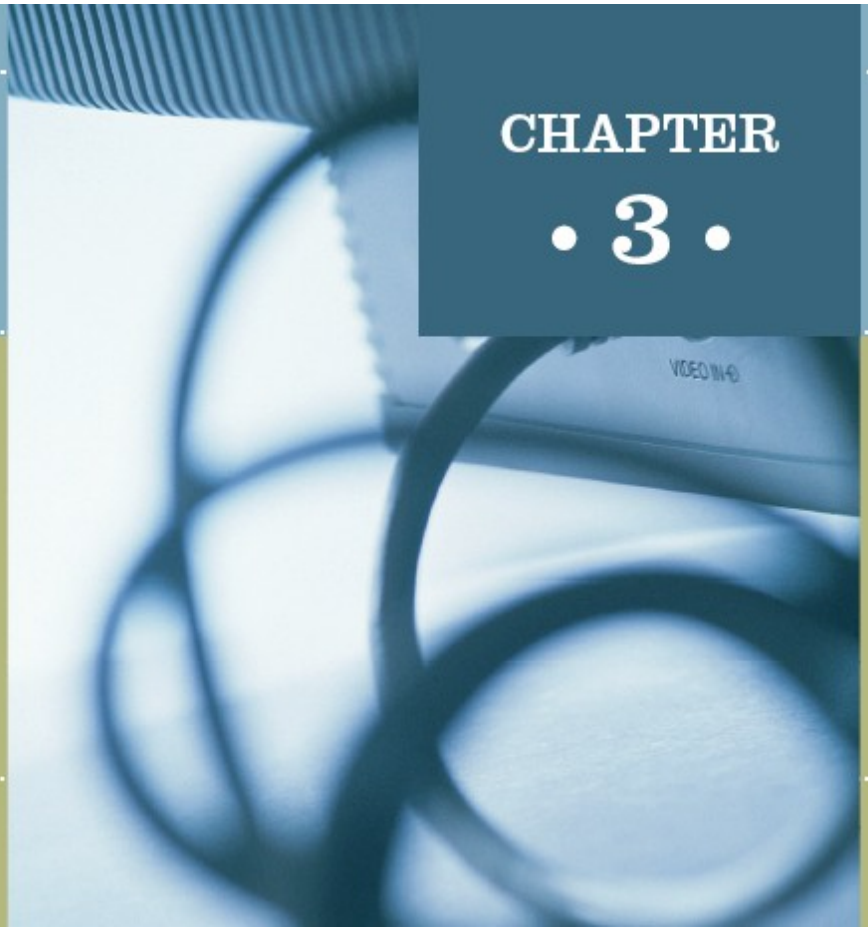
TỔ CHỨC DỮ LIỆU VÀ THÔNG TIN



CHAPTER

• 3 •

Organizing Data and Information

Fundamentals of Information Systems

Third Edition

Các mục tiêu

- Cách thức quản lý dữ liệu dựa trên phương pháp database có nhiều ưu điểm hơn phương pháp file-based.
  - Định nghĩa các khái niệm cơ bản về quản lý dữ liệu và các thuật ngữ có liên quan
  - Giới thiệu Mô hình dữ liệu quan hệ và liệt kê các tính năng cơ bản

Principles and Learning Objectives (continued)

- Một CSDL được thiết kế và quản lý tốt sẽ là một công cụ rất hữu ích cho việc ra quyết định trong kinh doanh
  - Xác định các chức năng cơ bản thực hiện bởi các hệ quản trị CSDL và mô tả một vài hệ quản trị CSDL thông dụng

Principles and Learning Objectives (continued)

- Các ứng dụng CSDL phát triển rất nhanh chóng và có những ảnh hưởng rất lớn đến hiệu quả kinh doanh
  - Xác định và miêu tả ngắn gọn về các ứng dụng CSDL hiện tại

Data Management: The Hierarchy of Data

- **Bit** (a binary digit): a circuit that is either on or off
- **Byte:** 8 bits
- **Character:** each byte represents a character; the basic building block of information
- **Field:** name, number, or characters that describe an aspect of a business object or activity

The Hierarchy of Data (continued)

- **Record:** a collection of related data fields
- **File:** a collection of related records
- **Database:** a collection of integrated and related files
- **Hierarchy of data**
  - Bits, characters, fields, records, files, and databases

The Hierarchy of Data (continued)

