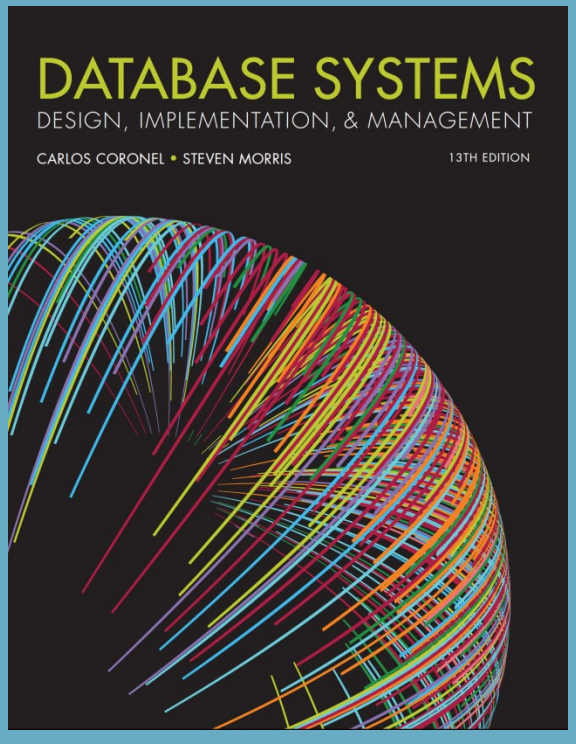
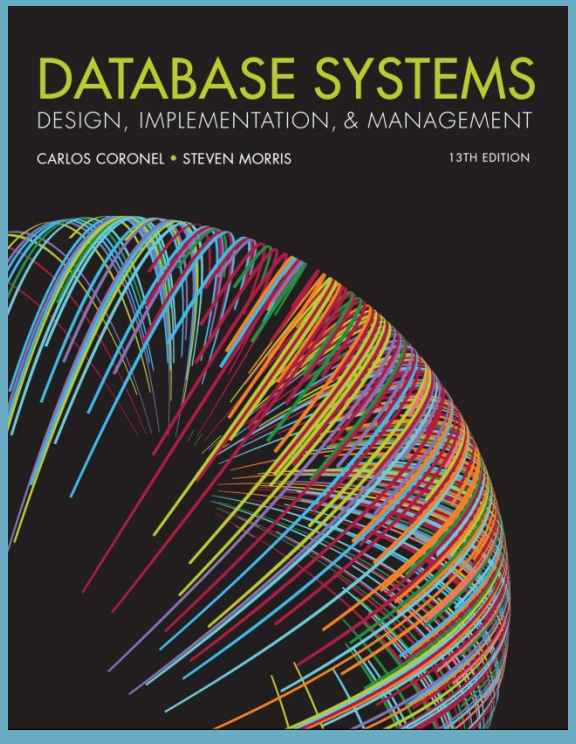


CISC 3810

Database Systems



Chapter 1

Database Systems

Learning Objectives

- After completing this chapter, you will be able to:
  - Define the difference between data and information
  - Describe what a database is, various types, and why they are valuable assets for decision making
  - Explain the importance of database design
  - See how modern databases evolved from file systems
  - Understand flaws in file system data management
  - Outline the main components of the database system
  - Describe the main functions of a database management system (DBMS)

Why Databases?

- Characteristics of data in today's world
  - Ubiquitous (i.e., abundant, global, and everywhere)
  - Pervasive (i.e., unescapable, prevalent, and persistent)
- Databases make data persistent and shareable in a secure way
  - Specialized structures that allow computer-based systems to store, manage, and retrieve data very quickly

Why Databases?

FIGURE 1.1 THE PERVASIVE NATURE OF DATABASES

A Day In Susan's Life

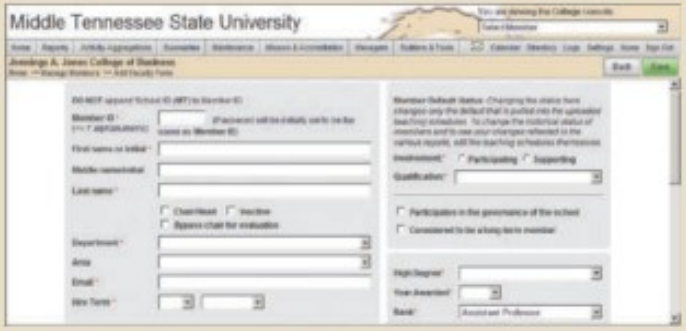
See how many databases she interacts with each day

Before leaving for work, Susan checks her Facebook and Twitter accounts	*On her lunch break, she picks up her prescription at the pharmacy*	*After work, Susan goes to the grocery store*	*At night, she plans for a trip and buys airline tickets and hotel reservations online*	*Then she makes a few online purchases*
				
