Database Principles: Fundamentals of Design, Implementation, and Management

Tenth Edition

Introduction to Database

Objectives

In this chapter, you will learn:

- The difference between data and information.
- What a database is, the various types of databases, and why they are valuable assets for decision making.
- How modern databases evolved from file systems.
- The importance of database design.
- About flaws in file system data management.
- The main components of the database system.
- The main functions of a database management system (DBMS).

Introduction

- Good decisions require good information derived from raw facts
- Data is managed most efficiently when stored in a database
- Databases evolved from computer file systems
- Understanding file system characteristics is important

Why Databases?

- Databases solve many of the problems encountered in data management
  - Used in almost all modern settings involving data management:
    - Business
    - Research
    - Administration
- Important to understand how databases work and interact with other applications

Data vs. Information

- Data are raw facts
- Information is the result of processing raw data to reveal meaning
- Information requires context 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.)

- Data: building blocks of information
- Information produced by processing data
- Information used to reveal meaning in data
- Accurate, relevant, timely information is the key to good decision making
- Good decision making is the key to organizational survival
- Data management: focuses on proper generation, storage, and retrieval of data

Introducing the Database

- Database: shared, integrated computer structure that stores a collection of:
  - End-user data: raw facts of interest to end user
  - Metadata: data about data
    - Provides description of data characteristics and relationships in data
    - Complements and expands value of data
- Database management system (DBMS): collection of programs
  - Manages structure and controls access to data

Role and Advantages of the DBMS

- DBMS is the intermediary between the user and the database
  - Database structure stored as file collection
  - Can only access files through the DBMS
- DBMS enables data to be shared
- DBMS integrates many users' views of the data

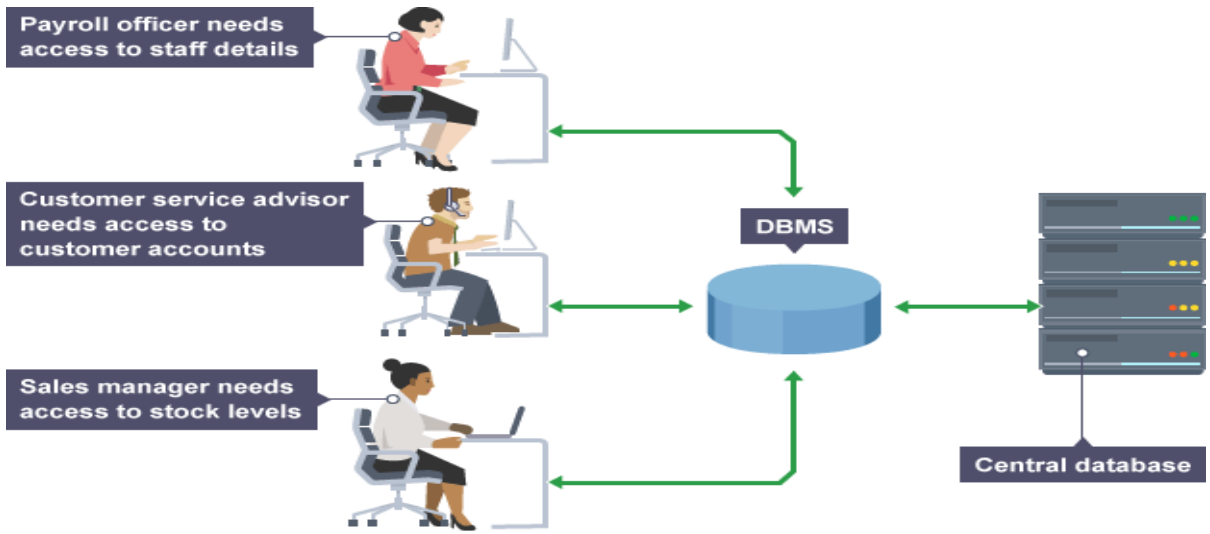
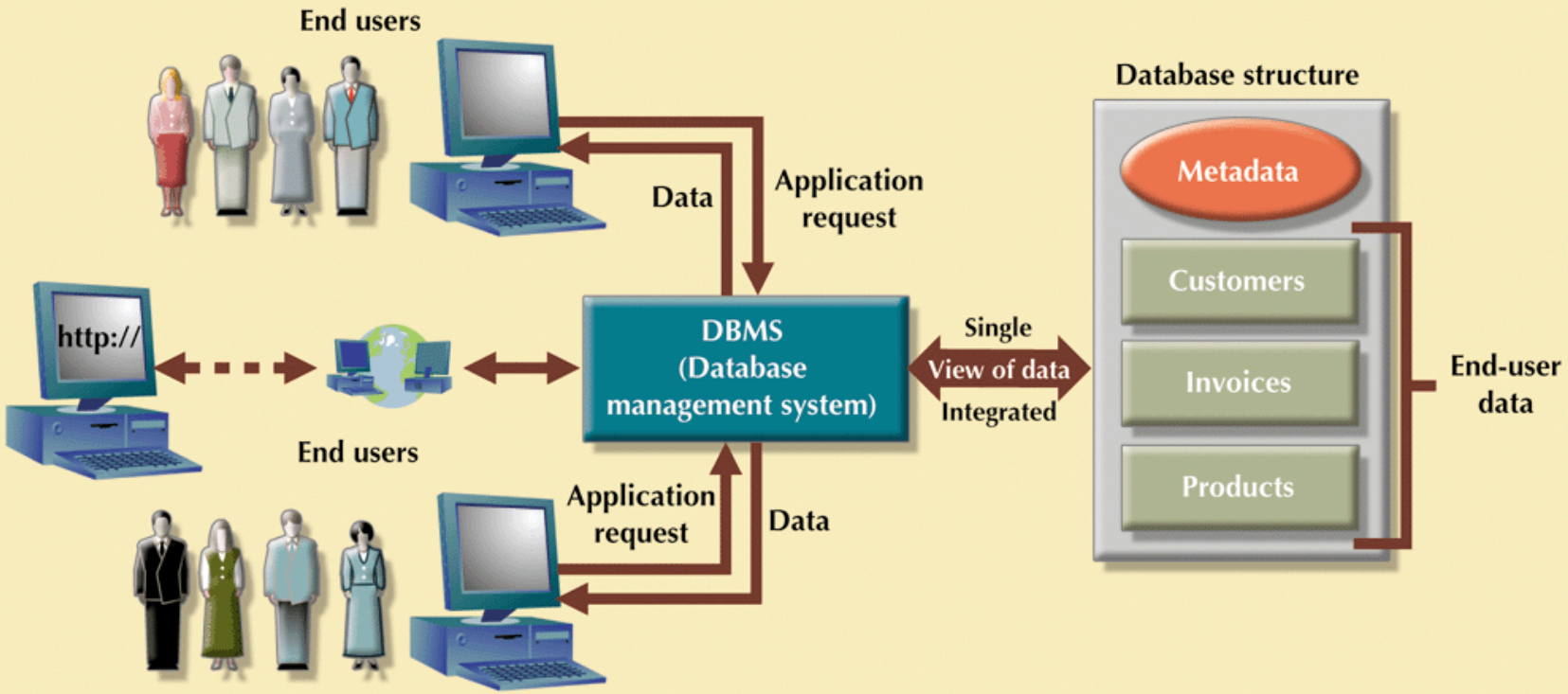


