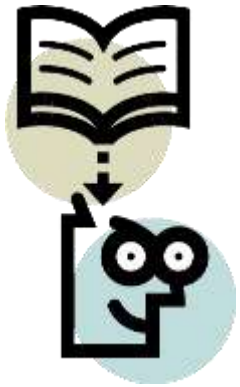
Database Concepts

Lecture 1

Learning Outcomes

- In this chapter, you will learn:
  - What is data and information?
  - What is a database?
  - What is a file system?
  - Main components of database system
  - What is a database management system (DBMS)

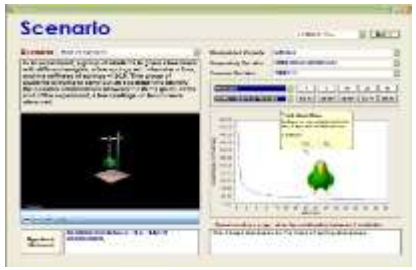


Introduction

- Good *decisions* require good *information* derived from *raw facts*
- Data can be managed most efficiently when *stored* in a database.

Why Databases?

- Databases solve many of the problems encountered in data management
  - ***Business***
  - ***Research***
  - ***Administration***
- E.g.,



Computer-based training materials

Knowledge Assessment Systems

Analysis of market's behaviour

Data vs. Information

- ***Data*** are raw facts
- ***Information*** is the result of processing raw data to reveal meaning
- Raw data must be *formatted* for *storage*, *processing*, and *presentation*
- Data are the foundation of information, which is the bedrock of knowledge.

Data vs. Information (cont'd.)

- Information is:
  - *produced by processing data*
  - *used to reveal meaning in data*
- Information is meant for helping in decision making.

