

CMPS 182: Database Systems I

Spring 2019

Instructor: Sheldon (Shel) Finkelstein

Textbook:

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

Science and Engineering Library has textbook on Reserve.
2cd edition is also available, but this course uses 3rd edition.

Course Syllabus and Chapter 1

CMPS 182, Spring 2019 Course Information

Classes: MWF 12:00PM-1:05PM, BE-152

Instructor: Sheldon (Shel) Finkelstein, shel@ucsc.edu

Office Hours: Fri, 2:00– 3:00 PM, E2-249B

or by appointment

Teaching Assistants:

Pengyang Zhou, pzhou15@ucsc.edu

Nitesh Kumar Singh, nsingh32@ucsc.edu

Labs Sections (all in BE-105)

01A: Fri, 6:00-7:35 PM (Nitesh)

01B: Wed, 4:00-5:35 PM (Pengyang)

01C: Thu, 8:00-9:35 AM (Nitesh)

Course Page (including Lecture Slides):

<https://piazza.com/ucsc/spring2019/cmps182/home>

Piazza Signup: <https://piazza.com/ucsc/spring2019/cmps182/>

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 2014, mainly Database courses

Teaching Assistants

Pengyang Zhou
pzhou15@ucsc.edu

Lab Section, in BE-105:
01B: Wed, 4:00-5:35 PM

Office Hours:
Wed, 6:00 – 7:00 PM, **BE-119**
(or by appointment)



Teaching Assistants

Nitesh Kumar Singh
nsingh32@ucsc.edu

Lab Sections in BE-105

01A: Fri, 6:00 - 7:35 PM

01C: Thurs, 8:00 - 9:35 AM

Office Hours:

Thu, 4:00 - 5:00 PM, **BE-119**
(or by appointment)



Course Description/Syllabus

CMPS 182 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: XML, JSON, etc.
- Cloud Databases (time permitting)
- Big Data and NOSQL (time permitting)

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

- 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	4/1	1
The Relational Data Model	2	4/3 – 4/4	2.1, 2.2
SQL: DDL, DML; defining relations & constraints, and writing queries	12	4/8 – 5/3	2.3, 2.5, 6.1-6.5, 7.1, 7.2
Relational Algebra and Query Execution	3	5/6, 5/10, 5/13	2.4, 2.5
Midterm	1	Wed 5/8	-
Database Application Development	2	5/15 – 5/17	9.1, 9.2, 9.6
Schema Refinement and Normal Forms	4	5/20 – 5/29	3.1-3.5
OLAP	1	5/31	10.6, 5.2.7
Semi-Structured Data Models: XML, JSON, etc.	2	6/3 – 6/5	11.1-11.3, 12.1-12.2
Big Data, NOSQL (time permitting)	0.5	6/7	
Review (time permitting)	0.5	6/7	
	'		
Final, in classroom		Mon 6/10, 4:00 - 7:00 PM	
(Holidays—no class)		Mon 5/27	

Course Evaluation

- Gradiance Homeworks 10%
- Lab Assignments 20%
- Midterm 30%
- Final Exam 40%

General Information for CMPS 182 appears in Piazza on Course Information page; see also the General Information file that's under
Resources → General Resources

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 be on the slides.
- Lecture slides will be posted on Piazza under Resources-->Lectures
 - Slides may be modified after initial posting.
 - Date shown on Piazza is date that a Lecture presentation began.
- You should learn the concepts, not just learn what's presented in the Lectures.
- Students who don't attend Lectures and Labs tend to do poorly in this course. It's your decision.

Homeworks: Labs

- 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).
- We'll also use Canvas for grading.
 - Login to Canvas at <https://canvas.ucsc.edu> using your CruzID and Gold password. CMPS 182 should be one of the classes available.
 - Info on Canvas is at <http://its.ucsc.edu/canvas/index.html>
 - If you disagree with a grade, check with your TA, and then with me. **Never contact the Reader** who graded your Lab Assignment.

Tentative Lab Assignment Dates

- **Lab Sections** are a required part of course, but we don't take attendance.
 - There will be Lab Sections during the first week of classes, introducing you to Gradiance and PostgreSQL.
- **Lab1** (which is much easier than the rest of the Lab Assignments)
 - Assigned Friday, April 5
 - Due Sunday April 14, 11:59pm
- **Lab2**
 - Assigned Monday April 15
 - Due Sunday April 28, 11:59pm
- **Lab3**
 - Assigned Monday April 29
 - Due Sunday May 19, 11:59pm (extra week--Midterm is **Wednesday, May 8**)
- **Lab4**
 - Assigned Monday, May 20
 - Due Sunday June 2, 11:59pm
- **Last day of classes:** Friday, June 7
 - During last week of classes, Lab Sections will be held, discussing answers to Lab4, as well as addressing questions about overall course before Final Exam.

Other Homeworks

- Non-lab homework assignments will be assigned via Gradiance, which has automated grading; see:
<http://www.gradiance.com/pub/stud-guide.html>
 - For CMPS182, go to <http://www.gradiance.com/services> to enroll in this class, using class code **DD2C8570**
 - Gradiance uses HTTP, rather than HTTPS, so be sure to use a password that you don't use for any other account!
- There will also be ungraded Practice Homeworks that you don't hand in.
 - You're responsible for those Practice Homeworks, even though we won't discuss most of them in class.

