

CSE 180: Database Systems I

Winter 2024

Course Syllabus

Instructor:

Sheldon (Shel) Finkelstein, shel@ucsc.edu

Office Hours in E2-249B:

Wed 2:00-3:00pm

(or by appointment)

CSE 180 Course Logistics

Classes: MWF 10:40AM - 11:45AM, Baskin Auditorium 101

- This is an **In-Person** class for both Lectures and Lab Sections, unless there's an emergency.
 - Lecture Capture recordings will be available for all students on Canvas under Yuja.
 - However, students learn more when they attend class, since you can't ask (or answer) questions while you play a recording.
 - Lecture Capture is not available for Lab Sections.
- Lecture Slides and other class information are posted on [Piazza](#) under Resources→Lectures.
- General Information for CSE 180 appears in the “Welcome to CSE 180” posting on Piazza, and in the General Information file that’s under Resources→General Resources on Piazza.
- Canvas will only be used for submitting Lab Assignment and for Grading, use Piazza (not Canvas) for questions.
 - Enrolled students can submit Lab Assignment solutions on Canvas; students on the wait list do not have access to Canvas.
 - Login to Canvas at <https://canvas.ucsc.edu> using your CruzID and Gold password. CSE 180 should be one of the classes available.
 - Info on Canvas is at <http://its.ucsc.edu/canvas/index.html>
- There is a [CSE 180 Discord](#), but we respond to questions on Piazza, not on Discord.

Teaching Assistants and Lab Sections

Teaching Assistants:

- Gavin Cooke, gcooke@ucsc.edu
- Utkarsh Gupta, utgupta@ucsc.edu
- Nayan Bhatia, nbhatia3@ucsc.edu

Lab Sections, all in BE-105 (In-Person, not recorded)

All Lab Sections begin during the first week of classes!

- CSE 180-01A Mon 12:30PM - 2:05PM **Utkarsh**
- CSE 180-01B Tue 9:50AM - 11:25AM **Gavin**
- CSE 180-01C Wed 3:30PM - 5:05PM **Nayan**
- CSE 180-01D Thu 1:30PM - 3:05PM **Utkarsh**
- CSE 180-01E Fri 1:30PM - 3:05PM **Nayan**

Holidays (no Lectures, no Lab Sections, no Office Hours):

- Martin Luther King Day, Mon, Jan 15
- Presidents Day, Feb 19

Lab Section Format

- Approximately the first 60 minutes of each Lab Section will be presentation by TA, reviewing current material and helping everyone understand the Lab Assignment (Public portion).
 - Remaining 35 minutes of each Lab Section (Private portion) will help individuals with questions, serving as a miniature Office Hour.
 - Lab Sections are In-Person only.
 - Lab Sections begin on the first day of classes.
 - Students in the 01A Monday Lab Section should be sure to attend on Monday, Jan 8, especially since Monday, Jan 15 is a holiday, to be introduced to use of PostgreSQL and Gradiance (and more).
- Yes, you must register for a Lab Section, but students may attend any of the Lab Sections.
 - We don't check attendance for Lectures or Lab Sections.
 - You might consider attending a different Lab Section during holiday weeks.

Textbook

Textbook: “A First Course in Database Systems”, Jeffrey Ullman and Jennifer Widom, Prentice-Hall, 3rd edition.

- Textbook can be ordered new/used, e.g., from the [UCSC Bay Tree Bookstore](#).
- [Kindle ebook](#) is available for a textbook (based on the 2cd edition of our textbook but covering other material as well), “Database Systems: The Complete Book”, Hector Garcia-Molina, Jeffrey Ullman and Jennifer Widom, Prentice-Hall, 2cd edition.

Some other good database books (not used in this course)

- “Database Management Systems”, 3rd edition, Ramakrishnan and Gehrke, McGraw-Hill.
- “Database System Concepts”, 7th edition, Silberschatz, Korth and Sundarshan, McGraw-Hill. [ebook](#)

[Science and Engineering Library has these textbooks [on Reserve](#).]

Some Background: Shel Finkelstein

- IBM Almaden Research: Database and Distributed Systems
- Tandem Computers: Transaction Management and System Managed Storage
- Illustra Information Systems: Productizing Postgres (Time Series, Administration)
- ADB/Matisse (Object Database): Chief Scientist
- Sun Microsystems: Managed Enterprise Java architecture and partnership relationships when Java Enterprise Edition was created
- SAP: VP, Research Fellow, Chief Tech Architect, focusing on Applications and Database
- Teaching at UC Santa Cruz since January 2015, mainly Database courses

Teaching Assistants

Gavin Cooke

gcooke@ucsc.edu

Lab Section in BE-105

01B Tue 9:50AM - 11:25AM

Office Hour:

Thu 11:00am - noon, BE-312C/D
(or by appointment)



Teaching Assistants

Utkarsh Gupta

utgupta@ucsc.edu

Lab Sections in BE-105

01A Mon 12:30PM - 2:05PM

01D Thu 1:30PM - 3:05PM

Office Hours:

Wed 9:00 – 10:00am in BE-153A

Thu 10:00 – 11:00am in BE-151

(or by appointment)



Teaching Assistants

Nayan Bhatia

nbhatia3@ucsc.edu

Lab Section in BE-105

01C Wed 3:30PM - 5:05PM

01E Fri 1:30PM - 3:05PM

Office Hours

Mon 2:00 – 3:00pm in BE-312C/D

Tue 2:00 – 3:00pm in E2-315

(or by appointment)



Group Tutor-1

Alex Asch

alasch@ucsc.edu

Tutoring Times:

- Tuesday: 1:00pm – 2:00pm
- Friday: 9:20am – 10:25am
- Friday: 1:30pm – 2:30pm

Tutoring will be in Jack's Lounge, on the first floor of Baskin Engineering.



Group Tutor-2

Yash Raj Singh

ysingh5@ucsc.edu

Tutoring Times:

- Monday: 100pm – 2:00pm
- Tuesday: 2:00pm – 3:00pm
- Wednesday: 1pm – 2:00pm

Tutoring will be in Jack's Lounge, on the first floor of Baskin Engineering.