Where is the data about the friends and groups stored? Where are the "likes" stored and what would they be used for?	Where is the pharmacy inventory data stored? What data about each product will be in the inventory data? What data is kept about each customer and where is it stored?	Where is the product data stored? Is the product quantity in stock updated at checkout? Does she pay with a credit card?	Where does the online travel website get the airline and hotel data from? What customer data would be kept by the website? Where would the customer data be stored?	Where are the product and stock data stored? Where does the system get the data to generate product "recommendations" to the customer? Where would credit card information be stored?
Users Friends Posts	Products Sales Customers	Products Sales Customers	Flights Hotels Customers	Products Sales Customers

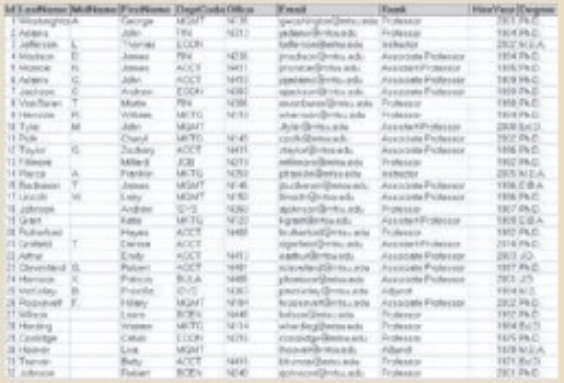
Data versus Information

FIGURE 1.2 TRANSFORMING RAW DATA INTO INFORMATION

a) Data entry screen



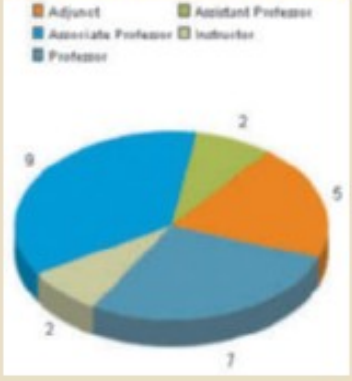
b) Raw data



c) Information in summary format

Rank	COUNT	%INFS	TOT/COL	%COL. TOT.	%COL. FAC.
Adjunct	5	20.00%	23	21.74%	3.27%
Assistant Professor	2	8.00%	28	7.14%	1.31%
Associate Professor	9	36.00%	37	24.32%	5.88%
Instructor	2	8.00%	18	11.11%	1.31%
Professor	7	28.00%	47	14.89%	4.58%

d) Information in graphical format



Data versus Information

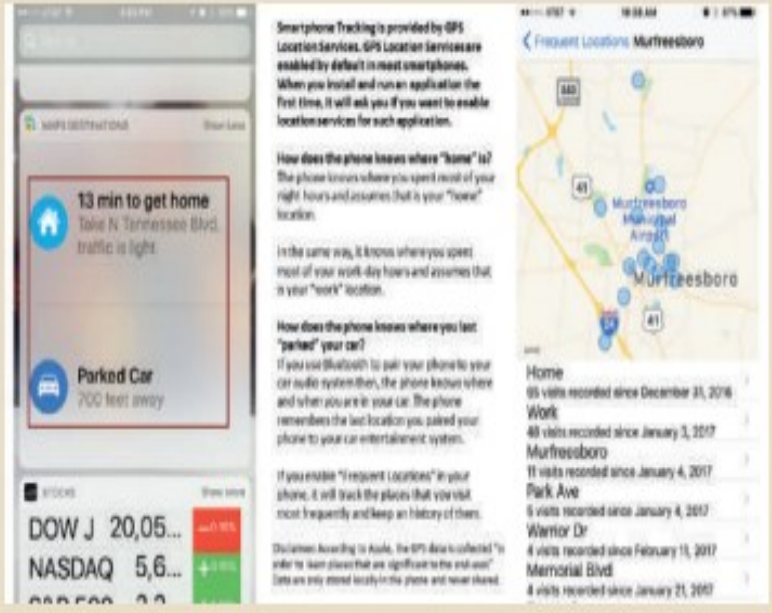
- Data consists of raw facts
  - Not yet processed to reveal meaning to the end user
  - Building blocks of information
- Information results from processing raw data to reveal meaning
  - Requires context
  - Bedrock of knowledge
  - Should be accurate, relevant, and timely

In this “information age,” production of accurate, relevant, and timely information is the key to good decision making.