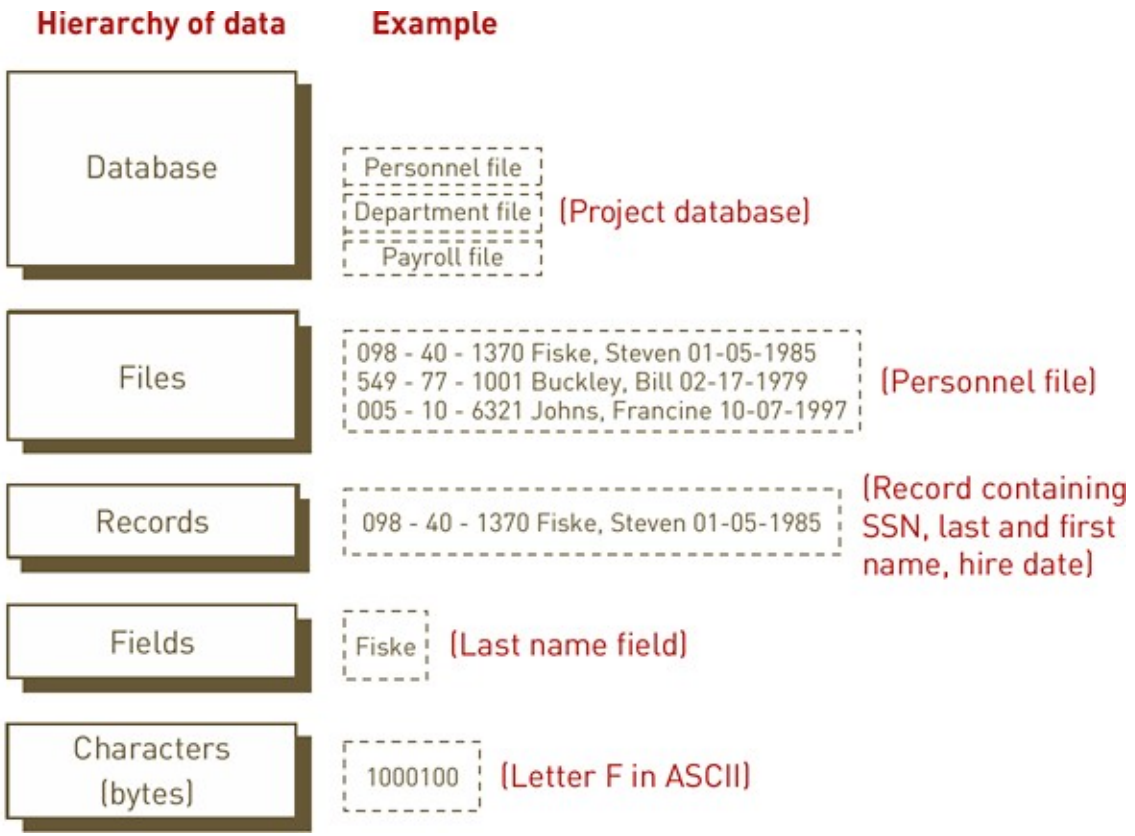


Figure 3.1: The Hierarchy of Data

Data Entities, Attributes, and Keys

- **Entity – Thực thể:** a generalized class of people, places, or things (objects) for which data is collected, stored, and maintained
- **Attribute – Thuộc tính:** a characteristic of an entity
- **Data item – Giá trị thuộc tính:** a value of an attribute
- **Key - Khóa:** field(s) that identify a record
- **Primary key – Khóa chính:** field(s) that *uniquely* identify a record

Data Entities, Attributes, and Keys (continued)

