

DBMS Overview

Chapter 1



Instructor: Atul Prakash
University of Michigan, Ann Arbor

Partly based on material by many other professors

What is a DBMS?



- Database: Large, integrated collection of data
- DBMS = Database Management System
= a software package designed to store and manage databases
e.g. Oracle, MS SQL server, SqlLite, Postgres, MySQL.
- **Goal:** Efficient and convenient access to data

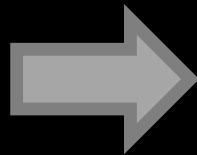
DBMS: Motivation

- Scenario:
 - ✧ Students taking classes, and obtaining grades
 - ✧ Q1: Find grade in class X
 - ✧ Q2: Find my GPA
 - ✧ ...

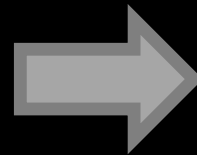
Ideas?



Think



Pair

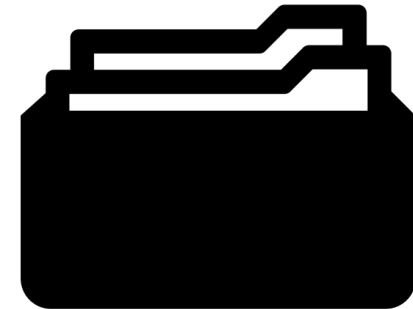


Share

Old-time Solution: Sorted Student Folders



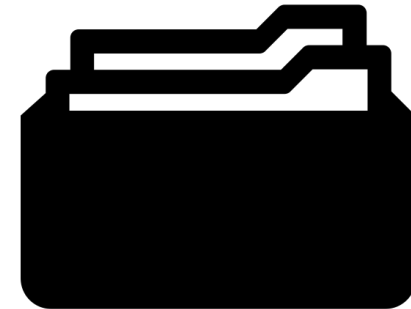
- Advantages?
- Disadvantages?



Old-time Solution: Sorted Student Folders



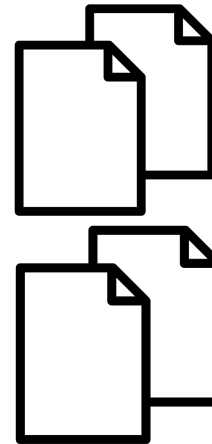
- Advantages?
 - ✧ cheap
- Disadvantages?
 - ✧ Large physical footprint
 - ✧ No sharing
 - ✧ No ad-hoc queries



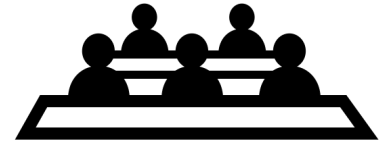
Other Solution: Flat Files



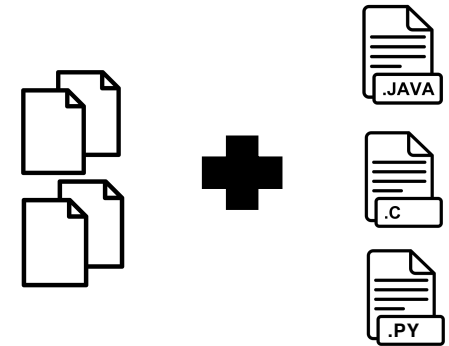
- Access?
 - ✧ using programs in C, Java, etc.
- Layout for the student records?



Other Solution: Flat Files



- Access?
 - ✧ using programs in C, Java, etc.
- Layout for the student records?



CSV:

Brown, Lisa, lbrown, db, A, os, B

Smith, Bart, bsmith

Tompson, Mary, mtom, vis, B+, db, A-

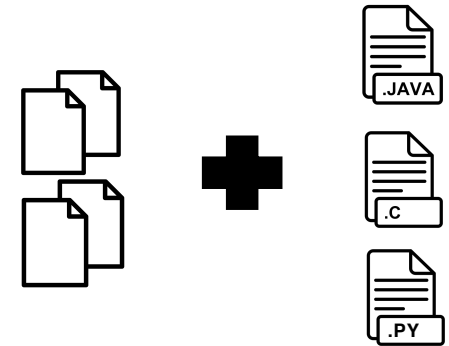
...