Course Description/Syllabus

CSE 180 covers concepts, approaches, tools, and methodology of database design and querying. Topics include:

- Relational Data Model and its history
- SQL language: Data Definition, Queries and Updates, Indexes, Views, Constraints, Rules
- Relational Algebra, Query Execution and Transaction Processing
- Database Application Development
- Design Theory: Schemas and Normal Forms
- On-Line Analytical Processing (OLAP)
- Semi-Structured Data Models (time permitting)
- Big Data, NOSQL and Cloud (time permitting)

Lots of practical material and logic ... and lots of concepts and theory, particularly in the second half of the quarter.

- Students who do not attend Lectures and Labs tend to do poorly in this course.
- Please keep that in mind if you want to pass this class.

Tentative Lecture Schedule

Topic	# Lectures	Dates	Chapters
History and Introduction	1	1/8	1
The Relational Data Model	2	1/10 – 1/12	2.1, 2.2
SQL: DDL, DML; Defining relations and constraints; writing queries, modifying data, transactions, views, indexes	12	1/17 – 2/12	2.3, 2.5, 6.1-6.5, 7, 8
Midterm	1	Wed 2/14	
Relational Algebra and Query Execution	3	2/16 – 2/23	9.1, 9.2, 9.6
Database Application Development	3	2/25 – 3/1	2.4, 2.5
Relational Database Design and Normalization	4	3/4– 3/11	3.1-3.5
OLAP	1	3/13	10.6, 5.2.7
Semi-Structured Data Models: XML, JSON	0.5	3/15	Augmented parts of 11, 12
Review (time permitting)	0.5	3/15	
Final		Tue 3/19 8:00am – 11:00am	
(U.S. Holidays—no classes)		Mon 1/15; Mon 2/19	

Lectures

- You are responsible for all material presented in Lectures, even though we don't take attendance.
 - Some material might not be in the textbook.
 - Some information might not appear on the slides.
- Lecture slides will be posted on Piazza under Resources → Lectures, usually before the Lecture.
 - Posted slides correspond to Topics, not dates.
 - Slides might be modified after initial posting.
- You should learn and understand the **concepts**, not just what's presented in the Lectures.

Course Evaluation

- Gradiance Homeworks: 10%
 - Lab Assignments: 10%
 - Midterm: 35%
 - Final Exam: 45%
-
- In F23, Lab Assignments didn't contribute to Course Score (due to concerns about ChatGPT), so many students didn't do them.
 - In W24, Lab Assignments do contribute to Course Score, but they're worth only 10% of your Course Score.

Homeworks: Lab Assignments

- This course involves database application development projects through 4 Lab Assignments.
 - You must attend a Lab Section, where the Teaching Assistant will help guide you through Lab Assignments and answer your questions.
 - Lab Assignments will be submitted via Canvas, as zip files.
 - You'll have to learn how to move files from unix to your computers, so that you can post them on Canvas.
 - **No late submissions.** Make sure that you post the correct file, and that you post it on time (by 11:59pm on the Due Date).
- Only enrolled students can submit Lab Assignment solutions on Canvas.
 - If you disagree with a grade, check with your TA, and then with me.
 - **Never** contact the Reader who graded your Lab Assignment.

Generative AI using Large Language Models (LLMs): Benefits and Disadvantages

Students are strongly encouraged to do Lab Assignments on their own, perhaps using Generative AI systems (such as OpenAI's [ChatGPT](#), Meta's [LLaMA](#) and Google's [Bard](#)) as learning aids, but not as substitutes for learning.

LLM-based systems can sometimes be very helpful tools for learning, especially since they can supply explanations justifying their answers.

- But these systems may provide **incorrect** answers, so they should be regarded as clever (somewhat unreliable) assistants which have terrific memories, not as perfect oracles.
- Detecting cheating using LLM-based systems is a huge waste of time, so we won't try to do that for Lab Assignments.

Students will not be able to use LLM-based systems on CSE 180 Exams, or during job interviews.

- Lab Assignments scores count 10% towards Course Scores, and students who use LLM-based systems on Lab Assignments to supply solutions might receive perfect scores ... **but such students are mostly cheating themselves.**
- CSE 180 Exams will be In-Person, except when there are severe illnesses ... and students who take Exams remotely **may be interviewed afterwards about their answers.**

Consequently, the most important feedback students receive from Lab Assignment grades is about their errors ,,,

- ... and students, not LLM-systems, are enrolled in CSE 180.

Tentative Lab Assignment Dates

- **Lab Sections** are a required part of course, but we don't take attendance.
 - All Lab Sections meet starting during the week of **January 8**, introducing you to Gradiance and PostgreSQL
- **Lab1** (which is much easier than the rest of the Lab Assignments)
 - Assigned Sunday, January 14 (or earlier)
 - Due Tuesday, January 23, 11:59pm
 - Monday, January 15 is a U.S. Holiday
- **Lab2**
 - Assigned Wednesday, January 24
 - Due Tuesday, February 6, 11:59pm
- **Lab3**
 - Assigned Wednesday February 14 (not February 7) due to Midterm on **Wednesday, February 14**
 - Due Tuesday, February 27 , 11:59pm
 - Monday, February 19 is a U.S. Holiday
- **Lab4** (which involves database access from a C program)
 - Assigned Wednesday, February 28
 - Due Tuesday, March 12, 11:59pm
- **Last day of classes:** Friday, March 15
 - During last week of classes, all Lab Sections will be held, discussing Lab4, as well as addressing questions about overall course before Final Exam, which is on Tuesday, March 19.

Gradiance Assignments and Practice Homeworks

- Non-lab homework assignments will be assigned via **Gradiance**, which has automated grading; see: <http://www.gradiance.com/pub/stud-guide.html>
 - For CSE180, go to <http://www.gradiance.com/services> to enroll in this class, using class token **EC7184D3**
 - Enter your name and your UCSC email, plus a login id and a password.
 - Please use your cruzid as your Gradiance login id, if you can.
 - You can reset your Gradiance password if you forget it, as long as you know your email and your Gradiance login id.
 - Gradiance assignments are automatically graded by the Gradiance system.
 - They will always be announced on Piazza (and in Lecture) when they are assigned.
- There will also be ungraded Practice Homeworks that you don't hand in.
 - You're responsible for those Practice Homeworks, even though we will not discuss most of them in class, and we will not post answers to them.
 - You can ask questions about Practice Homeworks in Lab Sections, during Office Hours, or on Piazza.