Exams

- **Midterm** on Wednesday, May 8 (65 minutes long) covers first half of term, potentially including material through Monday, May 7.
- **Final Exam** on Monday June 10, 4:00 – 7:00 PM (3 hours) is comprehensive, covering material from entire term, with somewhat greater emphasis on second half of term.
- For both the Midterm and the Final Exam, you may bring in an 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.
 - Sheets should have your name on top right. They will be collected at the end of the exam.
 - You **must** show UCSC ID when you hand in your exams.
- **No** Make-up Exams. **No** Early/Late Exams.

Learning Support Center and Modified Supplemental Instruction

- UCSC Learning Support Services: <http://lss.ucsc.edu/>
 - There will be a CMPS 182 Learning Assistant this term, who will introduce himself in class soon.
 - After introduction, students may sign-up for tutoring at the Slug Success website: <https://sserc.ucsc.edu/slug-success>
 - Slots tend to fill up, and are usually not available if you try to sign-up before Exams or due dates. If you're interested in getting tutoring, please consider signing up early.

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](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'll do it.***
- You are allowed to ask for some help when working on assignments, provided that you acknowledge the help that you received on the work that you turn in.
- Points will be deducted if it appears that labor has been divided among multiple students. There will be no penalty for small amounts of acknowledged assistance.
- If you have any questions about these rules, please discuss them with the instructor **immediately**.

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 privately during my office hours or by appointment, preferably within the first two weeks of the quarter.

At this 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](tel:831-459-2089), or by email at drc@ucsc.edu.

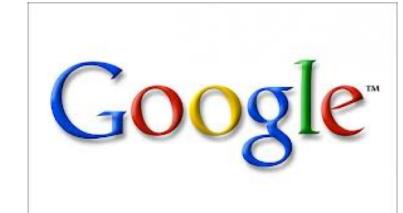
Student responsibilities are as follows:

1. Students contact the DRC to determine their eligibility for accommodations. Students will request and receive their accommodation letter. This letter is provided to the instructor. This is official notice of a request for accommodation.
2. Students then notify their instructor during office hours or after class of their accommodations, and provide their instructor with their Accommodation Letters, **preferably during the first two weeks of the term.**
3. Students will manage their own disclosure of disability status, which will be maintained confidentially according to UC Santa Cruz data practices.

What is a Database?

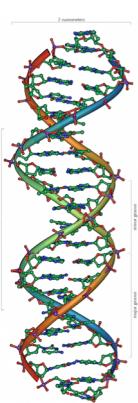
- A *database* is a collection of data.
 - Typically describes the activities of one or more related organizations over time.
- Data Management (including database) is *extremely* important.
 - Essential to every business organization.
 - Enterprises
 - Employee data, sales transactions.
 - Web data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

 amazon.com



Data...is Everywhere

MOST ENTERPRISES TODAY GENERATE MORE DATA THAN THEY CAN PROCESS



social



finance



health/
medical



learning



transportation

science



government

retail



entertainment



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

MIT COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE LABORATORY

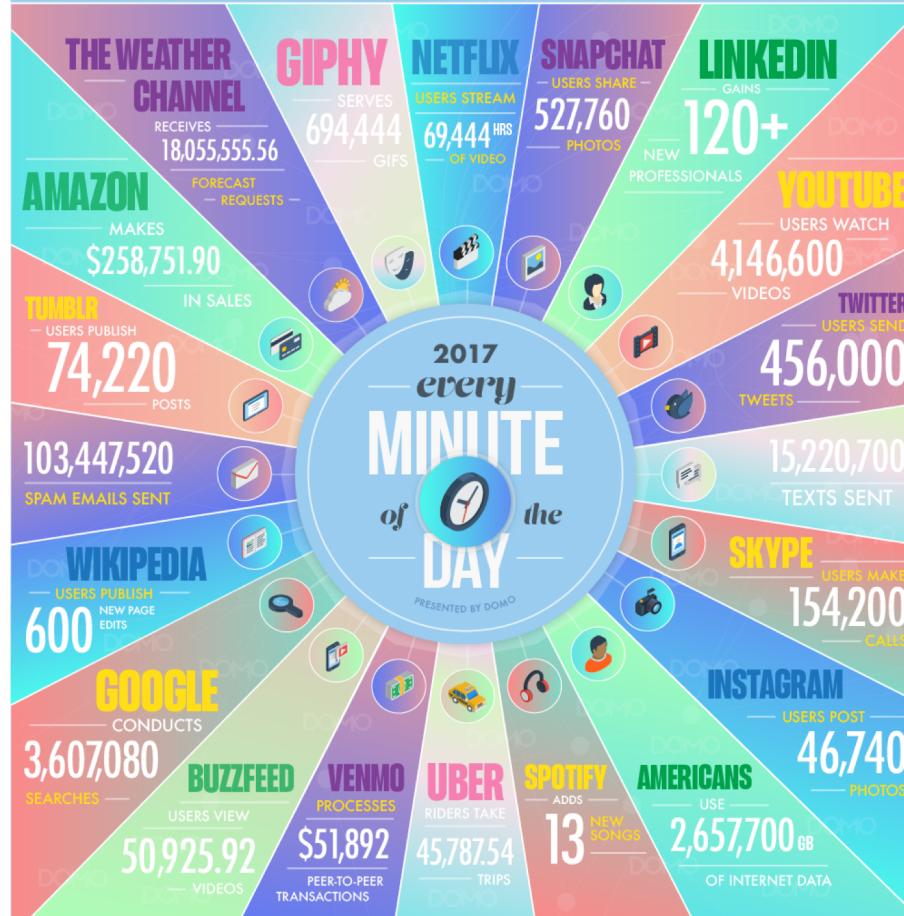
- according to IDC

DOMO

DATA NEVER SLEEPS 5.0

How much data is generated *every minute*?