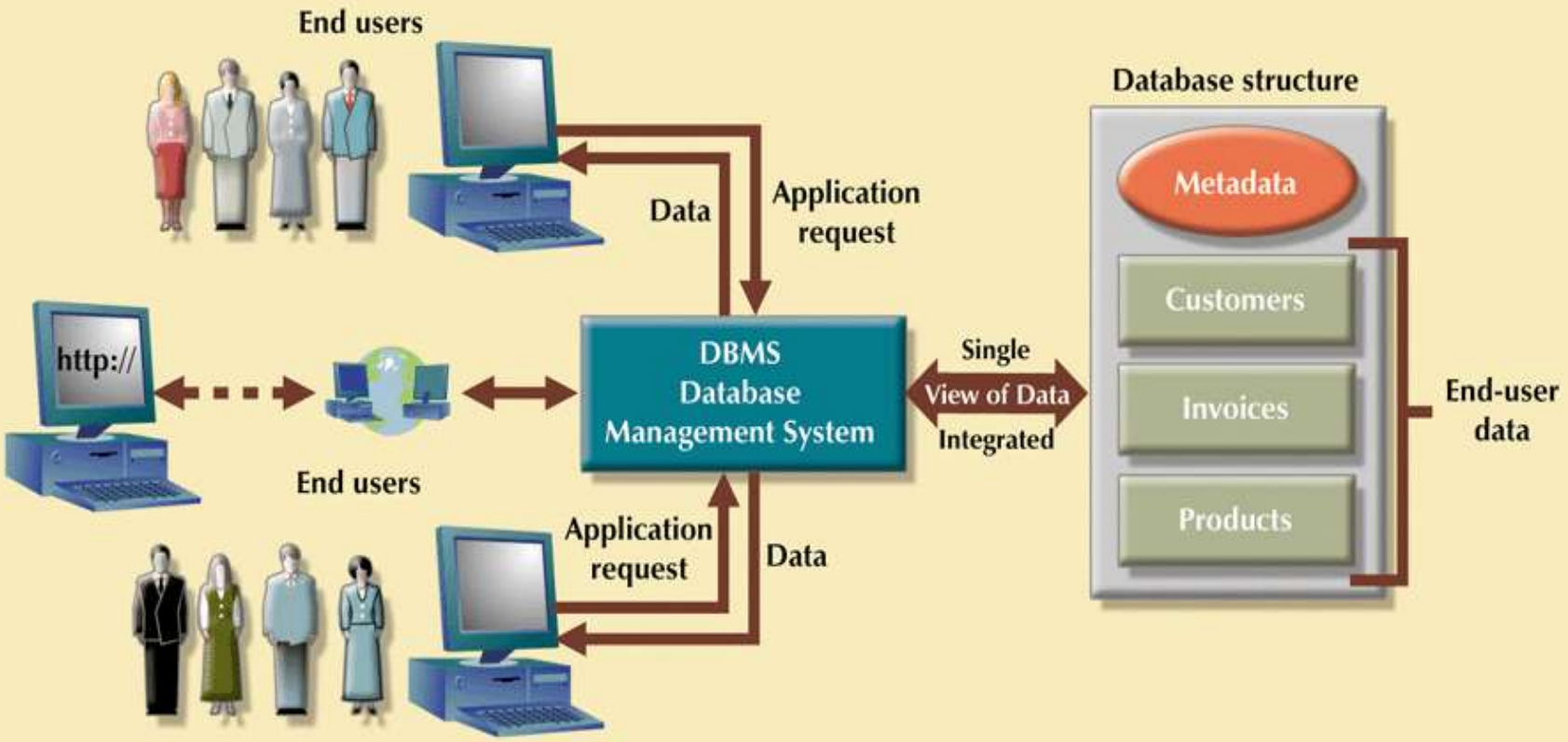
Introducing the Database

- **Database**: *shared* and *integrated* computer structure that stores a collection of:
  - ***End-user data***: raw facts
  - ***Metadata***: data about data
    - Provides description of data characteristics and relationships in data

Role and Advantages of the DBMS

- **Database management system (DBMS)**: collection of programs that manages *structure* and *control access* to data
- DBMS enables data to be shared and integrates many users' views of the data

FIGURE 1.2 The DBMS manages the interaction between the end user and the database



Role and Advantages of the DBMS (cont'd.)

- Advantages of a DBMS:
  - *Improved data sharing*
  - *Improved data security*
  - *Better data integration*
  - *Minimized data inconsistency*
  - *Improved data access*

Types of Databases

- Databases can be classified according to:
  - *Number of users*
  - *Database location(s)*
  - *Expected type and extent of use*
- ***Single-user*** database supports only one user at a time
  - Desktop database: *single-user*; runs on PC
- ***Multiuser*** database supports multiple users at the same time
  - *Workgroup and enterprise databases*

Types of Databases (cont'd.)

- ***Centralized*** database: data located at a single site
- ***Distributed*** database: data distributed across several different sites
- ***Operational*** database: supports a company's day-to-day operations
  - *Transactional or production database*
- ***Data warehouse***: stores data used for strategic decisions

Types of Data

- ***Unstructured*** data exist in their original state
- ***Structured*** data result from formatting
  - Structure applied based on type of processing to be performed
- ***Semi-structured*** data have been processed to some extent

Why Is Database Design Important?

- Well-designed database:
  - *Facilitates data management*
  - *Generates accurate and valuable information*
- Poorly designed database:
  - *Causes difficult-to-trace errors*

File Systems

- Collection of application programs that perform services for the end users (e.g. reports).
- Each program defines and manages its own data.
  - E.g. each department control and store its own data.

