



CSCC43H: Introduction to Databases

Lecture 1

Wael Abualsaadat

Acknowledgment: these slides are partially based on Prof. Garcia-Molina & Prof. Ullman slides accompanying the course's textbook.



Course Info...

Managing Data Then...





Managing Data: a historical perspective

Name ^	Size	Type	Date Modified
clients.txt	8 KB	Text Document	5/7/2008 6:51 AM
employees.txt	8 KB	Text Document	5/7/2008 6:51 AM
expenses.txt	8 KB	Text Document	5/7/2008 6:51 AM
income.txt	8 KB	Text Document	5/7/2008 6:51 AM
suppliers.txt	8 KB	Text Document	5/7/2008 6:51 AM
departments.txt	8 KB	Text Document	5/7/2008 6:51 AM

The screenshot shows a window titled "1 - COBA" with a menu bar (File, Edit, Transfer, Formats, Options, Macro, View, Window, Help) and a toolbar. The main area displays a table with the following data:

EMPNO	FIRSTNAME	MIDINIT	LASTNAME	WORKDEPT	PHONENO	HIREDATE
000010	CHRISTINE	I	HAAS	A00	3978	1965-01-01
000020	MICHAEL	L	THOMPSON	B01	3476	1973-10-10
000030	SALLY	A	KWAN	C01	4738	1975-04-05
000050	JOHN	B	GEYER	E01	6789	1949-08-17
000060	IRVING	F	STERN	D11	6423	1973-09-14
000070	EVA	D	PULASKI	D21	7831	1980-09-30
000090	EILEEN	W	HENDERSON	E11	5498	1970-08-15
000100	THEODORE	Q	SPENSER	E21	0972	1980-06-19
000110	VINCENZO	G	LUCCHESE	A00	3490	1958-05-16
000120	SEAN		O'CONNELL	A00	2167	1963-12-05
000130	DOLORES	M	QUINTANA	C01	4578	1971-07-28
000140	HEATHER	A	NICHOLLS	C01	1793	1976-12-15
000150	BRUCE		ADAMSON	D11	4510	1972-02-12
000160	ELIZABETH	R	PIANKA	D11	3782	1977-10-11
000170	MASATOSHI	J	YOSHIMURA	D11	2890	1978-09-15
000180	MARILYN	S	SCOUTTEN	D11	1682	1973-07-07
000190	JAMES	H	WALKER	D11	2986	1974-07-26
000200	DAVID		BROWN	D11	4501	1966-03-03
000210	WILLIAM	T	JONES	D11	0942	1979-04-11
000220	TFNNTFFR	K	IIT7	D11	0672	1968-08-29

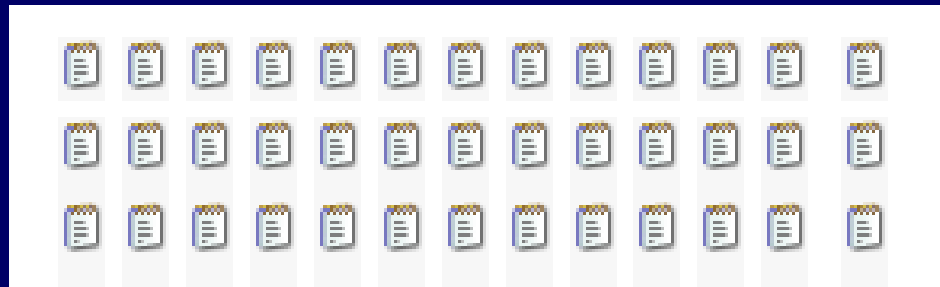
The status bar at the bottom shows "1 Sess-1 129.120.48.5 23/24".

Managing Data: a historical perspective

Operator



Data files



Managing Data: a historical perspective

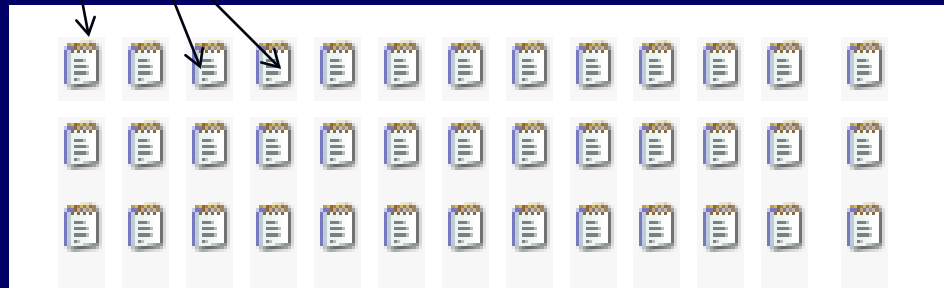
Operator



Custom Program(s)



