CSE 215: Database

Lecture 01

2023

Introductory Lecture

WHO AM I?

Ashikur Rahman

Professor, Ex-chairman, CSE, BUET

H.S.C.: FCC, 1991

B.Sc.: BUET, 1998 (Merit pos: 3rd)

(Started job as a Lecturer) (1998)

M.Sc.: BUET, 2001

Ph.D.: University of Alberta, Canada, 2006
(rejoined BUET as an Assistant Professor)

Postdoc: University of Calgary, Canada, 2011
State University of New York, USA, 2012
(rejoined BUET as an Associate Professor)

Became professor in 2014.

Became Chairman, CSE BUET in 2020.

Summary of profile

HSC → 1991
B.Sc. → 1998 (BUET)

Took 7 years

Ph.d. (start) → 2001
Ph.d. (end) → 2006

Took 5 years

B.Sc. Engg in BUET is much harder than Ph.D. in many many universities of the world

Teaching Staff

- Instructors:

Ashikur Rahman

Office: CSE 121

Md. Toufikuzzaman

Office: CSE 209,

Lectures: 1st half - from a user's perspective

1. **Foundations:** Relational data models & SQL
   - Weeks 1-3
   - How to manipulate data with SQL, a declarative language
     - *reduced expressive power but the system can do more for you*

2. **Database Design**: Design theory and constraints
   - Weeks 4-5
   - Designing relational schema to keep your data from getting corrupted

3. **E-R diagram & Relational Algebra:** Syntax & supporting systems
   - Week 6-7
   - A programmer's first step to data abstraction

Lectures: 2nd half - understanding how it works

4. Inside to database systems

- Indexing
- External Memory Algorithms (IO model) for sorting, joins, etc.
- Basics of query optimization (Cost Estimates)

5. Transactions: Syntax & supporting systems

✓A programmer's abstraction for data consistency

6. Data Recovery

Communications

- Web site:

  http://teacher.buet.ac.bd/ashikur/CSE215
  - Lectures available here (usually the morning before class)
  - Useful links to possibly useful reading materials
  - Class test marks and important announcements are posted !

- Email:

  {ashikur, toufikuzzaman}AT cse DOT buet DOT ac DOT bd

  {penguinswimming, md.toufikzaman} AT gmail DOT com

Textbook(s)

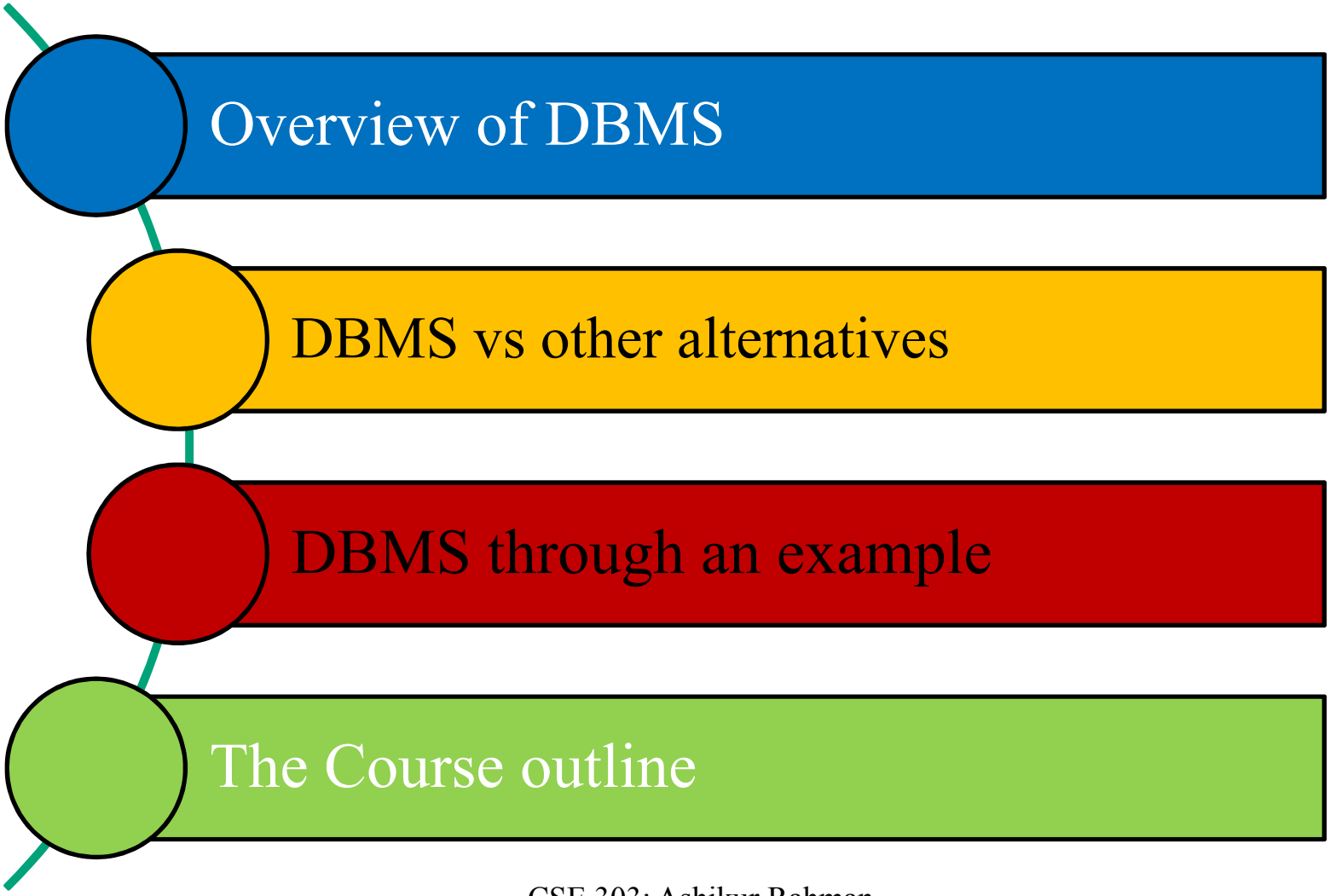
Main textbook:

- *Database Systems: The Complete Book*, Hector Garcia-Molina, Jeffrey Ullman, Jennifer Widom
  - Most chapters are good. Some are not as great (functional dependencies).
  - available at the Poroma Prokashoni
  - A free PDF version is available (linked from the course website)
- COME TO THE CLASS ! ASK QUESTIONS ! READ SLIDES !

