

Lecture 1

File Systems and Databases

Files and Databases

- File: A collection of records or documents dealing with one organization, person, area or subject (Rowley)
 - Manual (paper) files
 - Computer files
- Database: A collection of similar records with relationships between the records (Rowley)
 - Bibliographic, statistical, business data, images, etc.

Introducing the Database

- Major Database Concepts
 - Data and information
 - Data - Raw facts
 - Information - Processed data
 - Data management & Data Administrator
 - Database & Database Administrator
 - Metadata
 - Database management system (DBMS)

What is database?

- ◆ Database: An **organized** collection of **logically related data**
 - We assume that all databases are computer-based in this class
 - Just a collection of **data**
- ◆ Data: Known facts that could be recorded and stored on computer media
 - Facts could be in any format: text, graphics, images...
- ◆ Information: Processed data that could be used to increase the knowledge of the user
 - Closely related to data, but different
 - View data as “raw material” and information as “intermediate product”
 - The final product is “knowledge”

Applications of Databases

- ☐ Banking systems that keep information about customers and their account (large databases, real-time response to some queries, peak periods for transactions)
- ☐ Point of sales systems in shops that process the purchase of goods
- ☐ Reservation systems for booking airline flights and other travel, seats at the theatre or cinema, etc
- ☐ Inventory systems ranging from the books in a library to the stock in a furniture shop or car showroom
- ☐ Accounting systems that produce bills, statements etc.

Applications of Databases (cont.)

- ☐ Telephone-based insurance companies (real-time retrieval customer records)
- ☐ Historical records (read-only access, simplified management)
- ☐ Human resources or personnel systems that hold employee information
- ☐ Payroll systems that process salaries and benefits

Manual Filing Systems

- Before computer: Manual Filing Systems
 - ◆ Still used in many everyday activities
 - ◆ Usually organize data in a logical way
 - ◆ Example: Bank Statements, Bills, Recipes...
- Problems with Manual Filing Systems
 - ◆ Small-scale: Imagine Terra bytes of data!
 - ◆ Hard to cross-reference
 - ◆ Hard to analyze: "How much did I spend on grocery shopping last month?"

Contents of the CUSTOMER File

	C_NAME	C_PHONE	C_ADDRESS	C_ZIP	A_NAME	A_PHONE	TP	AMT	REN
▶	Alfred A. Ramas	615-844-2573	218 Fork Rd., Babs, TN	36123	Leah F. Hahn	615-882-1244	T1	\$100.00	3/12/99
	Leona K. Dunne	713-894-1238	Box 12A, Fox, KY	25246	Alex B. Alby	713-228-1249	T1	\$250.00	5/23/99
	Kathy W. Smith	615-894-2285	125 Oak Ln, Babs, TN	36123	Leah F. Hahn	615-882-2144	S2	\$150.00	1/5/98
	Paul F. Olowski	615-894-2180	217 Lee Ln., Babs, TN	36123	Leah F. Hahn	615-882-1244	S1	\$300.00	9/20/99
	Myron Orlando	615-222-1672	Box 111, New, TN	36155	Alex B. Alby	713-228-1249	T1	\$100.00	12/4/99
	Amy B. O'Brian	713-442-3381	387 Troll Dr., Fox, KY	25246	John T. Okon	615-123-5589	T2	\$850.00	8/29/99
	James G. Brown	615-297-1228	21 Tye Rd., Nash, TN	37118	Leah F. Hahn	615-882-1244	S1	\$120.00	3/1/99
	George Williams	615-290-2556	155 Maple, Nash, TN	37119	John T. Okon	615-123-5589	S1	\$250.00	6/23/99
	Anne G. Farriss	713-382-7185	2119 Elm, Crew, KY	25432	Alex B. Alby	713-228-1249	T2	\$100.00	11/9/99
	Olette K. Smith	615-297-3809	2782 Main, Nash, TN	37118	John T. Okon	615-123-5589	S2	\$500.00	2/18/99

C_NAME = Customer name	A_NAME = Agent name
C_PHONE = Customer phone	A_PHONE = Agent phone
C_ADDRESS = Customer address	TP = Insurance type
C_ZIP = Customer ZIP code	AMT = Insurance policy amount, in thousands of \$
	REN = Insurance renewal date

FIGURE 1.3 ■ CONTENTS OF THE CUSTOMER FILE

File Processing Systems

- ◆ File Processing Systems: A collection of application programs that perform services for the end-users such as production of reports.
 - Early attempt to computerize the manual filing approach
 - Developed in C, COBOL or other languages
 - Each program manages its own data
 - Usually managed by Data Processing (DP) staff