In turn, good decision making is the key to business survival in a global market

Data versus Information

FIGURE 1.3 SMARTPHONE TRACKING



Introducing the Database

- What is a Database.
- Shared, integrated computer structure that stores data
  - End-user data: raw facts of interest to end user
  - Metadata: data about data, through which the end-user data is integrated and managed

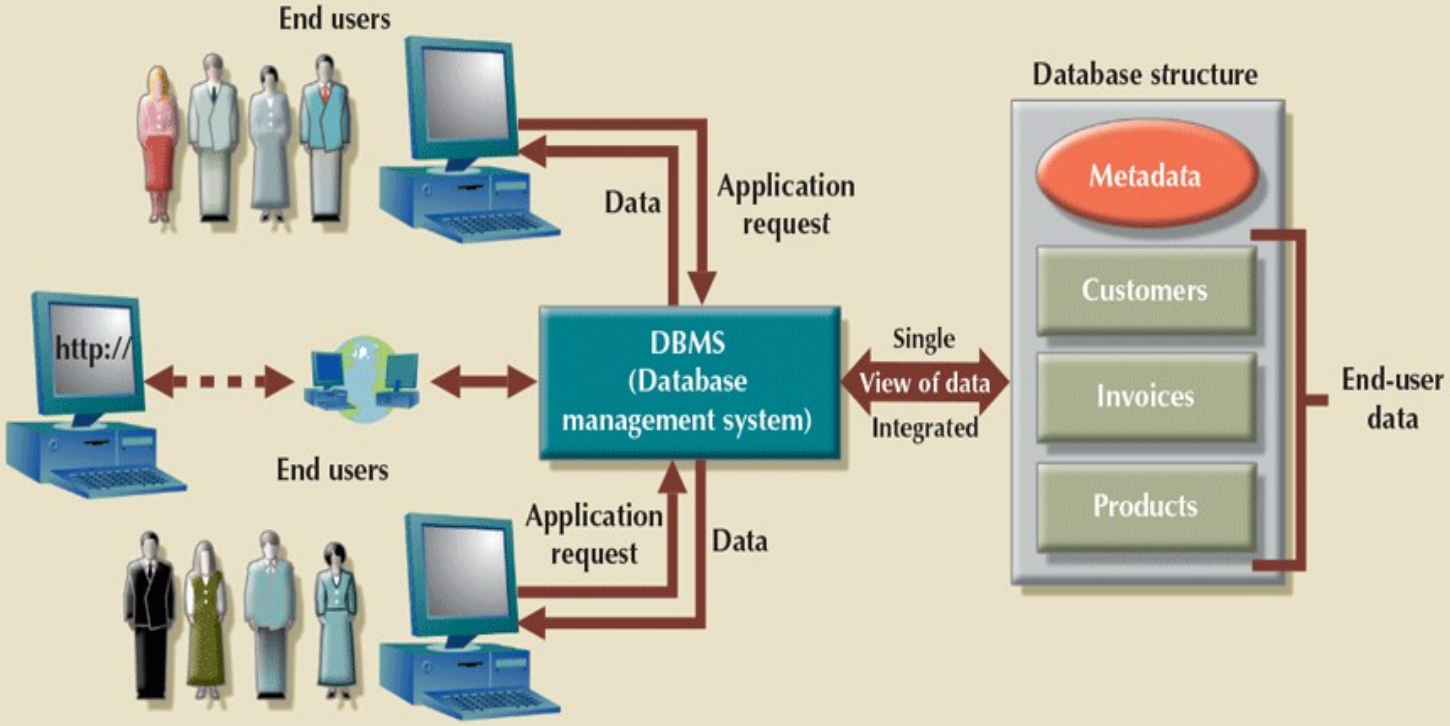
    -Describes data characteristics and relationships
- Database management system (DBMS)
  - Collection of programs
  - Manages the database structure
  - Controls access to data stored in the database

Role and Advantages of the DBMS (1 of 2)

- Database management system (DBMS): intermediary between the user and the database
  - Enables data to be shared
  - Presents the end user with an integrated view of data
  - Provides more efficient and effective data management
  - Improves sharing, security, integration, access, decision-making, productivity, etc.

Role and Advantages of the DBMS (2 of 2)

FIGURE 1.4 THE DBMS MANAGES THE INTERACTION BETWEEN THE END USER AND THE DATABASE



Types of Databases (1 of 5)

- Single-user database: supports one user at a time
  - Desktop database: single-user database on a personal computer
- Multiuser database: supports multiple users at the same time
  - Workgroup databases: supports a small number of users or a specific department
  - Enterprise database: supports many users across many departments

Types of Databases (2 of 5)

- Classification by location
  - Centralized database: data located at a single site
  - Distributed database: data distributed across different sites
  - Cloud database: created and maintained using cloud data services that provide defined performance measures for the database.