FIGURE 1.2 The DBMS manages the interaction between the end user and the database



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

- Advantages of a DBMS:
  - Improved data sharing
  - Improved data security – eg: password policies
  - Better data integration – eg: HR department data + Faculties data
  - Minimized data inconsistency – eg: less data appear in different places
  - Improved data access – eg: quick answers to ad hoc queries
  - Improved decision making
  - Increased end-user productivity

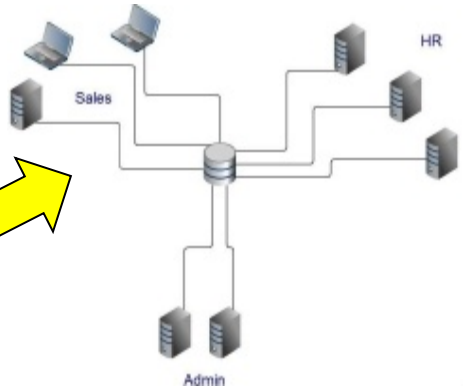
Types of Databases

- Databases can be classified according to:
  - Number of users
  - Database location(s)
  - Expected type and extent of use

1. Number of users:

- 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 database (supports less than 50 users at the same time)
  - Enterprise database (supports more than 50 users at the same time)

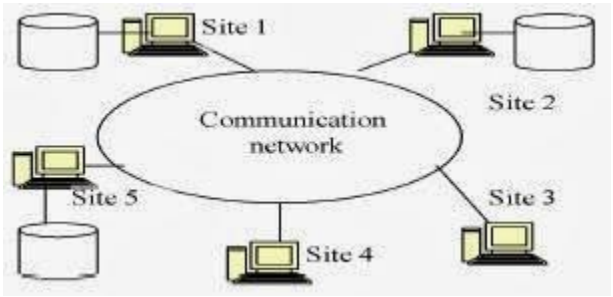
Types of Databases (cont'd.)