File Systems

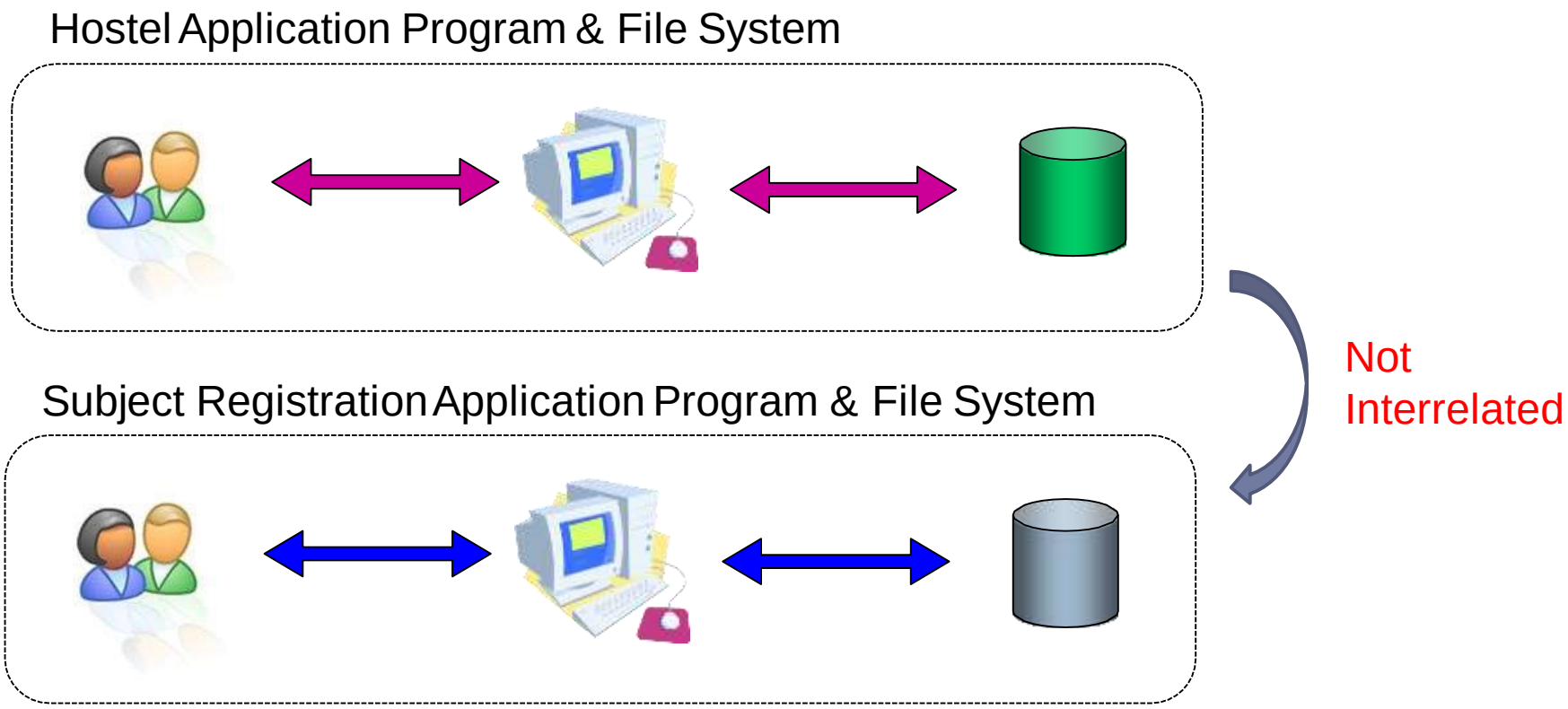


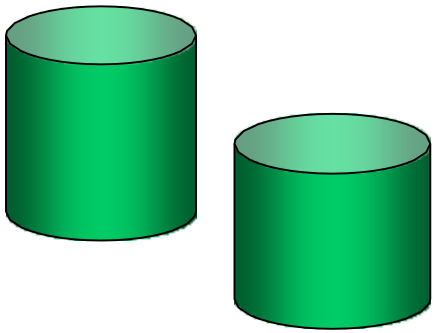
TABLE 1.2 Basic File Terminology

TERM	DEFINITION
Data	Raw facts, such as a telephone number, a birth date, a customer name, and a year-to-date (YTD) sales value. Data have little meaning unless they have been organized in some logical manner.
Field	A character or group of characters (alphabetic or numeric) that has a specific meaning. A field is used to define and store data.
Record	A logically connected set of one or more fields that describes a person, place, or thing. For example, the fields that constitute a customer record might consist of the customer's name, address, phone number, date of birth, credit limit, and unpaid balance.
File	A collection of related records. For example, a file might contain data about the students currently enrolled at Gigantic University.

Limitations

- *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.

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



Limitations

- *Data dependence*
  - File structure is defined in the program code.
- *Incompatible file formats*
  - Programs are written in different languages, and so cannot easily access each other's files.

Problems with File System

- File system limitations:
  - *Requires extensive programming*
  - *Cannot perform ad hoc queries*
  - *System administration is complex and difficult*
  - *Difficult to make changes to existing structures*
  - *Lack of security and limited data sharing*

Structural and Data Dependence

- ***Structural dependence***: access to a file is dependent on its own structure
  - All file system programs must be modified to conform to a new file structure
- ***Structural independence***: change file structure without affecting the application's ability to access data
- ***Data dependence***: data access changes when data storage characteristics change
- ***Data independence***: data storage characteristics do not affect data access

Data Redundancy

- ***Data redundancy***: same data stored unnecessarily in different places
  - Data stored in different locations is difficult to update consistently
- ***Data inconsistency***: different and conflicting versions of same data occur at different places

