Database Concepts and Technologies

Course Administration

- Lectures
- Labs
- Assignments/Quizes
- Mid Semester Examination
- Text Book:1. Fundamentals of Database Systems: 5th Edition by Elmasri/Navathe

  2. Database Systems: 4th Edition by Connolly/Begg

Lecture 1

Introduction To Databases

Outline

- Types of Databases and Database Applications
- Characteristics of file-based systems.
- Problems with file-based approach.
- Basic Definitions
- Typical DBMS Functionality
- Example of a Database (UNIVERSITY)
- Main Characteristics of the Database Approach
- Advantages of Using the Database Approach
- History of Databases

Basic Definitions

- **Data**
- **Database**
- **Mini-world**
- **Database Management System (DBMS)**
- **Database System**

Basic Definitions

- **Database:** A collection of related data.
- **Data:** Known facts that can be recorded and have an implicit meaning.
- **Mini-world:**
  - Some part of the real world about which data is stored in a database. For example, student grades and transcripts at a university.
- **Database Management System (DBMS):**
  - A software package/ system to facilitate the creation and maintenance of a computerized database.
- **Database System:**
  - The DBMS software together with the data itself. Sometimes, the applications are also included.

Restrictive:

- A database is a persistent, logically coherent collection of inherently meaningful data, relevant to some aspects of the real world.
- ***A database management system (DBMS) is a collection of programs that enables users to create and maintain a database. According to the ANSI/SPARC DBMS Report (1977), a DBMS should be envisioned as a multi-layered system:***

Why Learn Database?

- Databases and database systems are an essential component of life in modern society.
- Most of us encounter several activities every day that involve some interaction with a database.

Example:

- Bank: if we go to the bank to deposit or withdraw funds
- Reservation Systems: if we make a airline reservation
- Library: if we access a computerized library catalog to search for a bibliographic item
- Online Purchase: if we purchase something online—such as a book, toy, or computer
- In all the above cases chances are that our activities will involve someone or some computer program accessing a database

Why Learn Database?

- Databases are everywhere and so important that computer science or information technology graduates must know the concepts of database systems.
- It will help them in building their industry careers (Database Administrator, Database Designer or Developer).