2. Database location(s):

- Centralized database: data located at a single site

- Distributed database: data distributed across several different sites



3. Expected type and extent of use:

- Operational database: supports a company's day-to-day operations
  - Transactional or production database
- Data warehouse: stores data used for tactical or strategic decisions

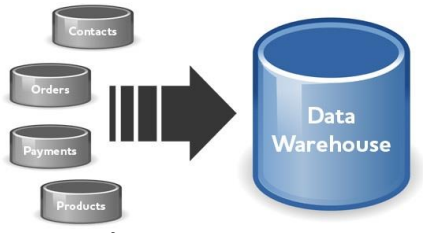


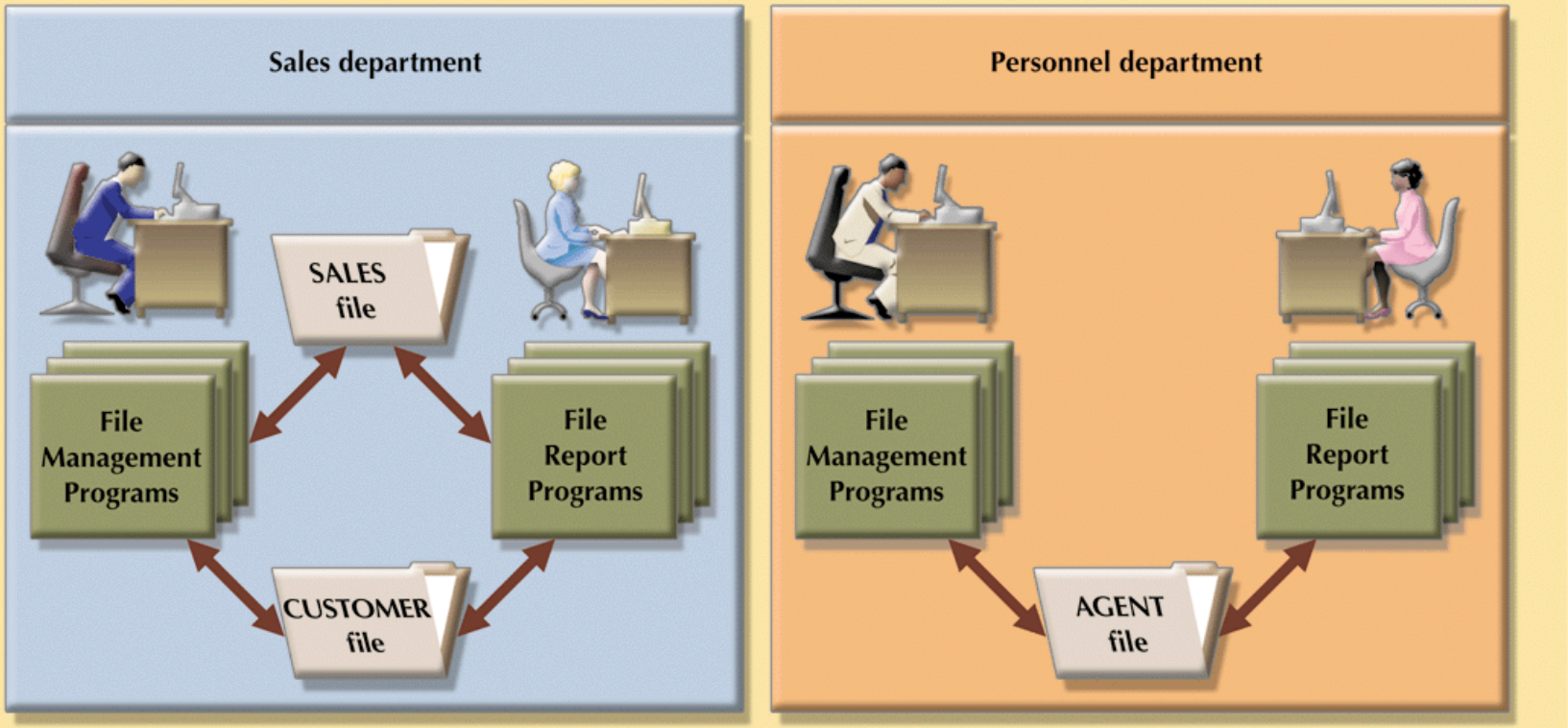
TABLE 1.1 Types of Databases

PRODUCT	NUMBER OF USERS			DATA LOCATION		DATA USAGE		XML
	SINGLE USER	MULTIUSER						
		WORKGROUP	ENTERPRISE	CENTRALIZED	DISTRIBUTED	OPERATIONAL	ANALYTICAL	
MS Access	X	X		X		X		
MS SQL Server	X[3]	X	X	X	X	X	X	X
IBM DB2	X[3]	X	X	X	X	X	X	X
MySQL	X	X	X	X	X	X	X	X
Oracle RDBMS	X[3]	X	X	X	X	X	X	X

File System Data Processing - Modern End-User Productivity Tools

- Use of personal productivity tools can introduce the same problems as the old file systems
- Microsoft Excel
  - Widely used by business users
  - Users have become so adept at working with spreadsheets, they tend to use them to complete tasks for which spreadsheets are not appropriate – database substitute

FIGURE 1.7 A simple file system



SOURCE: Course Technology/Cengage Learning

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.