Data Anomalies

- ***Data anomalies***: abnormalities when all changes in redundant data are not made correctly
  - ***Update anomalies***
  - ***Insertion anomalies***
  - ***Deletion anomalies***

Lack of Design and Data-Modeling Skills

- Most users lack the skill to properly design databases, despite multiple personal productivity tools being available
- Data-modeling skills are vital in the data design process
- Good data modeling facilitates communication between the designer, user and developer

Database Systems

- Database system consists of *logically* related data stored in a single logical data repository
  - User may view data repository as a single unit although they may be physically distributed among multiple storage locations
  - DBMS eliminates most of file system's problems
    - ***Data inconsistency***
    - ***Data redundancy***
    - ***Data/structural dependency***

Database Systems

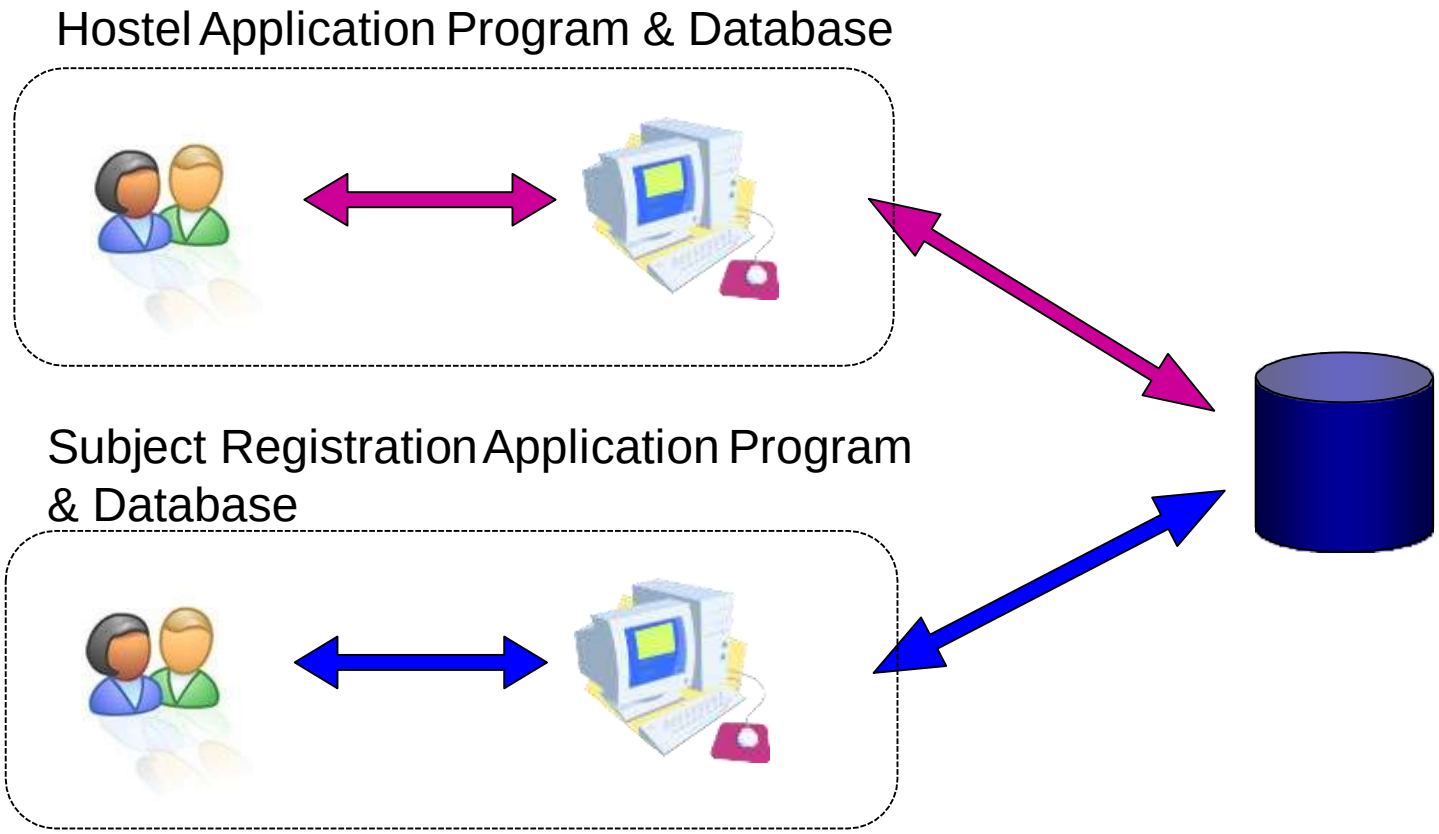
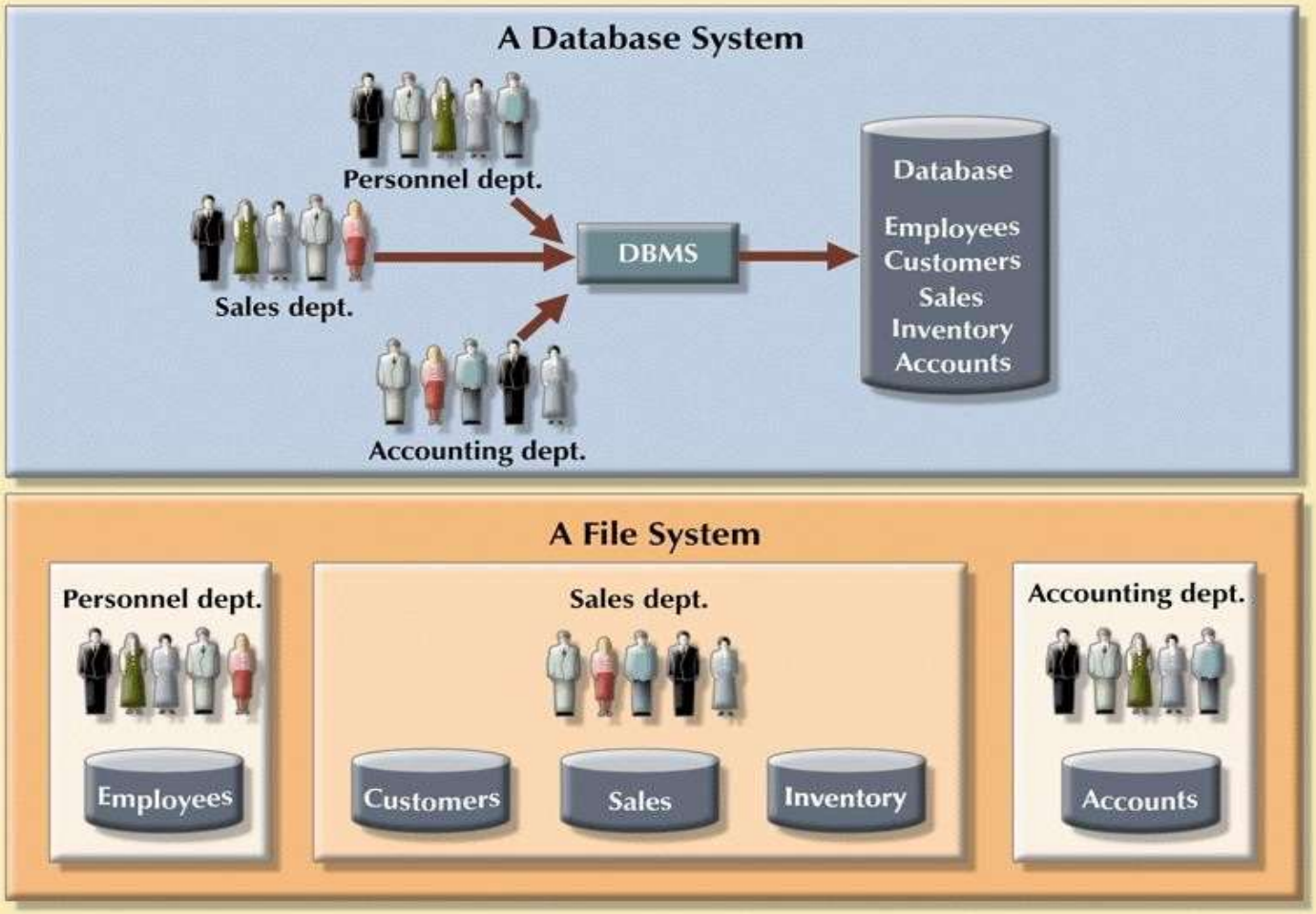