Exams

- Midterm on Wednesday, February 14 (65 minutes) covers first half of the quarter, potentially including material through Monday, February 12.
- Final Exam on Tuesday, March 19, 8:00am –11:00am (3 hours) is comprehensive, covering material from the entire quarter, with somewhat greater emphasis on the (more difficult) second half of the quarter.
- This is an In-Person class, so Midterm and Final will be taken In-Person in our classroom.
 - All students must take these exams In-Person, unless you have a specific severe medical reason not to do so.
 - For both the Midterm and the Final Exam, you may prepare a single 8.5" by 11" sheet of paper, with anything that you can read unassisted printed or written on both sides of the paper.
 - Sheets may not be shared during the exams.
- No Make-up Exams. No Early/Late Exams. No Devices during Exams. Both Exams are required.

CSE 180 Grading

- Curves
 - Midterms and Final may sometimes be curved a little.
 - There is no curve for Lab Assignments or for Gradiance.
- Course Grading is based on your overall Course Score, as in most other courses.
 - 90.00 to 93.33 A-
 - 86.67 to 89.99 B+
 - 83.33 to 86.66 B
 - ... and similarly for other grades.
- Getting a Course Score of 73.33 gives you a passing grade, C.
 - There is no curve for the Course Score.
- Canvas' evaluation of your Course Score is not correct.

Dealing with COVID

- Each individual at UC Santa Cruz should act in the best interests of everyone else in our community, even though masking is no longer required.
 - Face masks and rapid tests continue to be available for purchase at a variety of community locations. Additionally, the County of Santa Cruz provides [this list of locations that provide testing](#).
 - See [UCSC COVID Frequently Asked Questions](#) for information about UCSC policy on COVID, including [what to do if you have a positive COVID result](#).
- If you do have a positive COVID result or some other major illness, please let me know as soon as possible, so that we can determine ways for you to continue to make progress in this class.

Title IX Resources and Responsibilities

- Title IX prohibits gender discrimination, including sexual harassment, domestic and dating violence, sexual assault, and stalking. If you have experienced sexual harassment or sexual violence, you can receive confidential support and advocacy at the Campus Advocacy Resources & Education (CARE) Office by calling (831) 502-2273.
- In addition, Counseling & Psychological Services (CAPS) can provide confidential, counseling support, (831) 459-2628. You can also report gender discrimination directly to the University's Title IX Office, (831) 459-2462. Reports to law enforcement can be made to UCPD, (831) 459-2231 ext. 1. For emergencies call 911.
- Faculty and Teaching Assistants are required under the [UC Policy on Sexual Violence and Sexual Harassment](#) to inform the Title IX Office should they become aware that you or any other student has experienced sexual violence or sexual harassment.

Academic Integrity

- No form of academic dishonesty will be tolerated. You are encouraged to read the campus policies regarding academic integrity at https://www.ue.ucsc.edu/academic_misconduct. Violations may lead to penalties including (but not limited to) **failing this course**.
 - *Please be sure to take this very seriously. I don't like to penalize students for violations ... but I've done it before, and I'll do it again.*
- You are allowed to ask for some minor help when working on assignments, if your submission acknowledges the help that you received.
 - There will be no penalty for small amounts of acknowledged assistance.
- If you receive major help from other people (whether they are the class or outside the class), that's an Academic Integrity violation. **Don't!**
 - Help with concepts is okay, but help with solutions is not.
 - Instructors and TAs will also help explain concepts, but we will not do the assignments for you, and we will not tell you whether your solutions are correct.
- If you have any questions about these rules, please discuss them with the instructor **immediately**.

Intellectual Property and Copyright

Faculty have been advised to include the following statement in their course syllabus.

“My lectures and course materials, including presentations, tests, outlines, and similar materials, are protected by U.S. copyright law and by University policy. I am the exclusive owner of the copyright in the materials I create (except where I cite material created by others, for which they own the copyright).

You may take notes and make copies of course materials for your own use. You may also share those materials with other students who are enrolled in this course. You may not reproduce, distribute or display (post/upload) lecture notes or recordings or course materials in any other way — whether or not a fee is charged — without my express written consent.

You also may not allow others to do so. If you do so, you may be subject to student conduct proceedings under the UC Santa Cruz Student Code of Conduct, Section 102.23. Similarly, you own the copyright for your original papers and exam essays. If I am interested in posting your answers or papers on the course web site, I will ask for your written permission.”

Disability Resource Center

Special Accommodations:

UC Santa Cruz is committed to creating an academic environment that supports its diverse student body. If you are a student with a disability who requires accommodations to achieve equal access in this course, please submit your Accommodation Authorization Letter from the [Disability Resource Center \(DRC\)](#) to me, **preferably within the first two weeks of the quarter.**

- DRC Accommodation Letters are submitted privately by the DRC via the [Accommodate](#) system.
 - Access to Lecture Capture recordings is a **Service** that your DRC coordinator (not your instructor) may supply to you.
- At that time, I would also like us to discuss ways we can ensure your full participation in the course. I encourage all students who may benefit from learning more about DRC services to contact DRC by phone at [831-459-2089](#) or by email at drc@ucsc.edu.

A Word to the Unwise

- This is a tough class for some students since it involves a combination of theory and practice.
- Students who regularly attend Lectures and Labs often do well; students who don't regularly attend often do poorly.
 - We don't take attendance; you're responsible for your own choices.
- After the course ends, your course grade will be determined by your scores on Exams and Gradiance, as described on the “Course Evaluation” and “Grading” Slides.
 - You won't be able to do any additional work to improve your grade.
- Grades are based on what you've learned, so I hope that you'll all attend Lectures, Lab Sections, Office Hours and Tutoring, work hard and do well!