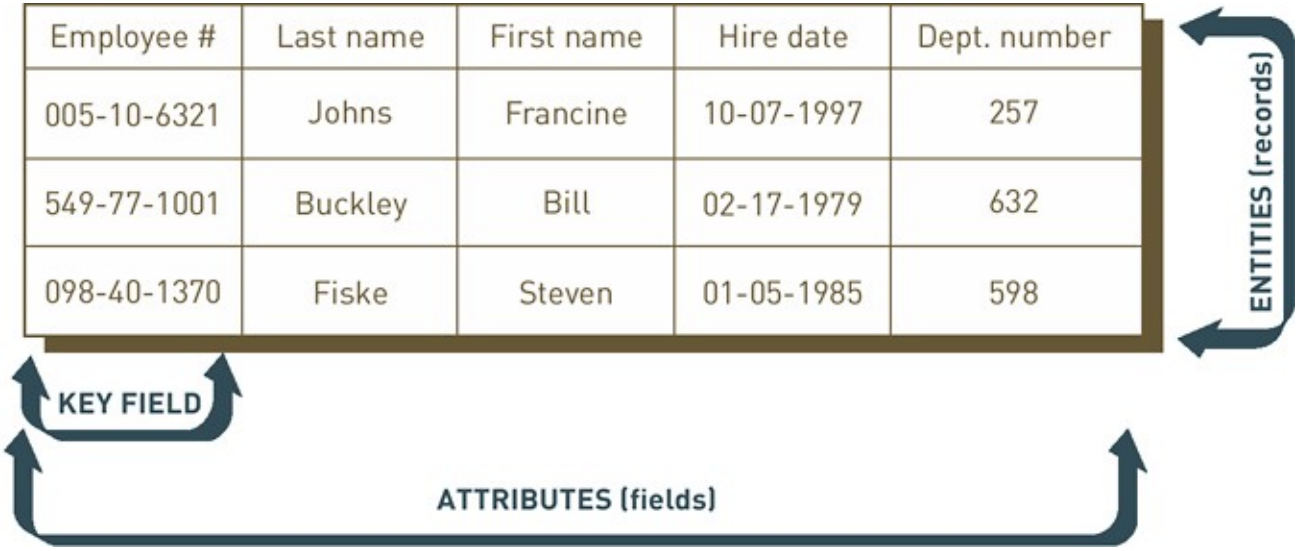


Figure 3.2: Keys and Attributes

The Traditional Approach Versus the Database Approach

- Traditional approach – Phương pháp truyền thống: separate data files are created for each application
  - Results in data redundancy (duplication) – Dư thừa dữ liệu
  - Data redundancy conflicts with data integrity – Toàn vẹn dữ liệu
- Database approach – Phương pháp dựa trên CSDL: pool of related data is shared by multiple applications
  - Significant advantages over traditional approach

The Traditional Approach Versus the Database Approach (continued)

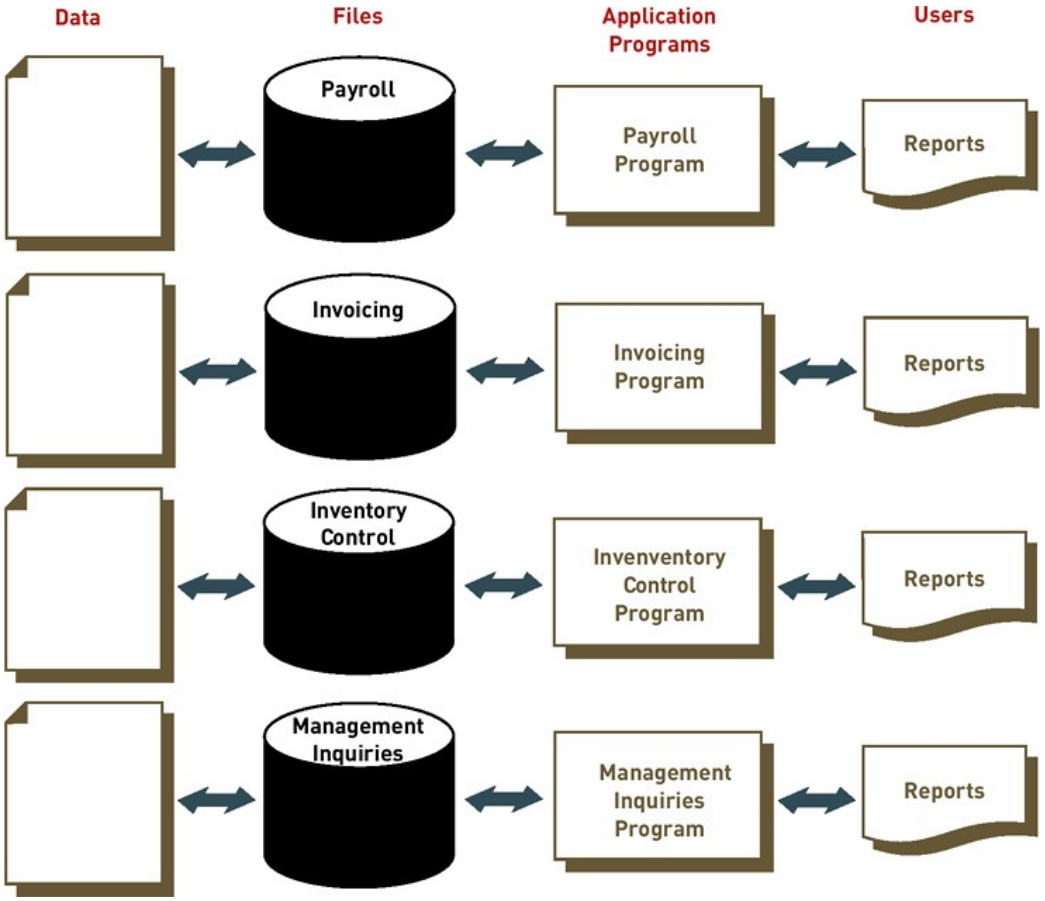


Figure 3.3: The Traditional Approach to Data Management

The Traditional Approach Versus the Database Approach (continued)