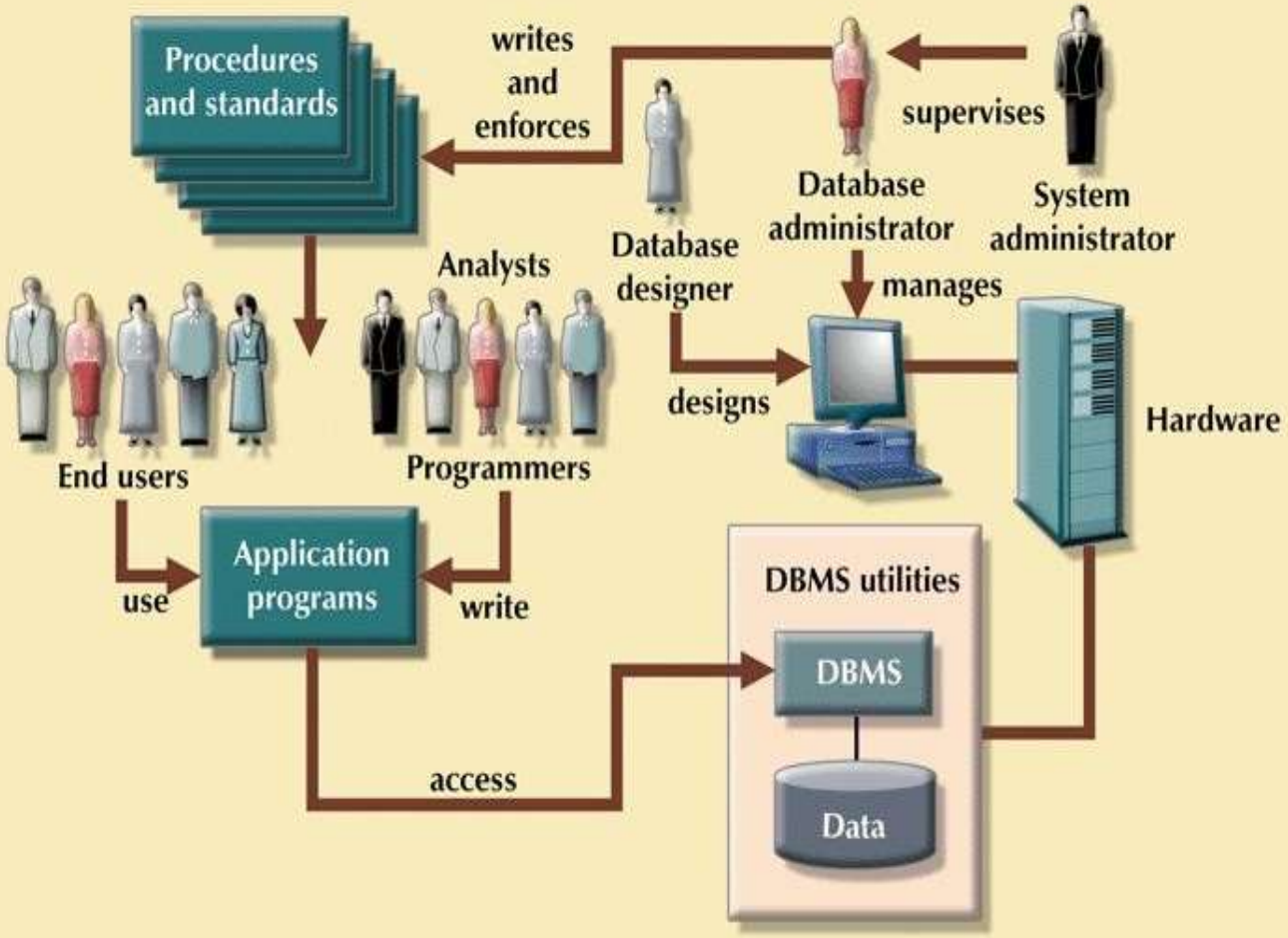
FIGURE 1.6 Contrasting database and file systems



The Database System Environment

- Five major parts of a database system:
  - ***Hardware***
  - ***Software***
  - ***People***
  - ***Procedures***
  - ***Data***

FIGURE 1.7 The database system environment



The Database System Environment (cont'd.)

- ***Hardware***: all the system's physical devices
- ***Software***: three types of software required:
  - Operating system software
  - DBMS software
  - Application programs and utility software

The Database System Environment (cont'd.)