90% of all data today was created in the last two years—that's 2.5 quintillion bytes of data per day. In our 5th edition of Data Never Sleeps, we bring you the latest stats on just how much data is being created in the digital sphere—and the numbers are staggering.



The world internet population has grown 7.5% from 2016 and now represents 3.7 billion people.



GLOBAL INTERNET POPULATION GROWTH 2012-2017
(IN BILLIONS)

With each click, swipe, share, and like, businesses are using data to make decisions about the future. Domo gives everyone in your business real-time access to data from virtually any data source in a single platform for smarter decision-making at any moment.

Learn more at domo.com

SOURCES: EXPANDEDRAMBLINGS.COM, WEARESOCIAL.COM, WIKIPEDIA, FORBES, ADWEEK.COM, FORTUNE.COM, BLOOMBERG.COM, ONEREACH.COM, IBM, BUZZFEED, INTERNET LIVE STATS, INTERNET WORLD STATS, BBC

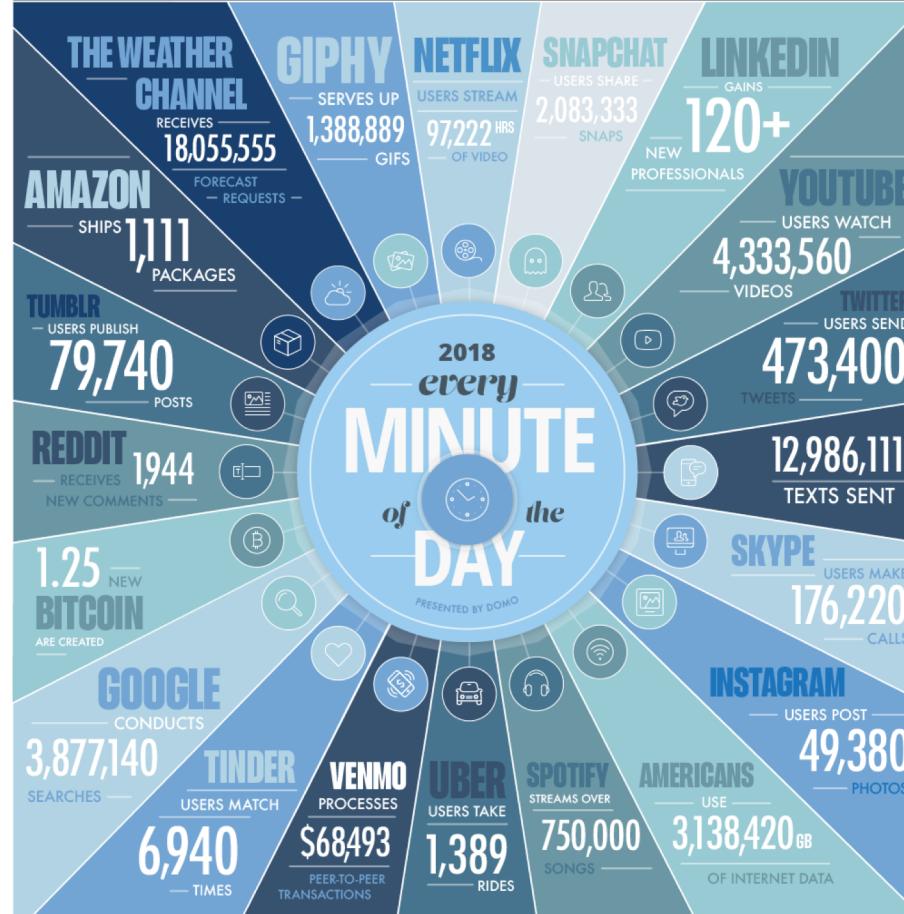


DOMO

DATA NEVER SLEEPS 6.0

How much data is generated *every minute*?

There's no way around it: big data just keeps getting bigger. The numbers are staggering, but they're not slowing down. By 2020, it's estimated that for every person on earth, 1.7 MB of data will be created every second. In our 6th edition of Data Never Sleeps, we once again take a look at how much data is being created all around us every single minute of the day—and we have a feeling things are just getting started.



The world's internet population is growing significantly year-over-year. In 2017, internet usage reached 47% of the world's population and now represents 3.8 billion people.



GLOBAL INTERNET POPULATION GROWTH 2012-2017
(IN BILLIONS)

The ability to make data-driven decisions is crucial to any business. With each click, swipe, share, and like, a world of valuable information is created. Domo puts the power to make those decisions right into the palm of your hand by connecting your data and your people at any moment, on any device, so they can make the kind of decisions that make an impact.

Learn more at domo.com

SOURCES: STATISTA, LINKEDIN, INTERNET LIVE STATS, EXPANDED RAMBLINGS, SLASH FILM, RIAA, BUSINESS OF APPS, INTERNATIONAL TELECOMMUNICATIONS UNION, INTERNATIONAL DATA CORPORATION



Can You Answer These Questions?

Not a homework; will answer in class later this week, but for CMPS182, you should understand how to figure out answers to most of them.

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.

