Overview of DBMS and Topics

Learning Objectives

- What is
  - A database? Why use them?
  - A database schema? Example?
  - Relational model?
  - Entity-Relationship (ER) model?
- What are the differences (if any) between:
  - Conceptual schema
  - Logical schema
  - Physical schema
  - External schema
  - Views
- What is data independence?

What Is a DBMS?

- Database: Large, structured collection of data.
- A database models data of some real-world enterprise
  - Entities (e.g., students, courses)
  - Relationships (e.g., Lisa Simpson is taking EECS 484)
- DBMS: Database Management System
  - A software package designed to store and manage databases
  - Often (loosely) called a database

Example Problem

Spreadsheets are commonly used to represent structured data. Try figure out the spreadsheet design to capture the following information about students for a Facebook-like system at a University:

- Students: Last Name, First name, Major, Home City
- Friends relationship: Student1, Student2

Old-time Solution: Sorted Student Folders

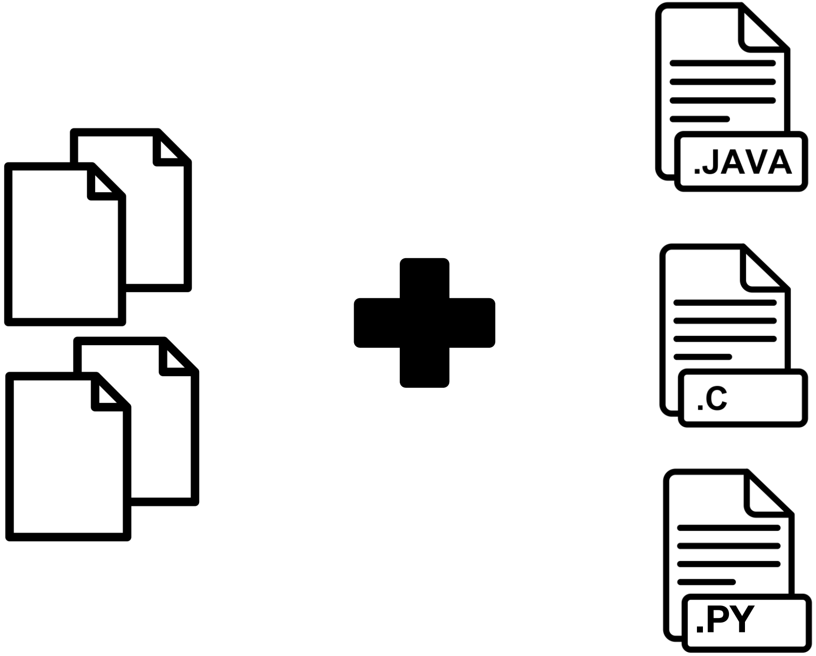
- Advantages?
- Disadvantages?

Old-time Solution: Sorted Student Folders

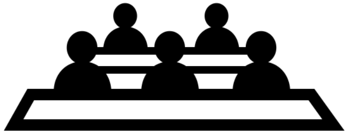
- Advantages?
  - cheap, universal, few dependencies
- Disadvantages?
  - Large weight & volume
  - Difficult to share
  - No ad-hoc queries

Other Solution: Flat Files

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



Other Solution: Spreadsheets?



- Maintain all student records in a giant spreadsheet?
- Maintain each student record in a different spreadsheet?

Other Solution: Flat Files

- Problems?
  - Inconvenient access to data
  - Potential data redundancy
  - Integrity problems
  - Atomicity problems (concurrent access issues)
  - Security problems

Why use a DBMS?

- It solves ALL these problems!
  - 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

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

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.
- Entity-Relationship (ER) model: A “semantic” data model, i.e., a higher-level more user-intuitive model
  - A (relational) DBMS understands only the relational model, so we will translate an ER schema to a relational schema

Relational and Other Data Models

- **DBMS using the relational DM** ('70s-'80s)
  - IBM DB2
  - Informix
  - Oracle
  - Sybase
  - Microsoft Access
  - Tandem
  - Teradata
  - …

- **Other data models**
  - Hierarchical (mid '60s-'70s)
    - IBM IMS
  - Network ('70s)
    - IDMS, IDS
  - Object-oriented (~'90s)
    - ObjectStore
  - Object-relational (relational model + object DB concepts)
    - Oracle
  - …

Relational (Data) Model

- The most widely-used model today
  - A collection of relations
  - Relation = set of records, naturally represented as a table with rows and (named, typed) columns
  - Each table row is also called a tuple or record