Lecture 1:

History and Introduction

Instructor: Shel Finkelstein

Reference:

*A First Course in Database Systems,
3rd edition, Chapter 1*

What is a Database?

- A *database* is a collection of data.
 - Typically describes the activities of one or more related organizations over time.
- Data Management (including database) is *extremely* important.
 - Essential to every business organization.
 - Enterprises
 - Employee data, sales transactions.
 - Web data
 - Amazon, Twitter/X, Facebook, Google, Apple, Microsoft, ...

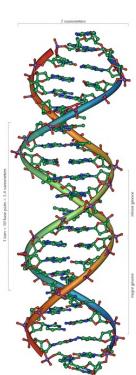
amazon.com



Microsoft

Data...is Everywhere

MOST ENTERPRISES TODAY GENERATE MORE DATA THAN THEY CAN PROCESS



social



finance



learning



health/
medical



transportation



government

entertainment



retail



...AND THE AMOUNT OF DATA IS GROWING AT 50% PER YEAR

MIT COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE LABORATORY

- according to IDC

Data Never Sleeps 9.0

(2021)

In every minute of every day, loads of data are being generated.

Just how much, you ask?

A lot happens in a minute ...

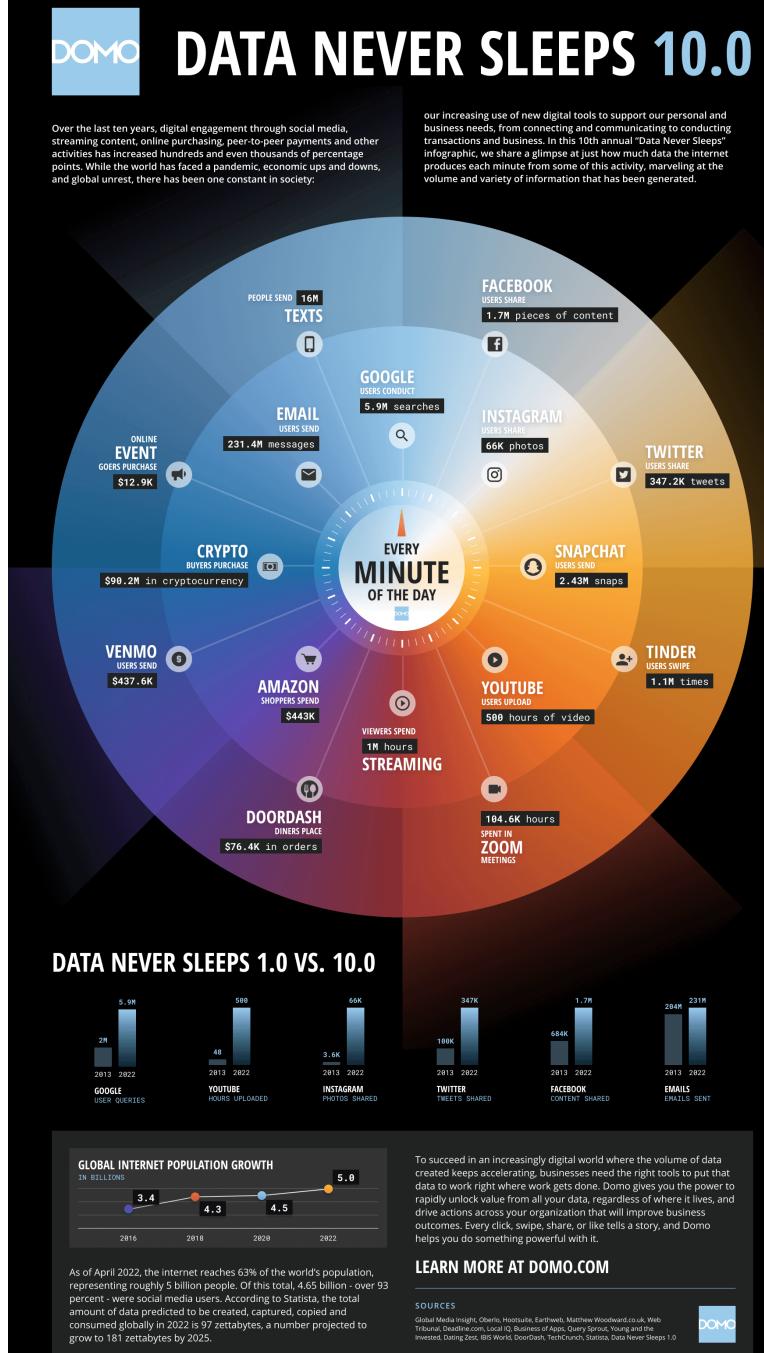


Data Never Sleeps 10.0 (2022)

In every minute of every day, loads of data are being generated.

Just how much, you ask?

A lot happens in a minute ...



Data Never Sleeps 11.0 (2023)

In every minute of every day, loads of data are being generated.

Just how much, you ask?

A lot happens in a minute ...



Data Never Sleeps 11.0

Domo has been keeping tabs on the world's data usage—in a minute—for over a decade now. What the numbers consistently show is that how we use data is always evolving—and that data isn't slowing down. We're also seeing some big changes. The rise of Artificial Intelligence (AI) is reshaping the way we communicate, work, and create. Digital payments continue to replace traditional transactions. Taylor Swift streams in countless headphones. And a rash of cybercrime grows alongside these digital experiences.

In Domo's 11th edition of Data Never Sleeps, we take the pulse of our digital age, where every click, swipe, and stream fuels an ever-expanding digital universe. These are not just numbers; they are the heartbeat of a world where data reigns supreme.



The world's internet population continues to grow significantly year-over-year. As of November 2023, the internet represents 5.2 billion people—approximately 64.6% of the global population. According to Statista, the total amount of data predicted to be created, captured, copied, and consumed globally in 2023 is 120 zettabytes, a number projected to grow to 181 zettabytes by 2025.

As data grows and evolves, businesses need to grow and evolve, too. Domo helps you harness the power of data so you can change as quickly as the world changes and make data-driven decisions that set you apart from the crowd. Let Domo help you make sense of all the clicks, swipes, and shares so you can see the big picture that a lot of small decisions make.

