# Introduction to Databases

## Introduction

DBMG

# Introduction to databases

- Information management
- Databases
- Data model
- Data independence
- Data access
- Advantages and disadvantages of DBMS

# Introduction to databases

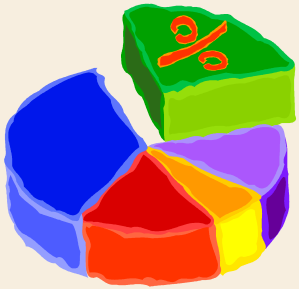
## Information management

DBMG

# Information management

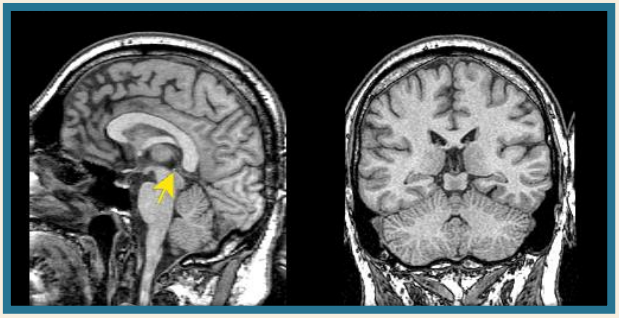
- Information is recorded and exchanged in different forms

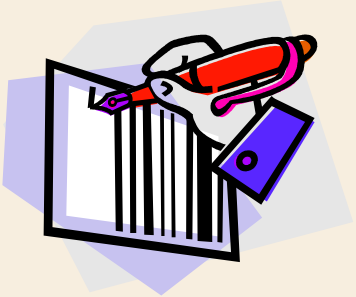




# Information management

- Information is recorded and exchanged in different forms





- Forms of information organization and codification have been introduced over time

BNCLCU79M44A488D

# Computer systems

- In computer systems information is represented by means of data
  - the data are raw symbols which have to be interpreted and correlated to provide information
  - example
    - data: "Mario Rossi" and 424242
    - information: result of looking up a telephone number in your personal telephone directory (e.g., list of contacts)

# Data characteristics

- Data are far more stable over time than the processes that manage them
  - Example
    - there have been no variations in the structure of bank applications data for decades
    - the procedures that manage the data vary from year to year
- Data are an important resource of the organization that manages them

# Introduction to the databases

## Databases

DBMG

# Database

- (General definition)
  - a database is a collection of data that represents information interesting for a computer system
- ("Technical" definition)
  - a database is a collection of data managed by a DBMS

# Data Base Management System - DBMS

- A DBMS (DataBase Management System) is a software system able to manage collections of data that are
  - large
  - shared
  - persistent

  ensuring their reliability and privacy

# DBMS characteristics

- Far greater dimensions than the central memory available
  - data management in secondary memory
- Data sharing between applications and users: a database is an *integrated* resource, shared by several company sectors
  - reduction of data redundancy
  - reduction of data inconsistency
  - competing access control mechanism

# DBMS characteristics

- Data persistence
  - lifetime not limited to execution of programmes that use them
- Data reliability in the case of hardware and software malfunction/failure
  - backup and recovery functionality
- Data privacy
  - authorization mechanisms to enable users

# DBMS characteristics

- Efficiency
  - capacity to carry out operations using a set of resources (time and space) acceptable for users
    - Adequately sized computer system
- Efficacy
  - capacity to render user activities productive

# DBMS or file system?

- "Simplified" approach to data: data stored in the persistent mode in the mass/secondary memory inside the file
  - it is possible to memorize and look for data
  - simple access mechanisms (sequential reading)
  - simple sharing mechanisms (read only sharing with writing options blocked)
- DBMS extends the functionalities of the file systems, providing more integrated services

# Introduction to databases

## Data model

DBMG

# Data model

- A data model is a set of concepts utilized for organizing data of interest and describing its structure in a way that is understood by a computer
  - elementary data types (integer, character...)
  - structuring mechanism for defining more complex structures (record builder, array,...)