Students

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

Courses

cid	cname	cred.
E-484	EECS484	4
E-584	EECS584	3

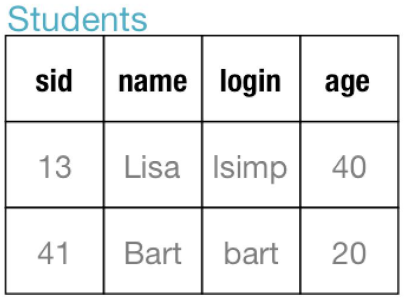
Enrolled

sid	cid	grade
41	E-484	A-
13	E-584	A+

Relational (Data) Model

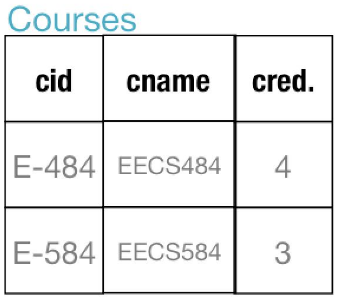
- Schema = a description of data in terms of a data model
  - Every relational database has a schema
  - Specifies the name of each relation, the name and type of the columns (or fields or attributes)

Students(sid:string, name:string, login:string, age:integer)
Courses(cid:string, cname:string, credits:integer)
Enrolled(sid:string, cid:string, grade:string)