Types of Databases (2 of 5)

- *Ex. **Microsoft Azure or Amazon AWS**. These services, provided by third-party vendors, provide defined performance measures (data storage capacity, required throughput, and availability) for the database,*
- *but do not necessarily specify the underlying infrastructure to implement it. For example, 3M Health Information Systems, the world's largest provider of health care analytics software in hospitals, used Amazon's AWS cloud database services to consolidate its multiple IT centers.*

Types of Databases (3 of 5)

- Classification by data type
  - General-purpose database: contains a wide variety of data used in multiple disciplines

for example, a census database that contains general demographic data and the LexisNexis

and ProQuest databases that contain newspaper, magazine, and journal articles for a variety of topics.

Types of Databases (3 of 5)

- Discipline-specific database: contains data focused on specific subject areas –

The data in this type of database is used mainly for academic or research purposes within a small set of disciplines

- Operational database: designed to support a company's day-to-day operations

Types of Databases (4 of 5)

- Analytical database: stores historical data and business metrics used exclusively for tactical or strategic decision making
- *Such analysis typically requires extensive "data massaging" (data manipulation) to produce information on which to base pricing decisions, sales forecasts, market strategies, and so on.*
  - *Typically, analytical databases comprise two main components: a data warehouse and an online analytical processing front end.*
  - Data warehouse: stores data in a format optimized for decision support .

Types of Databases (4 of 5)

The data warehouse contains historical data obtained from the operational databases as well as data from other external sources.

- Online analytical processing (OLAP): is a set of tools that work together to provide an advanced data analysis environment for retrieving, processing, and modeling data from the data warehouse. OLAP evolved into its own discipline business Intelligence
- Business intelligence: captures and processes business data to generate information that support decision making

Types of Databases (5 of 5)

- Databases can be classified to reflect the degree to which the data is structured
  - **Unstructured data** exists in its original (raw) state
  - **Structured data** results from formatting unstructured data to facilitate storage, use, and generation of information.
    - Structure is applied based on type of processing to be performed on the data.
  - **Semistructured data**: processed to some extent . *Most data you encounter is best classified as semistructured. For example, if you look at a typical webpage, the data is presented in a prearranged format to convey some information. Unstructured and semistructured data storage and management needs are being addressed through a new generation of databases known as XML databases.*

Types of Databases (5 of 5)

- **Extensible Markup Language (XML)** : Represents and manipulates data elements in textual format
- Some data might not be ready (unstructured) for some types of processing, but they might be ready (structured) for other types of processing.
- For example, the data value 37890 might refer to a zip code, a sales value, or a product code. If this value represents a zip code or a product code and is stored as text, you cannot perform mathematical computations with it.
- On the other hand, if this value represents a sales transaction, it must be formatted as numeric.

Types of Databases (5 of 5)

- Table 1.1 compares the features of several well-known database management systems

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*	X	X	X	X	X	X	X
IBM DB2	X*	X	X	X	X	X	X	X
MySQL	X	X	X	X	X	X	X	X
Oracle RDBMS	X*	X	X	X	X	X	X	X

Why Database Design Is Important

- Focuses on design of database structure that will be used to store and manage end-user data
  - Well-designed database: facilitates data management and generates accurate and valuable information
  - Poorly designed database: causes difficult-to-trace errors that may lead to poor decision making

Evolution of File System Data Processing (1 of 3)