Problems with File System Data Processing

- Summary of file system limitations:
  - Requires extensive programming
  - Cannot perform ad hoc queries
  - System administration is complex and difficult
  - Difficult to make changes to existing structures
  - Security features are likely to be inadequate

Data Redundancy

- File system structure makes it difficult to combine data from multiple sources
  - Vulnerable to security breaches
- Organizational structure promotes storage of same data in different locations
  - Islands of information
- Data stored in different locations is unlikely to be updated consistently
- Data redundancy: same data stored unnecessarily in different places

Data Redundancy (cont'd.)

- Data inconsistency: different and conflicting versions of same data occur at different places
- Data anomalies: abnormalities when all changes in redundant data are not made correctly
  - Update anomalies
  - Insertion anomalies / Insert anomalies
  - Deletion anomalies

Data Redundancy (cont'd.)

– Update anomalies

An **Update Anomaly** exists when one or more instances of duplicated data is updated, but not all. For example, consider Jones moving address - you need to update all instances of Jones's address.

StudentNum	CourseNum	Student Name	Address	Course
S21	9201	Jones	Edinburgh	Accounts
S21	9267	Jones	Edinburgh	Accounts
S24	9267	Smith	Glasgow	physics
S30	9201	Richards	Manchester	Computing
S30	9322	Richards	Manchester	Maths

Data Redundancy (cont'd.)

– Insertion anomalies / Insert anomalies

An **Insert Anomaly** occurs when certain attributes cannot be inserted into the database without the presence of other attributes. For example this is the converse of delete anomaly - we can't add a new course unless we have at least one student enrolled on the course.

StudentNum	CourseNum	Student Name	Address	Course
S21	9201	Jones	Edinburgh	Accounts
S21	9267	Jones	Edinburgh	Accounts
S24	9267	Smith	Glasgow	physics
S30	9201	Richards	Manchester	Computing
S30	9322	Richards	Manchester	Maths

Data Redundancy (cont'd.)

– Deletion anomalies

A **Delete Anomaly** exists when certain attributes are lost because of the deletion of other attributes. For example, consider what happens if Student S30 is the last student to leave the course - All information about the course is lost.

StudentNum	CourseNum	Student Name	Address	Course
S21	9201	Jones	Edinburgh	Accounts
S21	9267	Jones	Edinburgh	Accounts
S24	9267	Smith	Glasgow	physics
S30	9201	Richards	Manchester	Computing
S30	9322	Richards	Manchester	Maths

Why Database Design Is Important

- Database design focuses on design of database structure used for end-user data
  - Designer must identify database's expected use
- Well-designed database:
  - Facilitates data management
  - Generates accurate and valuable information
- Poorly designed database:
  - Causes difficult-to-trace errors