**Global Internet Population Growth
(IN BILLIONS)**



Learn more at domo.com

SOURCES: EARTHWEB, DUSTIN STOUT, DEMANDSAGE, HOOTSUITE, BUSINESSINSIDER, DOORDASH, SOCIALBAIT, X/TWITTER.COM, GITNIX, INVIGATE, THINKIMPACT, SIIMA.ORG, STRUTA, P/NWSIRE, NETSCOUT



Can You Answer These Questions?

Not a homework; will answer in class later this week, but CSE180 students should find these questions easy.

0-If set S is {1,3,5,7} and set T is {2,3,5,7}, what are S UNION T and S INTERSECT T?

1-If set A is {1,2,3} and set B is {u,v,w,x,y}, how many ways can you pick pairs of items, with the first from A and the second from B?

2-If you have a set of employees (with names and salaries) where John makes 10K, George makes 20K, Ringo makes 30K and Paul makes 40K, what are the names of the employee(s) who make less than the average salary?

3-Can there ever be an employee who makes more than every employee? If so, give an example. If not, explain why not.

4-Write the truth-table for p AND q, where p can be TRUE or FALSE and q can be TRUE or FALSE.

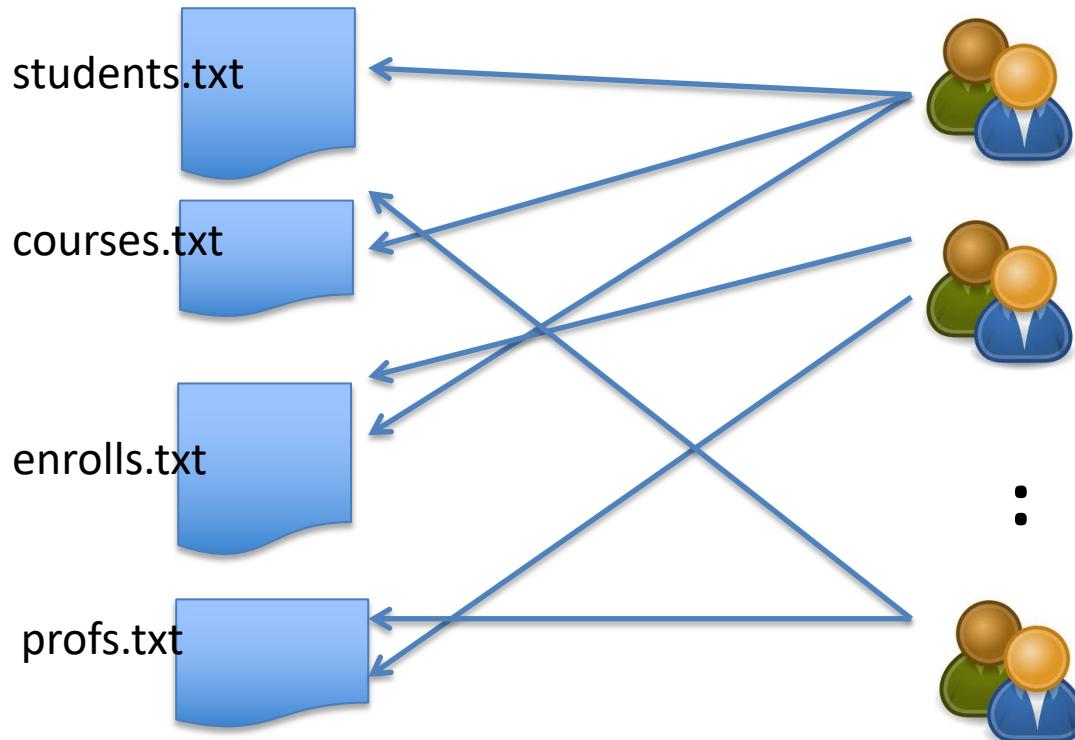
What is a Database Management System?

- A *database management system (DBMS)* is a software system designed to assist in creating, storing, accessing, and updating a database ... with enterprise qualities.
 - Transactions
 - High/continuous availability
 - Performance
 - Response time/latency
 - Throughput
 - Scalability
 - Security
 - Authorization, access control, etc.

Why Can't we Just Use a File System to Manage our Data?

- Suppose a company has a large database.
 - Data needs to be accessed *frequently* and *concurrently*.
 - Different queries need to be *posed easily* and answered *quickly*.
 - Updates to data by different users need to be managed and applied *consistently*.
 - New data needs to be *ingested speedily*.
 - Data must “*always*” be *accessible*.
 - Access to data need to be *restricted* to users *authorized* to see and change that data.
- Another answer, to be discussed later:
 - Sometimes, file systems are used to manage data
 - ... and **much more data** is kept in file systems than in DBs!

Simplified version of MyUCSC Campus Portal on a file system



Did Ann get an A for
CSE180 in Fall 2023?
Change Bob's grade to B.