Data files



Managing Data: a historical perspective

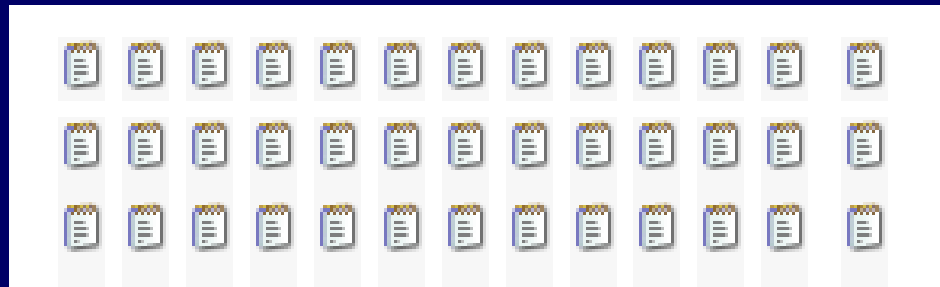
Operator



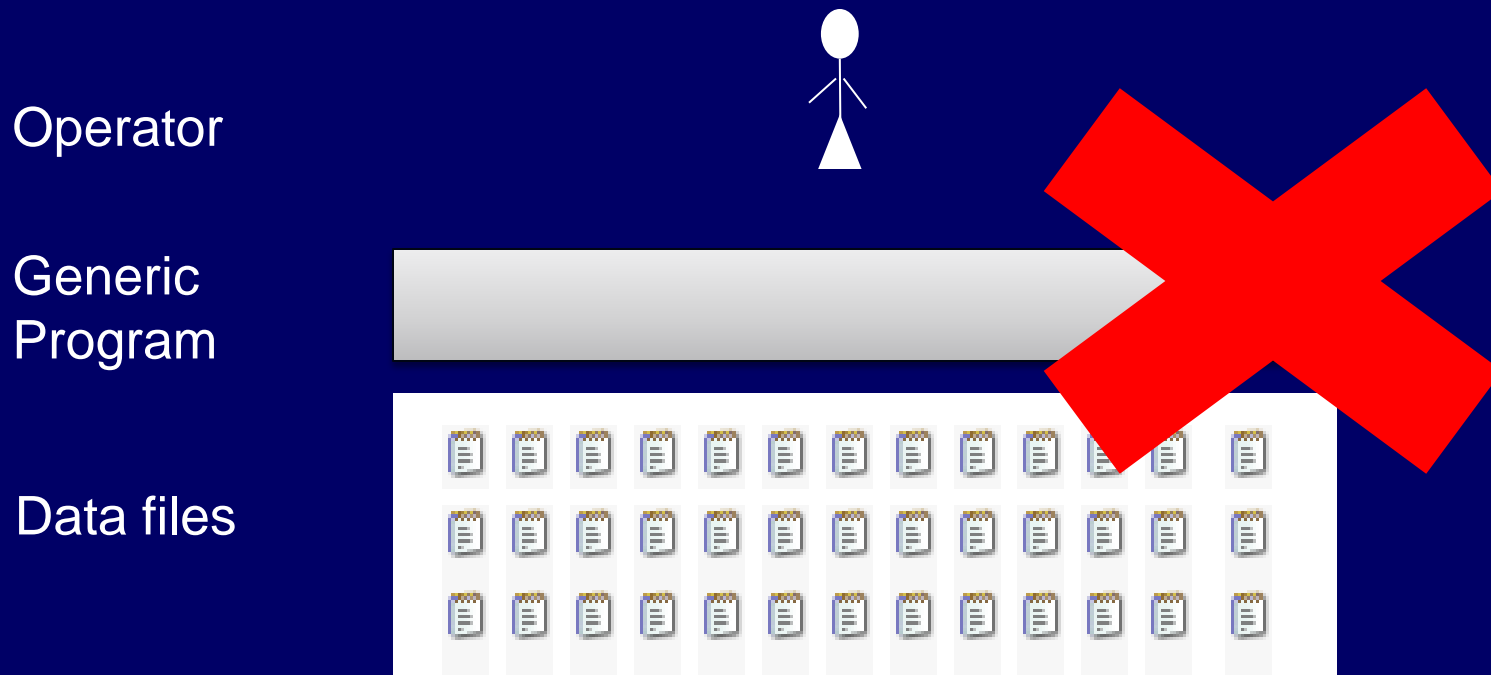
Generic
Program



Data files

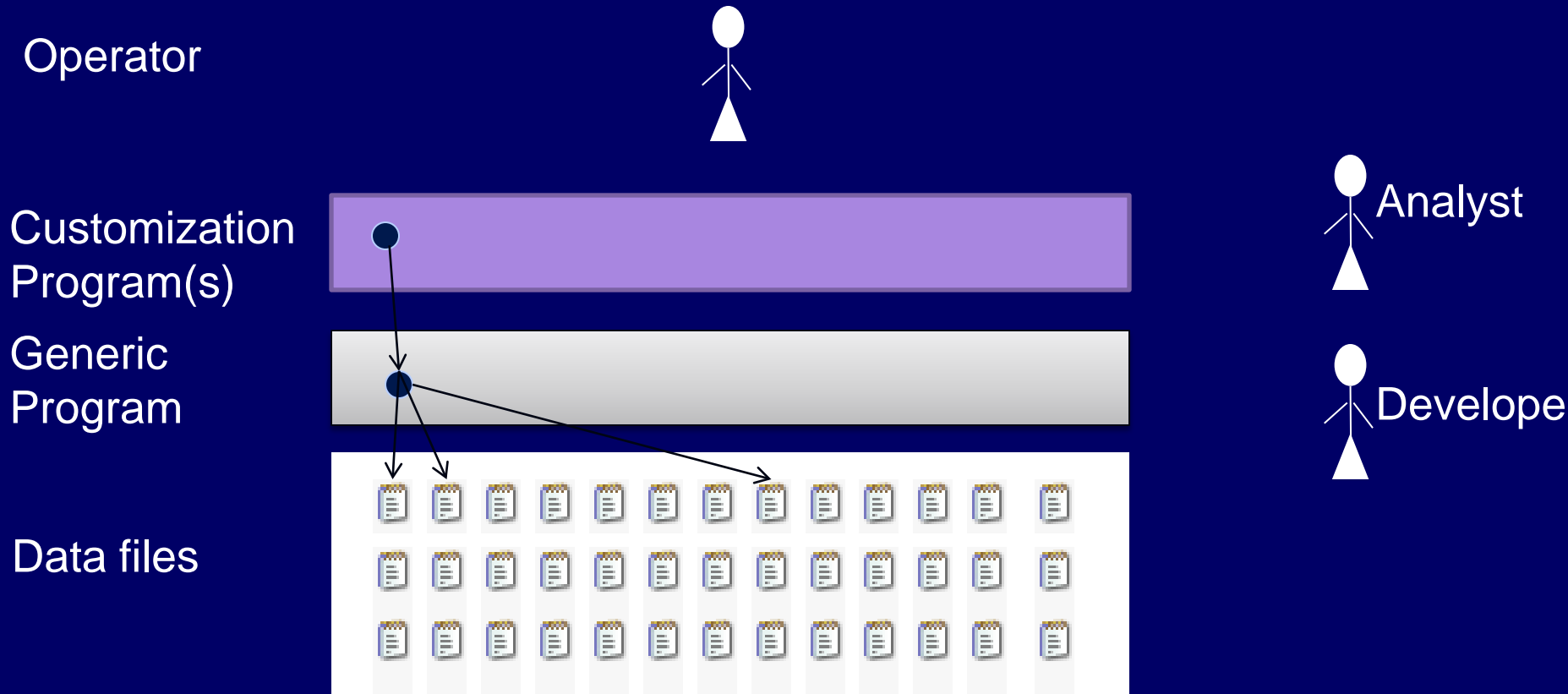


Managing Data: a historical perspective





Managing Data: a historical perspective





Managing Data: a historical perspective

Operator



Customization Program(s)



Generic Program



Data files



Database Management System (DBMS)