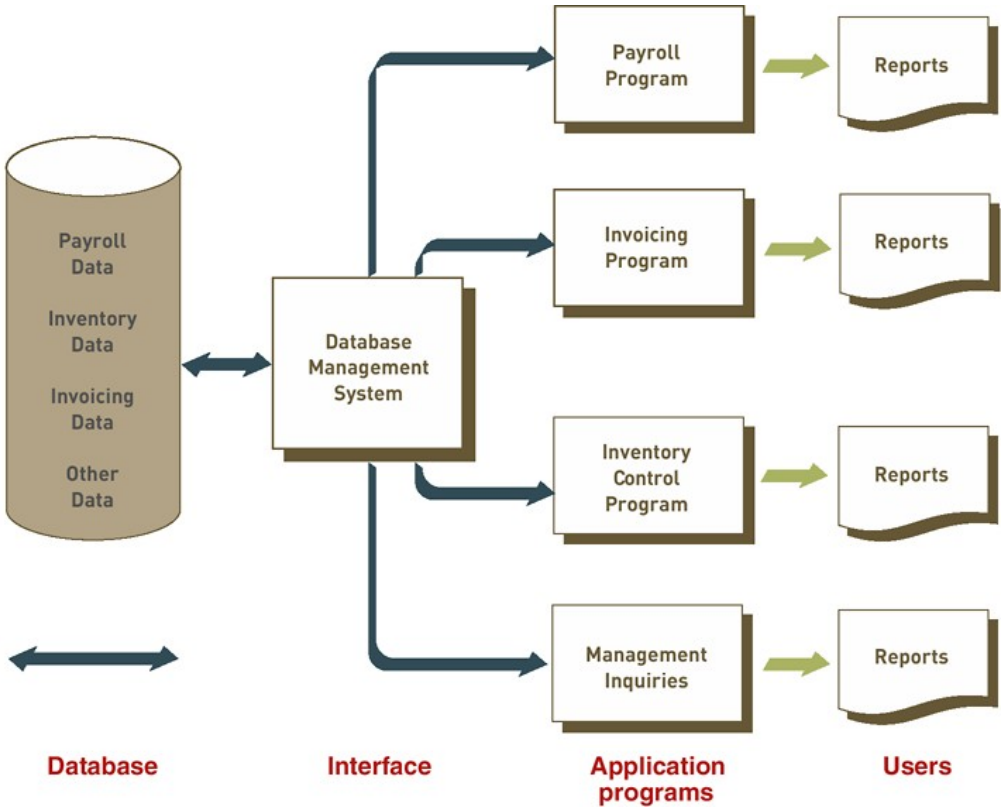


Figure 3.4: The Database Approach to Data Management

The Traditional Approach Versus the Database Approach (continued)

Advantages	Explanation
Improved strategic use of corporate data	Accurate, complete, up-to-date data can be made available to decision makers where, when, and in the form they need it.
Reduced data redundancy	The database approach can reduce or eliminate data redundancy. Data is organized by the DBMS and stored in only one location. This results in more efficient utilization of system storage space.
Improved data integrity	With the traditional approach, some changes to data were not reflected in all copies of the data kept in separate files. This is prevented with the database approach because there are no separate files that contain copies of the same piece of data.
Easier modification and updating	With the database approach, the DBMS coordinates updates and data modifications. Programmers and users do not have to know where the data is physically stored. Data is stored and modified once. Modification and updating are also easier because the data is stored at only one location in most cases.
Data and program independence	The DBMS organizes the data independently of the application program. With the database approach, the application program is not affected by the location or type of data. Introduction of new data types not relevant to a particular application does not require the rewriting of that application to maintain compatibility with the data file.

Table 3.1: Advantages of the Database Approach

The Traditional Approach Versus the Database Approach (continued)

Better access to data and information	Most DBMSs have software that makes it easy to access and retrieve data from a database. In most cases, simple commands can be given to get important information. Relationships between records can be more easily investigated and exploited, and applications can be more easily combined.
Standardization of data access	A primary feature of the database approach is a standardized, uniform approach to database access. This means that the same overall procedures are used by all application programs to retrieve data and information.
A framework for program development	Standardized database access procedures can mean more standardization of program development. Because programs go through the DBMS to gain access to data in the database, standardized database access can provide a consistent framework for program development. In addition, each application program need address only the DBMS, not the actual data files, reducing application development time.
Better overall protection of the data	The use of and access to centrally located data is easier to monitor and control. Security codes and passwords can ensure that only authorized people have access to particular data and information in the database, thus ensuring privacy.
Shared data and information resources	The cost of hardware, software, and personnel can be spread over a large number of applications and users. This is a primary feature of a DBMS.

