made at least a week prior to the date
- Late Assignment
  - Will be accepted within 3 days of due date with 10% penalty per day
  - 2 late date assignments per student

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 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_
*/
```

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

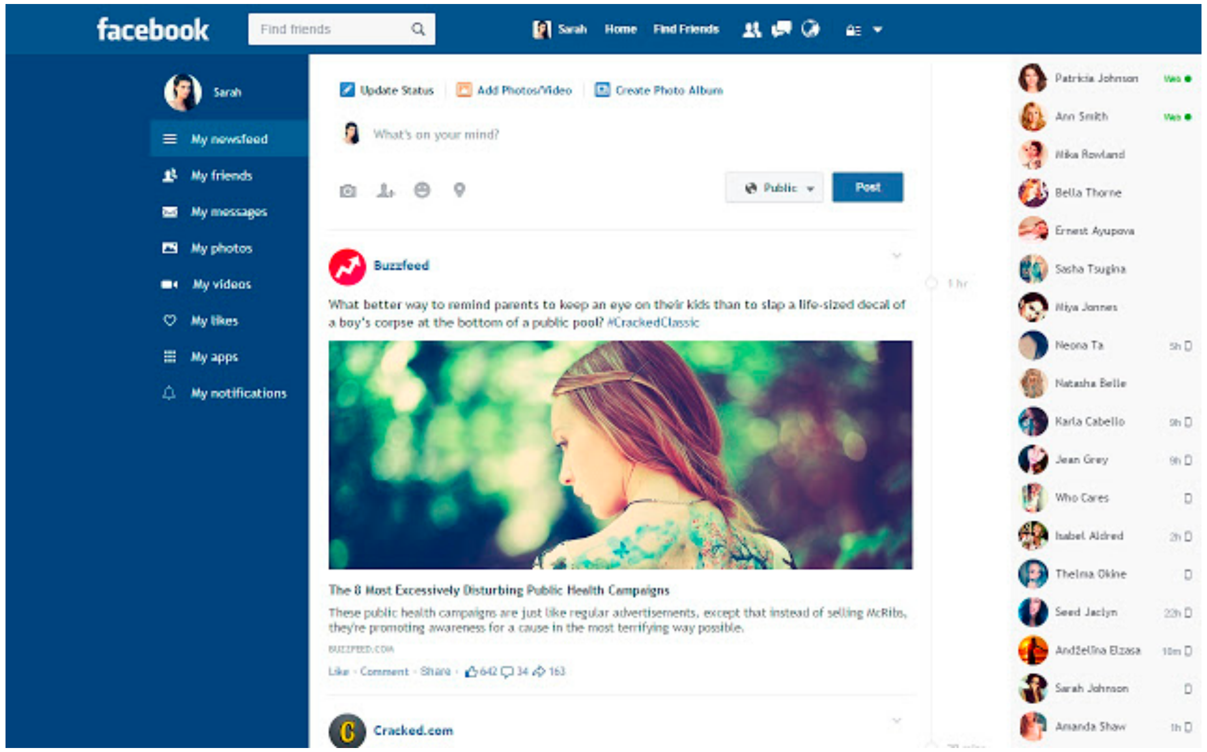
Examples of DBMS

- 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?