- ***People***: all users of the database system
  - System and database administrators
  - Database designers
  - Systems analysts and programmers
  - End users
- ***Procedures***: instructions and rules that govern the design and use of the database system
- ***Data***: the collection of facts stored in the database

The Database System Environment (cont'd.)

- Database systems are created and managed at different levels of complexity (eg., insurance claims system and restaurant booking system)
- Database solutions must be cost-effective as well as tactically and strategically effective
- Database technology already in use affects selection of a database system

DBMS Functions

- Data dictionary management
- Data storage management
- Data transformation and presentation
- Security management
- Multiuser access control
- Backup and recovery management
- Data integrity management
- Database access languages and application programming interfaces
- Data communication interfaces

DBMS Functions

- **Data dictionary management**
  - Data dictionary contains metadata (data about data)
    - Contains all the attribute names and characteristics for each table in the system
  - Any changes made are automatically recorded in the data dictionary without making changes to program structure

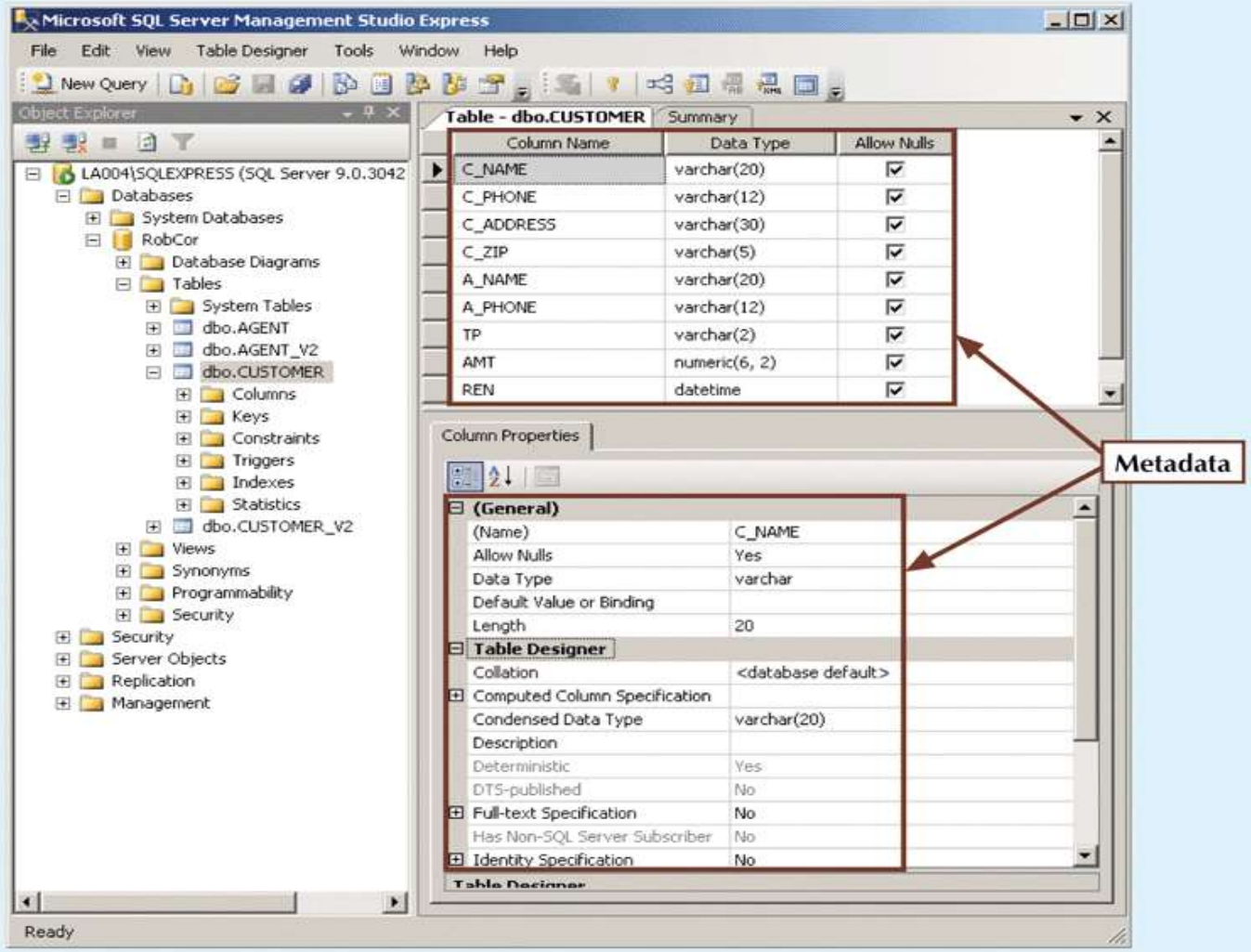
TABLE 3.6 **A Sample Data Dictionary**

TABLE NAME	ATTRIBUTE NAME	CONTENTS	TYPE	FORMAT	RANGE	REQUIRED	PK OR FK	FK REFERENCED TABLE
CUSTOMER	CUS_CODE	Customer account code	CHAR(5)	99999	10000–99999	Y	PK	
	CUS_LNAME	Customer last name	VARCHAR(20)	Xxxxxxxx		Y		
	CUS_FNAME	Customer first name	VARCHAR(20)	Xxxxxxxx		Y		
	CUS_INITIAL	Customer initial	CHAR(1)	X				
	CUS_RENEW_DATE	Customer insurance renewal date	DATE	dd-mmm-yyyy				
	AGENT_CODE	Agent code	CHAR(3)	999			FK	AGENT_CODE
AGENT	AGENT_CODE	Agent code	CHAR(3)	999		Y	PK	
	AGENT_AREACODE	Agent area code	CHAR(3)	999		Y		
	AGENT_PHONE	Agent telephone number	CHAR(8)	999-9999		Y		
	AGENT_LNAME	Agent last name	VARCHAR(20)	Xxxxxxxx		Y		
	AGENT_YTD_SLS	Agent year-to-date sales	NUMBER(9,2)	9,999,999.99		Y		

FK = Foreign key
PK = Primary key
CHAR = Fixed character length data (1–255 characters)
VARCHAR = Variable character length data (1–2,000 characters)
NUMBER = Numeric data (NUMBER(9,2)) are used to specify numbers with two decimal places and up to nine digits, including the decimal places. Some RDBMSs permit the use of a MONEY or CURRENCY data type.