Other Texts

- *Database System Concepts*,
  by Avi Silberschatz, Henry F. Korth, S. Sudarshan

Today's Outline



Quotation by “somebody” on the earth

“There is no point in having information unless it can be communicated from one point to another point or from one time to another time.”

Use of relay: an old concept of communication!

How to communicate data from one point to another point?

But with too many glitches.. Most of the data is exaggerated and fabricated ….. ;-)

Use of relay: an old concept of communication!

- Persian Royal Road ran some 2,857 km
- Used to take about 3 months (on foot of an army man!) to send a courier from one end to the other end

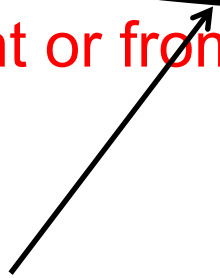
Horses creating ad hoc communication!

15

- Fresh horses and riders ready at each relay on the road
- Courier could be sent only in nine days.
- This system was used until development of effective optical telegraph systems in late 18th century.

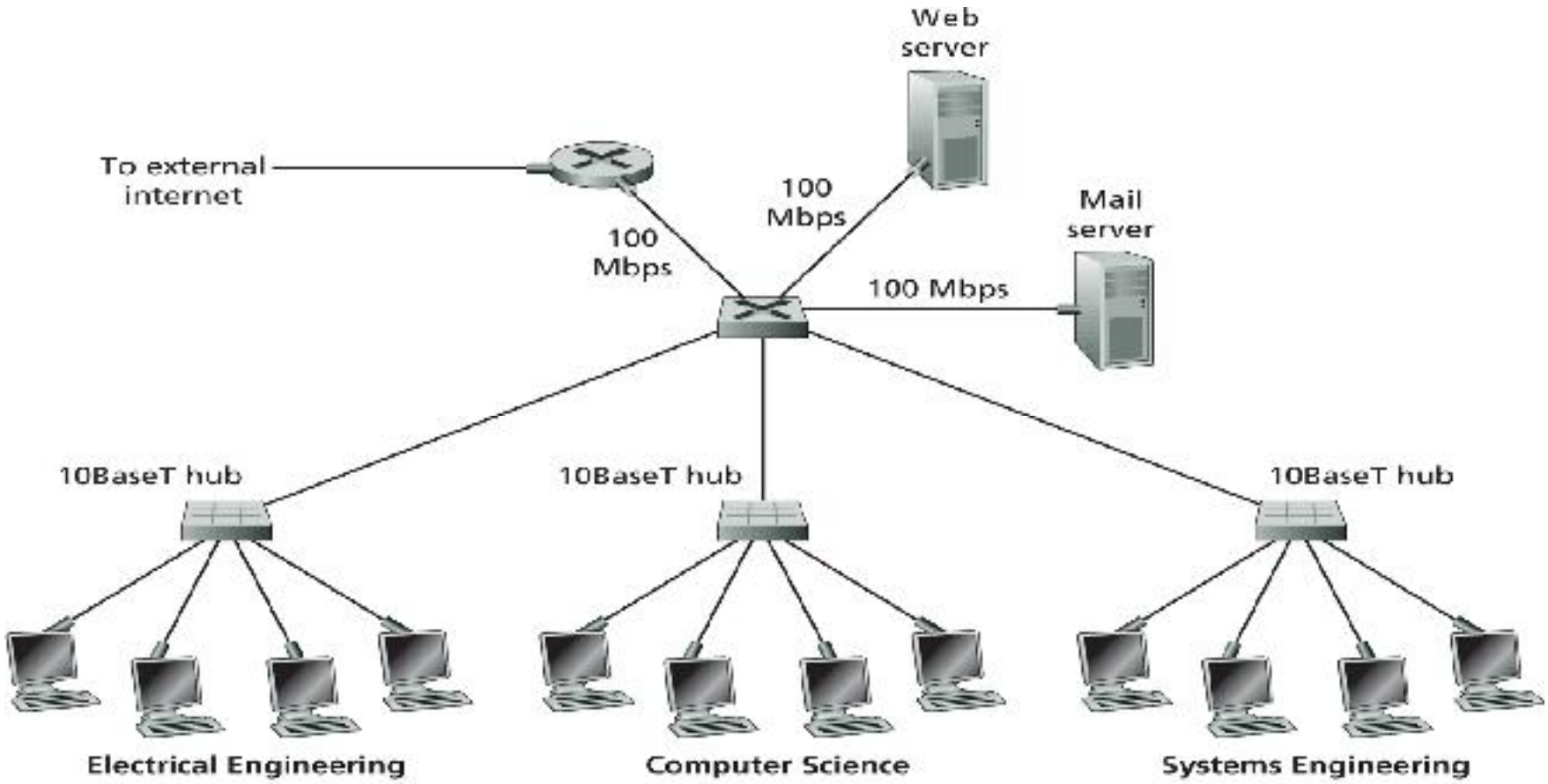
Quotation by “somebody” on the earth

“There is no point in having information unless it can be communicated from one point to another point or from one time to another time.”



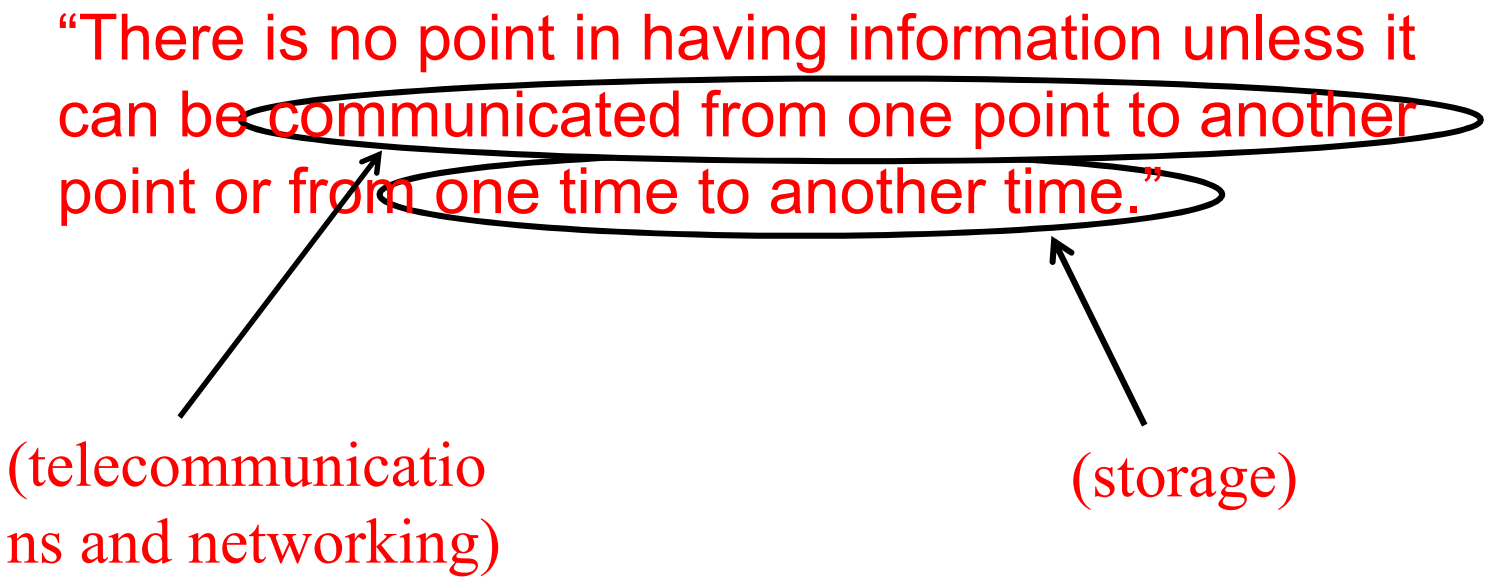
(telecommunicatio
ns and networking)