Managing Data: a historical perspective

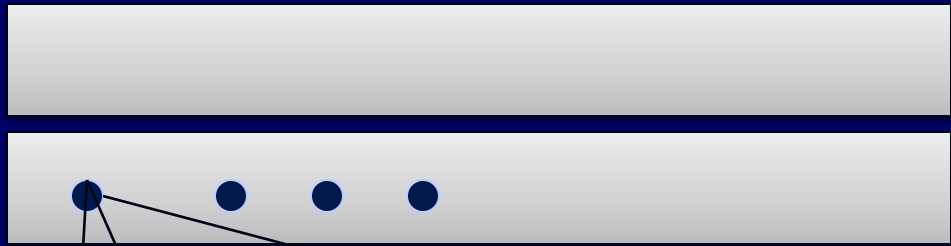
Database Administrator



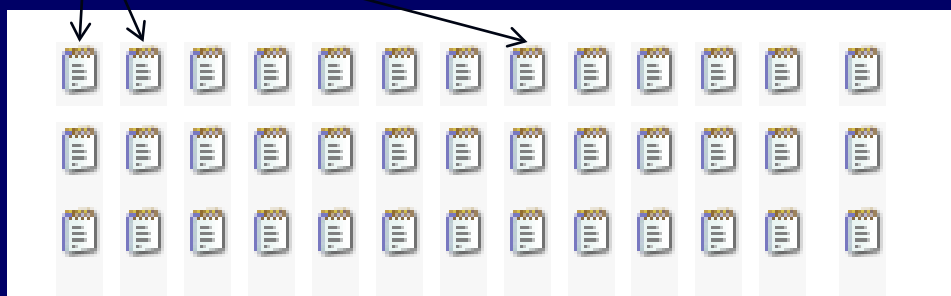
SQL Program(s)



SQL Interpreter
Generic Program



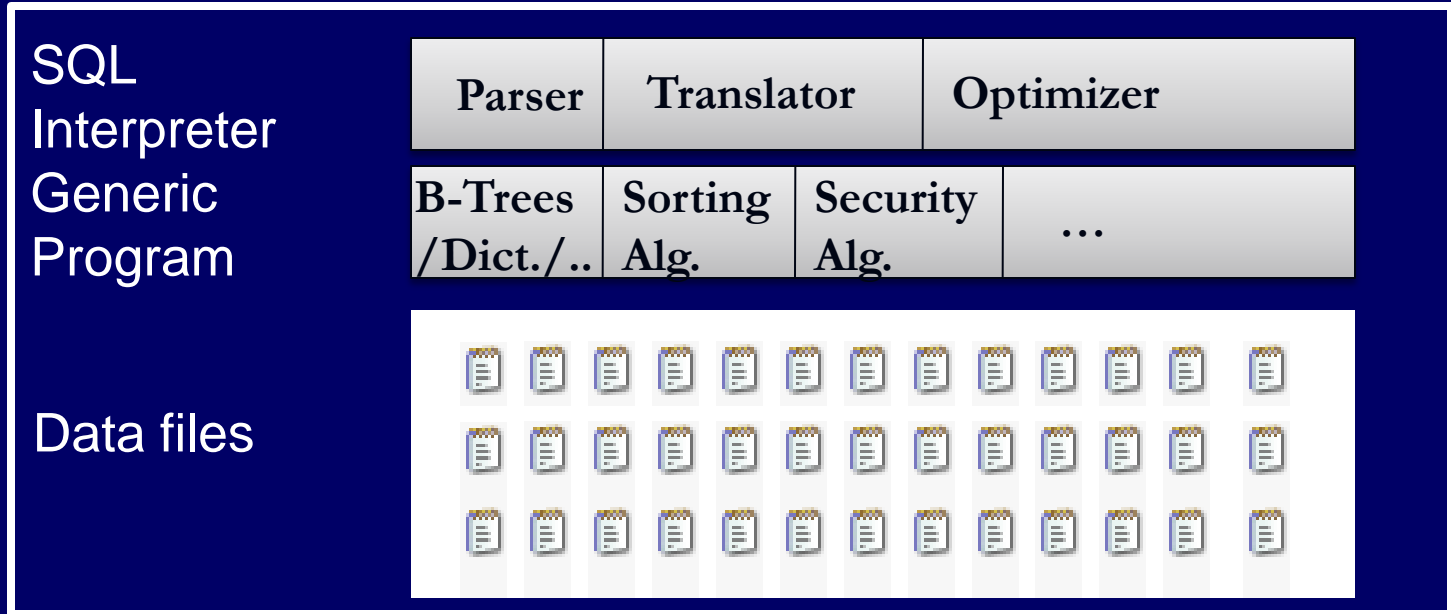
Data files



Database Management System (DBMS)