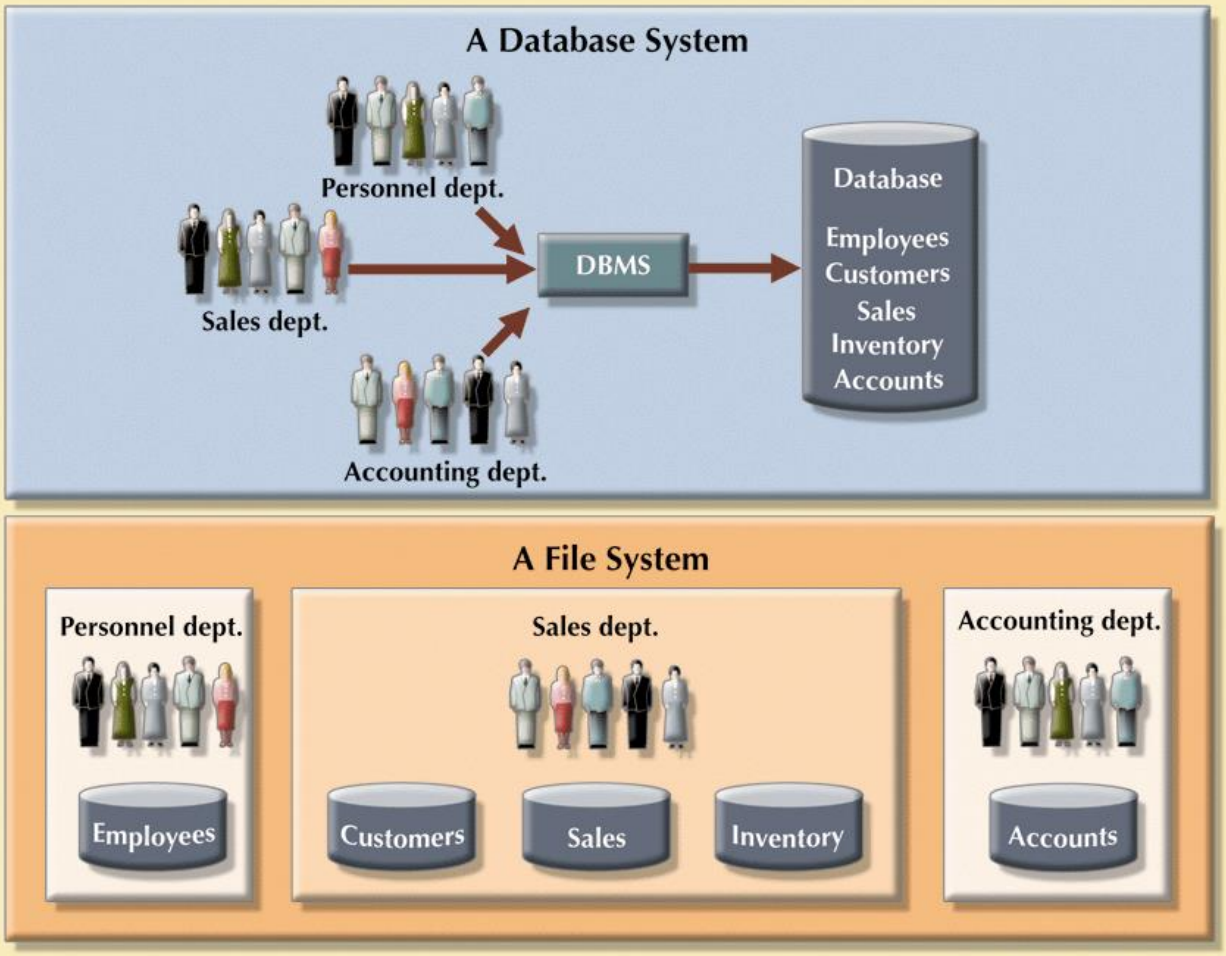
Lack of Design and Data-Modeling Skills

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

Database Systems

- Database system consists of logically related data stored in a single logical data repository
  - May be physically distributed among multiple storage facilities
  - DBMS eliminates most of file system's problems
  - Current generation stores data structures, relationships between structures, and access paths
    - Also defines, stores, and manages all access paths and components