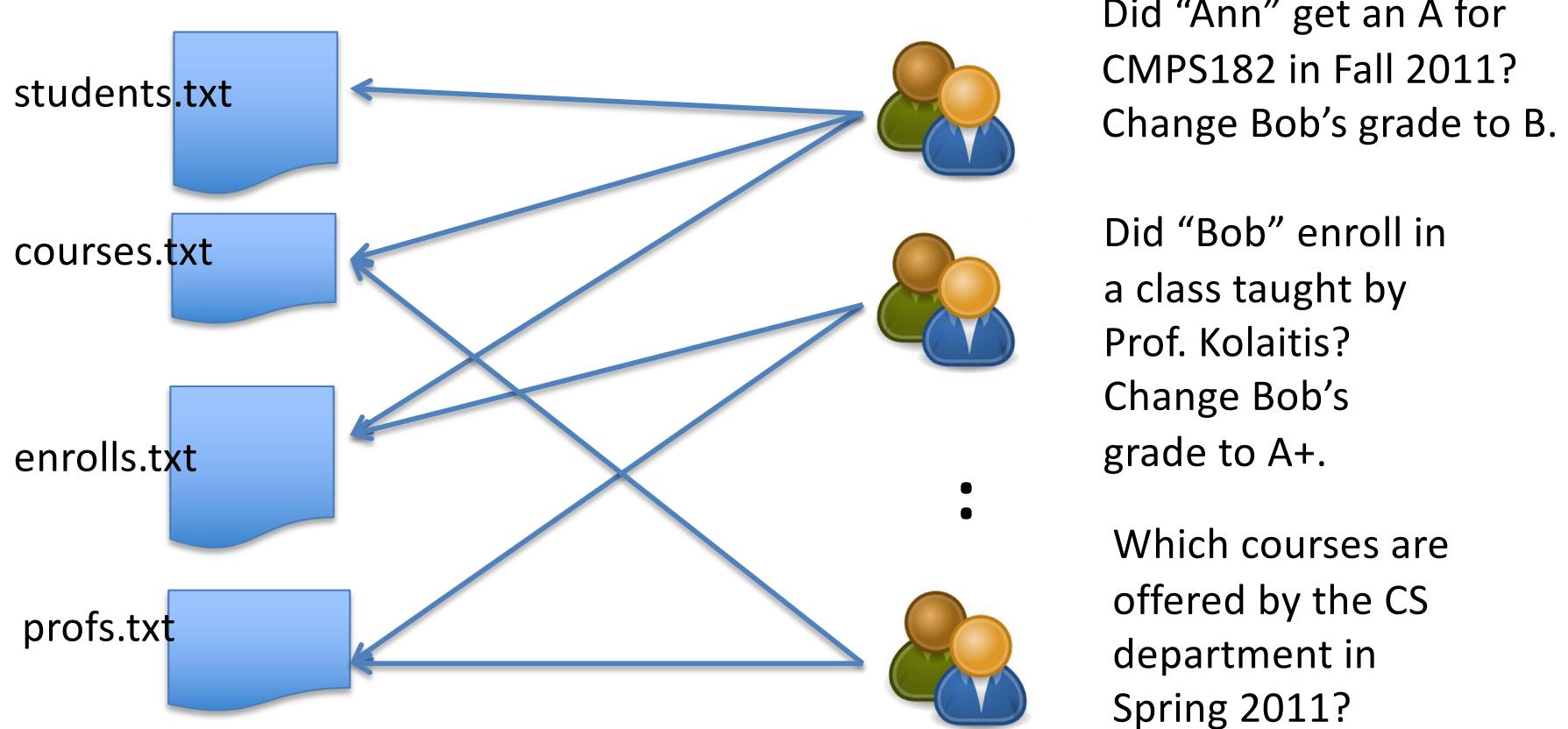
What is a Database Management System?

- 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 *posed easily* be answered *quickly*.
 - Updates to data by different users need to be managed and applied *consistently*.
 - Access to certain parts of the data by certain users need to be *restricted*.
- Another answer, to be discussed later
 - Sometimes we can use file system to manage data
 - ... and **much more data** is kept in file systems than in DBs!