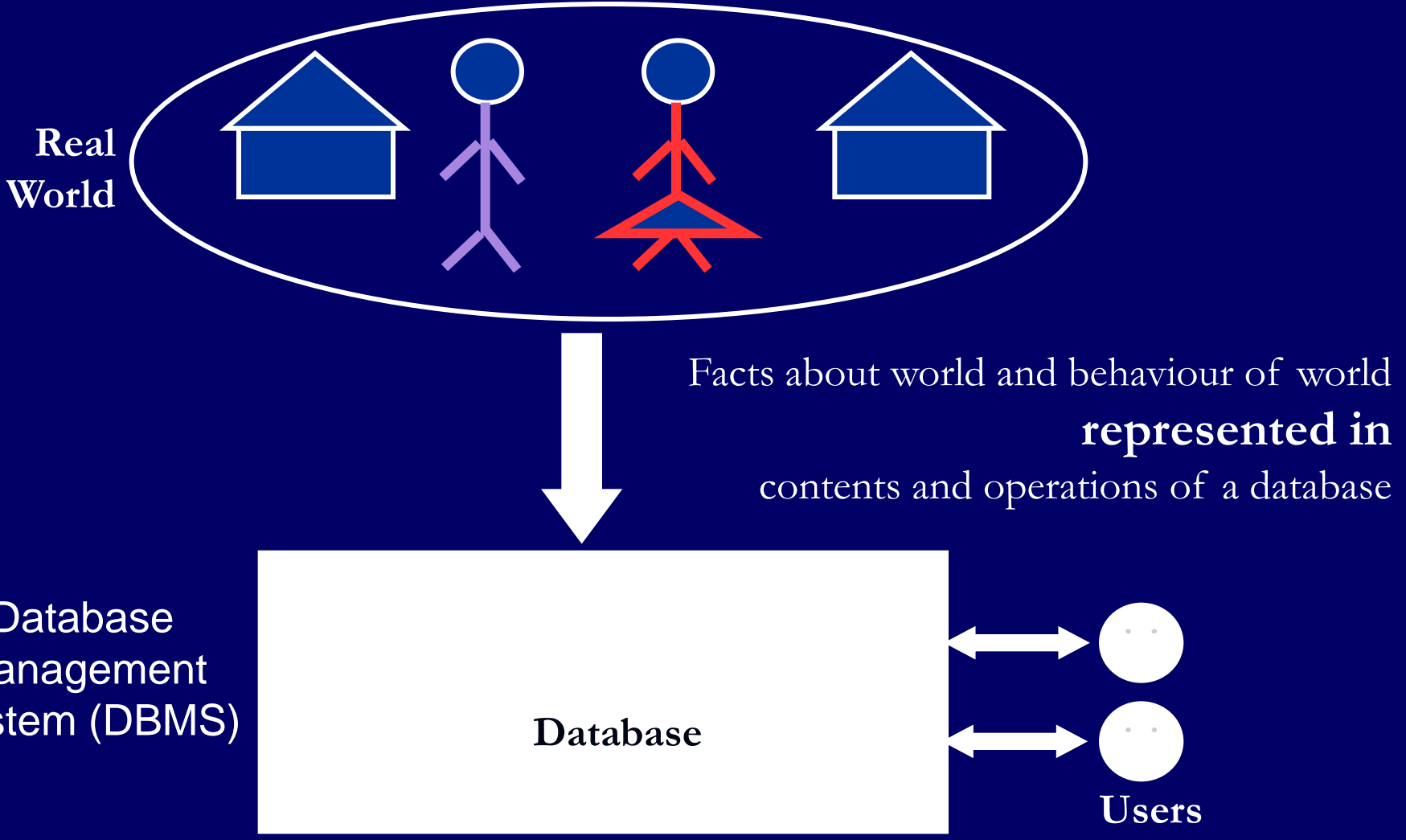
Database Management System (DBMS)



We will be using PostgreSQL



What Is a Database?





What: Is a File System a DBMS?

■ Thought Experiment 1:

- You and your project partner are editing the same file.
- You both save it at the same time.
- Whose changes survive?

A) Yours **B) Partner's** **C) Both** **D) Neither** **E) ???**

■ Thought Experiment 2:

- You're updating a file.
- The power goes out.
- Which changes survive?