...

Other Solution: Flat Files



- Access?
 - ✧ using programs in C, Java, etc.
- Layout for the student records?



Multiple files:

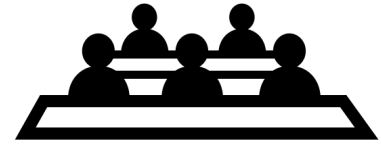


Brown, Lisa, lbrown
Smith, Bart, bsmith
Tompson, Mary, mtom
...
...

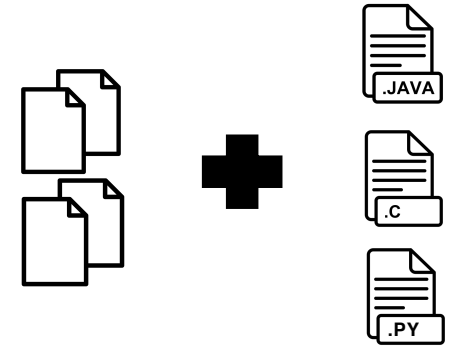


lbrown, db, A
lbrown, os, B
mtom, vis, B+
mtom, db, A-
...
...

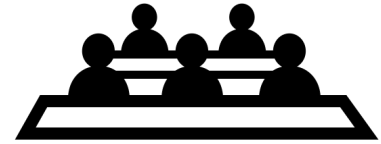
Other Solution: Flat Files



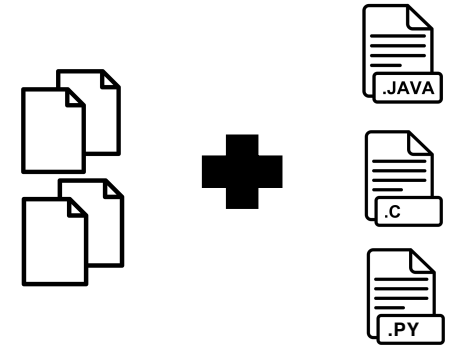
- Problems?



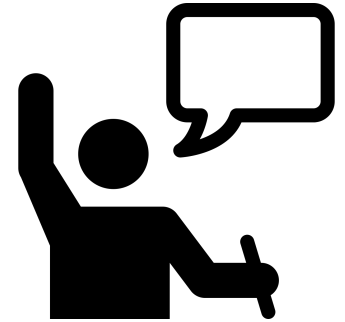
Other Solution: Flat Files



- Problems?
 - ✧ Inconvenient access to data
(requires programming experience
and knowledge of file layout)
 - ✧ Data redundancy
 - ✧ Integrity problems
 - ✧ Atomicity problems (concurrent
access issues)
 - ✧ Security problems

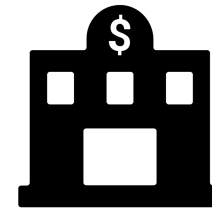


Who uses a DBMS?



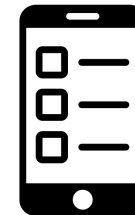
Who uses a DBMS?

- Everyone!
 - ✧ Your bank
 - ✧ Your university
 - ✧ Your coffee shop
 - ✧ Your favorite hotel
 - ✧ Your favorite website
 - ✧ Your phone
 - ✧ Your government
- How many databases have you used so far today?



amazon

 canvas



Why Study Databases?

- Data is useless without the tools to extract information (queries)

✧ “Optimal” pricing of an airline ticket

Select your departure to Cancun Fri, Jan 8

Prices are one way per person, include all taxes and fees, but do not include baggage fees.