# Relational model

- Most widespread data model
- Defines the relationships builder, which organizes the data into sets of homogeneous (fixed structure) records
  - The relationships are represented as tables

# Relational model

Courses

| Code | Name | TeacherID |
|---|---|---|
| M2170 | Information systems | D101 |
| M4880 | Computer Networks | D102 |
| F0410 | Databases | D321 |

Teacher

| ID | Name | Department | Phone# |
|---|---|---|---|
| D101 | Green | Computer Engeneering | 123456 |
| D102 | White | Telecommunications | 636363 |
| D321 | Black | Computer Engeneering | 414243 |

# Other data models

- Before the relational model, other models closer to the physical (not very abstract) structures of storing were used
  - hierarchical model
  - network model
- Since the relational model
  - Object model
  - XML
  - …

# Schema and instances

- Defined in the database are
  - the *schema*, which describes the structure of the data. The schema
    - is practically unvarying over time
    - is represented by the heading of each table (table name and column names)
- Example
  - schema of the database

| | | | | |
|---|---|---|---|---|
| Courses | Code | Name | TeacherID | |
| Teacher | ID | Name | Department | Phone# |

# Schema and instances

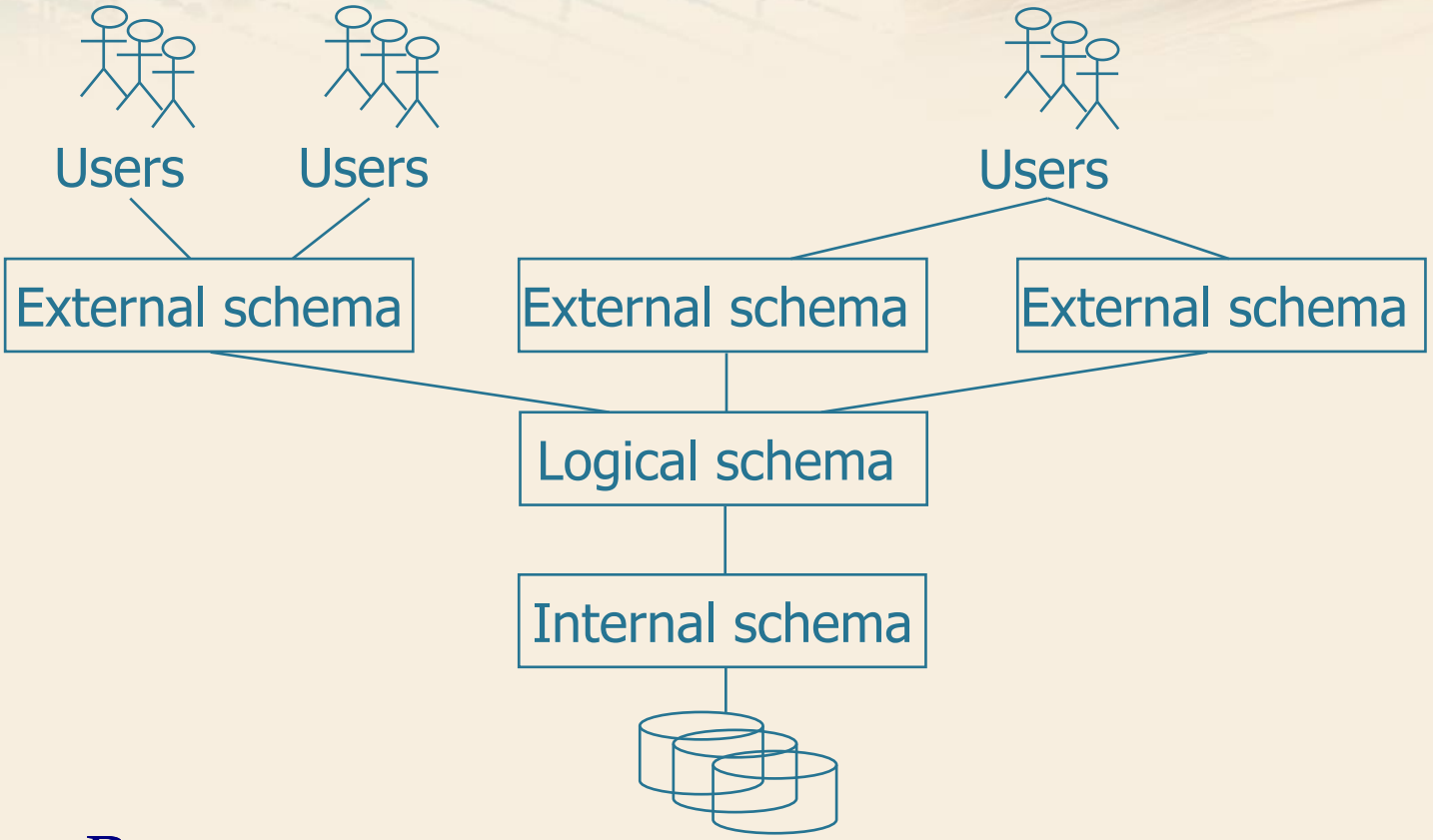
- Defined in a database are
  - the *instance*, composed of the content of each table, i.e. of the data effective values which are
    - variable over time, also very rapidly
    - represented by the rows in the tables
- Example
  - instance of the Teacher table