Types of Databases and Database Applications

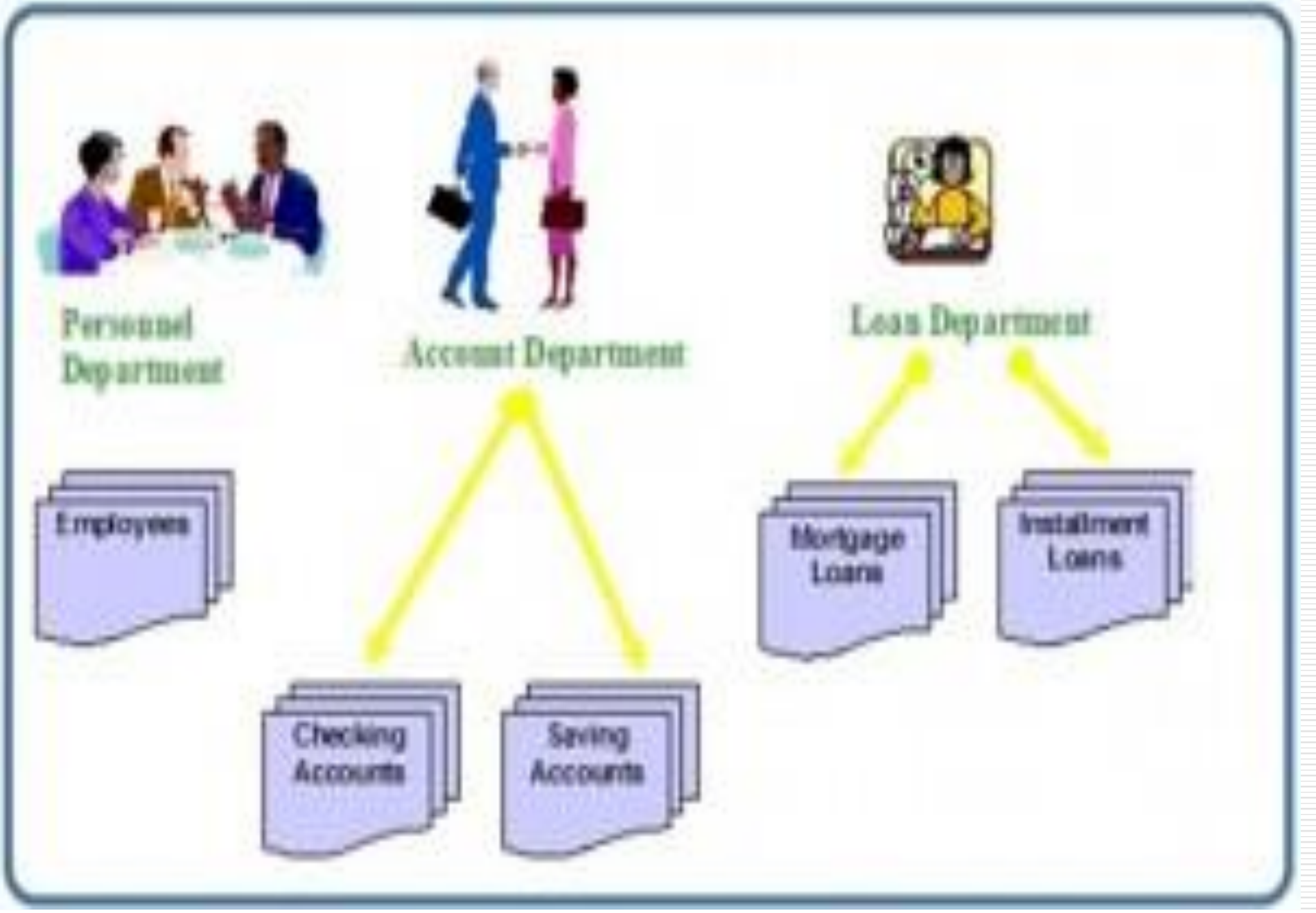
- Traditional Applications:
  - Numeric and Textual Databases
- More Recent Applications:
  - Multimedia Databases
  - Geographic Information Systems (GIS)
  - Data Warehouses
  - Real-time and Active Databases
  - Many other applications

File-Based Systems

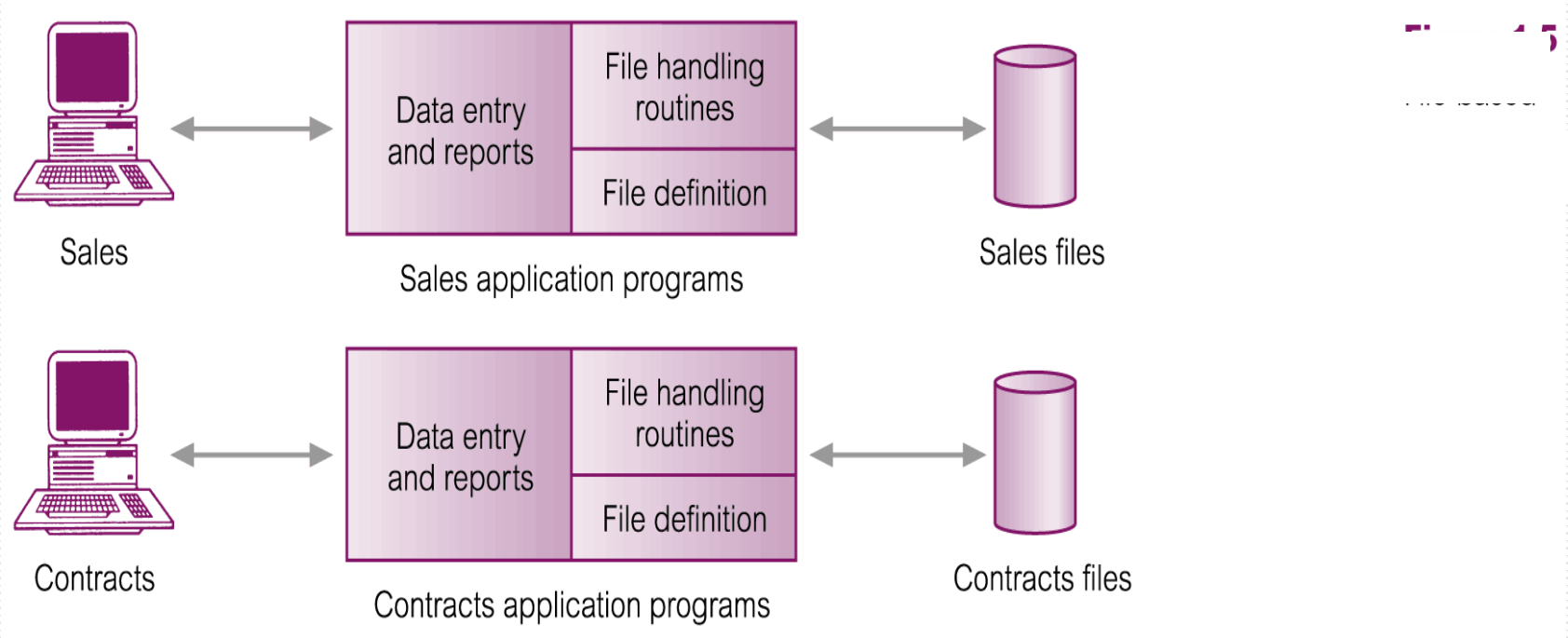
- One way to keep information on a computer is to store it in permanent files.
- A company system has a number of application programs; each of them is designed to manipulate data files.
- These application programs have been written at the request of the users in the organization.
- New applications are added to the system as the need arises.