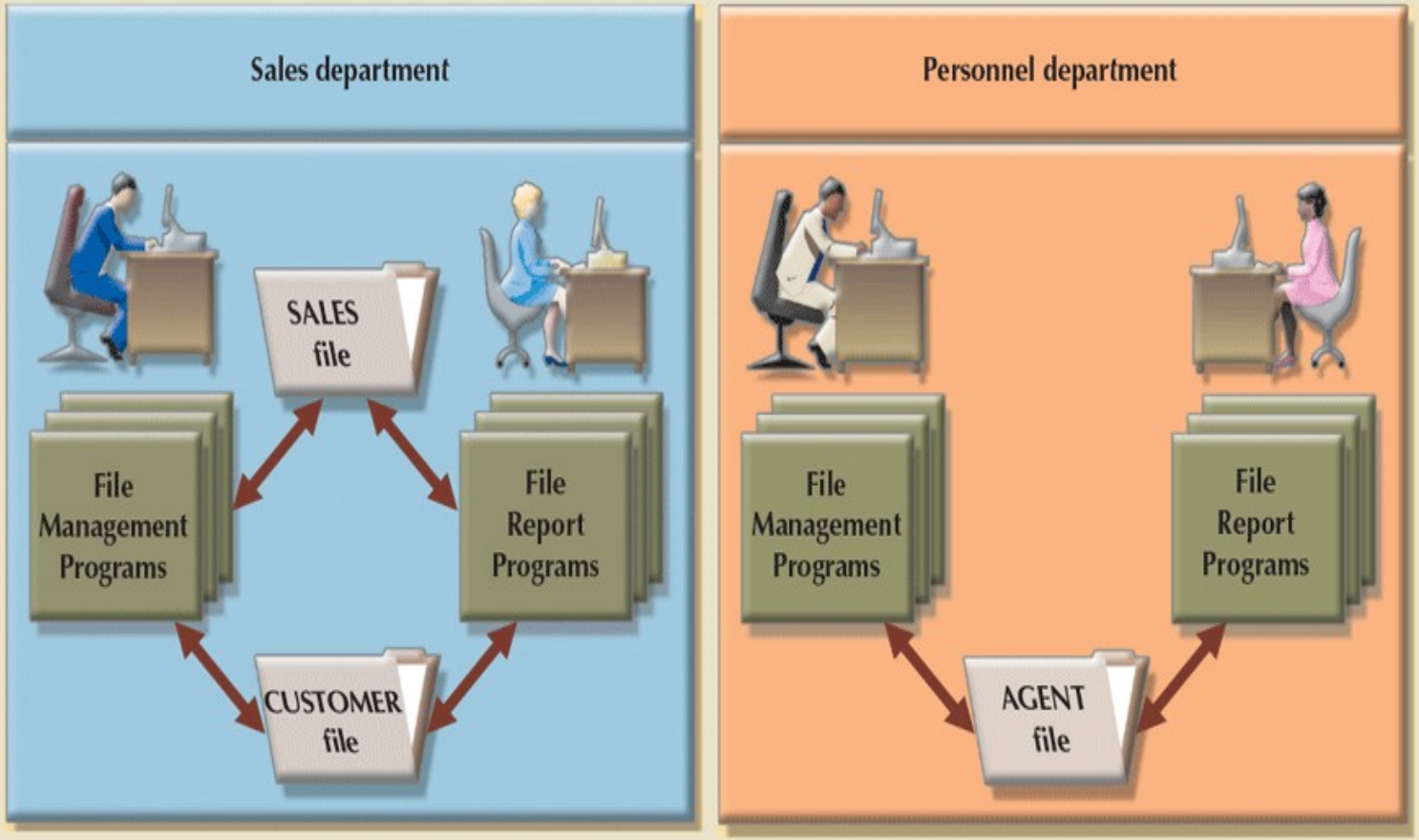
- Manual file systems
  - Accomplished through a system of file folders and filing cabinets
- Computerized file systems
  - Data processing (DP) specialist created a computer-based system to track data and produce required reports
- File system redux: modern end-user productivity tools
  - Includes spreadsheet programs such as Microsoft Excel

Evolution of File System Data Processing (2 of 3)

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 has little meaning unless it has 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 record for a customer 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.

Evolution of File System Data Processing (3 of 3)

FIGURE 1.9 A SIMPLE FILE SYSTEM



Problems with File System Data Processing

- Problems with file systems challenge the types of information that can be created from data as well as information accuracy
  - **Lengthy development times** : Problem with the file system approach is that even the simplist data reteieving task requires extensive programing. With the older file systems,a specification must be done and how to do it. Mordern database use a non procedural data manipulation language that allows the user to specify what must be done without specifying how.
  - **Difficulty of getting quick answers** : The need to write programs to reduce even the simplist reports makes adhoc queries impossible.

Problems with File System Data Processing

- **Complex system administration** :

System administration becomes more difficult as the number of files in the system expands.

Even a simple file system with a few files requires creating and maintaining several file management programs.

Each file must have its own file management programs that allow the user to add,modify and delete records ,to list the file contents and to generate reports.

Because adhoc queries are not possible, the file reporting programs can multiply quickly.

Problems with File System Data Processing

- **Lack of security and limited data sharing** :

Sharing data among multiple geographically dispersed users introduce a lot of security risks. Spreadsheet programs have insufficient security guarantees for robust data sharing among users.

- **Extensive programming**:

Changes to existing file structure in file system environment requires programs that reads, transforms,writes and repeats preceeding steps in original file.

Structural and Data Dependence (1 of 2)

- Structural dependence
  - File systems exhibit structural dependence , Access to a file is dependent on its own structure.
  - All file system programs are modified to conform to a new file structure
- Structural independence

  Exists when File structure is changed without affecting the application's ability to access the data .

  Example changes to data charcteristics such as integer to decimal, requires changes in all programs that acces the file.

Structural and Data Dependence (1 of 2)

The practical significance of data dependence

Is the difference between the logical data format (how the human being views the data) and

The physical data format (how the computer must work with the data).