FIGURE 1.8 Contrasting database and file systems

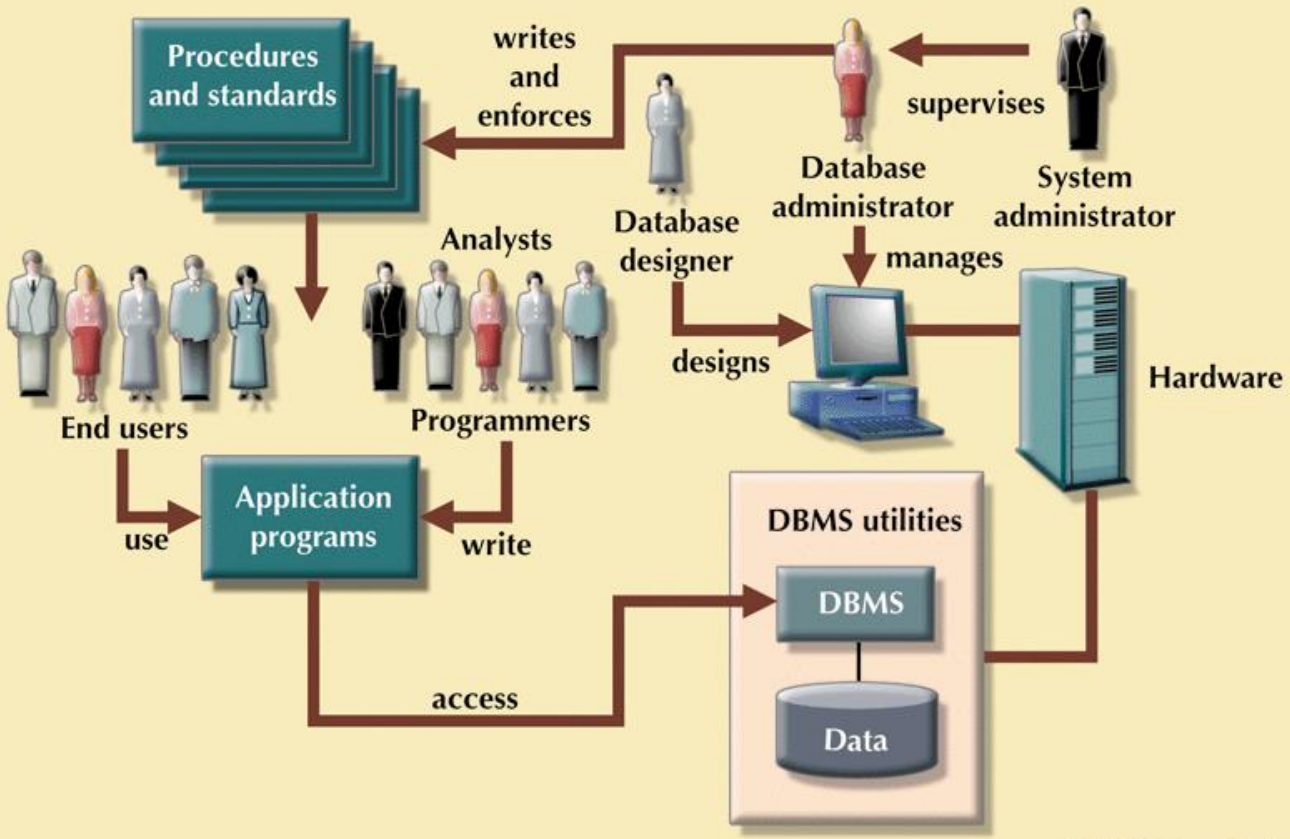


SOURCE: Course Technology/Cengage Learning

The Database System Environment

- Database system: defines and regulates (control) the collection, storage, management, use of data
- Five major parts of a database system:
  - Hardware
  - Software
  - People
  - Procedures
  - Data

FIGURE 1.9 The database system environment



SOURCE: Course Technology/Cengage Learning

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.)

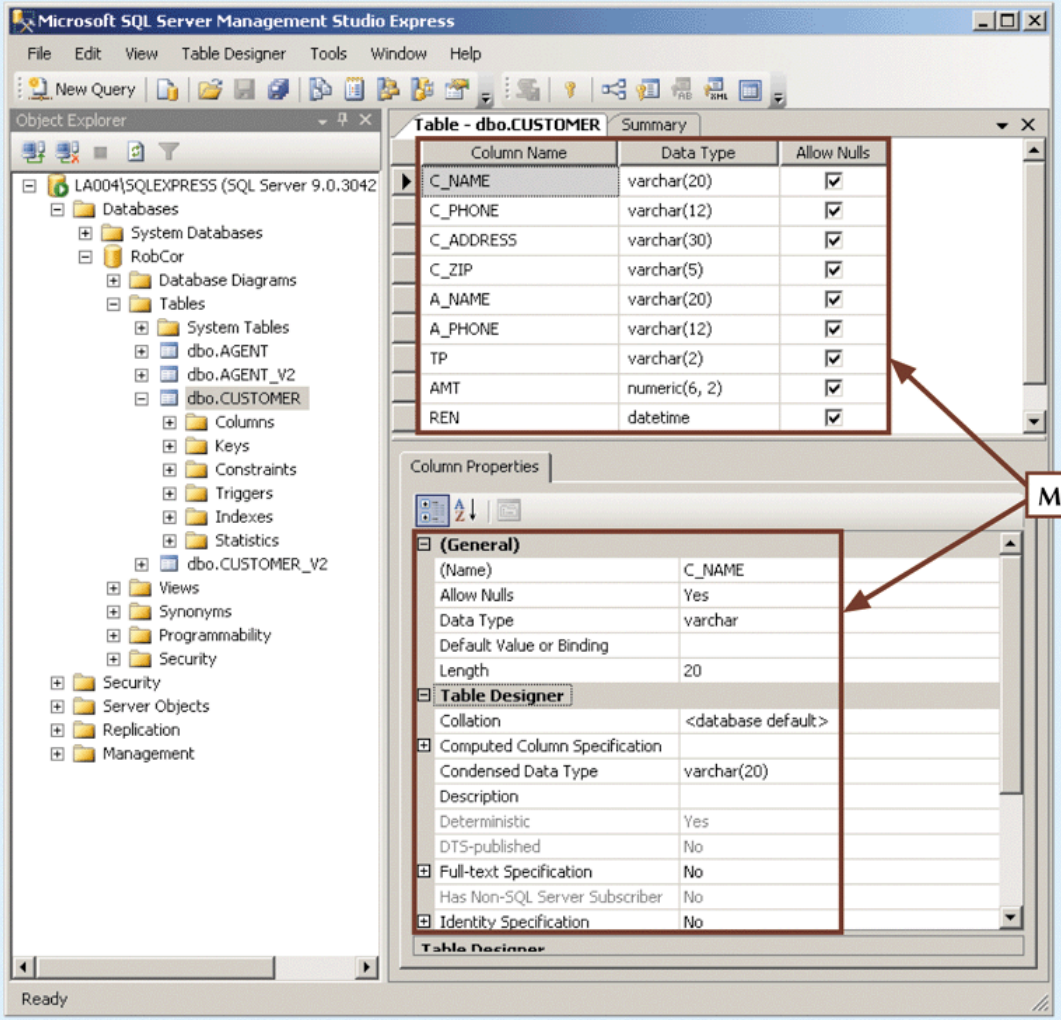
- Database systems are created and managed at different levels of complexity
- Database solutions must be cost-effective as well as tactically and strategically effective
- Database technology already in use affects selection of a database system