Connecting devices

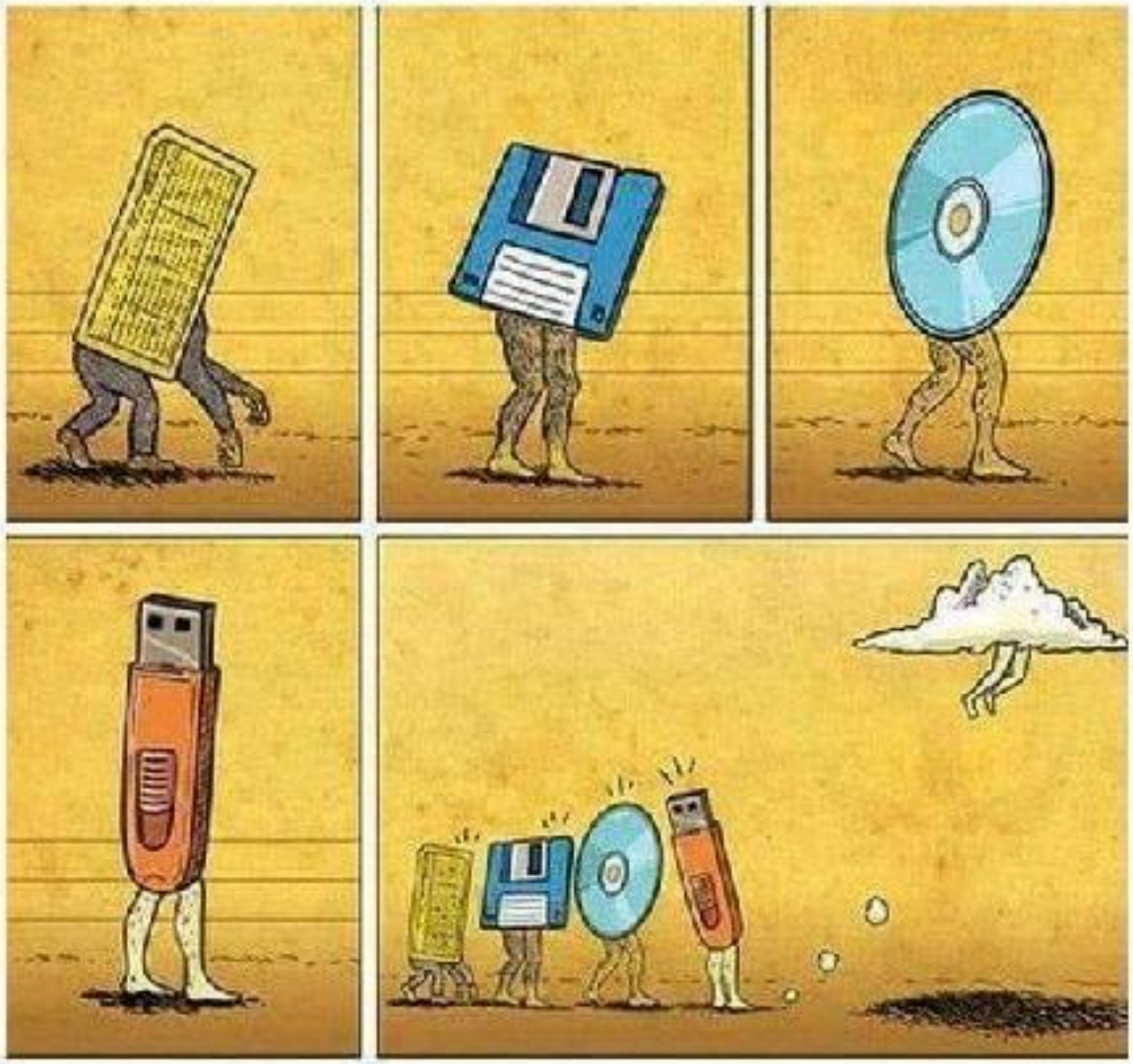


Figur itutional network using a combination of hubs, et switches, and a router

Quotation by “somebody” on the earth



The Evolution of Data Storage



“Data is the Future”

- Quotation by somebody on the earth

What is a database ?

FACEBOOK definition

Data BASE

“Data is the Future”
- Quotation by somebody on the earth

- What is a database?
- Any collection of related information
  - Phone book
  - Shopping list
  - Todo list
  - Your 5 best friends
  - Facebook’s user base
- Database can be stored in different ways
  - On paper
  - In your mind
  - On a computer

“Data is the Future”

- Quotation by somebody on the earth

Some commercial databases

- Accounts database; payroll database; BIIS students database

Increasingly many companies see themselves as **data driven**.

This is what happens in just ONE MINUTE on INTERNET

Link

Computers + Databases = <3

Amazon.com

- Keeps track of Products, Reviews, Purchase Orders, Credit Cards, Users, Media, etc
- Trillions of pieces of information need to be stored and readily available
- Information is extremely valuable and critical to Amazon.com's functioning
- Security is essential, Amazon stores peoples personal information
  - Credit card #, SSN, Address, phone
- Information is stored on a computer

vs

Shopping List

- Keeps track of consumer products that need to be purchased
- 10-20 pieces of information need to be stored and readily available
- Information is for convenience sake only and not necessary for shopping
- Security is not important
- Information is stored on a piece of paper, or even just in someone's memory

Millennia of Knowledge in Libraries

The world's scientific knowledge is **accessible**.