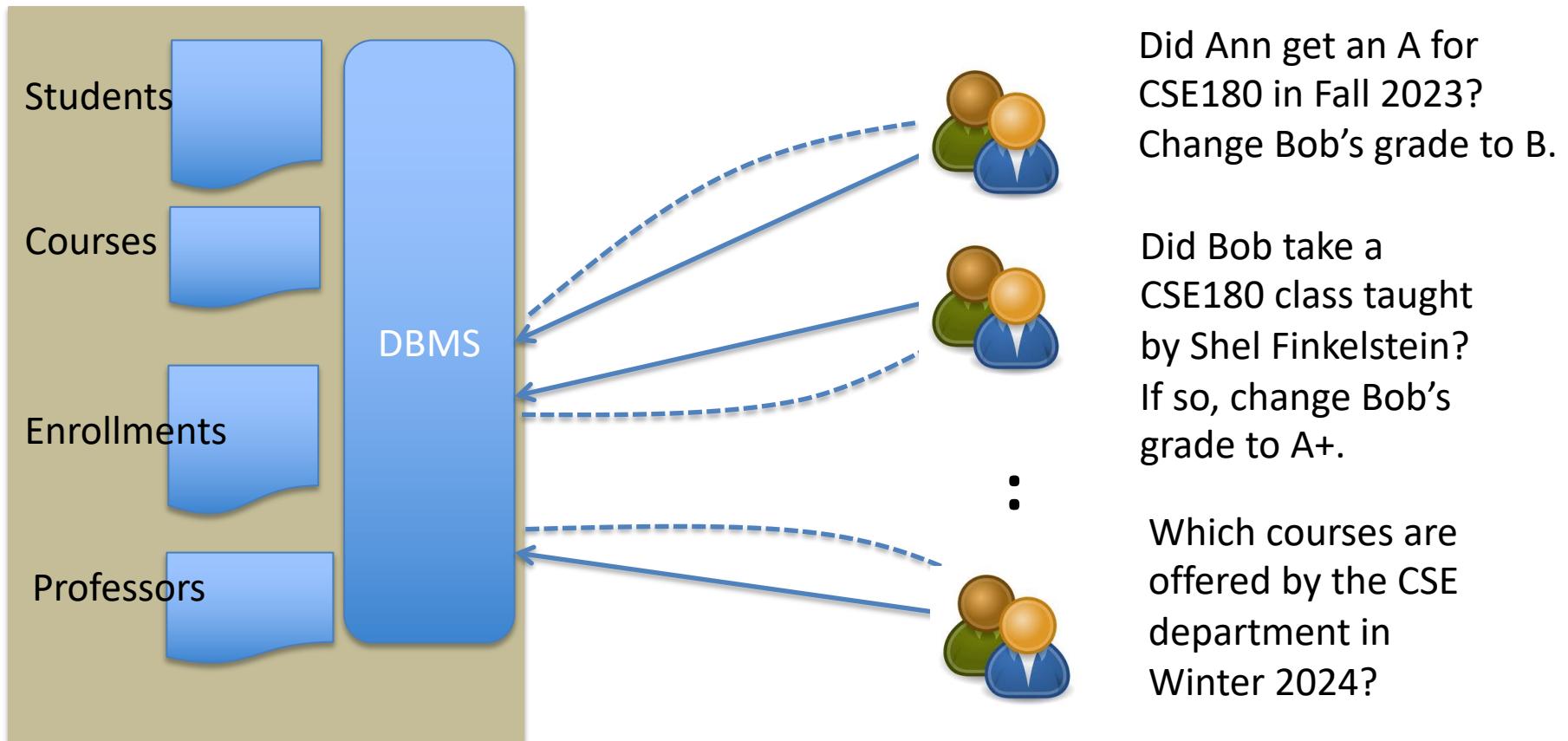
Did Bob take a
CSE180 class taught
by Shel Finkelstein?
If so, change Bob's
grade to A+.

Which courses are
offered by the CSE
department in
Winter 2024?

Key Characteristics of a DBMS

- Data Model
 - Provides an abstraction of the underlying data.
- High-level language for manipulating data
 - For defining, updating, retrieving and processing data.
- Transaction Processing
 - Concurrent access and updates, crash recovery.
- Access control
 - Limit access of certain data to certain users.

Key Characteristics of a DBMS



Advantages of a DBMS

- Users only need to understand the data model and high-level language for manipulating data.
 - Users focuses on *what* data is to be accessed and not *how* data is accessed.
 - Users are not aware of how data is actually stored or laid out on disks.
- Illusion that they are the only users of the DBMS.
- Data integrity is not compromised by system failures.
 - Withdraw from Checking Account: $\text{balance} = \text{balance} - 500$
 - ... and deposit to Savings Account: $\text{balance} = \text{balance} + 500$
 - System crashes after Withdrawal but before Deposit
 - Did you just lose 500?

Advantages of a DBMS (cont'd)

- Queries are automatically optimized for efficiency.
- Integrity of data is automatically enforced. For example:
 - Student ID is unique in any database instance.
 - age < 200
 - Courses in which students are enrolled must exist.
- Many different applications use data.
 - Fast development of new applications and change of existing applications.
- Well-understood database administration capabilities.
 - Utilities, or automated administration capabilities

Transactions have the ACID properties (to fill in)

- A(tomicity)
- C(onsistency)
- I(solation)
- D(urability)

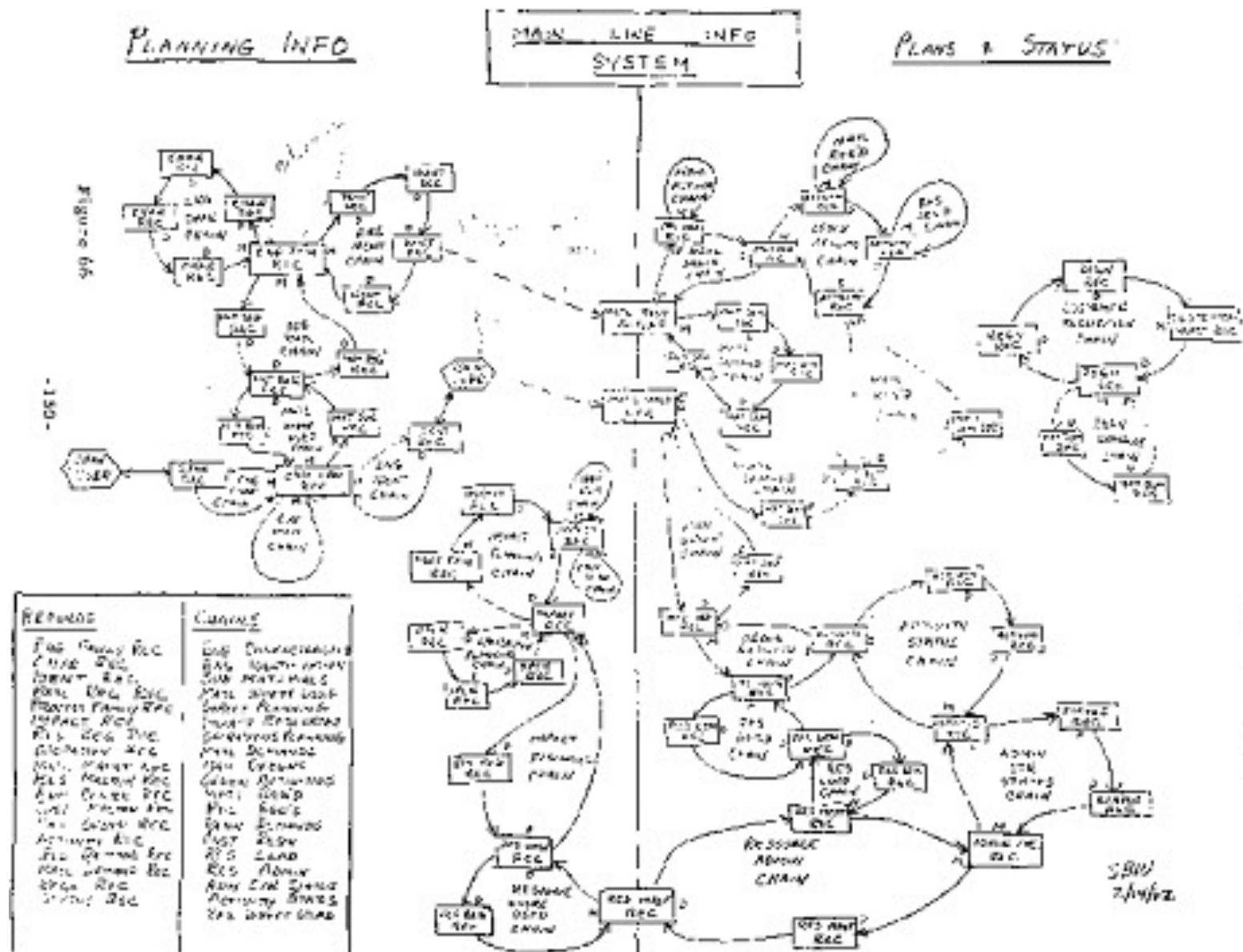
But What is a **Relational** Database Management System (RDBMS)?