DBMS Functions

- Most functions are transparent to end users
  - Can only be achieved through the DBMS

1. Data dictionary management
   - DBMS stores definitions of data elements and relationships (metadata) in a data dictionary
   - DBMS looks up required data component structures and relationships
   - Changes automatically recorded in the dictionary
   - DBMS provides data abstraction and removes structural and data dependency

FIGURE 1.10 Illustrating metadata with Microsoft SQL Server Express

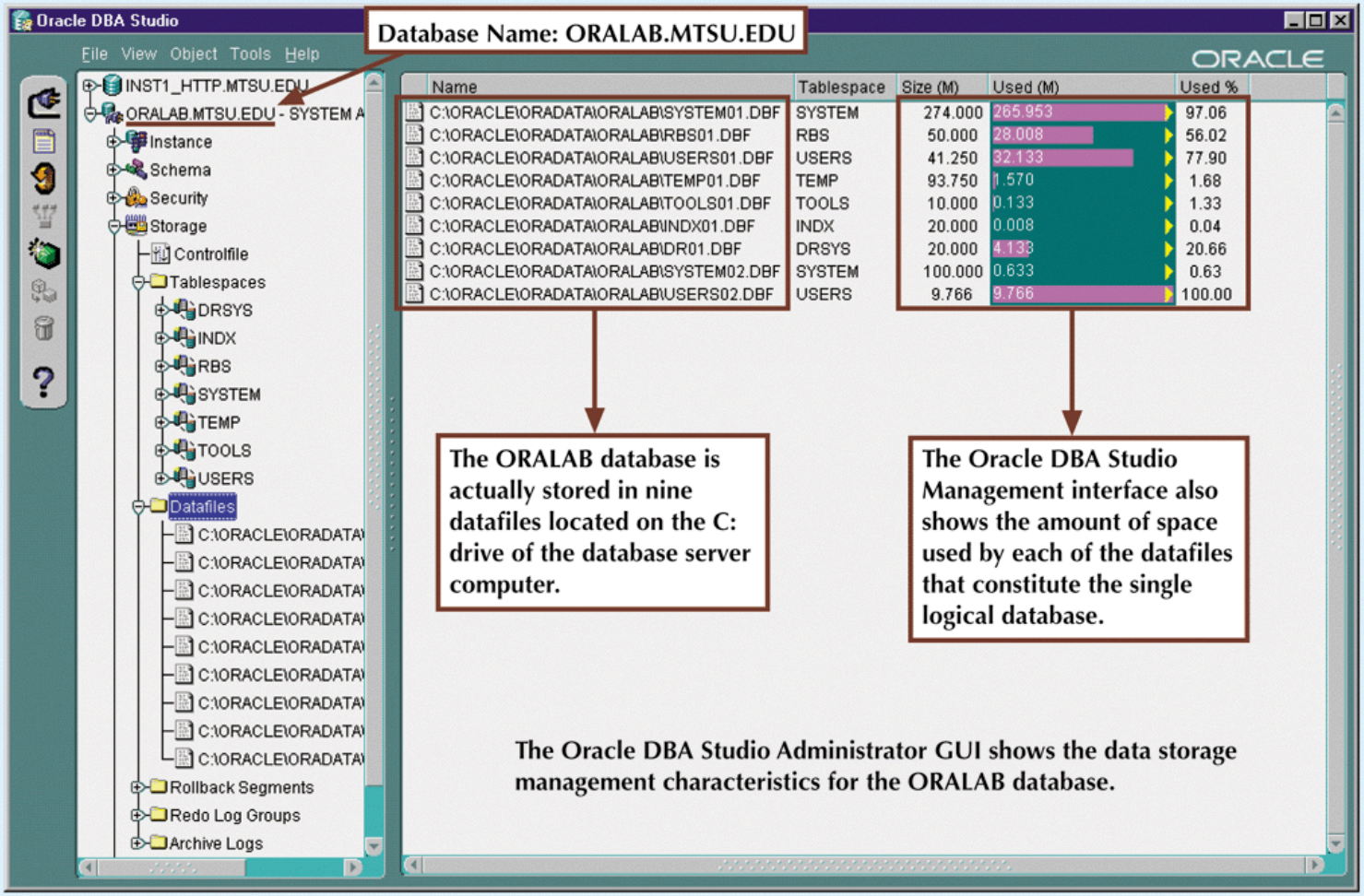


SOURCE: Course Technology/Cengage Learning

DBMS Functions (cont'd.)

2. Data storage management
   - DBMS creates and manages complex structures required for data storage
   - Also stores related data entry forms, screen definitions, report definitions, etc.
   - Performance tuning: activities that make the database perform more efficiently
   - DBMS stores the database in multiple physical data files

FIGURE 1.11 Illustrating data storage management with Oracle



SOURCE: Course Technology/Cengage Learning

DBMS Functions (cont'd.)

3. 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
4. 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.)

5. Multiuser access control
   - DBMS uses sophisticated algorithms to ensure concurrent access does not affect integrity
6. 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.)

7. Data integrity management
   - DBMS promotes and enforces integrity rules
     - Minimizes redundancy
     - Maximizes consistency
   - Data relationships stored in data dictionary used to enforce data integrity
   - Integrity is especially important in transaction-oriented database systems

DBMS Functions (cont'd.)

8. Database access languages and application programming interfaces
   - DBMS provides access through a query language
   - Query language is a nonprocedural language
   - Structured Query Language (SQL) is the de facto query language
     - Standard supported by majority of DBMS vendors

DBMS Functions (cont'd.)

9. Database communication interfaces

- Current DBMSs accept end-user requests via multiple different network environments
- Communications accomplished in several ways:
  - End users generate answers to queries by filling in screen forms through Web browser
  - DBMS automatically publishes predefined reports on a Web site
  - DBMS connects to third-party systems to distribute information via e-mail

Managing the Database System: A Shift in Focus

- Database system provides a framework in which strict procedures and standards enforced
  - Role of human changes from programming to managing organization's resources
- Database system enables more sophisticated use of the data