Table 3.1: Advantages of the Database Approach (continued)

The Traditional Approach Versus the Database Approach (continued)

Disadvantages	Explanation
More complexity	Database management systems can be difficult to set up and operate. Many decisions must be made correctly for the database management system to work effectively. In addition, users have to learn new procedures to take full advantage of a database management system.
More difficult to recover from a failure	With the traditional approach to file management, a failure of a file only affects a single program. With a database management system, a failure can shut down the entire database.
More expensive	Database management systems can be more expensive to purchase and operate. The expense includes the cost of the database and specialized personnel, such as a database administrator, who is needed to design and operate the database.

Table 3.2: Disadvantages of the Database Approach

Mô hình hóa dữ liệu và Mô hình Cơ sở dữ liệu quan hệ

- Khi xây dựng một cơ sở dữ liệu, cần xem xét các khía cạnh
  - Nội dung: What data should be collected, at what cost?
  - Truy cập: What data should be provided to which users, and when?
  - Cấu trúc logic: How should data be arranged to make sense to a given user?
  - Tổ chức vật lý: Where should data be physically located?

Mô hình dữ liệu - Data Modeling

- Building a database requires two types of design
  - Logical design (Thiết kế logic)
    - Shows an abstract model of how data should be structured and arranged to meet an organization's information needs
  - Physical design (Thiết kế vật lý)
    - Fine-tunes the logical database design for performance and cost considerations

Data Modeling (continued)

- **Mô hình dữ liệu - Data model:** Lược đồ mô hình các thực thể dữ liệu và mối quan hệ giữa chúng
- **Lược đồ Quan hệ Thực thể - Entity-relationship (ER) diagrams:** Các mô hình dữ liệu sử dụng các ký hiệu đồ họa để biểu diễn tổ chức dữ liệu và mối quan hệ giữa các dữ liệu

Data Modeling (continued)

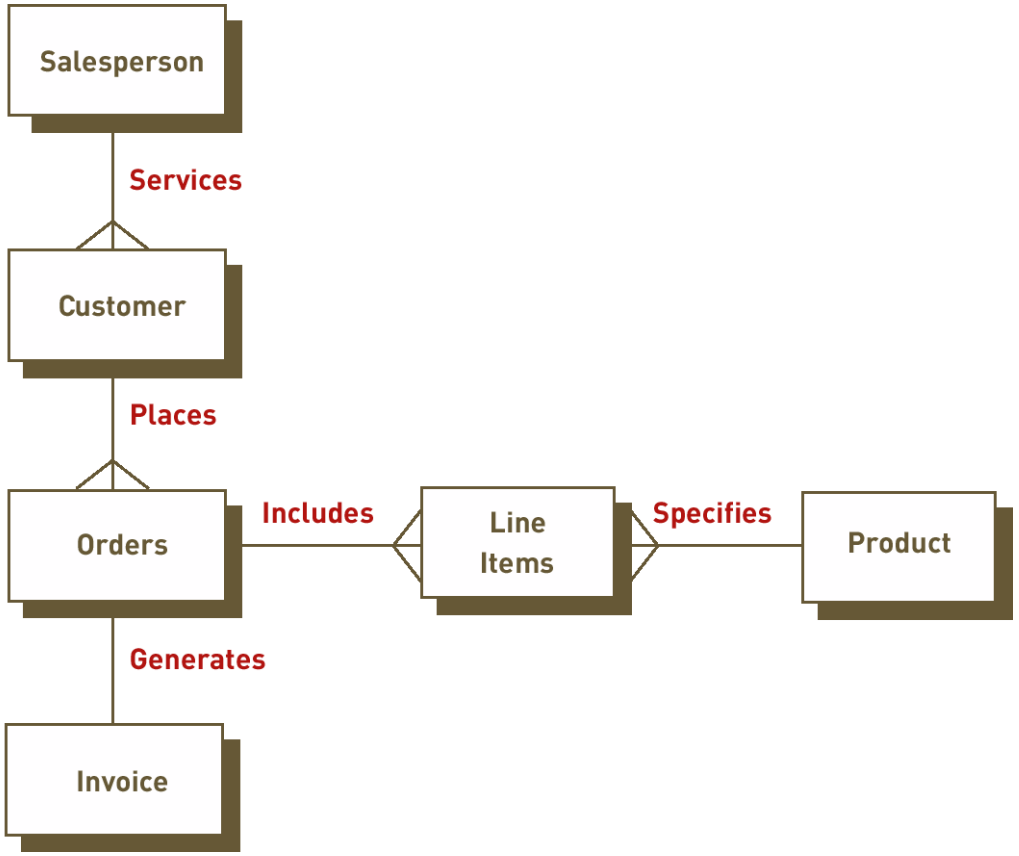


Figure 3.5: An Entity-Relationship (ER) Diagram for a Customer Order Database

The Relational Database Model

- **Mô hình quan hệ - Relational model:** all data elements are placed in two-dimensional tables (relations), which are the logical equivalent of files
- In the relational model:
  - Each row of a table represents a data entity
  - Columns of the table represent attributes
- **Miền - Domain:** the allowable values for data attributes