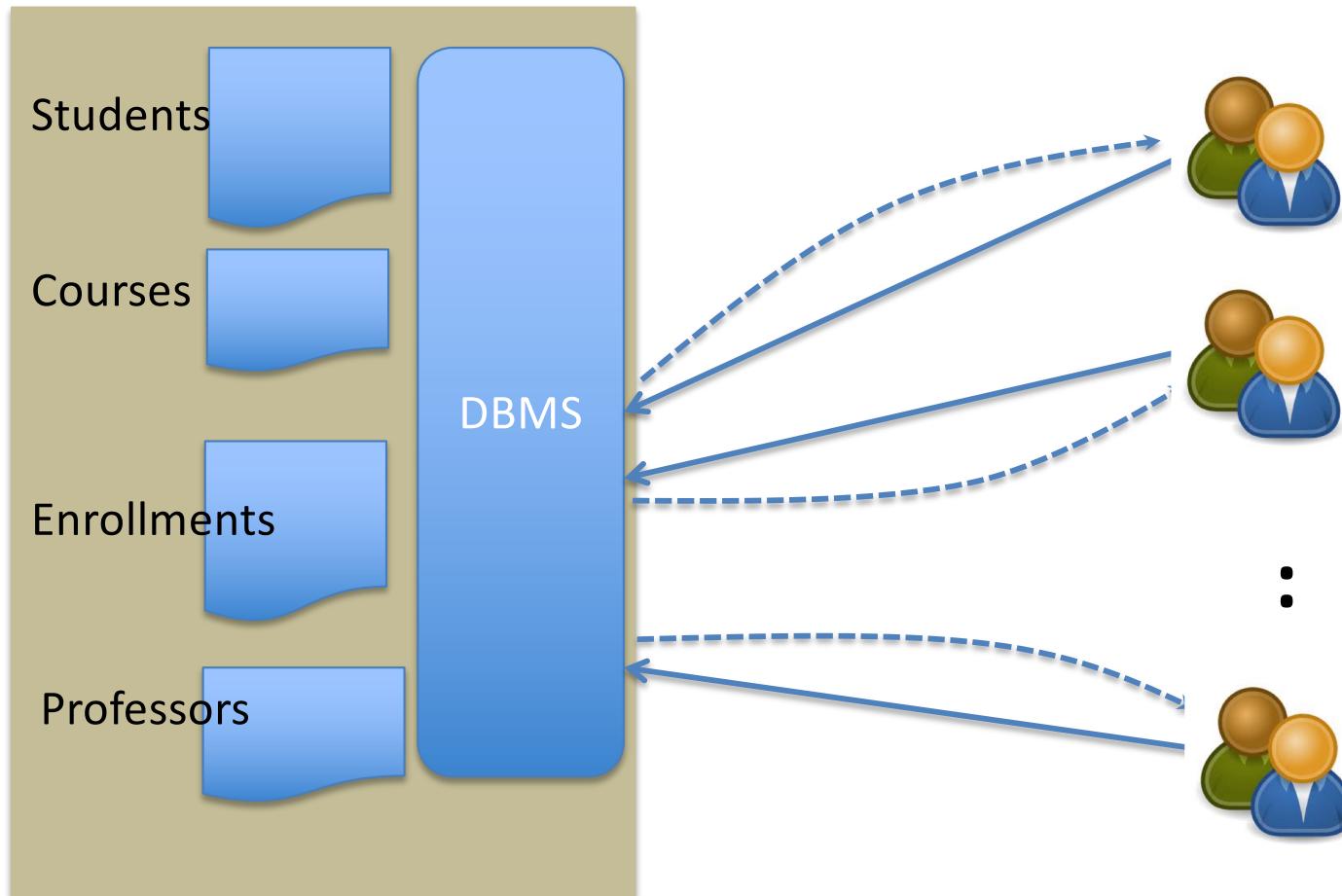
Simplified version of MyUCSC Campus Portal on a file system



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



Did “Ann” get an A for CMPS182 in Fall 2011?
Change Bob’s grade to B.

Did “Bob” enroll in a class taught by Prof. Kolaitis?
Change Bob’s grade to A+.

Which courses are offered by the CS department in Spring 2011?

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.
 - Deposit: $\text{balance} = \text{balance} + 500;$
 - In parallel, a withdrawal for your monthly car payment:
 $\text{balance} = \text{balance} - 300;$
 - System crashes... What is the balance?

Advantages of a DBMS (cont'd)

- Queries are automatically optimized for efficiency.
- Integrity of data is automatically enforced.
 - E.g., Employee id is unique, age < 200, courses that students are enrolled in must exist.
- Ease of data administration.
 - Well-developed user interfaces.
- Fast application development.
 - Available APIs and libraries.
- Data is managed centrally.
 - Costs are shared across applications.

Transactions have the ACID properties (to fill in)

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

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

- Network Model
- Hierarchical Model
- Relational Model

Network Data Model

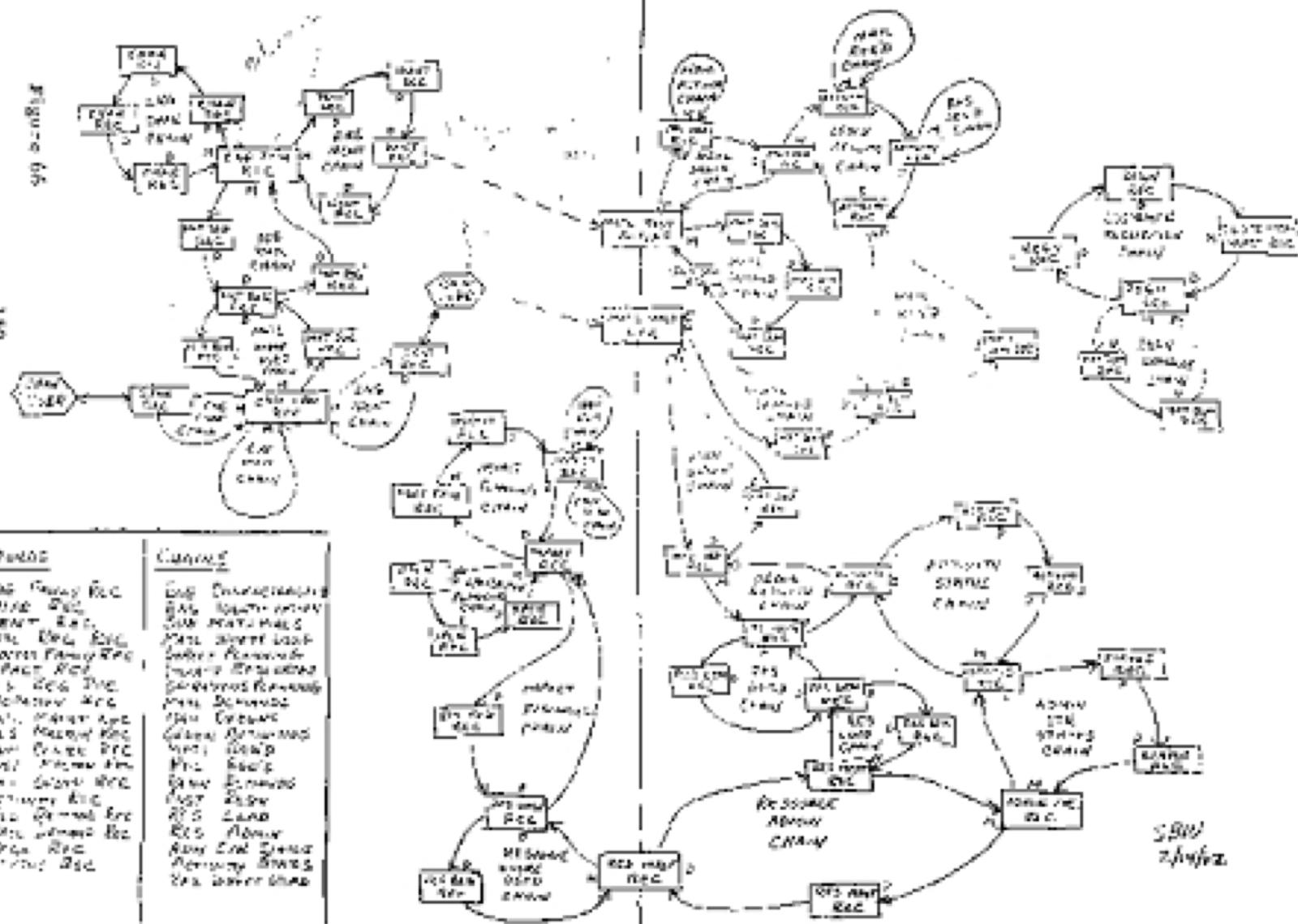
- 1960s
 - First general purpose DBMS was built.
 - Integrated data store (IDS)
 - by Charles Bachman of General Electric.
 - Network Data Model
 - The computer navigates through a space of data records connected by pointers. A graph-based data structure.
 - A user needs to formulate the process of navigating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