Any program that access a file system format file must tell the computer not only what to do but also how to do it.

Each program must contain lines that specify the opening of a specific file type, its record specifications ,and its field definitions.

Data dependence makes the file system extremely cumbersome from the point of view of a programmer and database manager.

Data Redundancy (1 of 2)

- The file system structure makes it difficult to combine data from multiple sources ,and its lack of security renders the file system vulnerable to security breaches. The organizational structure promotes the storage of the same basic data in different locations .
- Unnecessarily storing the same data at different places
  - Islands of information (i.e., scattered data locations)
  - Increases the probability of having different versions of the same data .

Data Redundancy (1 of 2)

- Example :

In a file system entire sales department will share Access to SALES data file through the data management and reporting programs.

And with the use of spread sheets each member of the sales department can create his or her own copy of the sales data.

Because stored data at different locations may not be updated consistently,the islands of informations may contain deffirent versions of the same data (data redundancy).

Data Redundancy (2 of 2)

- Possible results of uncontrolled data redundancy
  - **Poor data security** : Multiple copies of data increases the chances for a copy of data to be susceptible to unauthorized access.
  - **Data inconsistency** : Exits when different and conflicting versions of the same data appears in different places. Example changing Agents phone number in the AGENT file and you forget to make corresponding changes in the CUSTOMER file , the files will contain different data for the same agent .Reports will yield inconsistent results depending on what version of the data is used.

Data Redundancy (2 of 2)

- **Data-entry errors** : Data entry errors are more likely to occur when complex entries (such as 10 digit phone numbers) are made in several different files or recur frequently in one or more files.
- **Data integrity problems** : It is possible to enter a non existent sales agents's name and phone number into the CUSTOMER file ,but customers are not likely to be impressed if the insurance agency supplies the name and phone numbers of an agent who does not exist.
- Also data entry error such incorrectly spelt name or an incorrect phone number yields the same kind of data integrity problems.

Data Anomalies

- Develop when not all of the required changes in the redundant data are made successfully
  - Update anomalies
  - Insertion anomalies
  - Deletion anomalies

Database Systems (1 of 2)

- Logically related data stored in a single logical data repository
  - Physically distributed among multiple storage facilities
  - DBMS eliminates most of file system's data inconsistency, data anomaly, data dependence, and structural dependence problems

Database Systems (1 of 2)

- Current generation DBMS software
  - -Stores data structures, the relationships between structures, and the access paths to those structures. All in a central location.
  - - Also takes care of Defining, storing, and managing all required access paths to those components.

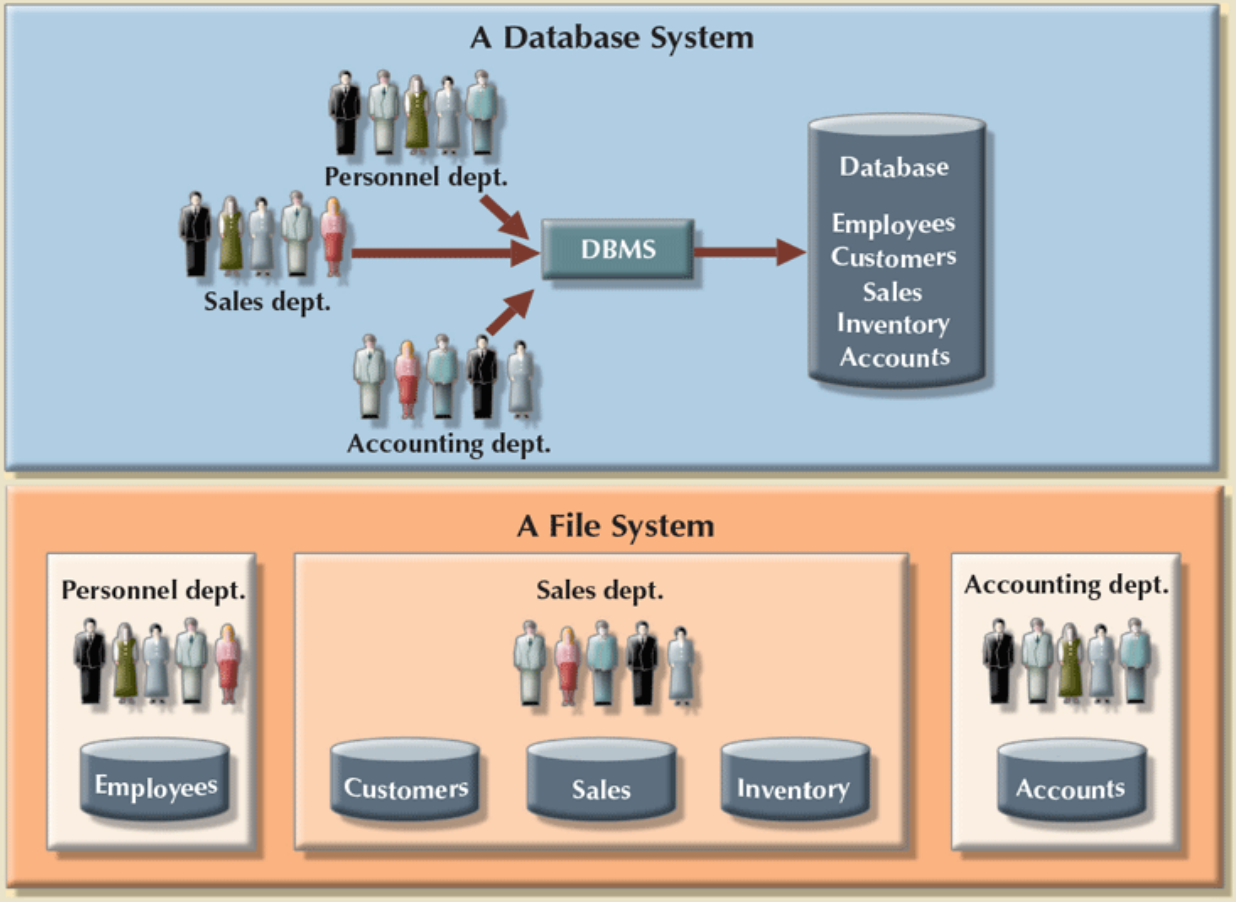
Note: DBMS is just one of several crucial components of a database system.