Filter your results by

Stops

☐ Nonstop (5)

☐ 1 Stop (54)

☐ 2+ Stops (1)

Airlines included

☐ American Airlines (17)

☐ Delta (14)


☐ Aeromexico (12)

Show all


Departure time

☐ Morning (5:00a - 11:59a)

☐ Afternoon (12:00p - 5:59p)

From:	\$315		10:00a - 7:00p	9h 0m	1 stop	\$230.07
	\$231		Air Canada	DTW - CUN	3h 40m in YYZ	✓ Live one way
	\$679		Air Canada 8022 operated by Air Canada Express - Jazz			
			Air Canada 1812 operated by Air Canada Rouge			

[Flight details and baggage fees ↗](#)

From:	\$247		7:05a - 7:00p	11h 55m	1 stop	\$230.07
	\$542		Air Canada	DTW - CUN	6h 35m in YYZ	✓ Live one way
	\$351		Air Canada 7281 operated by Air Canada Express - Air Georgian			
			Air Canada 1812 operated by Air Canada Rouge			

[Flight details and baggage fees ↗](#)

- Datasets increasing in diversity and volume

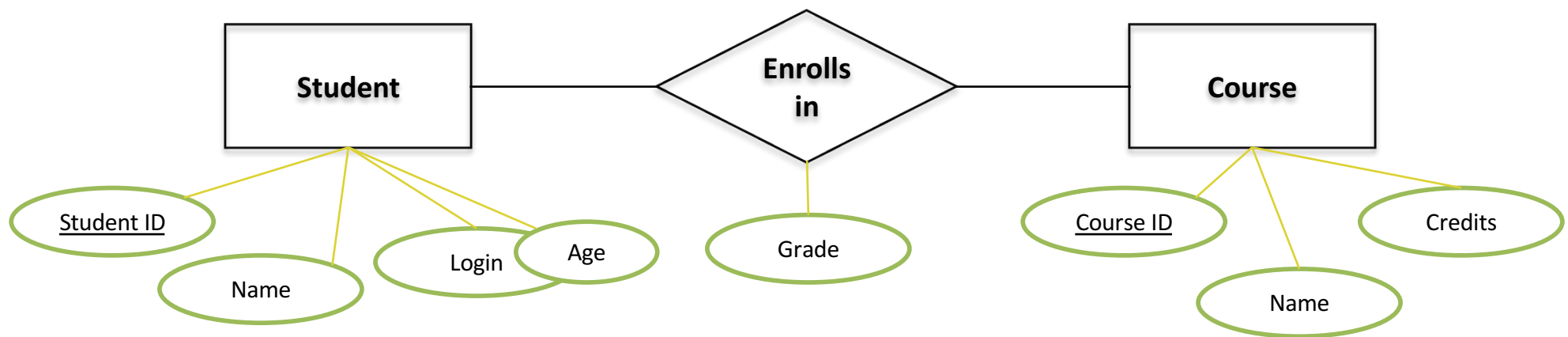
- ✧ Websites, digital libraries, interactive video, human genome project, mobile applications



- Databases touch most of CS

✧ OS, languages, theory, AI, multimedia, logic, ...

University Database Example



This is called an Entity-Relationship (ER) Model
Entities: Student, Course
Relationship: Enrolled_in

University Database

- **Relational representation.** Store data in simple tables or *relations*.
 - ✧ Students(sid:integer, name:string, login:string, age:integer)
 - ✧ Courses(cid: integer, cname:string, credits:integer)
 - ✧ Enrolled(sid:integer, cid:integer, grade:string)

Students

sid	name	login	age
13	Lisa	lsimp	40
41	Bart	bart	20

Courses

cid	cname	Cr.
1	EECS484	4
2	EECS584	3

Enrolled

sid	cid	Grade
41	1	A-
13	2	A+

CYU

- In the previous slide, what course(s) has Lisa taken?
- In the previous slide, what course(s) has Bart taken?
- If Bart were to take both courses, what row would you add to the Enrolled table?

Data Models

- **Data model**: a collection of concepts for describing data.
- **Schema**: a description of a particular collection of data, using a given data model.
- **Relational model**: the most widely-used model today.
 - ✧ Data model: Database is a collection of **relations**
 - ✧ A relation is a table with rows and columns.
- **Entity-Relationship (ER) model**: A “semantic” data model, i.e. a higher-level more user-intuitive model
 - ✧ A (relational) DBMS only understands the relational model →
Must translate an ER schema to a relational schema