PLANNING INFO

MAIN LINE INFO
SYSTEM

Plans & Status

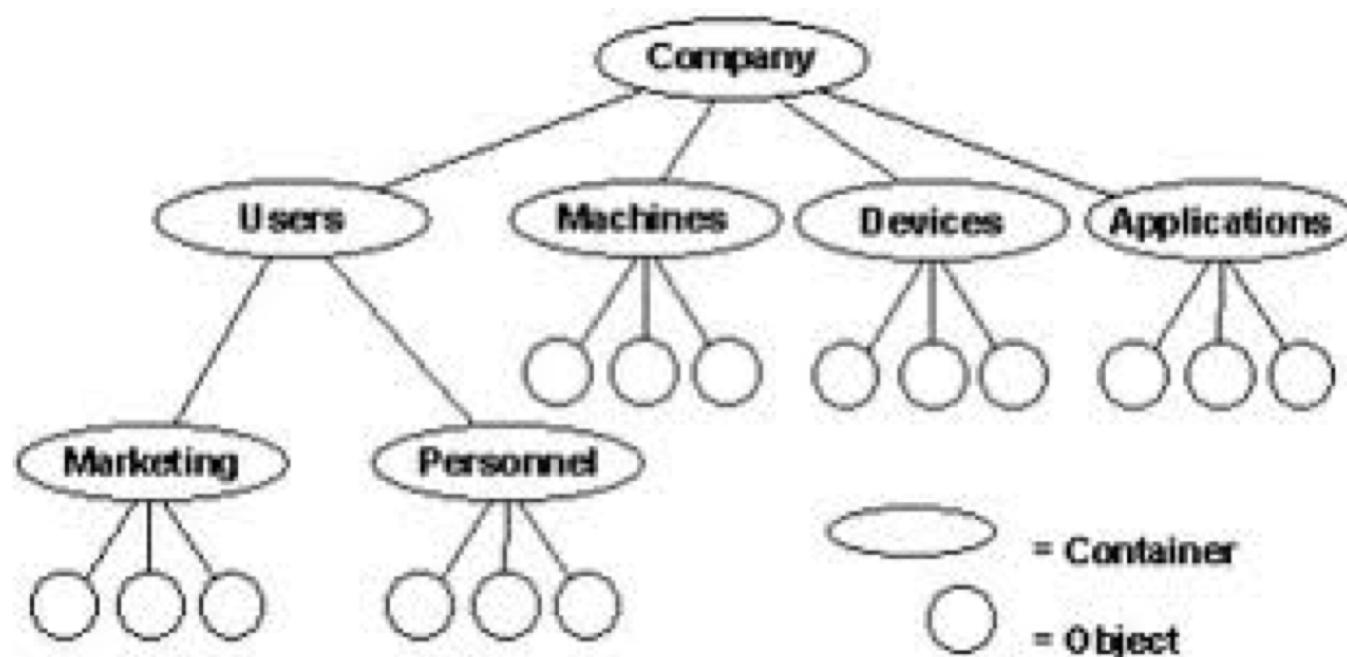


C.R.I. 125-1-26

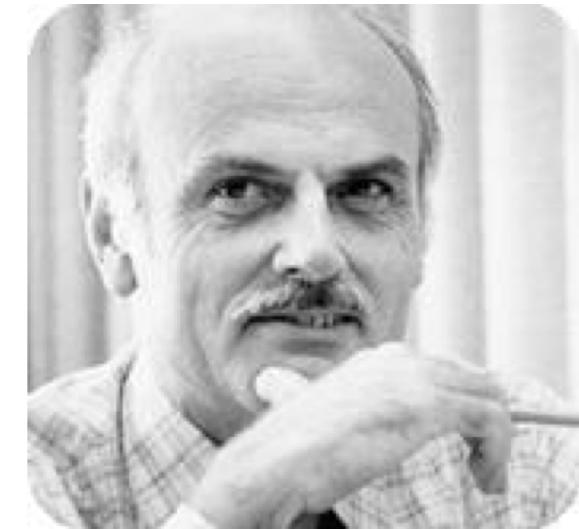
Source: Fifty years of databases <http://wp.sigmod.org/?p=688>