Table 1.1 Basic File Terminology

Data	“Raw” facts that have little meaning unless they have been organized in some logical manner. The smallest piece of data that can be “recognized” by the computer is a single character, such as the letter A, the number 5, or some symbol such as; ‘ ? > * +. A single character requires one byte of computer storage.
Field	A character or group of characters (alphabetic or numeric) that has a specific meaning. A field might define a telephone numbers, a birth date, a customer name, a year-to-date (YTD) sales value, and so on.
Record	A logically connected set of one or more fields that describes a person, place, or thing. For example, the fields that comprise a record for a customer named J. D. Rudd might consist of J. D. Rudd’s name, address, phone number, date of birth, credit limit, unpaid balance, and so on.
File	A collection of related records. For example, a file might contain data about ROBCOR Company’s vendors; or, a file might contain the records for the students currently enrolled at Gigantic University.

Limitation of Computerized File-based Approach

Separation and Isolation of data:

Each application maintains its own set of data. Users of one application may be unaware of potentially useful data held by other applications.

Duplication of data (data redundancy):

Same data is held by different files.

Limited Data Sharing:

No centralized control of data

Lengthy Development Times:

Programmers must design their own file formats

Excessive Program Maintenance:

80% of information systems budget

Limitation of File-based Approach

Accessibility

Data is stored using programs that are written in different languages and format.

Cannot easily access other files due to incompatible file format.

No control over access and manipulation of data beyond that imposed by application programs.

Hard Coded Queries (Data dependence)

File structure and all definition of data are defined in the application program code.

Any new requirement needs a new program.

SOLUTION:

Database Processing
System



Data Administrator

Data Administrator is responsible for the entire data of an organization. He normally develops the overall functional requirements for the databases being used in offices.

Database Administrator

The Database Administrator is responsible for the design, implementation, operations, management and maintenance of the database.

Database Management System (DBMS)

A **collection of programs** that enables users to define, create and maintain and control access to the database.

General-purpose software system that facilitates the processes of **defining**, **constructing** and **manipulating** databases for various applications.

Defining - specifying data types, structures, constraints.

Constructing - storing of data

Manipulation - querying, updating and reporting.