A) All **B) None** **C) All Since Last Save** **D) ???**



What: Is a File System a DBMS?

■ Thought Experiment 1:

Q: How do you write programs over a subsystem when it promises you only “???” ?

A: Very, very carefully!!

–Which changes survive?

A) All B) None C) All Since Last Save D) ???



Database vs. File Systems

- What more could we want than a file system?
 - Simple, efficient *ad hoc*¹ queries
 - concurrency control
 - recovery
 - benefits of good data modeling
- S.M.O.P.²?
 - as we'll see this semester...
 - in fact, the OS often gets in the way!

¹ad hoc: formed or used for specific or immediate problems or needs

²SMOP: Small Matter Of Programming



Database Systems Today

MySpace

http://www.myspace.com/ Google

MySpace | People | Web | Music | Music Videos | Blogs

myspace.com a place for friends

Home | Browse | Search | Invite | Film | Mail | Blog | Favorites | Forum | Groups | Events | Videos | Music | Comedy | Classifieds

Cool New Videos 48,591 uploaded today!

Insane Ladder Race Future Airforce
Iceberg Rolling Over Name
Dance Book Eddie Uehara
Transformers Vs. Potters Robo-Joni

Help us find the next Hip-Hop Star! myspace presents

Member Login
E-Mail :
Password :
Remember Me
LOGIN SIGN UP!
Forgot your password?