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 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 the 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

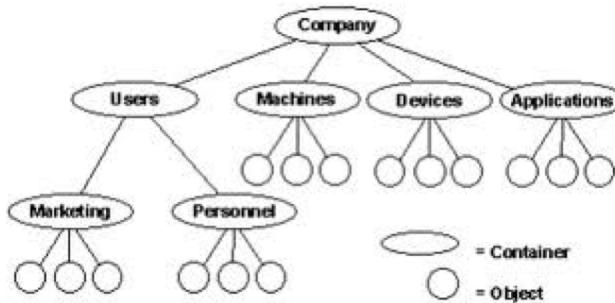
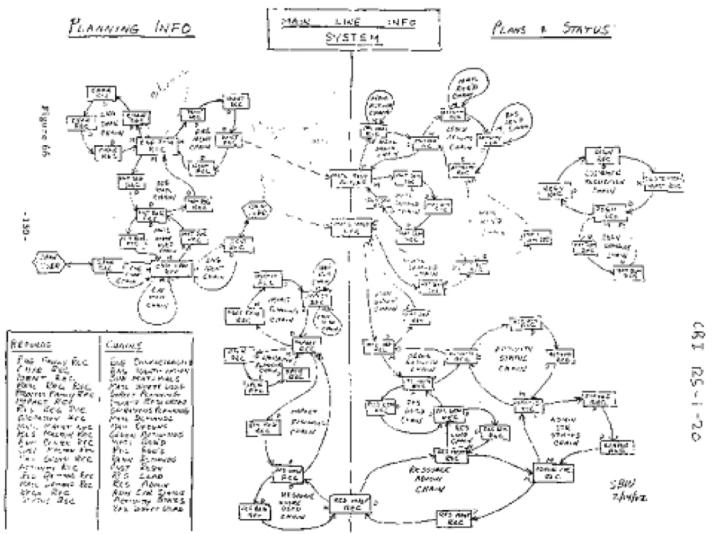
Some Relational Projects and Products

- 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": "3546813312000001",
          "secret": "4479bcbf9",
          "server": "2451",
          "farm": 3,
          "title": "Mexican train dominoes with Brian and Michelle",
          "ispublic": 1,
          "isfriend": 0,
          "isfamily": 0
        },
        {
          "id": "3891661852",
          "owner": "1864803010007",
          "secret": "7fde602257",
          "server": "2590",
          "farm": 3,
          "title": "Peaches",
          "ispublic": 1,
          "isfriend": 0,
          "isfamily": 0
        }
      ]
    }
}
```

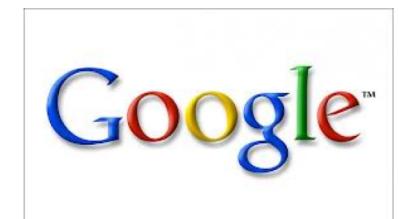
2010s: JSON

```
<Books>
  <Book ISBN="0553212419">
    <title>Sherlock Holmes: Complete Novels...
    <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)

Today, Data Also Resides Outside Databases

- Before the Web
 - Business data typically resided in databases on Premises
 - Typically sensitive enterprise information, such as bank accounts, employee data, sale transactions
- 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, analysis of social media and mail/text, sensor data, ...
 - Data entry and data access is via the Web
 - More and more data is stored in the Cloud, rather than on Premises