File-Based Systems

- The system just described is called the file-based system.
- Collection of application programs that perform services for the end users (e.g. reports).
- In traditional file processing, each user defines and implements the files needed for a specific software application as part of programming the application
- Each program defines and manages its own data.


Personnel Department
Account Department
Loan Department
Employees
Checking Accounts
Saving Accounts
Mortgage Loans
Installment Loans

File-Based Processing



LIMITATIONS/DISADVANTAGES OF THE FILE-BASED APPROACH

1. Separation Isolation of Data cont'd

- Separation and isolation of data
  - Each program maintains its own set of data.
  - Users of one program may be unaware of potentially useful data held by other programs.
- Data are scattered in various files, and the files may be in different format, writing new application program to retrieve data is difficult

Separation Isolation of Data

- *Data isolation* is a property that determines when and how changes made by one operation become visible to other concurrent users and systems.
- This issue occurs in a concurrency situation.
  - This is a problem because:
    - It is difficult for new applications to retrieve the appropriate data, which might be stored in various files.

2. Duplication of Data

- **Duplication of data**
  - **Same data is held by different programs.**
  - **Wasted space and potentially different values and/or different formats for the same item.**

3. INTEGRITY PROBLEMS

- It refers to the maintenance and assurance that the data in a database are correct and consistent. Factors to consider when addressing this issue are:
  - Data values must satisfy certain consistency constraints that are specified in the application programs.
    - It is difficult to make changes to the application programs in order to enforce new constraints

4. CONCURRENCY ACCESS

- *Concurrency* is the ability of the database to allow multiple users access to the same record without adversely affecting transaction processing.
- A file-based system must manage, or prevent, concurrency by the application programs.

CONCURRENCY ACCESS CONT'D

- Typically, in a file-based system, when an application opens a file, that file is locked. This means that no one else has access to the file at the same time.
- In database systems, concurrency is managed thus allowing multiple users access to the same record.
- This is an important difference between database and file-based systems.

6. SECURITY PROBLEMS

- Security can be a problem with a file-based approach because:
- There are constraints regarding accessing privileges.
- Application requirements are added to the system in an ad-hoc manner so it is difficult to enforce constraints.

7. DIFFICULT IN ACCESSING DATA

- It is not easy to retrieve information using a conventional file processing system.
- Convenient and efficient information retrieval is almost impossible using conventional file processing system.

Other Limitations of File-Based

8. Data dependence

- File structure is defined in the program code.

9. Incompatible file formats

- Programs are written in different languages, and so cannot easily access each other's files.

10. Fixed Queries/Proliferation of application programs

- Programs are written to satisfy particular functions.
- Any new requirement needs a new program.