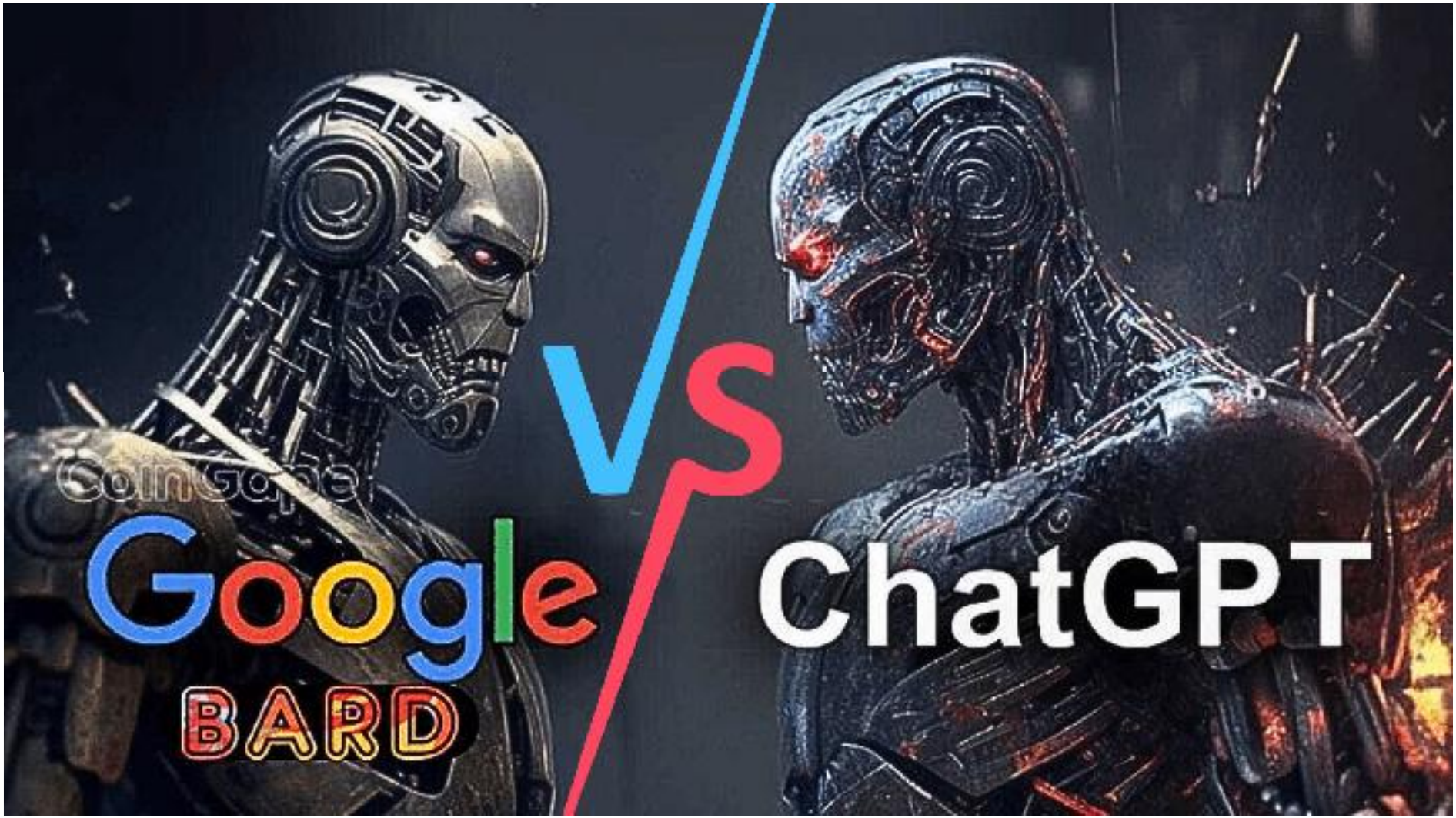
But we're still human…

The world's scientific knowledge is **accessible**, but not **readable**.

Could we build a machine to **read** for us?

Building database is the first step towards machine readable data

Thanks to Generative AI …



Database Management System

What is a DBMS ?

- *A big C/C++ program written by someone else that allows us to manage efficiently a large database and allows it to persist over long periods of time*

Database Management System

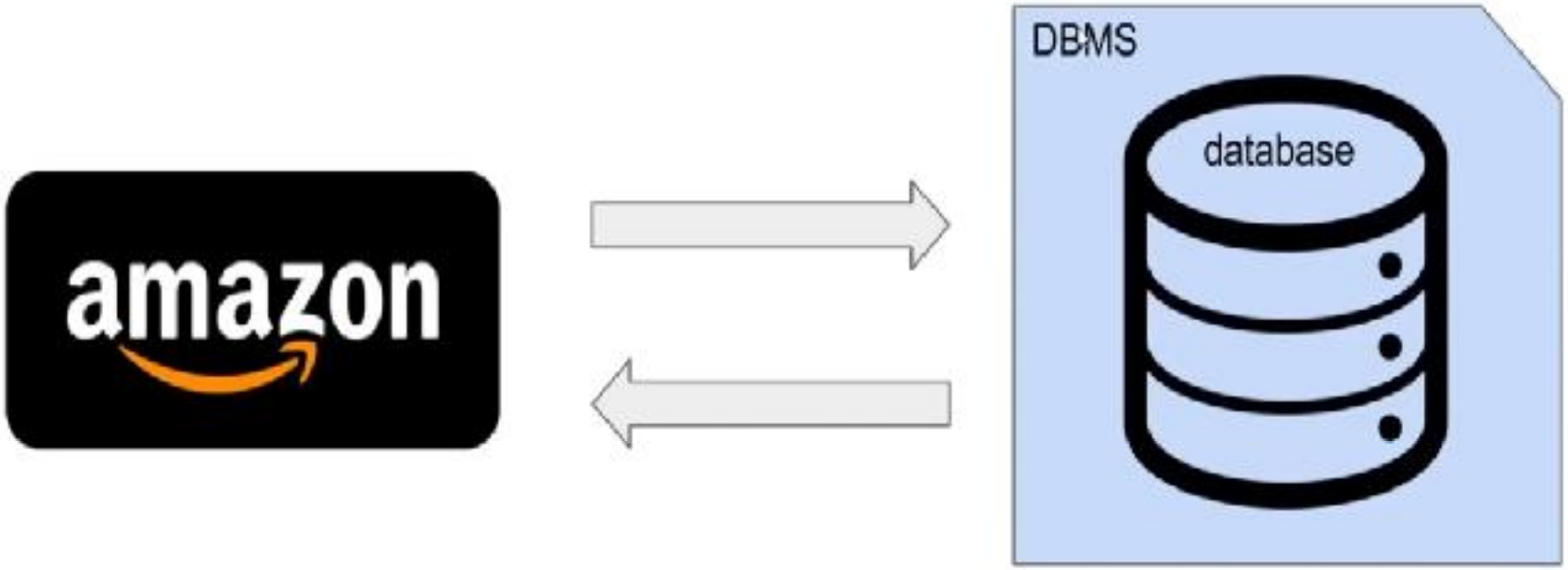
What is a DBMS(formal definition)?