Note: Telephone area codes are always composed of digits 0–9. Because area codes are not used arithmetically, they are most efficiently stored as character data. Also, the area codes are always composed of three digits. Therefore, the area code data type is defined as CHAR(3). On the other hand, names do not conform to some standard length. Therefore, the customer first names are defined as VARCHAR(20), thus indicating that up to 20 characters may be used to store the names. Character data are shown as left-justified.

FIGURE 1.8 Illustrating metadata with Microsoft SQL Server Express



DBMS Functions (cont'd.)

- ***Data storage management***
  - DBMS creates and manages complex structures required for data storage
    - Also stores related data entry forms, screen definitions, report definitions, etc.
  - DBMS stores the database in multiple ***physical data files***

DBMS Functions (cont'd.)

- ***Data transformation and presentation***
  - DBMS transforms data entered to conform to required data structures
  - DBMS transforms physically retrieved data to conform to user's logical expectations
- ***Security management***
  - DBMS creates a security system that enforces user security and data privacy
  - Security rules determine which users can access the database, which items can be accessed, etc.

DBMS Functions (cont'd.)

- ***Multiuser access control***
  - DBMS uses sophisticated algorithms to ensure concurrent access does not affect integrity
- ***Backup and recovery management***
  - DBMS provides backup and data recovery to ensure data safety and integrity
  - Recovery management deals with recovery of database after a failure
    - Critical to preserving database's integrity

DBMS Functions (cont'd.)

- ***Data integrity management***
  - DBMS promotes and enforces integrity rules
    - Minimizes redundancy
    - Maximizes consistency

DBMS Functions (cont'd.)

- ***Database access languages and application programming interfaces***
  - ***Structured Query Language (SQL)*** is the de facto query language
    - Standard supported by majority of DBMS vendors

DBMS Functions (cont'd.)

- ***Database communication interfaces***
  - Current DBMSs accept end-user requests via multiple different network environments
  - Communications accomplished in several ways:
    - Through Web browser
    - Automatically publishes predefined reports on a Web site
    - Connects to third-party systems to distribute information via e-mail

Disadvantages of Database Systems

- *Increased costs*
- *Management complexity*
- *Vendor dependence*
- *Frequent upgrade/replacement cycles*