- Network Model
- Hierarchical Model
- Relational Model

Network Data Model

- 1960s
 - First general purpose DBMS was built.
 - Integrated data store (IDS)
 - by Charles Bachman of General Electric.
 - **Network Data Model**
 - The computer navigates through a space of data records connected by pointers. A graph-based data structure.
 - A user needs to formulate the process of navigating through records and pointers to compute an answer for a query.
 - 1973 Turing award lecture.
 - “The Programmer as Navigator”

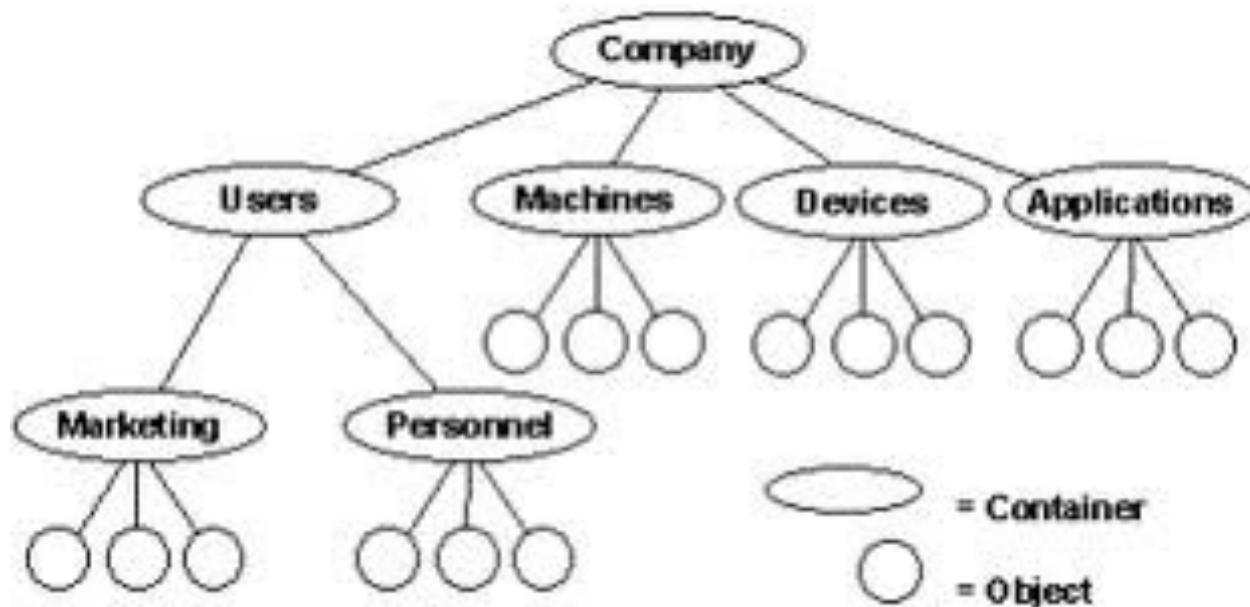




Source: Fifty years of databases <http://wp.sigmod.org/?p=688>

Hierarchical Data Model

- Also in the 1960s:
 - Hierarchical Data Model proposed IBM (IMS product)
 - A tree-based data structure.



Relational Data Model



- 1970s
 - The beginning of *relational* database management systems.
 - Edgar (Ted) F. Codd at the IBM San Jose Research Laboratory (now called IBM Almaden Research Center) published a seminal paper:

“A relational model for data for large shared data banks”
Communications of the ACM, 1970.