•A **Database Management System (DBMS)** is a piece of software designed to create, store and manage databases

Functionalities of DBMS

- Makes it easy to manage large amounts of information
- Handles Security
- Backups
- Importing and exporting data
- Concurrency
- Interacts with software application

Amazon.com Database Diagram



Amazon.com will interact with the DBMS in order to create, read, update and delete information

Database Management System

Give examples of DBMS

- The big commercial database vendors:
  - Oracle
  - IBM (with DB2) bought Informix recently
  - Microsoft (SQL Server)
- Some open source (free) database systems:
  - Postgresql
  - Mysql
  - Predator

DBMS Market Shares

In 2006, www.gartner.com

- ORACLE Oracle: 47% market share, $7.2BN in sales
- IBM: 21% market share with $3.2BN in sales
- Microsoft: 17% market with $2.6BN in sales

DBMS Market Shares (Cloud based)

2017		2018		2019		2020		2021	
Vendor	Share	Vendor	Share	Vendor	Share	Vendor	Share	Vendor	Share
Oracle	36.1%	Oracle	31.1%	Oracle	27.4%	Microsoft	24.3%	Microsoft	24.0%
Microsoft	21.5%	Microsoft	23.6%	Microsoft	24.7%	Oracle	23.8%	AWS	23.9%
IBM	12.7%	AWS	13.5%	AWS	17.1%	AWS	20.6%	Oracle	20.6%
AWS	9.2%	IBM	10.4%	IBM	8.8%	IBM	6.8%	Google	6.5%
SAP	7.4%	SAP	6.9%	SAP	6.5%	SAP	5.6%	IBM	5.6%

DBMS Latest ranking http://db-engines.com/

Rank May 2022	Rank Apr 2022	Rank May 2021	DBMS
1.	1.	1.	Oracle
2.	2.	2.	MySQL
3.	3.	3.	Microsoft SQL Server
4.	4.	4.	PostgreSQL
5.	5.	5.	MongoDB
6.	6.	↑ 7.	Redis
7.	↑ 8.	↓ 6.	IBM Db2
8.	↓ 7.	8.	Elasticsearch
9.	9.	↑ 10.	Microsoft Access
10.	10.	↓ 9.	SQLite

An Example

The Internet Movie Database
http://www.imdb.com

- Entities:
  Actors (800k), Movies (400k), Directors, …

- Relationships:
  who played where, who directed what, …

Tables

Directors:

id	fName	lName
15901	Francis Ford	Coppola
. . .		

Movie_Directors:

id	mid
15901	130128
. . .	

Movies:

mid	Title	Year
130128	The Godfather	1972
. . .		

A Motivating, Running Example

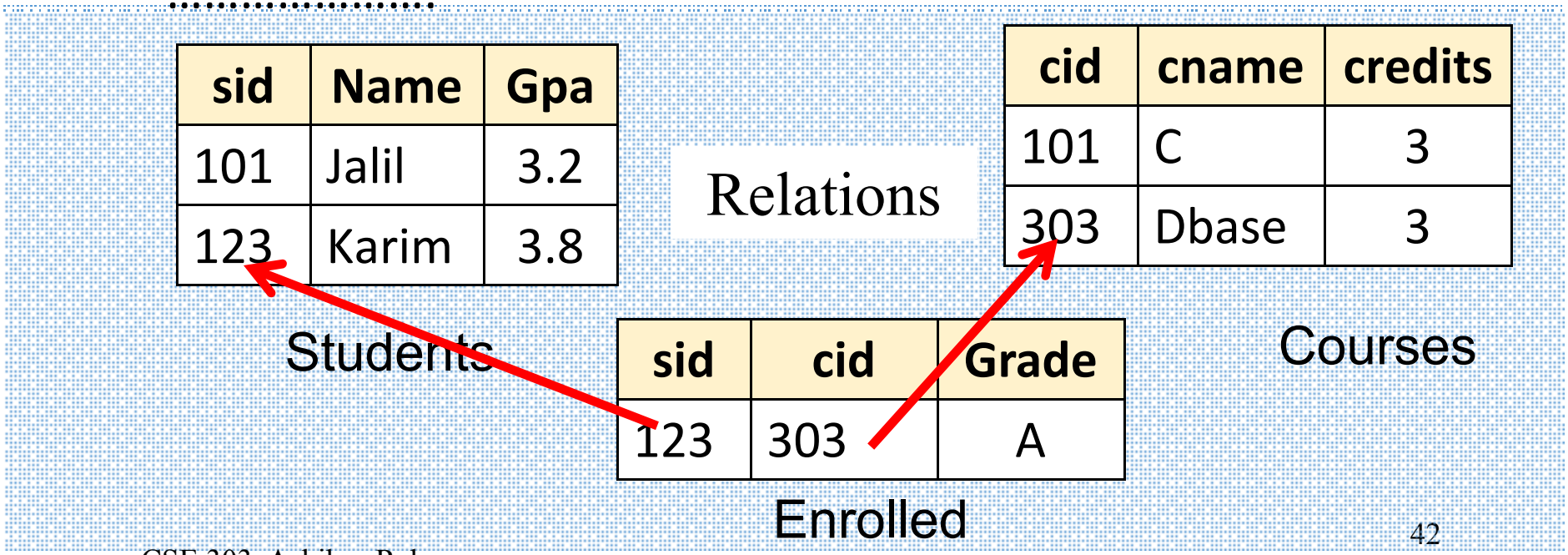
- Consider building a course management system (**CMS**):
  - Students
  - Courses
  - Professors
  - Who takes what
  - Who teaches what