| | | | |
|---|---|---|---|
| D101 | Green | Computer Engeneering | 123456 |
| D102 | White | Telecommunications | 636363 |
| D321 | Black | Computer Engeneering | 414243 |

# Model types

- *Conceptual model*
  - It is possible to represent data independently from the logical
    - describes real world concepts
    - used in the designing phase
  - example: entity-relationship model
- *Logical model*
  - Describes the data structure in the DBMS
    - used by the programmes accessing the data
    - independent from the physical structures
  - Example: relational model

# Abstraction levels in a DBMS



# Standard three-level ANSI/SPARC architecture for DBMS

- Logical schema
  - description of the database using the logical model of the DBMS
- Internal schema
  - representation of the logical schema using physical storing structures
- External schema
  - description of parts of the database, called "views", which reflect the point of view of particular users
  - defined on the logical model

# Introduction to databases

## Data independence

# Data independence

- Data independence guarantees that users and application programmes which utilize a database can ignore the designing details used in the construction of the database
- It is a consequence of the subdivision into levels of abstraction

# Data independence

- Physical independence
  - enables interaction with the DBMS independently from the physical structure of the data
  - access to a (logical or external level) relationship always takes place in the same way, independently from the means of memorization
  - it is possible to change the way the data is physically memorized without affecting the programmes utilizing the data

# Data independence

- Logical independence
  - enables interaction with the external level independently from the logical level
  - it is possible to change the logical level maintaining the external structures unaltered (as long as the correspondences are unaltered)
  - it is possible to add new views or alter existing views without changing the logical schema

# Introduction to databases

## Data access

DBMG

# Data access languages

- User-friendly interfaces that enable specific queries without using a textual language
- Interactive languages (SQL)
- Commands similar to interactive commands introduced into traditional programming languages (C, C++, COBOL, Java, …), so-called host languages
- Commands similar to interactive commands introduced into ad hoc development languages, often with specific functionalities (generation of graphics, complex prints, screens)

# Data access languages

- Languages are divided into two categories
  - *Data Definition Languages* (DDL) used to define the logical, external and physical schemas, and access authorizations
  - *Data Manipulation Languages* (DML) used for querying and updating database instances

# Users

- Database administrator: in charge of (centralized) control and management of the database
  - guarantees sufficient performance
  - ensures system reliability
  - manages authorizations and access to data

# Users

- Designers and programmers: they define and realize
  - the structure of the database
  - the programmes accessing the database
- Users: utilize the database for their activities
  - end users: they use transactions, i.e. programmes that carry out predefined activities
  - casual users: they formulate queries (or updates) which are not predefined by the interactive access languages of the database

# Transactions

- Programmes that carry out frequent predefined activities
- Examples
  - flight bookings
  - bank transfers
- Generally realized by introducing SQL into a host language

# Introduction to databases

## Advantages and disadvantages of DBMS

DBMG

# DBMS advantages

- Data as a common resource of the whole organization
  - reduction of redundancies and inconsistencies
- Unified and precise data model of facts of interest to the organization
- Possible centralized control of data
  - standardization, economies of scale
- Data independence

# DBMS disadvantages

- These are expensive, complex products that require
  - direct investment
    - purchase of the product
  - indirect investments
    - purchase of the necessary hardware and software resources
    - conversion of the applications
    - training of personnel
- They provide a set of services in an integrated form
  - it is not possible to separate out unused services that cause a reduction in performance