Observations and Conclusions

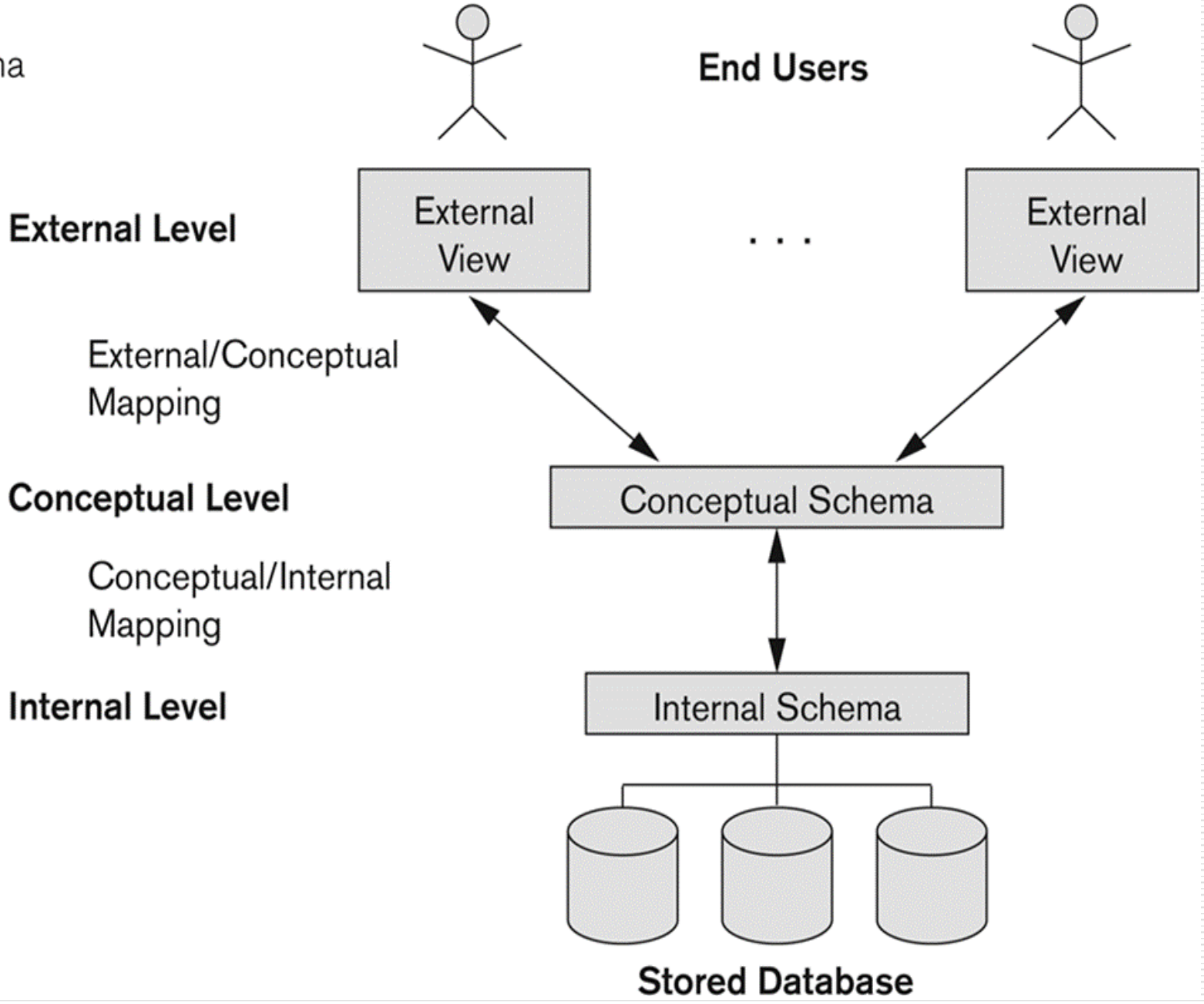
- Data Redundancy may leads to Data inconsistency , if redundant data are not updated simultaneously. Data inconsistency leads the system into an inconsistent state, since the operations based on inconsistent data results in to more inconsistency.

Database Approach

- **Arose because:**
  - Definition of data was embedded in application programs, rather than being stored separately and independently.
  - No control over access and manipulation of data beyond that imposed by application programs.

- **Result:**
  - the database and Database Management System (DBMS).

Figure 2.2
The three-schema architecture.



