DATABASES AND INFORMATION SYSTEMS

Introduction to concepts, examples, terminology

Examples of Information Systems

- **Payroll system**: handles payments of employees in an organization

- **Sales order processing system**: handles supply/orders of products

- **Project planning and control system**: handles time/resource requirements of activities related to a project

- **Geographic information system**: handles geo-spatial information (maps, satellite images, statistical data, etc. associated with a given geographic area)

Types of Information Systems

- Transaction processing systems: process individual operations in a business (e.g., marketing, financial operations, etc.)

- Decision-support systems: aid decision making of management (e.g., where to build a factory, which products to sell etc.)

- Expert (knowledge-based) systems: attempt to simulate the role of the human "expert"; use knowledge base of a given domain to provide solutions (e.g., diagnose reasons for failure of a given business)

- Office automation systems: handle office applications such as email, meeting management, fax transmission, etc.

- Etc.

Information System

- Component of an organization that manages (gets, processes, stores, communicates) the information of interest

    - each organization has an information system

    - usually, the information system operates in support to other components of the organization

Management of Information

Information is handled and recorded according to various techniques:

- – informal ideas
- – natural language (written or spoken)
- – drawings, diagrams,
- – numbers
- – codes

Structured information

- Several different forms of organization and codification for information have been devised

- Example: information about people

  – in most countries a structure for the name has been introduced in the last few centuries
  – later, it was realized that it could be useful to keep track of birthdate and birthplace (and use them in order to identify people, together with the name)
  – social security numbers (or tax codes) have been introduced in order to obtain unique identification

Information and data

• In computer- based systems information is represented by means of data

   – raw facts, to be interpreted and correlated in order to
     provide information

• An example:

   – "John Smith" and 25755 are a name (or, better, a string) and a number: two pieces of data

   – if they are provided as a reply to a request:
     "Who is the head of department, and what is his/her telephone extension,"
     then we get information out of those two pieces of data

Components of Computer-Based Information Systems (IS)

- Database (DB)

- Database software

- Application software

- Computer hardware (e.g., storage media)

- Personnel using and developing the system

NOTE: the DB is a fundamental component

DB and DBMS

- DB: collection of logically related data of interest to the IS

- DB is managed by the **Database Management System (DBMS)** = software component that interacts with the DB and the user application programs

Database Management System (DBMS)

Software system able to
1. manage collections of data that are:

- <u>large</u> (bigger, often much bigger, than the main memory available)

- <u>shared</u> (used by various applications and users)

- <u>persistent</u> (with a lifespan that is not limited to single executions of the programs that use them)

2. ensure their reliability (so preserving the database in case of hardware or software failure) and privacy (controlling accesses and authorizations). Like any software product, a DBMS must be <u>efficient</u> (using the appropriate amount of resources, such as time and space) and <u>effective</u> (supporting the productivity of its users).

Sharing

• Most organizations have a structure (departments, divisions, …) and each component is interested in a portion of the information system

• The data of interest of the various components often overlap

• A database is an <u>integrated resource</u>, shared by various components

• Integration and sharing allow a reduction of <u>redundancy</u> and the consequent possibility of <u>inconsistency</u>

• Since sharing is never complete, DBMS provide support for privacy of data and access authorizations

• Sharing also requires that multiple accesses to data are suitably organized: <u>concurrency control</u> techniques are used

Example: Online Bookseller

Data about **books** (title, authors, categories, etc.), **customers, orders, order histories, trends and preferences**, etc.

<u>Very large</u> amounts of data (too big for memory)
− 100s of gigabytes, more if we keep all order histories over time, clickstream logs (for user preferences), images of book covers and sample pages , etc.

<u>Persistent</u> data
− Data outlives programs that operate on it

<u>Safe</u> data
− From hardware failures, software failures, power outages, ...
− From malicious users

<u>Multi-user</u> access
− Many people/programs access same database/data simultaneously (concurrency)

File-Based System

- Collection of application programs that perform services for end-users (e.g., production reports)

- Each program defines and manages its own data (stored in a file)

- Used as first attempt to computerize the traditional manual filing system within an organization (e.g., office, etc.)

File Systems

- Allow to manage large and persistent sets of data

- Do not provide much support for data sharing

  ⇨ Replaced by DB systems

NOTE: DB systems use files to store the data but they provide users with abstract view of data, trying to hide storage and manipulation details

Database languages

Various forms:

1. Interactive textual languages, such as SQL

2. Interactive commands <u>embedded</u> in a <u>host</u> language (C, Java, etc.)

3. Interactive commands <u>embedded</u> in a <u>ad-hoc</u> development language, usually with additional features (for the production of forms, menus, reports, ...)

4. By means of non- textual <u>user-friendly</u> interfaces

A distinction

- Data definition language (DDL):
    Used to define the logical, external and physical schemas and access authorizations

- Data manipulation language (DML):
    Used for querying and updating database instances