What is a Database?

The Lay View

- "A database is any collection of information"
- Might be a simple file
- Might not even be on a computer!
- Far too broad, but it's hard to fight everyday usage

An extreme operational definition

- "A database is a structured collection of information created, managed, and accessed using one of a number of software packages commonly accepted as Database Management Systems."
- Too limited, but on the right track.
- "structured" and "using a DBMS" are fairly essential

Common Characteristics

These are not found in every DB, but most are found in most DBs

- Large size
- Multiple files
- Well-structured data
  - not free-form text, not large binary blobs
- Many instances of a few structures
- A snapshot of the real world

More Characteristics

- Many users
  - both end-users and programmers
  - not viewed directly by end-users
- Under administrative control
  - separate from end-users and programmers
- “Data processing” rather than “computer science” applications

Characteristic applications are large, long-lived, and critical to the organization

The Database Management System (DBMS)

- Definition: *Software components for creating, administering, and managing access to the database.*
- "Below" the DBMS: file system, operating system, networking, hardware
- "Above" the DBMS: DB administrator, application programs.
- "Above" the application programs: end-users.

PCs vs Mainframes

- PCs: The same person may act as administrator, programmer, end-user, etc.
- Mainframes (or big applications generally):
  - normally a sharp division of human roles
  - DB and the DBMS may be part of an even more complex software/hardware arrangement
    - multiple computers, multiple platforms, client/server, networks, tie-ins to other software systems...

Some Classic DB Topics

- Data modeling
  - Look at the real world and the application, identify pieces of data needed and their relationships
  - *We will study "E/R modeling"*
- Database design
  - Turning the model into a structure supported by the DBMS
  - *We will focus on the "relational" model*

Classic Topics (continued)

- Query Languages
  - Languages designed to support "queries" (requests for data) and other DB operations
  - *We will study the formal "relational algebra" and the practical language SQL.*
- Transaction processing
  - Managing user-activated programs, to guarantee service and preserve DB integrity.
  - *We will learn the key issues and terminology.*

Classic topics (concluded)

- File storage and structures
  - How data is organized on disk for efficient access.
  - *In particular we will learn about a data structure called a "B-tree."*
- Query processing (time permitting)
  - How complex queries can be internally reorganized for efficiency

At Least One Modern Topic

- Object-oriented DB
  - Complex data types (sound, images, time-series, etc.)
  - "Persistent" objects
  - Integration with OO languages like C++, Java
  - *Theory and practice are both in flux*

Database Soup

- 4GL RAD RAID R/3 Datamart ODMS CORBA COM JDC SAP MDM OLAP...
- *What's going on out there?!* Schedule permitting, we'll use a little class time to sort through some current commercial developments.
- Your papers could also stir through the soup
- Your projects might use some of that stuff...

As you've probably noticed...

- DB has links with many other areas of Computer Science:
  - Operating Systems
  - Software Engineering
  - Computer Architecture
  - Programming Languages
  - etc., etc.

Try to spot the connections!