Relational and Other Data Models

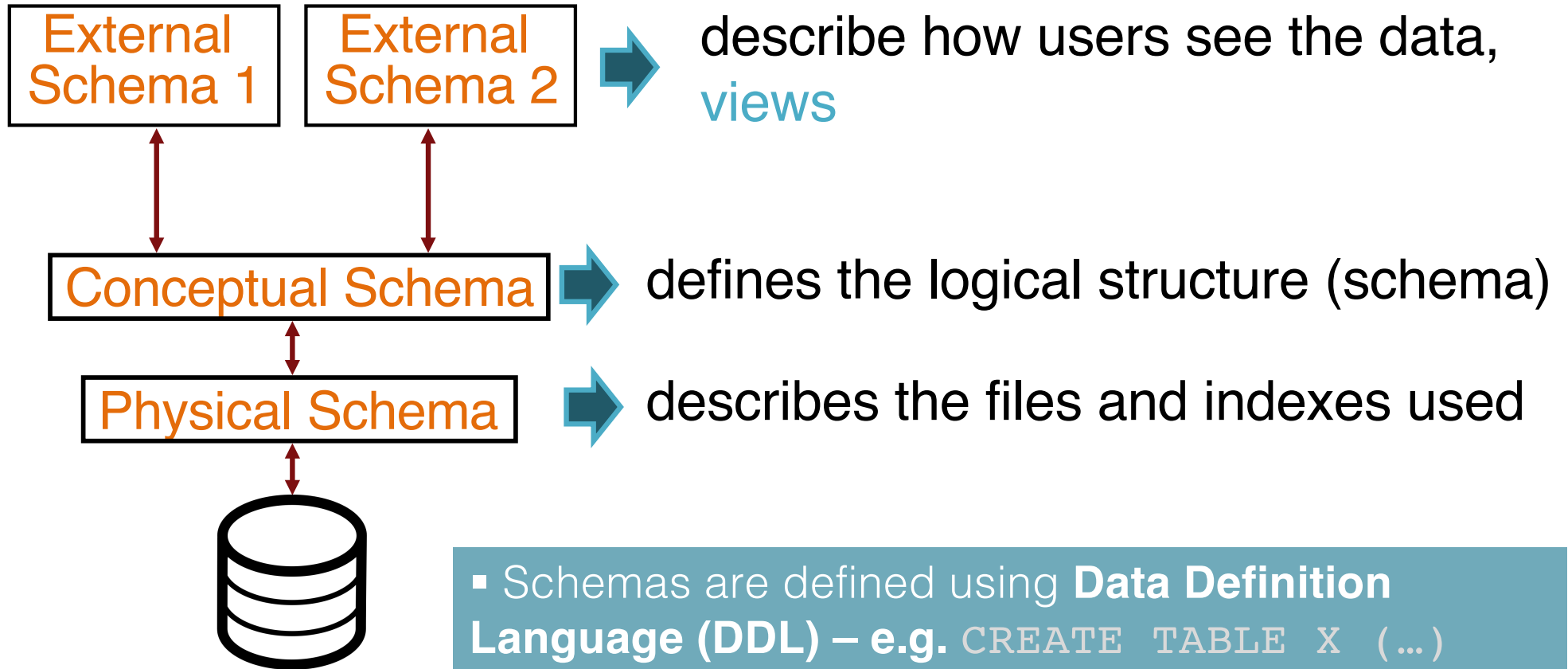
- **DBMS using the relational DM**
 - ✧ IBM DB2
 - ✧ Informix
 - ✧ Oracle
 - ✧ Sybase
 - ✧ Microsoft Access
 - ✧ Tandem
 - ✧ Teradata
 - ✧ ...
- **Other data models**
 - ✧ Hierarchical
 - IBM IMS
 - ✧ Network
 - IDMS, IDS
 - ✧ Object-oriented
 - ObjectStore
 - ✧ Object-relational
 - Oracle
 - ✧ ...

Why use a DBMS?