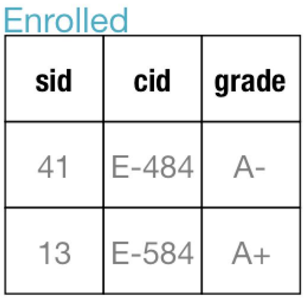
Students

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



Courses

cid	cname	cred.
E-484	EECS484	4
E-584	EECS584	3



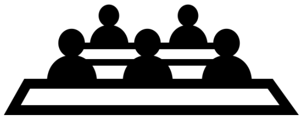
Enrolled

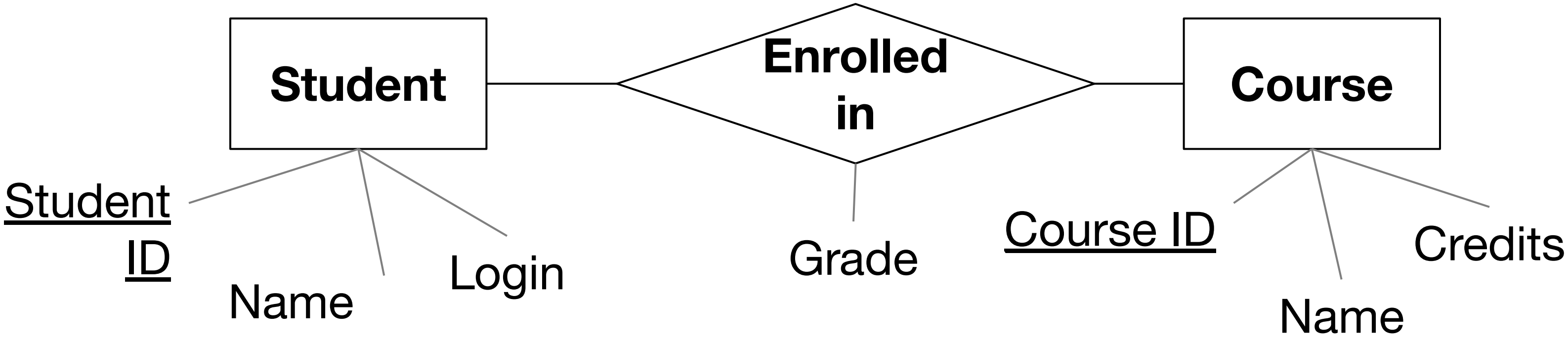
sid	cid	grade
41	E-484	A-
13	E-584	A+

484

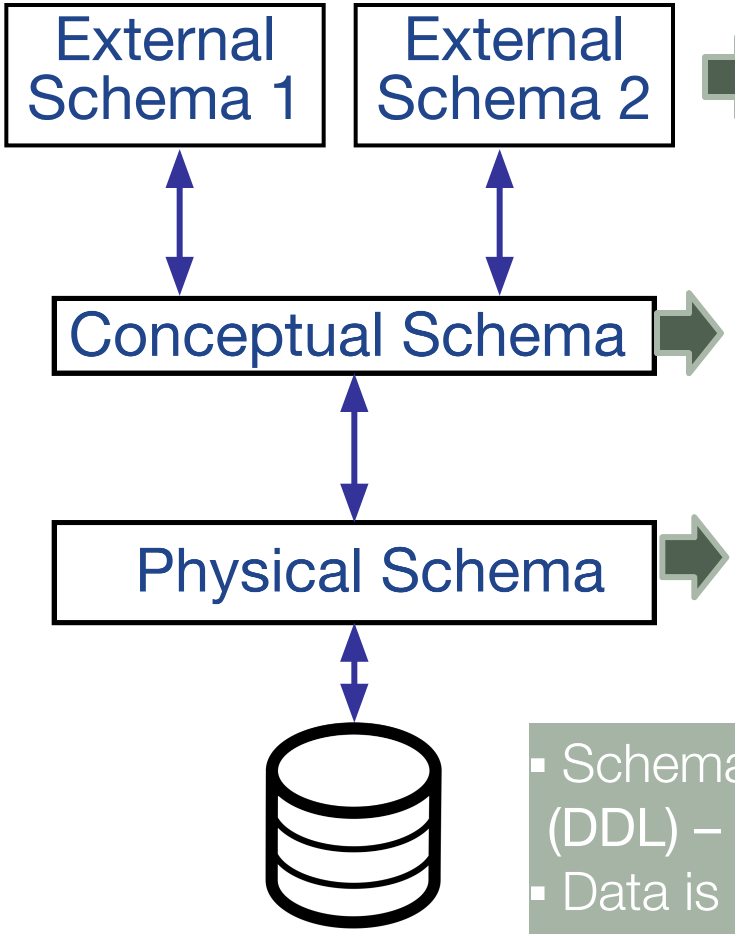
Entity-Relationship (ER) Model

- A "semantic" data model
  - a higher-level, user-intuitive model
- Entity-Relationship diagram:
  - Entities: Student, Course
  - Relationship: Enrolled_in





Levels of Abstraction



External Schema 1

External Schema 2

Conceptual Schema

Physical Schema

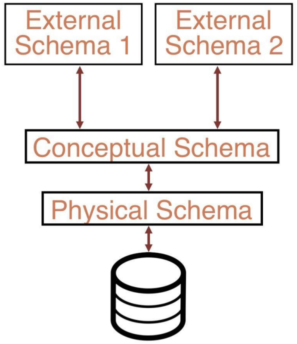
describe how users see the data. Also called **views.**

defines the logical structure. Also called **logical schema**.

describes the files and indexes used

- 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 …`

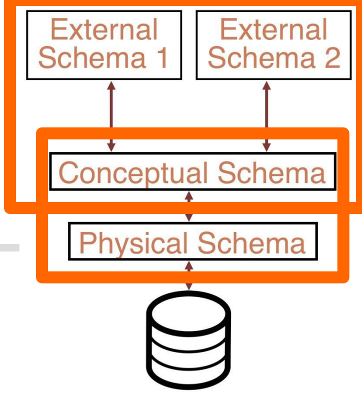
Example



- Logical schema (1):
  - Students(sid:string, name:string, login:string)
  - Courses(cid:string, cname:string, credits:integer)
  - Enrolled(sid:string, cid:string, grade:string)
- Physical schema (1):
  - Relations stored as unordered files.
  - Index on first column of Students.
- External Schema or View (≥ 1):
  - Course_info(cid:string, cname:string, enrollment:integer)
  - Student_Credits_Enrolled(sid:string, totalcredits:integer)

Views can be computed from the relations in the logical Schema

Data Independence



- Applications insulated from data format and storage details
- Logical data independence: Protection from changes in logical structure of data
  - External / Logical schema interface
- Physical data independence: Protection from changes in physical structure of data
  - Logical / Physical schema interface

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

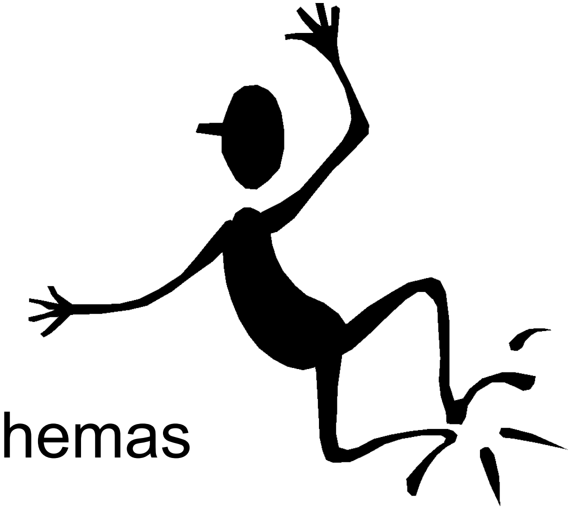
- Let's say UM provides you access to a relational table that gives just your grades in various courses. Does the design of that relation represent:
  a) An external schema?
  b) A conceptual schema?
  c) A physical schema?
  d) A logical schema?

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?
  d) A logical schema?
  e)

Lots of People use DBMS ...

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



DBA 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 jobs and are well-paid! 🤑
- DBMS R&D is one of the most exciting areas in CS. 🤭

Please read Chapter 2 (26 pp)