Simplified database system environment

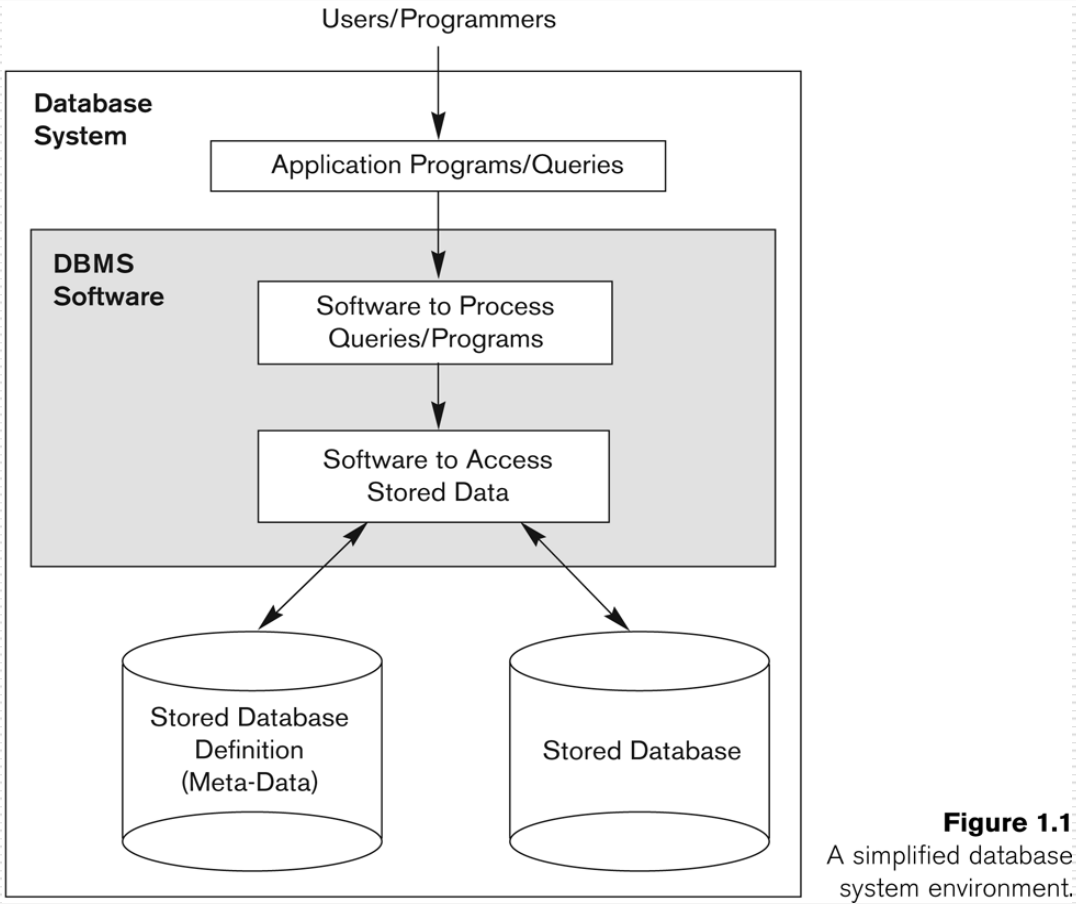


Figure 1.1
A simplified database system environment.

Typical DBMS Functionality

- *Define* a particular database in terms of its data types, structures, and constraints
- *Construct* or Load the initial database contents on a secondary storage medium
- *Manipulating* the database:
  - Retrieval: Querying, generating reports
  - Modification: Insertions, deletions and updates to its content
  - Accessing the database through Web applications
- *Processing* and *Sharing* by a set of concurrent users and application programs – yet, keeping all data valid and consistent

Typical DBMS Functionality

- Other features:
  - Protection or Security measures to prevent unauthorized access
  - "Active" processing to take internal actions on data
  - Presentation and Visualization of data
  - Maintaining the database and associated programs over the lifetime of the database application
    - Called database, software, and system maintenance

Example of a Database (with a Conceptual Data Model)

- **Mini-world for the example:**
  - Part of a UNIVERSITY environment.
- **Some mini-world *entities*:**
  - STUDENTs
  - COURSEs
  - SECTIONs (of COURSEs)
  - (academic) DEPARTMENTs
  - INSTRUCTORs