The Relational Database Model (continued)

Data table 1: Project table

Project number	Description	Dept. number
155	Payroll	257
498	Widgets	632
226	Sales manual	598

Data table 2: Department table

Dept.	Dept. name	Manager SSN
257	Accounting	005-10-6321
632	Manufacturing	549-77-1001
598	Marketing	098-40-1370

Data table 3: Manager table

SSN	Last name	First name	Hire date	Dept. number
005-10-6321	Johns	Francine	10-07-2005	257
549-77-1001	Buckley	Bill	02-17-1989	632
098-40-1370	Fiske	Steven	01-05-1995	598

Figure 3.6: A Relational Database Model

Các thao tác với dữ liệu – Manipulating Data

- **Phép chọn - Selecting:** eliminates rows according to criteria
- **Phép chiếu - Projecting:** eliminates columns in a table
- **Phép nối - Joining:** combines two or more tables
- **Phép liên kết - Linking:** relates or links two or more tables using common data attributes

Manipulating Data (continued)

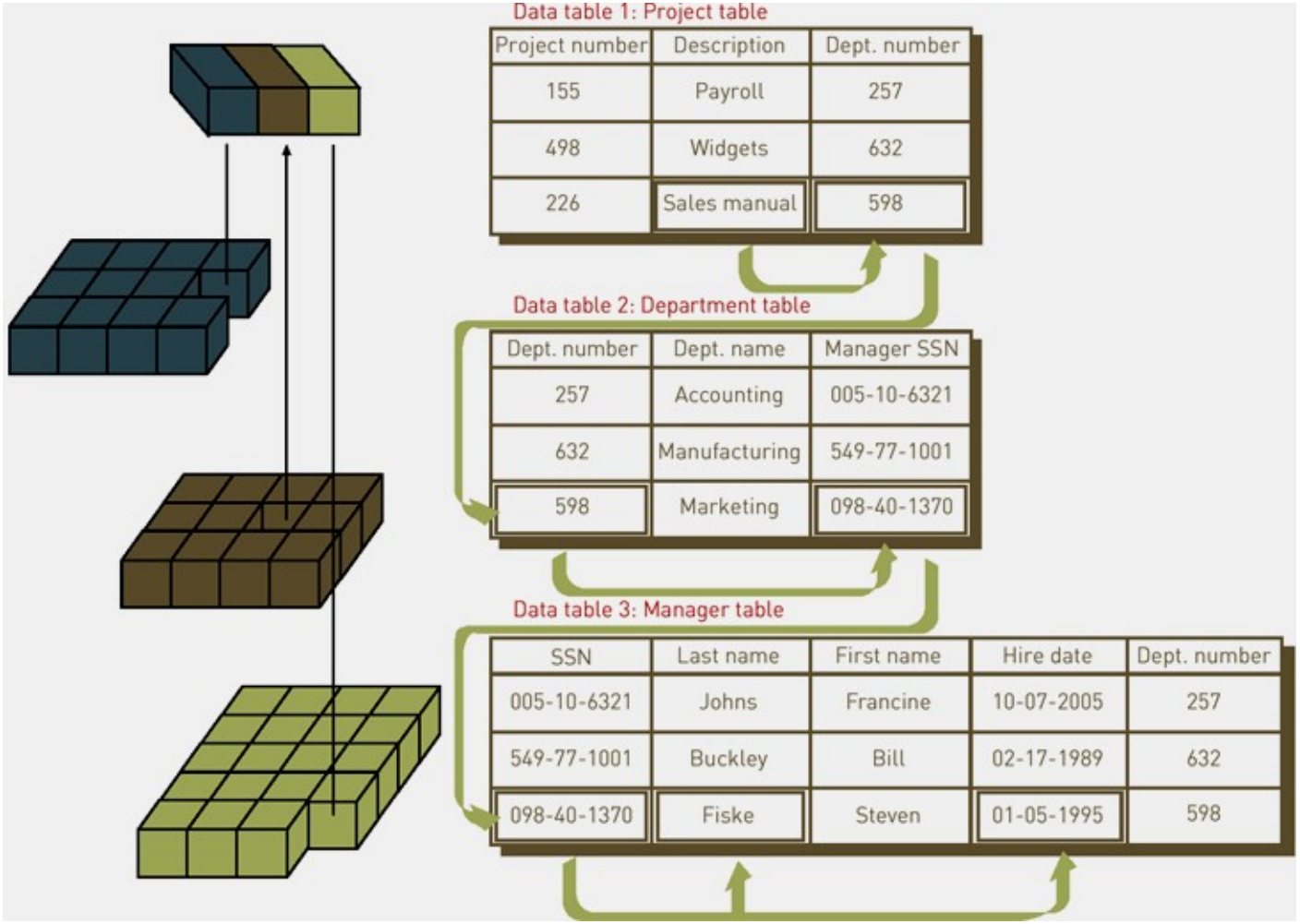


Figure 3.8: Linking Data Tables to Answer an Inquiry

Hệ Quản trị Cơ sở dữ liệu - Database Management Systems (DBMS)

- Interface between
  - Database and application programs
  - Database and the user
- Database types
  - Flat file
  - Single user
  - Multiple users

Providing a User View

- **Schema:** description of the entire database
- **User view:** user-accessible portion of the database
- **Subschema**
  - Contains a description of a subset of the database
  - Identifies which users can view and modify the data items in the subset
  - Is used to create different user views

Providing a User View (continued)