What better way to remind parents to keep an eye on their kids than to slap a life-sized decal of a boy's corpse at the bottom of a public pool? #CrackedClassic

The 8 Most Excessively Disturbing Public Health Campaigns

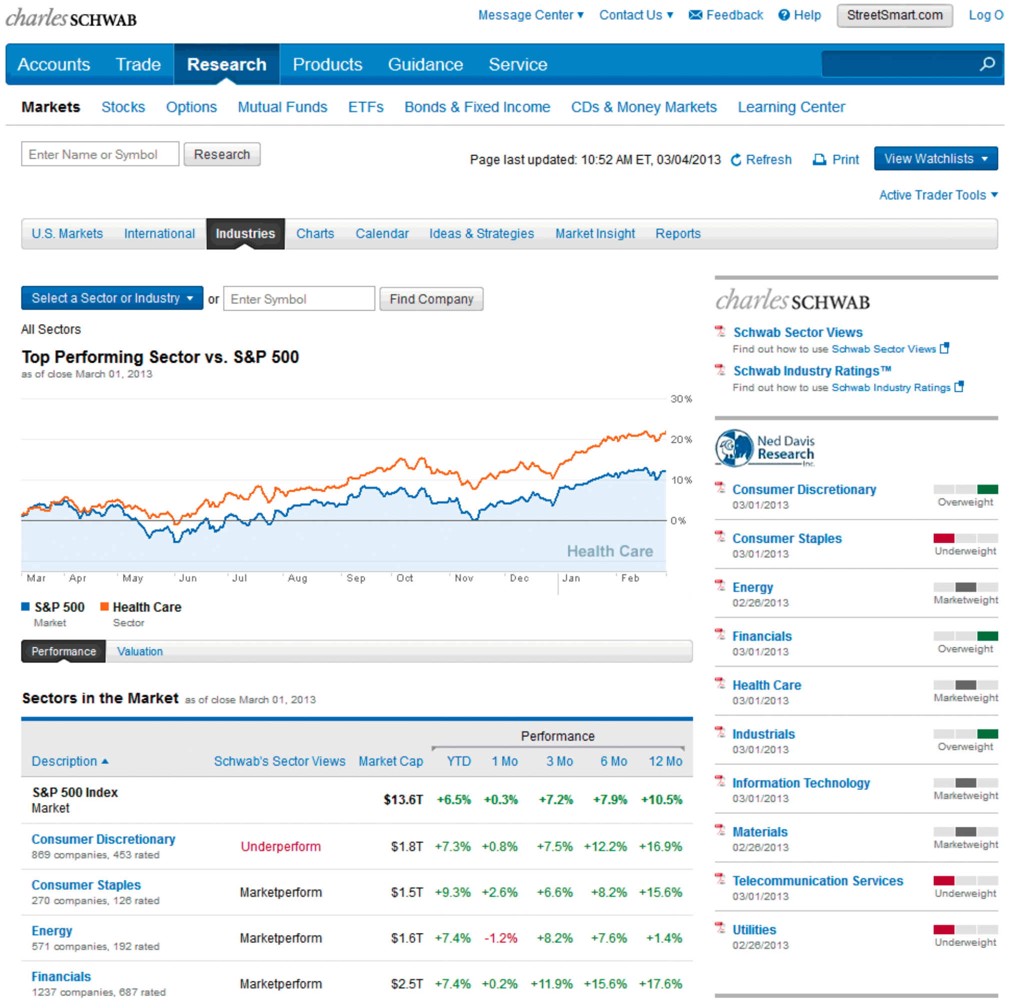
These public health campaigns are just like regular advertisements, except that instead of selling McRibs, they're promoting awareness for a cause in the most terrifying way possible.

BUZZFEED.COM

Facebook uses MySQL

Example: Charles Schwab

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



Page last updated: 10:52 AM ET, 03/04/2013

All Sectors

Top Performing Sector vs. S&P 500
as of close March 01, 2013

Sectors in the Market as of close March 01, 2013

			Performance				
Description ▲	Schwab's Sector Views	Market Cap	YTD	1 Mo	3 Mo	6 Mo	12 Mo
S&P 500 Index Market		$13.6T	+6.5%	+0.3%	+7.2%	+7.9%	+10.5%
Consumer Discretionary 869 companies, 453 rated	Underperform	$1.8T	+7.3%	+0.8%	+7.5%	+12.2%	+16.9%
Consumer Staples 270 companies, 126 rated	Marketperform	$1.5T	+9.3%	+2.6%	+6.6%	+8.2%	+15.6%