Entities

Relationships

Modeling the CMS using DBMS

- *Logical Schema*
  - Students(sid: *string*, name: *string*, gpa: *float*)
  - Courses(cid: *string*, cname: *string*, credits: *int*)
  - Enrolled(sid: *string,* cid*: string,* grade*: string*)
  - ……



sid	Name	Gpa
101	Jalil	3.2
123	Karim	3.8

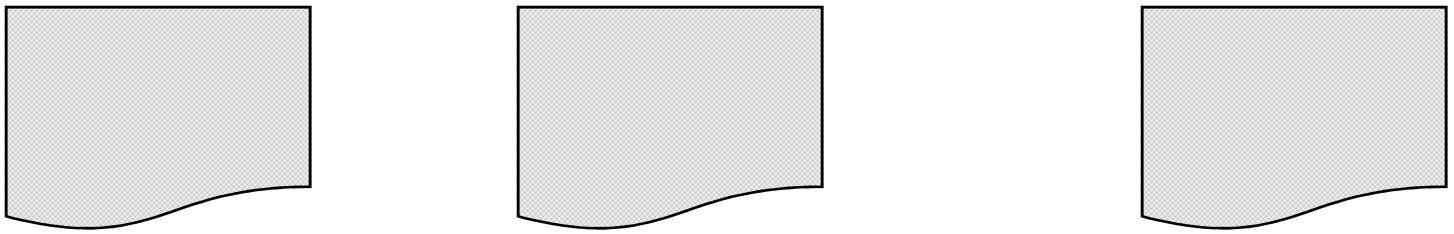
cid	cname	credits
101	C	3
303	Dbase	3

sid	cid	Grade
123	303	A

Can we do it without a DBMS ?

Sure we can! Start by storing the data in files:

students.txt courses.txt professors.txt



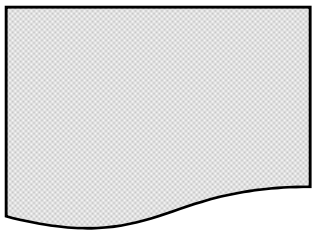
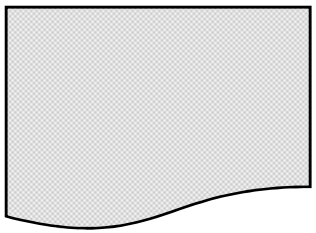
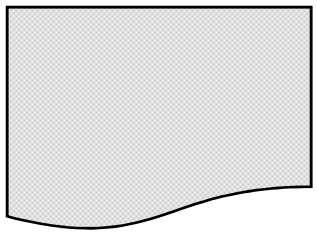
You can store data using text editors, OR

You can write C or Java programs to implement specific tasks

Can we do it without a DBMS ?

Another way: store the data in EXCEL files:

students.xls courses.xls professors.xls

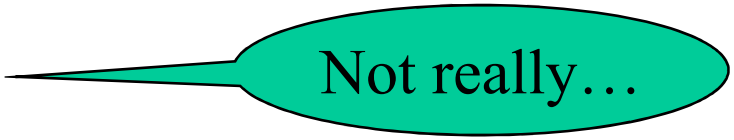
Possible Organizations

- Files
- Spreadsheets
- DBMS

1. Create/store Large Datasets

- Files

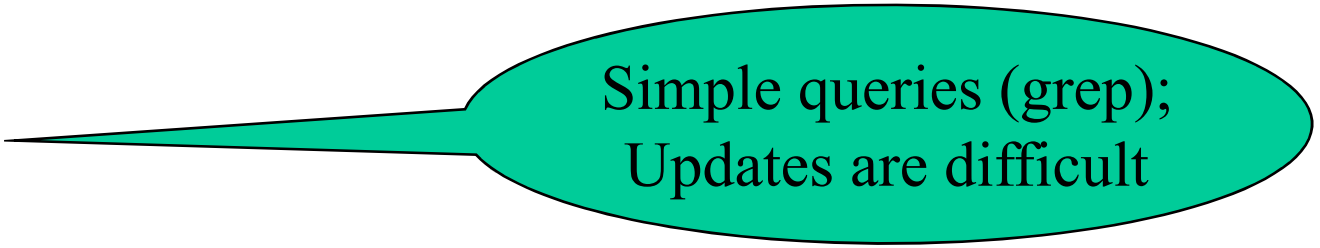
- Spreadsheets



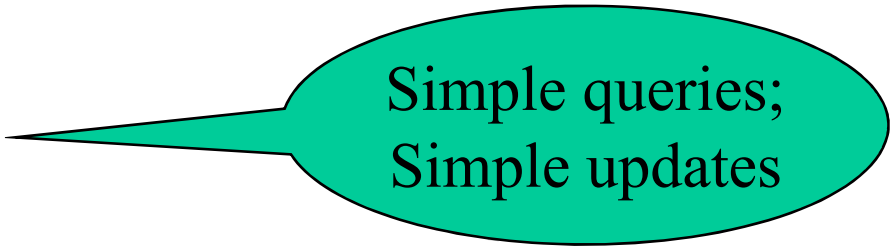
- DBMS

2. Search/Query/Update

- Files



- Spreadsheets

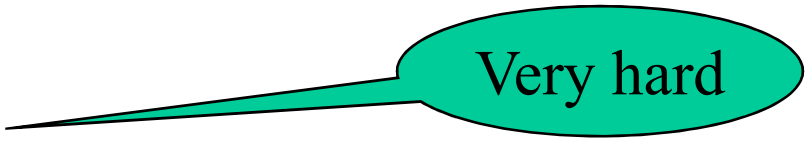


- DBMS

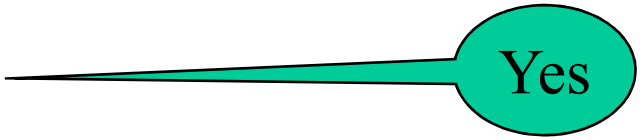
3. Change the Structure

Add *Address* to each Actor

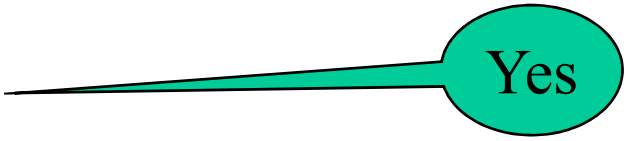
- Files



- Spreadsheets



- DBMS



4. Concurrent Access

Multiple users access/update the data concurrently

- What can go wrong ?