- Data structures created within the database and their relationships determine effectiveness

Managing the Database System: A Shift in Focus (cont'd.)

- Disadvantages of database systems:
  - Increased costs – hardware, software, highly skilled personnel, training etc.
  - Management complexity – many different technologies.
  - Maintaining currency – must keep your system current.
  - Vendor dependence – heavy investment in technology and personnel training, companies might be reluctant to change database vendors.
  - Frequent upgrade/replacement cycles – DBMS vendors frequently upgrade their products by adding new functionality. Some of the versions require hardware upgrades. Cost money to train database users and administrators to properly use and manage the new features.

Preparing for Your Database Professional Career

TABLE 1.3 Database Career Opportunities

JOB TITLE	DESCRIPTION	SAMPLE SKILLS REQUIRED
Database developer	Creates and maintains database-based applications	Programming, database fundamentals, SQL
Database designer	Designs and maintains databases	Systems design, database design, SQL
Database administrator	Manages and maintains DBMS and databases	Database fundamentals, SQL, vendor courses
Database analyst	Develops databases for decision support reporting	SQL, query optimization, data warehouses
Database architect	Designs and implements database environments (conceptual, logical, and physical)	DBMS fundamentals, data modeling, SQL, hardware knowledge
Database consultant	Helps companies leverage database technologies to improve business processes and achieve specific goals	Database fundamentals, data modeling, database design, SQL, DBMS, hardware, vendor-specific technologies
Database security officer	Implements security policies for data administration	DBMS fundamentals, database administration, SQL, data security technologies

Summary

- Data are raw facts
- Information is the result of processing data to reveal its meaning
- Accurate, relevant, and timely information is the key to good decision making
- Data are usually stored in a database
- DBMS implements a database and manages its contents

Summary (cont'd.)

- Metadata is data about data
- Database design defines the database structure
  - Well-designed database facilitates data management and generates valuable information
  - Poorly designed database leads to bad decision making and organizational failure
- Databases evolved from manual and computerized file systems

Summary (cont'd.)

- Database management systems were developed to address file system's inherent weaknesses
- DBMS present database to end user as single repository
  - Promotes data sharing
  - Eliminates islands of information
- DBMS enforces data integrity, eliminates redundancy, and promotes security