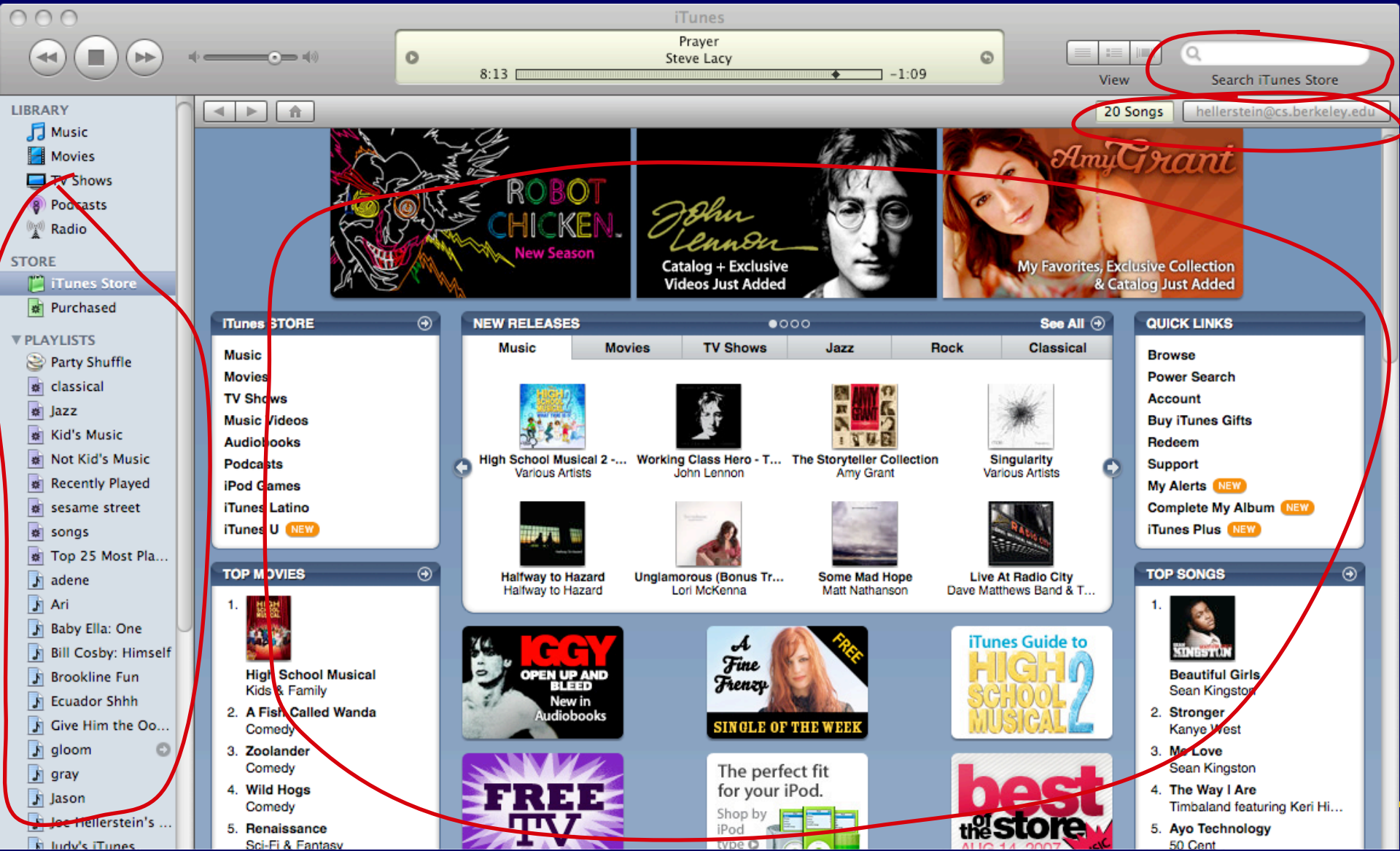
Cool New People
CvRDE LL Andrea Josh

Help us find the next Hip-Hop Star! myspace presents

Books Forum Mobile Profile Editor NEW!
Blogs Grade My Prof. Movies Ringtones NEW!
ChatRooms Groups Music Schools
Comedy Horoscopes Music Videos TV On Demand
Downloads Impact NEW! MySpaceIM Videos
Filmmakers Jobs News NEW! Weather NEW!



What: Database Systems Today





What: Database Systems Today

Accounts | Bill Pay | Transfers | Brokerage | Account Services | Messages & Alerts **A**

Account Summary | Account Activity

Account Summary **Try it!** [Enroll in Online Statements](#) **?** Help

Related Services
[Related Services](#)
[Add Accounts to View](#)
[Open a New Account](#)

Cash Accounts

Account	Account Number	Available Balance
CHECKING	111-2006xxx	\$7,289.46
SAVINGS	557-2911xxx	\$186.46
SAVINGS	111-1535xxx	\$1,262.65
Total		\$8,738.57

Investment Accounts

Account	Account Number	Total Account Value
BROKERAGE	VV674xxxxxx	\$15,866.56
<small>• Not FDIC Insured • No Bank Deposits • May Lose Value</small>		
Total		\$15,866.56

Credit Accounts

Account	Account Number	Outstanding Balance	Available Credit
MASTERCARD	5490-9600-0008-xxxx	\$1,631.79	\$6,668.21
Total		\$1,631.79	\$6,668.21

Loan Accounts

Account	Account Number	Outstanding Principle Balance
STUDENT LOAN	70004xxxx	\$5,000.00
Total		\$5,000.00



What: Database Systems Today

Loading "Yahoo! Maps, Driving Directions, and Traffic"

http://maps.yahoo.com/broadband#mvt=m&tp=1&tt=uc+berkeley&trf=0&lon=-122.256203&lat=37.871936

ncbi genome database

Yahoo! LOCAL Maps Sign In New User? Sign Up

Maps Home - Dial-Up Map (Original) - Help

GET MAP AND DIRECTIONS

Address, City, State

FIND A BUSINESS ON THE MAP

uc berkeley Search

Browse by Category

Search Results: uc berkeley Refine

- 1 UC Berkeley
- 2 UC Berkeley Botanical Garden
- 3 UC Berkeley Art Museum
- 4 UC Berkeley Foundation
- 5 Weinstein, Rhona S PhD - UC Berkeley Pysch Clinic
- 6 Capri Motel Berkeley
- 7 University Art Museum
- 8 Enterprise Rent A Car
- 9 Faculty Club
- 10 California Engineering Company
- 11 Campus Flowers



What: Database Systems Today

The screenshot shows the NCBI Map Viewer interface. At the top, the browser address bar displays the URL: http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=9606. The page title is "Entrez Genome view". The main navigation bar includes links for PubMed, Nucleotide, Protein, Genome, Gene, Structure, PopSet, Taxonomy, and Help. Below this, there is a search bar with the text "Search for" and a dropdown menu set to "on chromosome(s)". A "Find" button and an "Advanced Search" link are also present.

The main content area is titled "Homo sapiens (human) genome view" and includes links for "Build 36.2 statistics" and "Switch to previous build". A link for "BLAST search the human genome" is also visible. The central part of the page displays a karyotype of the human genome, with chromosomes arranged in two rows. The first row contains chromosomes 1 through 13, and the second row contains chromosomes 14 through 22, X, Y, and HT. Each chromosome is represented by a vertical bar with a centromere.

