Databases

Data collection: a set of data collected for a specific purpose
Data: facts that can be recorded and have explicit meaning
Entity: an object, person, event, system or place about which we need to store data. We describe an entity in terms of its **attributes** (representative features).
Record: the set of data about one entity

Data requirements: data should be

- *Well organized*
- *Accurate*: data must be correct in order to be useful.
- *Up-to-date*
- *Secure*: measures should be taken so that data can be restored from backup copies in case it has been lost or damaged.
- *Accessible*: only authorized users should be able to access data, and they should be able to access it at any time.

Data management operations

Data management: a set of procedures used to ensure that data satisfies the above requirements. Data management procedures include:

- design of structures to hold data
- creating the structures
- entering the data
- accessing the data
- modifying the structures
- updating the data (adding, deleting, changing)
- processing the data
- destroying the structures

Data organization:

- Manual
- Computer-based

Computer data organization

- Files
- Databases

Database: a collection of related data

A typical database:
- represents some aspect of the real world
- is used for specific purposes
- by one or more groups of users.

Examples of databases:
1. Personnel database
   Files: Employees, Departments, Projects
2. Library database
   Files: Books, Borrowers, Loans, Librarians
3. National census database
   Files: Person, Dwelling, Household
4. Airline reservations
   Flights, Airports, Planes, Crew

Data Base Management System (DBMS): a collection of programs for creating and using a database

Database System = database + software
Database software = DBMS + application programs

Database schema: description of a database

Organization of databases

- Logical organization: the organization of data perceived by users
- Physical organization: the organization of data on disk files

DBMS features

- Handle huge quantities of data
- Flexible (Data dictionary & program-data independence)
- Query languages
- Interfaces
- Security
- Recovery
- Utilities (creating files, monitoring performance)

Classification of DBMSs according to the number of end users:

- single-user
- multi-user

Classification of DBMSs according to the data model used:
- 1st generation (1950s) simple file systems
- 2nd generation (1960s) Hierarchical & network models
- 3rd generation Relational databases
- 4th generation (1990s) Post-relational models (OO, WWW, e-commerce)

Database users:

- End-users
- Database designers
- Database administrator (DBA)

  DBA is the person with highest responsibilities and privileges in the database world. DBA responsibilities: participating in database design, user coordination, backup and recovery, system security and performance monitoring.
- Applications developers
- DBMS developers

Database management operations

- **Access – browse**: look at the content of a database.
- **Access – query**: select data from the database that match the user's request.
- **Design**
  Database design is the process of analyzing the area in which the database is needed and the typical functions of the database in order to identify the appropriate structures to hold data. It is a complex, time-consuming and important process.
- **Create**: create the tables and additional structures (i.e. indexes) to hold data.
- **Populate**: enter the data into tables.
- **Update**: change the context of the database. Types of update operations:
  - Add new tuples
  - Modify the value of an attribute
  - Delete tuples
- **Alter**: Change the definition of the database (its schema/model)
- **Develop applications:** Develop specific applications that automate some of the tasks (data entry, report printing, retrieval etc).

Database applications

- General business applications
- Multimedia, time-series and historical data management
- Engineering databases
- GIS: Geographical Information Systems
- Human Genome Initiative
- Electronic commerce
- Data mining (AI)

Advantages of databases

- Centralised control (Improved data integrity and security)
- Flexibility (Program-data independence)
- Reduced data redundancy
- Reduced development time
- Sharing of data among programs (concurrency)
- Multiple user interfaces

Disadvantages of databases

- complexity
- size (require more memory space)
- cost (both hardware and software cost)
- reduced speed
- higher impact of failure

When not to use databases:

- Small, well-defined, stable data sets
- Strict real-time requirements
- Multi-user access unnecessary
- When no DBMS may suffice
    - Complex data
    - Special operations required