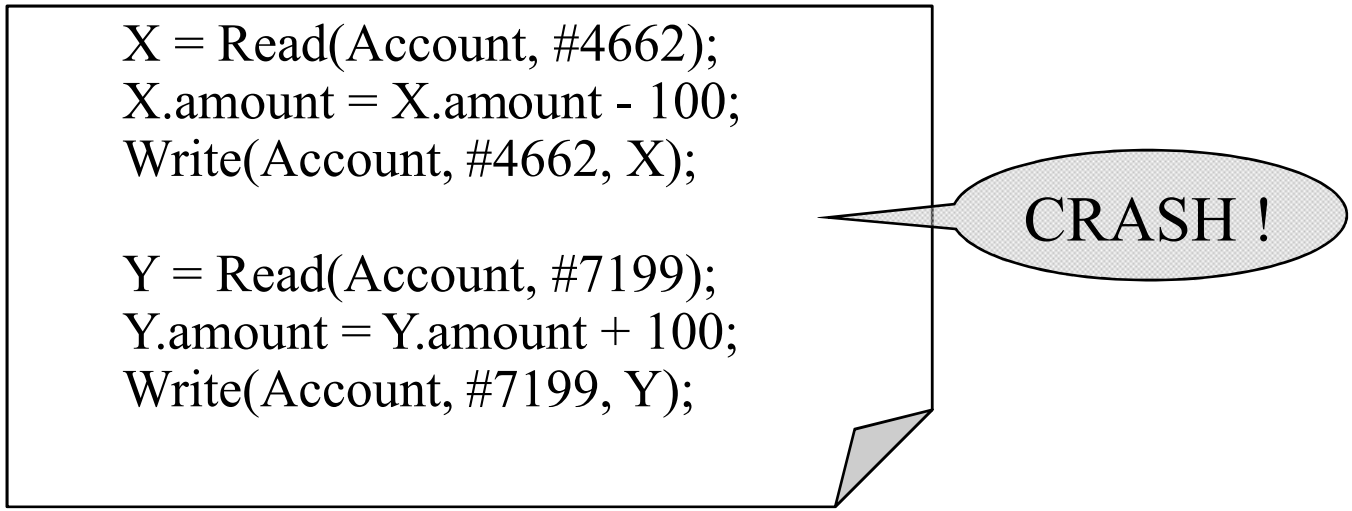
  Lost updates; inconsistent reads,…

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

- How do we protect against that in OS ? locks

5. Recover from crashes

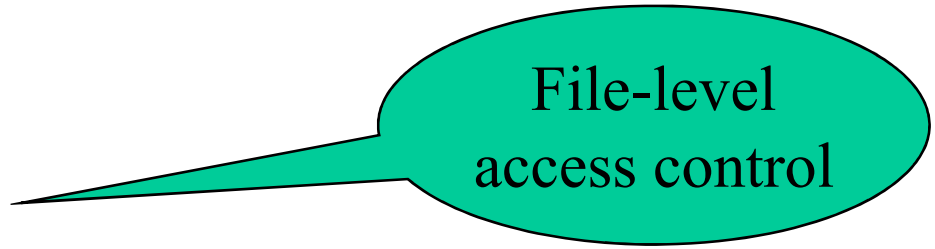
- Transfer $100 from account #4662 to #7199:

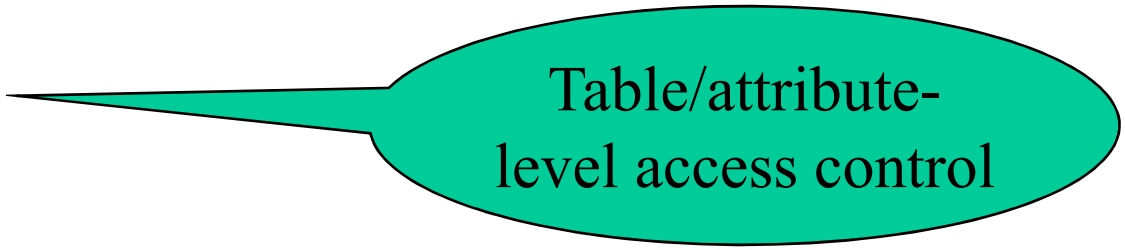


What is the problem ?

