IS 300 — Lecture 9

- What is the terminology of data management?
- What is the traditional file environment?
- What is the modern database environment?
- Where should databases be physically stored?
- What are some recent developments in the database area?

What is the terminology of data management?

Terminology

Byte

Field

Key Field

Record

File

Database

What is the traditional file environment?

Definition: a set of separate, application-oriented systems and support files. Functional orientation.

Student Information System – Student file
Payroll System – Payroll file

Why is this a common approach?

Major problems:

Data Redundancy Problem. The same field or record is stored in two or more files. Results in

- Excess data storage
- Multiple updates which, if done wrong, leads to data inconsistencies

Data Integration Problem. Data needed to answer a query is scattered in several different files and it is difficult or impossible to integrate it.

- "ad hoc" queries
- Give me a list of courses currently being taken by student employees

Key field problems:
ASUW in student file
SSN in payroll file

Program/Data Dependence Problem. Programs describe data they use.

Example:

```
EmpRec = record
        Name string[30]
        Address
                Street string [20]
                City string [12]
                State string [2]
                Zip string [5]
```

If data changes (zip changes to 9 characters), all programs that use the data need to be changed too.

Maintenance costs and potential for errors increase.

What is the modern database environment?

Philosophy: A collection of data organized to serve many applications. A common pool of data. Generally subject (not function) oriented, e.g., employees, customers, suppliers, and components. Data models used to understand how to organize the data.

Technology: DBMS. Software used to create and maintain database and to enable individual business applications to extract the data they need.

Supporting technologies.

- Data Definition Language – relationships and cardinality
- Data Manipulation Language – SQL or QBE
- Data Dictionary – defines the data elements

Where should databases be physically stored?

Centralized database: All data stored in one location.

Distributed database: The database is stored in more than one physical location. Options include:

Partitioned: Split the database into unique subsets and store subsets at different locations.

Replicated: duplicate the database at multiple locations.

Tradeoffs

- Storage requirements
- Data communication costs
- Keeping data current and consistent
- Reliability (vulnerability to data loss)
- Access time

What are some recent developments in the database area?

Data Warehouses. A relational DB designed to support managerial decision making.

Contain historical operational data and data from external sources.

Data Marts. Smaller, more specialized versions of data warehouses.

Data Mining. Automated discovery of patterns and relationships within the data.

Uses statistical techniques and machine learning.

Example applications:

- Fraud detection
- Market basket analyses
- Trend analyses

OLAP. On-Line Analytical Processing. Tool to support high-speed analyses of data involving complex relationships.

Drill down
Multidimensional (Product by Region by Year by Salesperson)