The DBMS may even be refered to as the database systems heart.

However, just as it takes more than a heart to make a human being function,it takes more than a DBMS to make a database system function.

Database Systems (2 of 2)

FIGURE 1.10 CONTRASTING DATABASE AND FILE SYSTEMS



The Database System Environment (1 of 2)

- Database system: organization of components that define and regulate the collection, storage, management, and use of data within a database environment.
- From general management point of view , the database system is composed of five major parts:
  - Hardware
  - Software
  - People
  - Procedures
  - Data

The Database System Environment (1 of 2)

- **Software :** Three types of software are needed to make the database system function fully:
  - *Operating system software* manages all hardware components and makes it possible for all other software to run on the computers.
  - DBMS software manages the database within the database system. Examples , Microsoft SQL sever, Oracle Corporation's Oracle, and Oracles's MySQL and IBM's DB2.
  - *Application programs and utilities software* are used to access and manipulate data in the DBMS and manage the computers environment in which data access and manipulation takes place.
  - *Utilities* are the software tools used to help manage the database systems computer components .
  - Examples: DBMS vendors provide GUI to help create database structures,control database access and monitor database operations.

The Database System Environment (1 of 2)

- **People** : All users of the database system.

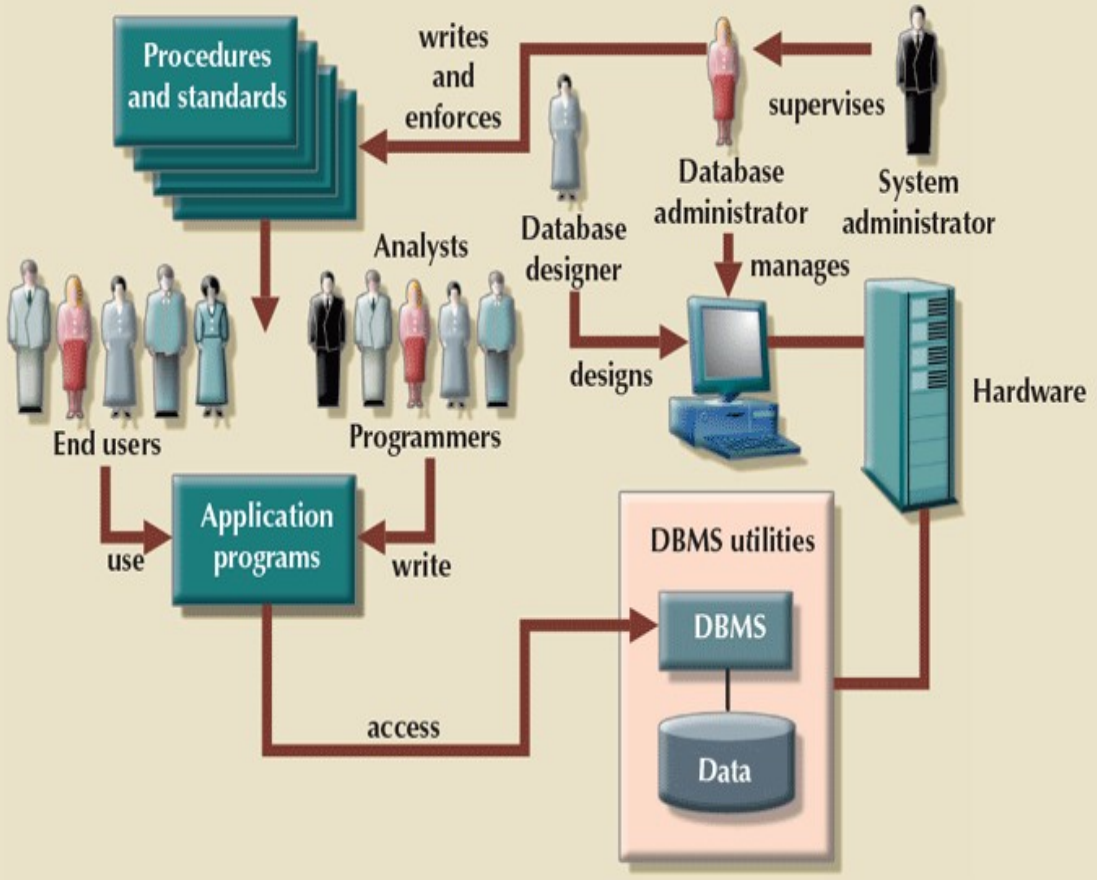
  Five type of users can be identified.