- It solves ALL the problems with other ad hoc solutions such as files of records, coding data structures, etc.!
 - ✧ Data independence
 - Apps need a view of the data, not info about internal representation and storage
 - ✧ Efficient storage and access
 - ✧ Centralized data administration
 - ✧ Data integrity and security
 - ✧ Concurrent access, recovery from crashes
 - ✧ Reduced application dev time

Levels of Abstraction

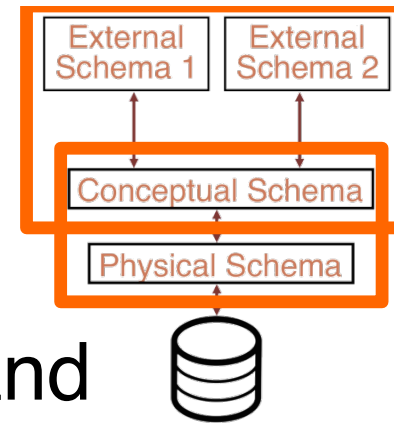


- Schemas are defined using **Data Definition Language (DDL)** – e.g. `CREATE TABLE X (...)`
- Data is modified/queried using **Data Manipulation Language (DML)** – e.g. `SELECT FROM X WHERE ...`

Example: University Database

- Conceptual schema:
 - ✧ Students(sid: string, name: string, login: string, age: integer)
 - ✧ Courses(cid: string, cname: string, credits: integer)
 - ✧ Enrolled(sid: string, cid: string, grade: string)
- Physical schema:
 - ✧ Relations stored as unordered files.
 - ✧ Index on first column of Students.
- External Schema:
 - ✧ View: Course_info(cid: **string**, enrollment: **integer**)
 - ✧ View: Class_rank(sid: **string**, gpa: **real**, rank: **integer**)

Data Independence



- Applications insulated from data format and storage details
- **Logical data independence:** Protection from changes in *logical* structure of data
 - ✧ External / Conceptual schemas
- **Physical data independence:** Protection from changes in *physical* structure of data
 - ✧ Conceptual / Physical schemas

Other key benefits:

- Declarative query processing
- Transactions

CYU

- Which of these are more suitable for storing in a DBMS rather than files in an OS?
 - (a) Grades for students at the university
 - (b) Source code for a program
 - (c) Contents of a textbook



CYU

- Let's say UM provides you access to a relational table that gives just your grades in various courses. Does that relation represent:
 - a) An external schema?
 - b) A conceptual schema?
 - c) A physical schema?

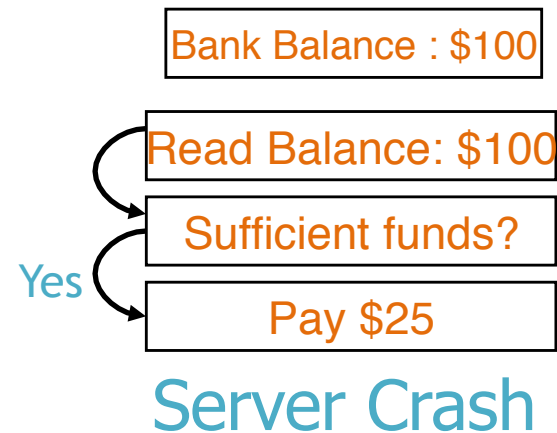
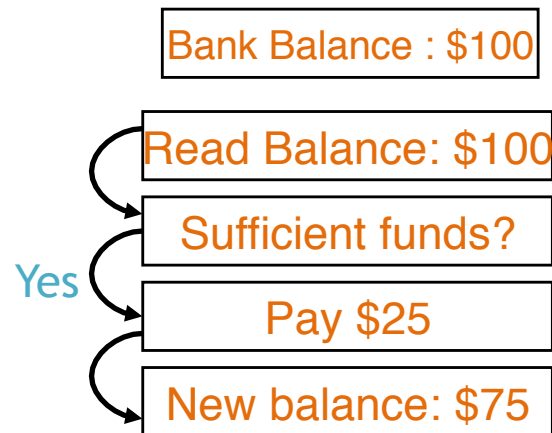
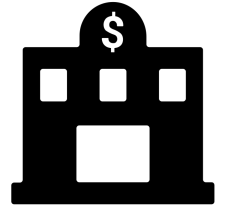


CYU

- The relational table with student grade information is very large and stored on multiple servers for performance. Does the storage scheme represent:
 - a) An external schema?
 - b) A conceptual schema?
 - c) A physical schema?