Energy 571 companies, 192 rated	Marketperform	$1.6T	+7.4%	-1.2%	+8.2%	+7.6%	+1.4%
Financials 1237 companies, 687 rated	Marketperform	$2.5T	+7.4%	+0.2%	+11.9%	+15.6%	+17.6%

charles SCHWAB

Schwab Sector Views
Find out how to use Schwab Sector Views

Schwab Industry Ratings™
Find out how to use Schwab Industry Ratings

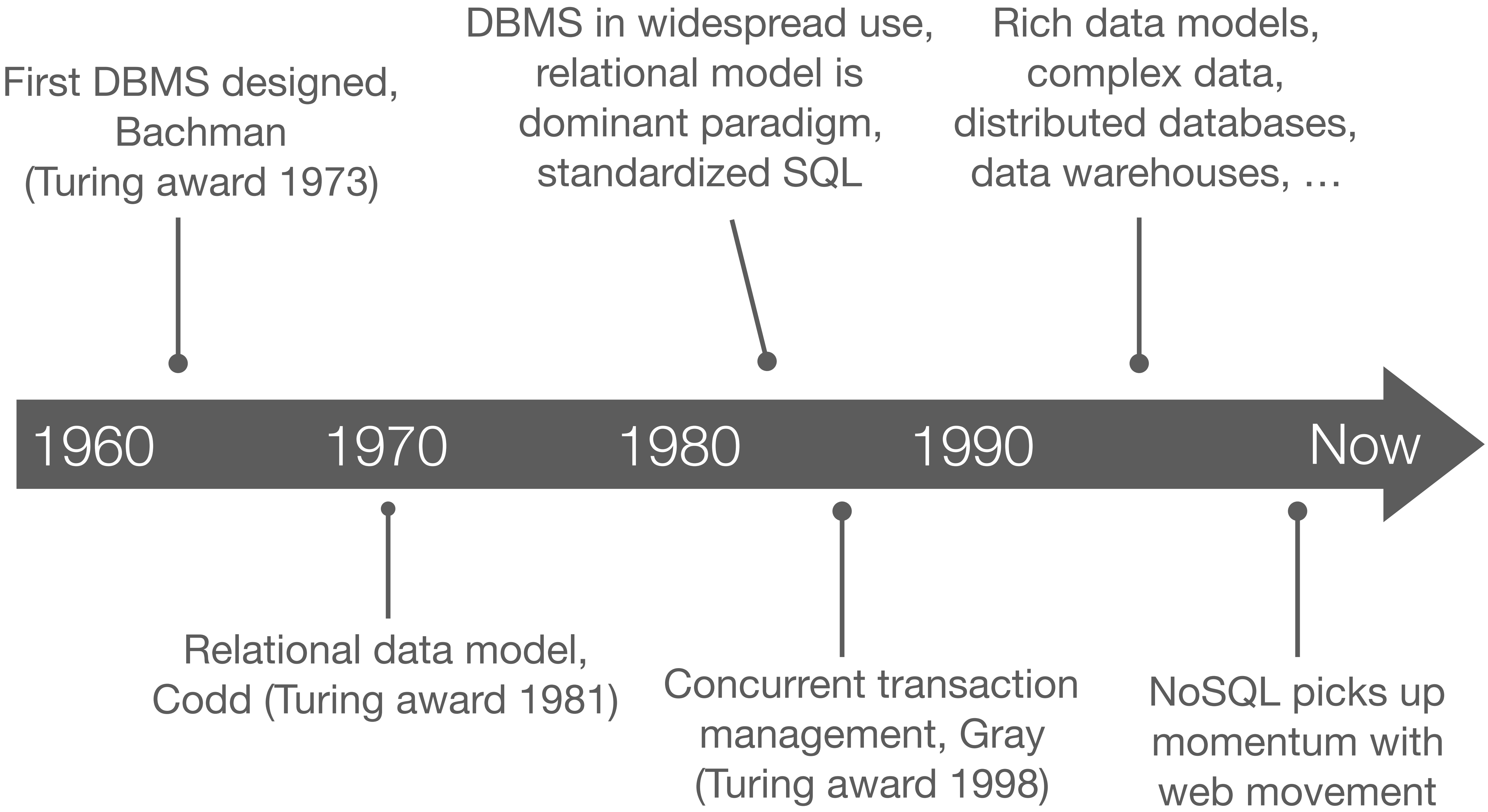
Consumer Discretionary 03/01/2013	Overweight
Consumer Staples 03/01/2013	Underweight
Energy 02/26/2013	Marketweight
Financials 03/01/2013	Overweight
Health Care 03/01/2013	Marketweight
Industrials 03/01/2013	Overweight
Information Technology 03/01/2013	Marketweight
Materials 02/26/2013	Marketweight
Telecommunication Services 03/01/2013	Underweight
Utilities 02/26/2013	Underweight

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



1960

1970

1980

1990

Now

First DBMS designed, Bachman (Turing award 1973)

Relational data model, Codd (Turing award 1981)

DBMS in widespread use, relational model is dominant paradigm, standardized SQL