6. Security

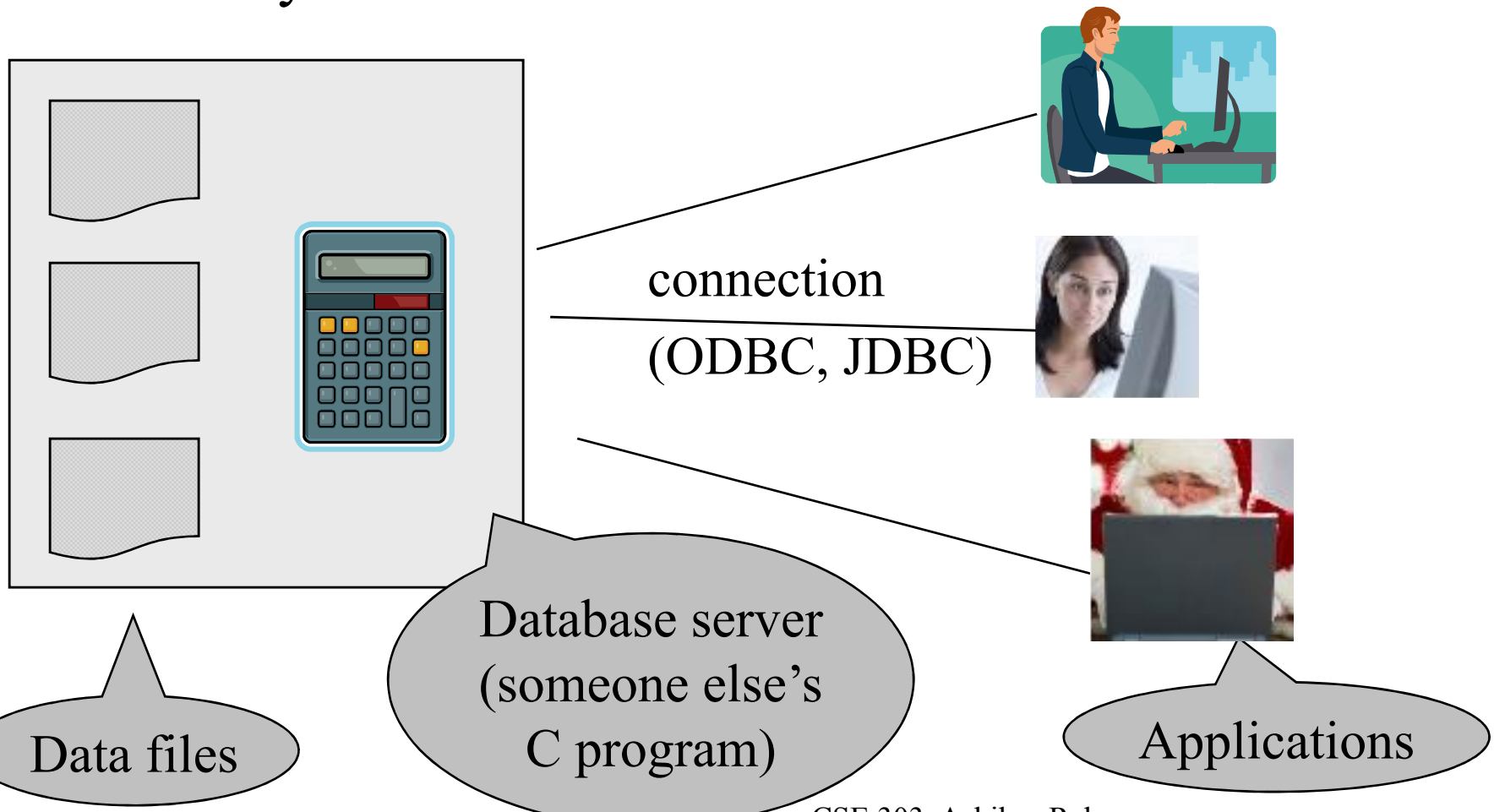
- Files
- Spreadsheets
- DBMS





How DBMS does all those things

"Two tier system" or "client-server"



Data Independence

Logical view

Directors:

id	fName	lName
15901	Francis Ford	Coppola
. . .		

Movie_Directors:

id	mid
15901	130128
. . .	

Movies:

mid	Title	Year
130128	The Godfather	1972
. . .		

Directors_file

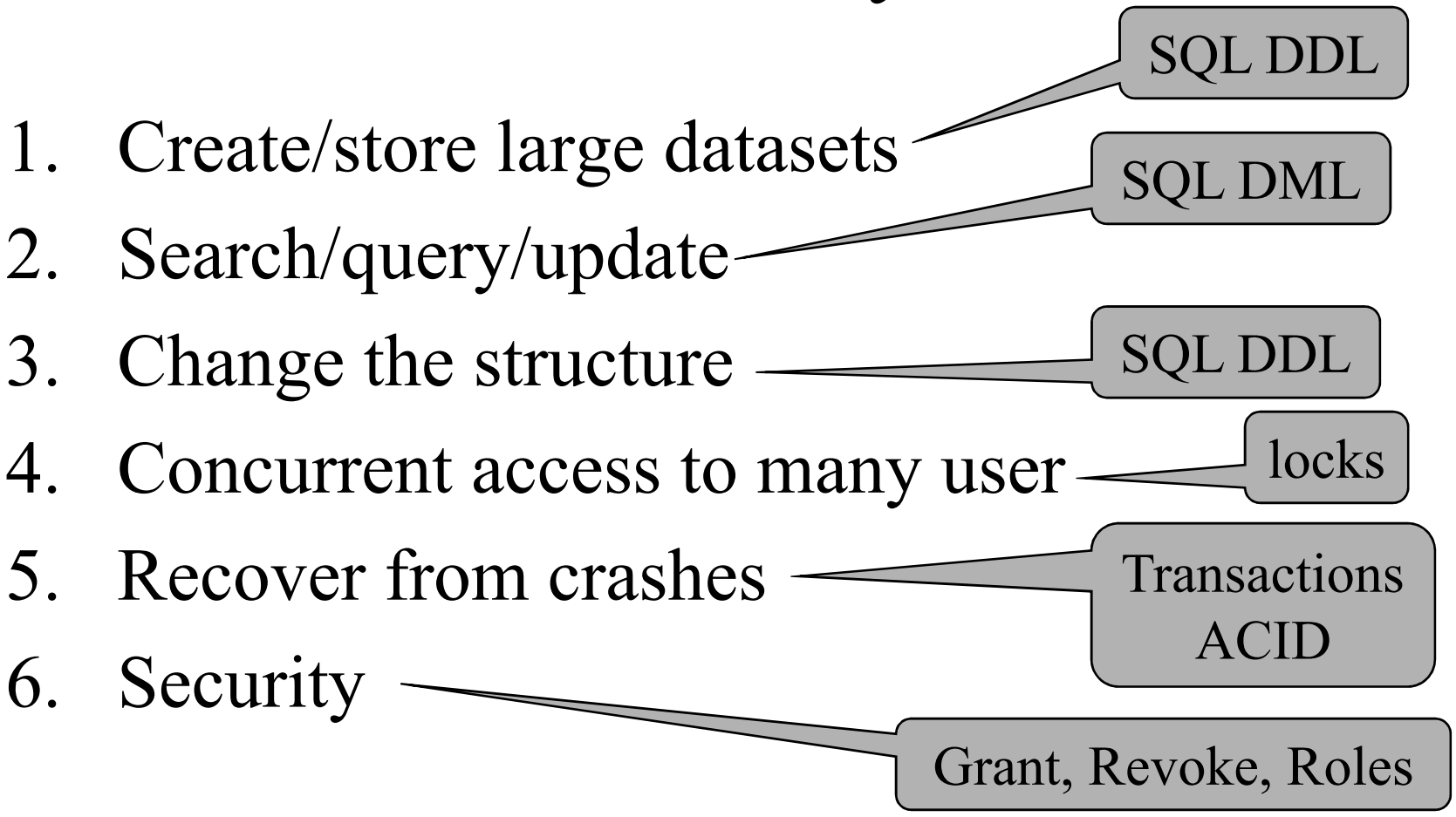
Moviews_title_index_file

Directors_fname_index_file

Movies_file

Physical view

What the Database Systems Does



Course Outline (first half) - TENTATIVE !!

1. Introduction (chapter 1)
2. Relational model (chapter 2)
3. SQL (chapter 6)
4. Constraints (chapter 7)
5. Design theory (chapter 3)
6. E-R diagram (chapter 4)
7. Relational Algebra (chapter 2)

Until Next Time…

- Go to the course website
- Install Oracle Xpress Edition 11/12
- Start using sqlplus, and sql developer
- Read the guideline on the course website to use sqlplus
- Find a partner for the project in the lab
- Start reading about SQL online and in the book