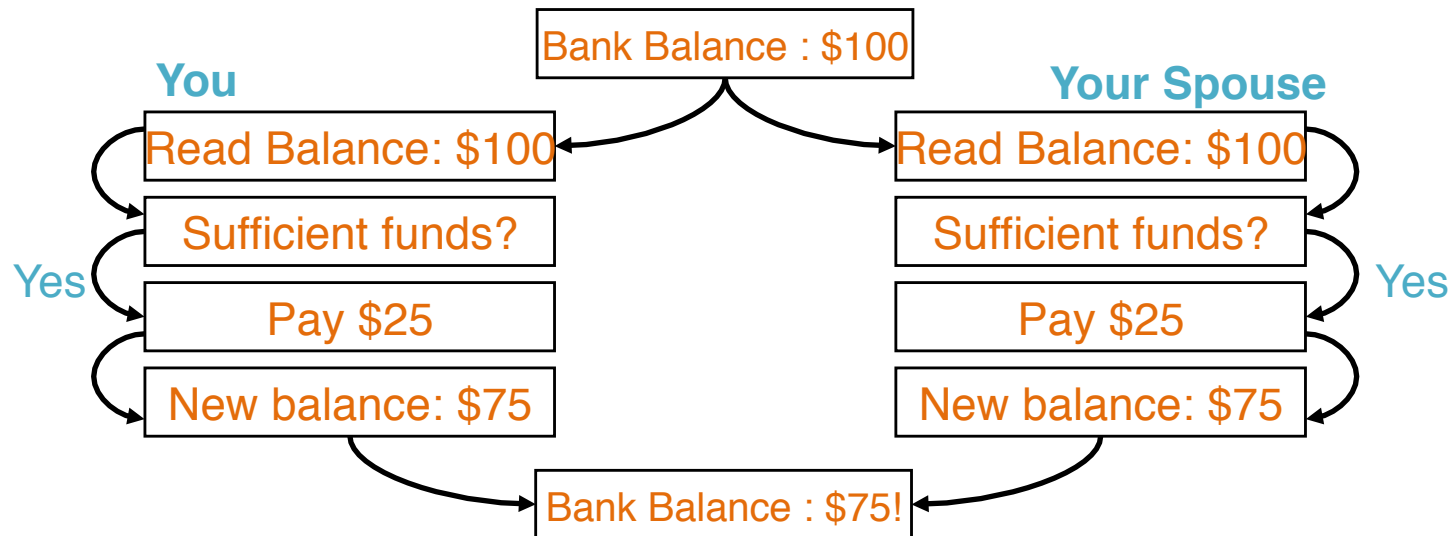
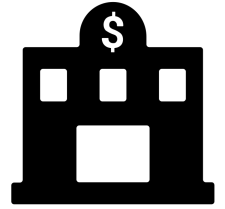


Transactions (I)



- Transaction: any one execution of a user program in a DBMS
- Inconsistency caused by incomplete operations
- DBMS ensures atomic operations!
 - ✧ i.e. all or nothing!
 - ✧ Automatic recovery from crashes!

Transactions (II)



- Inconsistency caused by interleaving actions of different user programs
- DBMS provides the illusion of a “single-user” system
 - ✧ Key concept: **Transaction**, an atomic sequence of R/W
 - ✧ Concurrency control, transaction management

Lots of People use DBMS ...

- DBMS vendors
- DB application programmers
 - ✧ e.g. smart webmasters
- **Database administrator (DBA)**
 - ✧ Designs logical /physical schemas
 - ✧ Handles security and authorization
 - ✧ Data availability, crash recovery
 - ✧ Database tuning as needs evolve



Must understand how a DBMS works!

Summary

- DBMS used to maintain & query large datasets
- **Benefits** include
 - ✧ recovery from system crashes,
 - ✧ concurrent access,
 - ✧ quick application development,
 - ✧ data integrity and
 - ✧ security
- Levels of abstraction give data independence
- DBAs hold responsible, interesting, well-paid jobs



Next

- The Entity-Relationship (ER) Model