Today

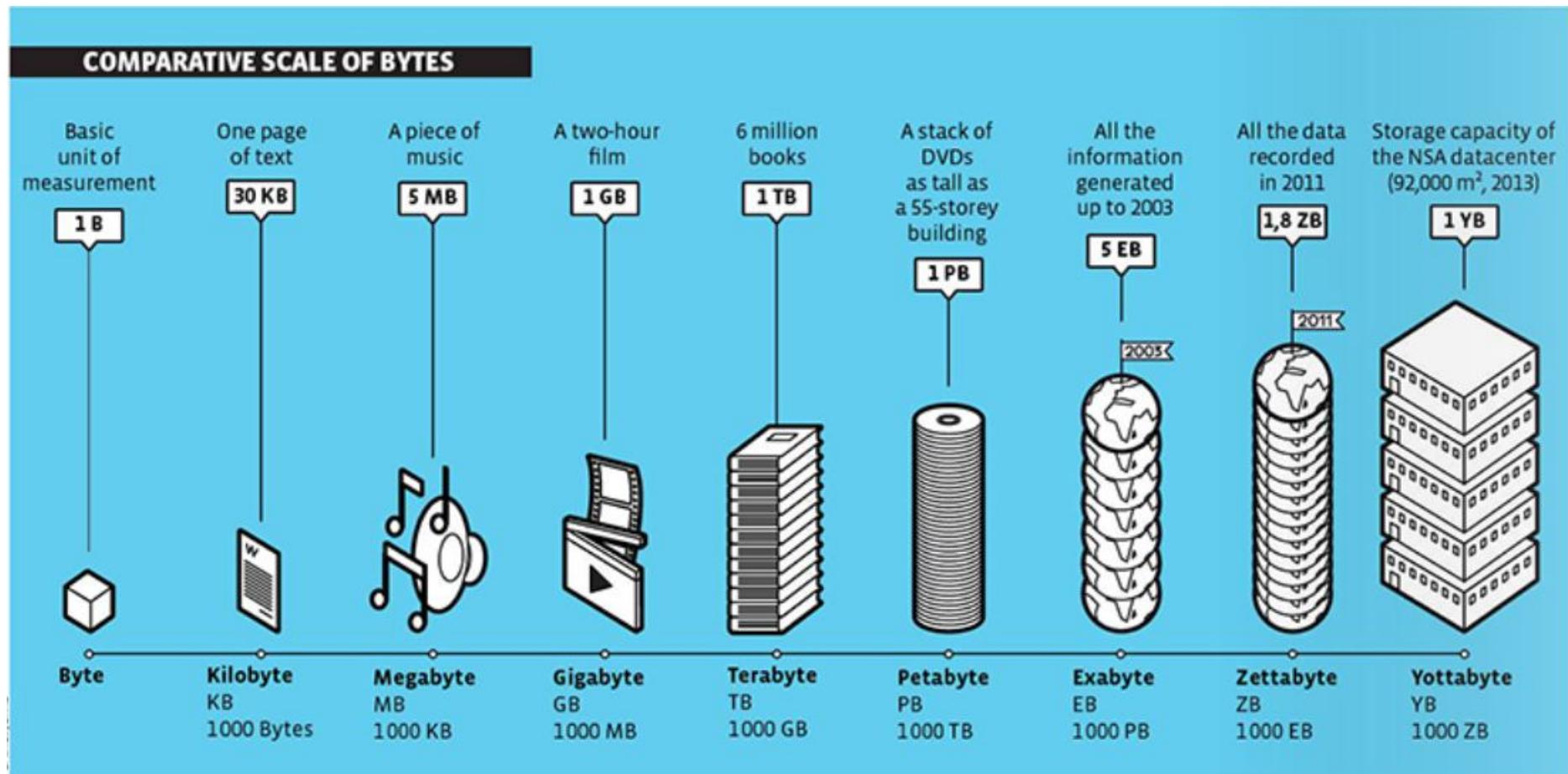
- NOSQL (“Not Only SQL”) systems
 - Map/Reduce (Hadoop, Spark)
 - Column store (HBase, Cassandra)
 - Graph databases (Neo4J, Virtuoso)
 - Document databases (MongoDB)
- Simplicity of design, easy scale-out
- Compromise consistency in favor of availability and partition tolerance
- Lack of full ACID support, use of low-level query languages, lack of standardized interfaces (**changing**)

From Bytes to Yottabytes

Multiples of bytes			V · T · E	
SI decimal prefixes		Binary usage	IEC binary prefixes	
Name (Symbol)	Value		Name (Symbol)	Value
kilobyte (kB)	10^3	2^{10}	kibibyte (KiB)	2^{10}
megabyte (MB)	10^6	2^{20}	mebibyte (MiB)	2^{20}
gigabyte (GB)	10^9	2^{30}	gibibyte (GiB)	2^{30}
terabyte (TB)	10^{12}	2^{40}	tebibyte (TiB)	2^{40}
petabyte (PB)	10^{15}	2^{50}	pebibyte (PiB)	2^{50}
exabyte (EB)	10^{18}	2^{60}	exbibyte (EiB)	2^{60}
zettabyte (ZB)	10^{21}	2^{70}	zebibyte (ZiB)	2^{70}
yottabyte (YB)	10^{24}	2^{80}	yobibyte (YiB)	2^{80}

See also: Multiples of bits · Orders of magnitude of data

Comparative Size of Bytes – credit to Testyotta



Big Data & Real Business Issues

Source: <https://www.waterfordtechnologies.com/big-data-interesting-facts/>, 2017

- According to estimates, the volume of business data worldwide, across all companies, doubles every 1.2 years.
- Poor data can cost businesses 20%–35% of their operating revenue.
- According to execs, the influx of data is putting a strain on IT infrastructure. 55 percent of respondents reporting a slowdown of IT systems and 47 percent citing data security problems, according to a global survey from Avanade.
- A quarter of decision-makers surveyed predict that data volumes in their companies will rise by more than 60 per cent by the end of 2014, with the average of all respondents anticipating a growth of no less than 42 per cent.
- 40% projected growth in global data generated per year vs. 5% growth in global IT spending.

2018:

- Data volume is set to grow 800% over the next 5 years and 80% of it will reside as Unstructured Data – Gartner
- IDC estimate that global Data will grow to 163 zettabytes (ZB) by 2025

Some Supplementary Reading Material

- Database management systems:
http://en.wikipedia.org/wiki/Database_management_system
- Fifty years of databases:
<http://wp.sigmod.org/?p=688>
- IDC White Paper (sponsored by Seagate):
“Data Age 2025: The Evolution of Data to Life-Critical”
<https://www.seagate.com/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.pdf>

Estimates that in 2025, the world will create and replicate 163 ZB of data, representing a tenfold increase from the amount of data created in 2016.

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 (CMPS 181)