Concurrent transaction management, Gray (Turing award 1998)

Rich data models, complex data, distributed databases, data warehouses, …

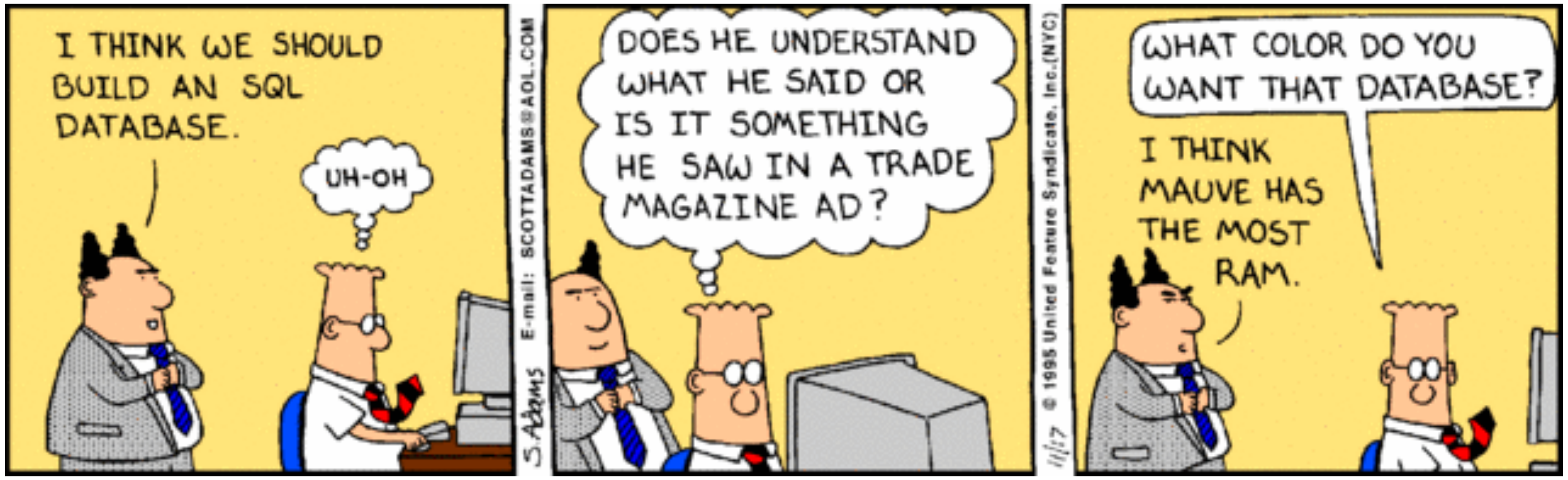
NoSQL picks up momentum with web movement

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

Course Content

- Basic concepts
  - Data independence, database system components, 3 level database architecture
- Data modeling and design
  - Entity-Relationship Model
  - Relational Data Model
  - Relational Database design theory - normal functions, functional dependencies
- Database programming
  - Relational algebra
  - SQL (Structured Query Language)
  - Oracle JDBC and PHP
- Selected Topics
  - Database internals (indexes, query optimization, transactions)
  - NoSQL


I THINK WE SHOULD BUILD AN SQL DATABASE.
UH-OH
DOES HE UNDERSTAND WHAT HE SAID OR IS IT SOMETHING HE SAW IN A TRADE MAGAZINE AD?
WHAT COLOR DO YOU WANT THAT DATABASE?
I THINK MAUVE HAS THE MOST RAM.
E-mail: SCOTTADAMS@AOL.COM
S. Adams
© 1995 United Feature Syndicate, Inc.(NYC)
11/17