Popular DBMS In The Market

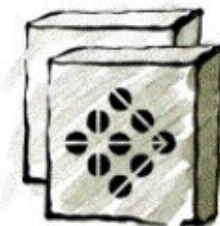


Informix

▼ Informix Dynamic Server



Paradox

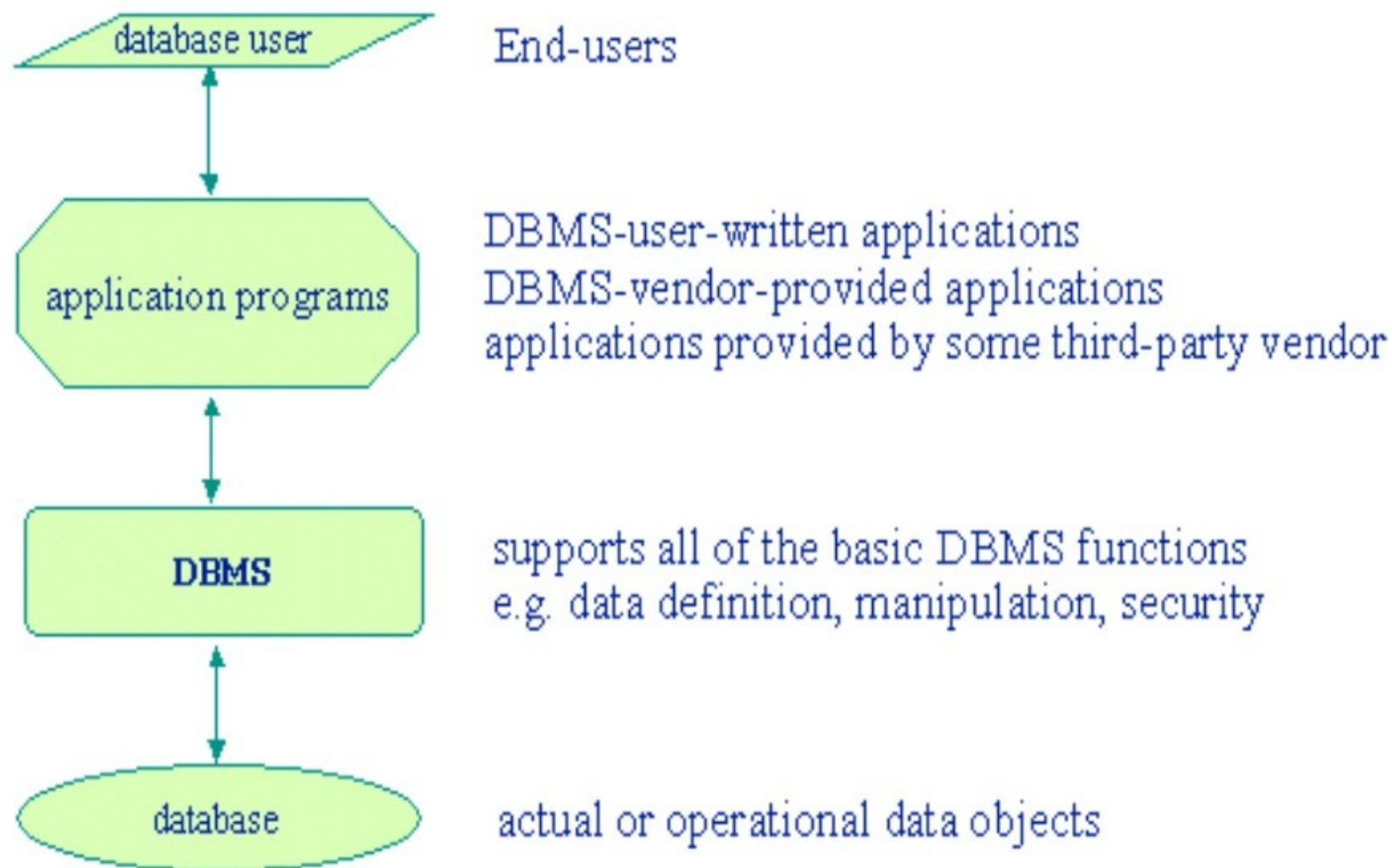


▼ Borland Interbase

Database Management Systems

- ➡ Computer software with the capability i) to store data in an integrated, structured format and ii) to enable users to retrieve, manipulate and manage the data
- ➡ There are many different types of DBMS, and even more DBMS products on the markets