On the left side, there is a sidebar with the following sections:

- Map Viewer**: Map Viewer Home, Map Viewer Help, Human Maps Help, Release Notes
- NCBI Resources**: Genome Project, TaxPlot, Consensus Coding Sequence (CCDS), Human Genome Resources, NCBI Handbook, RefSeq, Whole Genome Association (WGA)
- Organism Data in GenBank**

At the bottom of the main content area, there is a text box with the following text:

Lineage: [Eukaryota](#); [Metazoa](#); [Chordata](#); [Craniata](#); [Vertebrata](#); [Euteleostomi](#); [Mammalia](#); [Eutheria](#); [Euarchontoglires](#); [Primates](#); [Haplorrhini](#); [Catarrhini](#); [Hominidae](#); [Homo](#); [Homo sapiens](#)

September 2006: NCBI released an annotation update for the human genome (NCBI Build 36.2); this update does not change the genome assembly. The previous version of the genome assembly, [NCBI Build 35.1](#), can still be accessed for Map Viewer display and for BLAST. For additional information about changes, statistics, and the status of the CGDS project, please refer to...



Data vs. Information

- Data is used to refer to what is **actually stored** in the database
- Information is used to refer to **meaning** of that data as understood by some user.



Data in DBMS will be integrated and shared.

■ Integrated

- Mean the database can be thought of as a unification of several distinct files, with any redundancy among those files partly or wholly eliminated

■ Shared

- Mean the database can be shared among different users, in the sense that different users can have access to the same data, possible even at the same time (“**Concurrent access**”).



Database Management System (DBMS)

- A collection of programs that enable:
 - **Defining** (describing the structure),
 - **Populating** by data (Constructing),
 - **Manipulating** (querying, updating),
 - **Preserving** consistency,
 - **Protecting** from misuse,
 - **Recovering** from failure, and
 - **Concurrent** using
of a database.



Why take this class?

A. Database systems are the core of CS

- Shift from computation to information
 - True in corporate computing for years
 - Web made this clear for “the rest of us” by the end of 90’s
 - Increasingly true of scientific computing
- Need for DB technology has exploded in the last years
 - **Corporate**: retail swipe/clickstreams, “customer relationship mgmt”, “supply chain mgmt”, “data warehouses”, etc.
 - **Web: not** just “documents”. Search engines, maps, e-commerce, blogs, wikis, social networks. Web 2.0.
 - **Scientific**: digital libraries, genomics, satellite imagery, physical sensors, simulation data
 - **Personal**: Music, photo, & video libraries. Email archives. File contents (“desktop search”).

Why take this class?

B. DBs are incredibly important to society

- “Knowledge is power.” -- Sir Francis Bacon
- “With great power comes great responsibility.” -- Spiderman's Uncle Ben



- Policy-makers should understand technological possibilities.
- Informed Technologists needed in public discourse on usage.



Why take this class?

C. The topic is intellectually rich.

- representing information
 - data modeling
- languages and systems for querying data
 - complex queries & query semantics*
 - over massive data sets
- concurrency control for data manipulation
 - controlling concurrent access
 - ensuring transactional semantics
- reliable data storage
 - maintain data semantics even if you pull the plug

* semantics: the meaning or relationship of meanings of a sign or set of signs



Why take this class?

~~D. It isn't that much work.~~

- Bad news: It is a fair bit of work.
 - varies from year to year

- Good news: the course is front loaded
 - Most of the hard work is in the first half of the semester
 - Load balanced with most other classes



Why take this class?

E. Looks good on my resume.

- Yes, but why? This is not a course for:
 - Oracle administrators
 - IBM DB2 engine developers
 - Though it's useful for both!
- It is a course for well-educated computer scientists
 - Database system concepts and techniques increasingly used “outside the box”
 - Ask your friends at Microsoft, Yahoo!, Google, Apple, etc.
 - A rich understanding of these issues is a basic and (un?)fortunately unusual skill.



Database Management System (DBMS)

- A collection of programs that enable:
 - **Defining** (describing the structure),
 - **Populating** by data (Constructing),
 - **Manipulating** (querying, updating),
 - **Preserving** consistency,
 - **Protecting** from misuse,
 - **Recovering** from failure, and
 - **Concurrent** using
of a database.