Ted Codd's Relational Model

- Advocates a radically different data model, called the *relational* data model.
 - All data must be stored in flat, table-like relations.
 - No pointers, no hierarchy!
 - Two database query languages:
 - Relational Algebra and Relational Calculus

Employees

EmpNo	FirstName	LastName	Department
100	Sally	Baker	10-L
101	Jack	Douglas	10-L
102	Sarah	Schultz	20-B
103	David	Drachmeier	20-B

Equipment

SerialNum	Type	UserEmpNo
3009734-4	Computer	100
3-23-283742	Monitor	100
2-22-723423	Monitor	101
232342	Printer	100

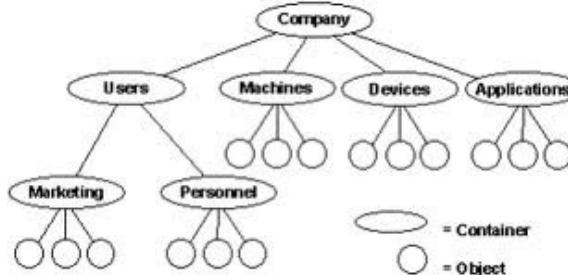
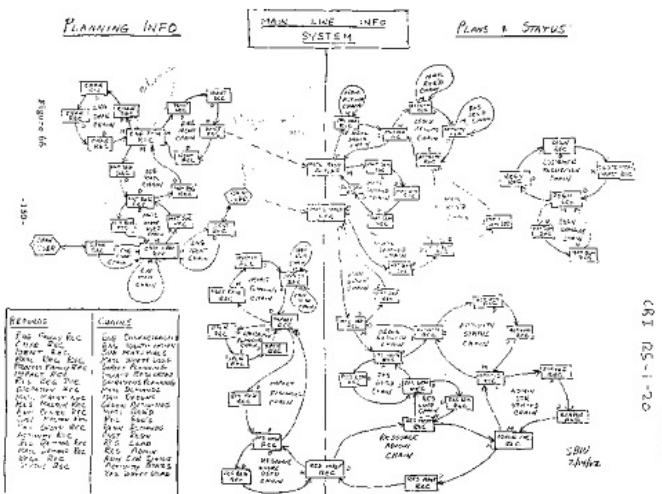
Some Relational Projects and Products

- System R project started at IBM Research in 1974.
 - System R became today's DB2.
 - 1981 Turing Award to Edgar F. Codd, "Relational Database: A Practical Foundation for Productivity".
 - 1998 Turing Award to Jim Gray for Transaction Processing.
- Michael Stonebraker and Eugene Wong at UC Berkeley started INGRES project based on Codd's papers.
 - Relational Technology Inc. became company, INGRES.
 - Later POSTGRES project led to company, Illustra Information Technologies and became open-source PostgreSQL.
 - 2014 Turing Award to Mike Stonebraker.
- Larry Ellison founded Relational Storage Inc., which became Oracle. First Oracle RDBMS was released in 1979 (and was called v2).

RDBMS Today

- Lots of relational database management systems
 - http://en.wikipedia.org/wiki/List_of_relational_database_management_systems
- Examples of open-source RDBMS:
 - MySQL, PostgreSQL
- Examples of proprietary RDBMS:
 - Oracle, IBM DB2, Microsoft SQL Server, SAP HANA

1960s: Network data model and Hierarchical data model



1970s: Relational data model

EmpNo	First Name	Last Name	Dept. Num	Serial Num	Type	User EmpNo
100	Sally	Baker	10-L	3009734-4	Computer	100
101	Jack	Douglas	10-L	3-23-283742	Monitor	100
102	Sarah	Schultz	20-B	2-22-723423	Monitor	100
103	David	Drachmeier	20-B	232342	Printer	100

```

{
  "photos": [
    {
      "page": 1,
      "pages": 94276,
      "perpage": 15,
      "total": "1414129",
      "photo": [
        {
          "id": "3891667770",
          "owner": "354681331200005",
          "secret": "4479beef9",
          "server": "2451",
          "farm": 3,
          "title": "Mexican train dominoes with Brian and Michelle",
          "ispublic": 1,
          "isfriend": 0,
          "isfamily": 0
        },
        {
          "id": "3891661852",
          "owner": "106483010MB7",
          "secret": "79de502257",
          "server": "2590",
          "farm": 3,
          "title": "Peaches",
          "ispublic": 1,
          "isfriend": 0,
          "isfamily": 0
        }
      ]
    }
  ]
}
  
```

2010s: JSON

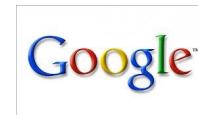
```

<Books>
  <Book ISBN="0553212419">
    <title>Sherlock Holmes: Complete Novels...</title>
    <author>Sir Arthur Conan Doyle</author>
  </Book>
  <Book ISBN="0743273567">
    <title>The Great Gatsby</title>
    <author>F. Scott Fitzgerald</author>
  </Book>
  <Book ISBN="0684826976">
    <title>Undaunted Courage</title>
    <author>Stephen E. Ambrose</author>
  </Book>
  <Book ISBN="0743203178">
    <title>Nothing Like It In the World</title>
    <author>Stephen E. Ambrose</author>
  </Book>
</Books>
  
```

1990s: XML (Semi-structured data model)

Lots of Data Resides Outside Databases

- Before the Web
 - Business data typically resided in databases on Premises
 - Sensitive enterprise information, such as bank accounts, employee data, and sale transactions was typically in databases.
- Today
 - Data resides in both databases and in less structured formats in files
 - Consumer data is a huge business: Advertising, Influencing
 - Webpage clicks, location data, social media posts, mail/text, sensor data, photos/images, voice, ...
 - Data entry and data access is via the Web
 - More data is stored in the Cloud, rather than on Premises.
 - “Data Lakes”, rather than separate siloed data for each business unit of a company.



Today

- NOSQL (“Not Only SQL”) systems—a few examples
 - Map/Reduce (Apache Hadoop, Apache Spark)
 - Column stores (Apache HBase, Apache Cassandra, Google BigTable, Amazon DynamoDB)
 - Key/Value stores (RocksDB, Couchbase)
 - Graph databases (Neo4J, OpenLink Virtuoso)
 - Document databases (MongoDB)
 - Data Lakes (DataBricks Delta Lake)
- Simplicity of design, great scale-out
- May compromise Consistency in favor of Scalability

**So ... what is a Relational Database
Management System (RDBMS)?**

RDBMS Components

- Upper half of system: Similar to Language Processing
 - Parsing SQL
 - Query Optimization
 - Plan Generation
 - Security
- Lower half of system: Similar to Operating Systems
 - Storage
 - Plan Execution
 - Concurrency Control and Scheduling
 - Logging and Recovery

DBMS Structure (CSE 181)

