

Lecture 1 : Introduction.

See Folk + Zoellick chs. 1, 2.1, 3.4

This course covers data processing from a computer science perspective:

- Storage of data,
- Organization of data,
- Access to data,
- Processing of data.

Motivation: Most computers are used for data processing (over \$60 billion/year)
A big growth area in the "information age."

Course Outline

- Storage devices (disk & tape)
- File systems
- File organization
 - sequential (tape)
 - direct (hashing)
 - indexed sequential (B-trees)
 - multi-key (secondary indices)
- Intro to database systems

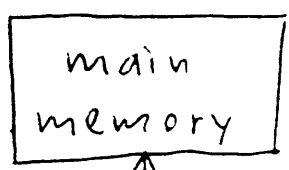
Most of
this course.

Examples of file systems (non-computerized):

- Telephone book (primary index on name)
- Library (primary index on number, and secondary indices on author, title, subject)

Computer Architecture

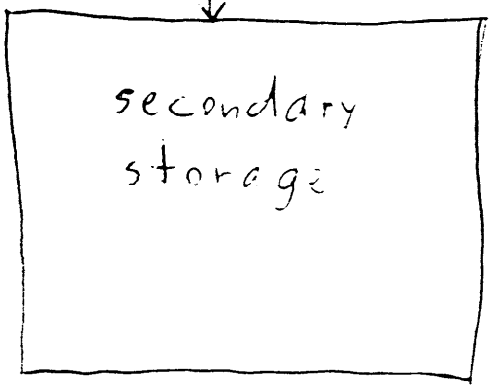
Data is manipulated here.



- semiconductors
- fast expensive, volatile small

data transfer

Data is stored here.



- disks, tape
- slow, cheap, stable, large.

Programs are executed in main memory (MM) because it is fast

Advantages

- (1) Main memory is fast.
- (2) Secondary storage is big (because it is cheap).
- (3) Secondary storage is stable (non-volatile).
ie, Data is not lost during power failures.

Disadvantages

- (1) Main memory is small (because it is expensive). Many databases are too large to fit in M.M.
- (2) Main memory is volatile.
ie, Data is lost during power failures.
- (3) Secondary storage (SS) is slow.
(10,000 times slower than MM).

Normal Arrangement

- Secondary storage (SS) provides reliable, long-term storage for large volumes of data.
- At any given time, we are usually interested in only a small portion of the data.
- This data is loaded temporarily into main memory, where it can be rapidly manipulated and processed.
- As our interests shift, data is transferred automatically between MM & SS, so the data we are focussed on is always in MM.

Problem

- Transferring data between MM & SS is slow (because SS is slow).
- An important goal of database systems is to minimize the amount of data transfer.
- i.e., to eliminate unnecessary transfers.

File Systems

- Data is not scattered hither & thither on disk.
- Instead, it is organized into files.
- Files are organized into records.
- Records are organized into fields.

Example

1-9

- A student file may be a collection of student records, one record for each student.
- Each student record may have several fields, such as
 - Name
 - Address
 - Student Number
 - Sex
 - Age
 - Grade Point Average (GPA)
- Typically, each record in a file has the same fields.

- Files are large and are stored in SS.
 - Records we are currently interested in are copied into MM.
 - Organizing the records of a file, and getting at the records we are interested in, are the main topics of this course.
-

- Files have three main properties that make them suitable for databases.

...

Properties of Files

- (1) Persistence: Data written into a file persists after the program stops, so the data can be used later.
- (2) Sharability: Data stored in files can be shared by many programs and users simultaneously.
- (3) Size: Data files can be very large. Typically, they cannot fit into M.M.