Example of a Database (with a Conceptual Data Model)

- **Some mini-world *relationships*:**
  - SECTIONs *are of specific* COURSEs
  - STUDENTs *take* SECTIONs
  - COURSEs *have prerequisite* COURSEs
  - INSTRUCTORs *teach* SECTIONs
  - COURSEs *are offered by* DEPARTMENTs
  - STUDENTs *major in* DEPARTMENTs

Example of a simple database

COURSE

Course_name	Course_number	Credit_hours	Department
Intro to Computer Science	CS1310	4	CS
Data Structures	CS3320	4	CS
Discrete Mathematics	MATH2410	3	MATH
Database	CS3380	3	CS

SECTION

Section_identifier	Course_number	Semester	Year	Instructor
85	MATH2410	Fall	04	King
92	CS1310	Fall	04	Anderson
102	CS3320	Spring	05	Knuth
112	MATH2410	Fall	05	Chang
119	CS1310	Fall	05	Anderson
135	CS3380	Fall	05	Stone

GRADE_REPORT

Student_number	Section_identifier	Grade
17	112	B
17	119	C
8	85	A
8	92	A
8	102	B
8	135	A

PREREQUISITE

Course_number	Prerequisite_number
CS3380	CS3320
CS3380	MATH2410
CS3320	CS1310

Figure 1.2
A database that stores student and course information.

Main Characteristics of the Database Approach

- **Self-describing nature of a database system:**
  - A DBMS **catalog** stores the description of a particular database (e.g. data structures, types, and constraints)
  - The description is called **meta-data**.
  - This allows the DBMS software to work with different database applications.
- **Insulation between programs and data:**
  - Called **program-data independence**.
  - Allows changing data structures and storage organization without having to change the DBMS access programs.

Example of a simplified database catalog

RELATIONS

Relation_name	No_of_columns
STUDENT	4
COURSE	4
SECTION	5
GRADE_REPORT	3
PREREQUISITE	2

COLUMNS

Column_name	Data_type	Belongs_to_relation
Name	Character (30)	STUDENT
Student_number	Character (4)	STUDENT
Class	Integer (1)	STUDENT
Major	Major_type	STUDENT
Course_name	Character (10)	COURSE
Course_number	XXXXNNNN	COURSE
….	….	…..
….	….	…..
….	….	…..
Prerequisite_number	XXXXNNNN	PREREQUISITE

Note: Major_type is defined as an enumerared type with all known majors. XXXXNNNN is used to define a type with four alpha characters followed by four digits

Figure 1.3
An example of a database catalog for the database in Figure 1.2.

Main Characteristics of the Database Approach (continued)

- **Data Abstraction:**
  - A **data model** is used to hide storage details and present the users with a conceptual view of the database.
  - Programs refer to the data model constructs rather than data storage details
- **Support of multiple views of the data:**
  - Each user may see a different view of the database, which describes **only** the data of interest to that user.

Main Characteristics of the Database Approach (continued)

- **Sharing of data and multi-user transaction processing:**
  - Allowing a set of **concurrent users** to retrieve from and to update the database.
  - *Concurrency control* within the DBMS guarantees that each **transaction** is correctly executed or aborted
  - *Recovery* subsystem ensures each completed transaction has its effect permanently recorded in the database
  - **OLTP** (Online Transaction Processing) is a major part of database applications. This allows hundreds of concurrent transactions to execute per second.

What Does a DBMS Do?

- Database management systems provide several functions in addition to simple file management:
- allow concurrency
- control security
- maintain data integrity
- provide for backup and recovery

What Does a DBMS Do?

- control redundancy
- allow data independence
- provide non-procedural query language
- perform automatic query optimization

Who Interacts with a DBMS?

Many different individuals are involved with a database management system over its life:

- systems analysts
- database designers
- database administrators
- application developers
- users

Historical Development of Database Technology

- Early Database Applications:
  - The Hierarchical and Network Models were introduced in mid 1960s and dominated during the seventies.
  - A bulk of the worldwide database processing still occurs using these models, particularly, the hierarchical model.
- Relational Model based Systems:
  - Relational model was originally introduced in 1970, was heavily researched and experimented within IBM Research and several universities.
  - Relational DBMS Products emerged in the early 1980s.

End