System Administrators , Database administrators, database designers, system analyst and programmers and end users.

- **Procedures**: Instructions and rules that govern the design and use of db systems.
- **Data** : Collection of facts stored in the database.

The Database System Environment (2 of 2)

FIGURE 1.11 THE DATABASE SYSTEM ENVIRONMENT



DBMS Functions (1 of 3)

- DBMS performs several functions that guarantee the integrity and consistency of the database. Most of those functions are transparent to end users, can be achieved only through the use of the DBMS. The functions includes:
- Data dictionary management
  - Data dictionary: stores definitions of data elements and their relationships
- Data storage management
  - Performance tuning ensures efficient performance
- Data transformation and presentation
  - Data is formatted to conform to logical expectations
- Security management : Enforces user security and data privacy

DBMS Functions (2 of 3)

- Multiuser access control
  - Sophisticated algorithms ensure that multiple users can access the database concurrently without compromising its integrity
- Backup and recovery management
  - Enables recovery of the database after a failure
- Data integrity management
  - Minimizes redundancy and maximizes consistency

DBMS Functions (3 of 3)

- Database access languages and application programming interfaces
  - Query language: lets the user specify what must be done without having to specify how
  - Structured Query Language (SQL): de facto query language and data access standard supported by the majority of DBMS vendors
- Database communication interfaces
  - Accept end-user requests via multiple, different network environments

Managing the Database System: A Shift in Focus

- Disadvantages of database systems
  - Increased costs
  - Management complexity
  - Maintaining currency
  - Vendor dependence
  - Frequent upgrade/replacement cycles

Preparing for Your Database Professional Career

TABLE 1.3	DATABASE CAREER OPPORTUNITIES	
JOB TITLE	DESCRIPTION	SAMPLE SKILLS REQUIRED
Database Developer	Create and maintain database-based applications	Programming, database fundamentals, SQL
Database Designer	Design and maintain databases	Systems design, database design, SQL
Database Administrator	Manage and maintain DBMS and databases	Database fundamentals, SQL, vendor courses
Database Analyst	Develop databases for decision support reporting	QL, query optimization, data warehouses
Database Architect	Design and implementation of database environments (conceptual, logical, and physical)	DBMS fundamentals, data modeling, SQL, hardware knowledge, etc.
Database Consultant	Help companies leverage database technologies to improve business processes and achieve specific goals	Database fundamentals, data modeling, database design, SQL, DBMS, hardware, vendor-specific technologies, etc.
Database Security Officer	Implement security policies for data administration	DBMS fundamentals, database administration, SQL, data security technologies, etc.
Cloud Computing Data Architect	Design and implement the infrastructure for next-generation cloud database systems	Internet technologies, cloud storage technologies, data security, performance tuning, large databases, etc.
Data Scientist	Analyze large amounts of varied data to generate insights, relationships, and predictable behaviors	Data analysis, statistics, advanced mathematics, SQL, programming, data mining, machine learning, data visualization

Summary

- Data consists of raw facts and is usually stored in a database
  - Database design defines the database structure
    - Can be classified according to the number of users, location, as well as data usage and structure
  - Databases evolved from manual and computerized file systems
    - There are some limitations of file system data management
    - DBMSs were developed to address the file system's inherent weaknesses