Architecture for a Database System



The DBMS Manages the Interaction Between the End User and the Database

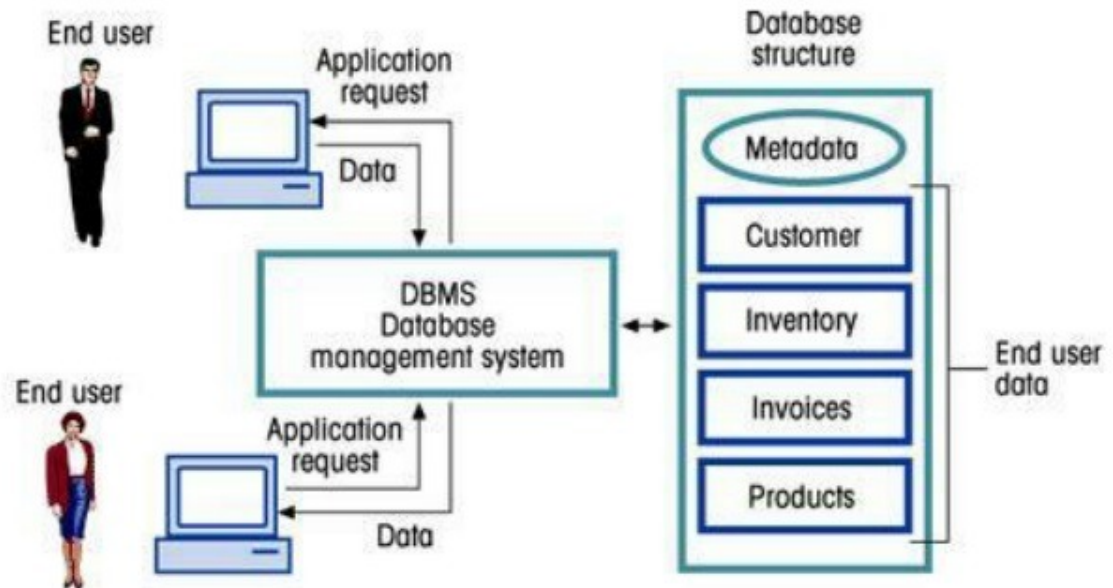


FIGURE 1.2 THE DBMS MANAGES THE INTERACTION BETWEEN THE END USER AND THE DATABASE

Database Systems

Types of Database Systems

- **Location**

- Centralized
- Distributed

Database Systems

Types of Database Systems

- **Location**

- Centralized
- Distributed

Database Systems

Types of Database Systems

- **Location**

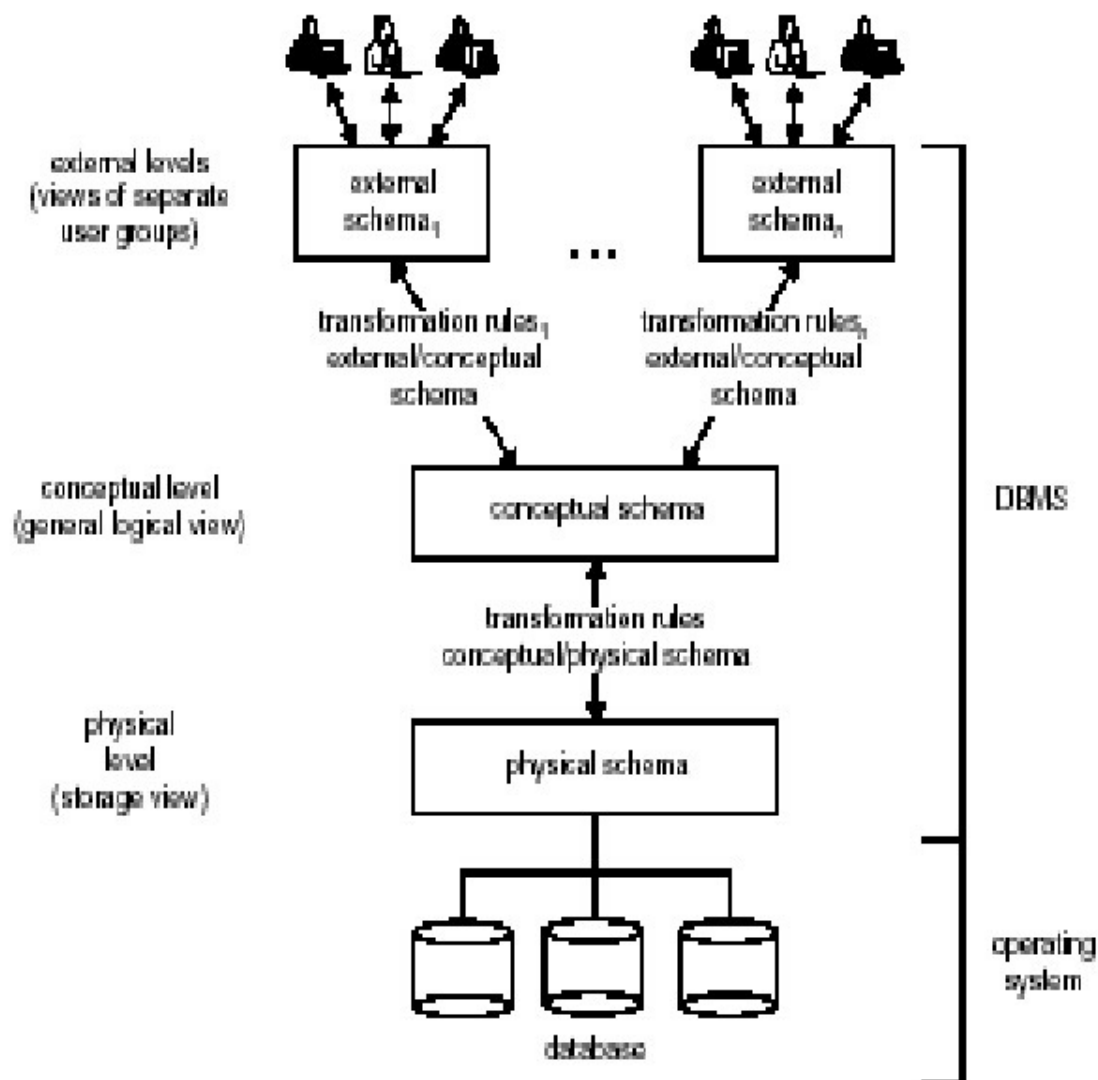
- Centralized
- Distributed

DBS has several abstraction levels

- external/view levels describe the part of the DB, which is relevant for the user
- conceptual/logical level gives information about existing data and relationships in the DB
- physical/internal level describes how data are physically stored

Database schema and state

- a schema describes the structure/the design of a DB
- a state describes a concrete instance of a DB



Data Independence

Logical Data Independence:

The capacity to change the conceptual schema without having to change the external schemas and their associated application programs.

Physical Data Independence:

The capacity to change the internal schema without having to change the conceptual schema. For example, the internal schema may be changed when certain file structures are reorganized or new indexes are created to improve database performance