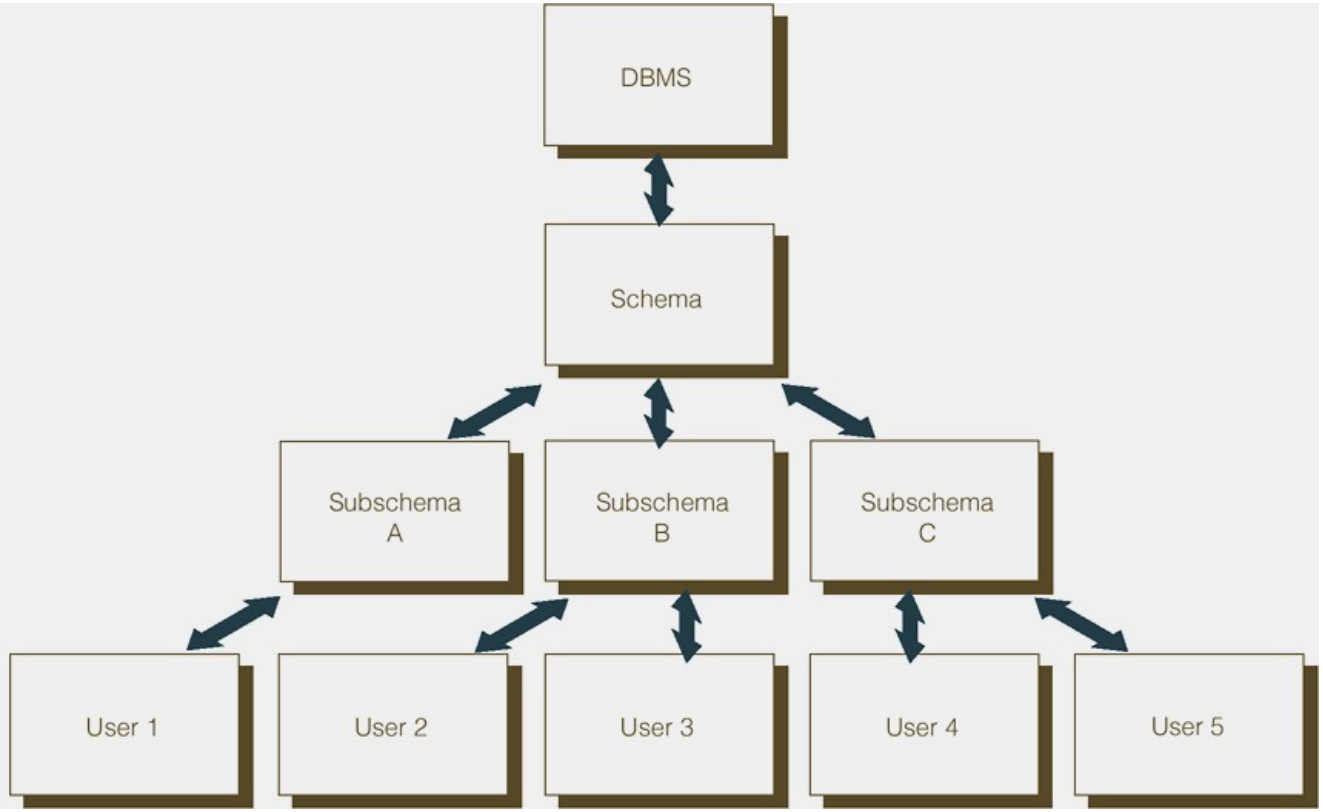


Figure 3.10: The Use of Schemas and Subschemas

Creating and Modifying the Database

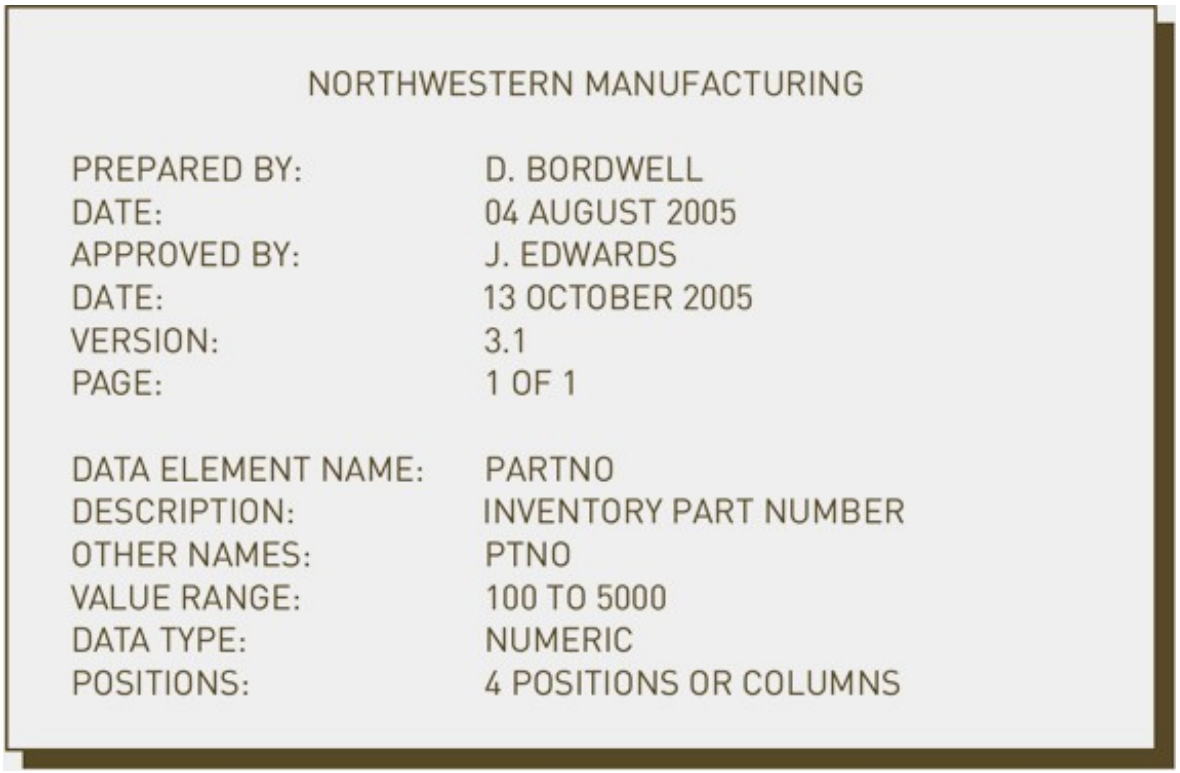
- **Data definition language (DDL)**
  - Collection of instructions/commands that define and describe data and data relationships in a database
  - Allows database creator to describe the data and the data relationships that are to be contained in the schema and the subschemas
- **Data dictionary:** a detailed description of all the data used in the database

Creating and Modifying the Database (continued)

```
SCHEMA DESCRIPTION
SCHEMA NAME IS XXXX
AUTHOR        XXXX
DATE          XXXX
FILE DESCRIPTION
      FILE NAME IS XXXX
        ASSIGN XXXX
      FILE NAME IS XXXX
        ASSIGN XXXX
AREA DESCRIPTION
      AREA NAME IS XXXX
RECORD DESCRIPTION
      RECORD NAME IS XXXX
      RECORD ID IS XXXX
      LOCATION MODE IS XXXX
      WITHIN XXXX AREA FROM XXXX THRU XXXX
SET DESCRIPTION
      SET NAME IS XXXX
      ORDER IS XXXX
      MODE IS XXXX
      MEMBER IS XXXX
      .
      .
      .
```

Figure 3.11: Using a Data Definition Language to Define a Schema

Creating and Modifying the Database (continued)



NORTHWESTERN MANUFACTURING

PREPARED BY:	D. BORDWELL
DATE:	04 AUGUST 2005
APPROVED BY:	J. EDWARDS
DATE:	13 OCTOBER 2005
VERSION:	3.1
PAGE:	1 OF 1
DATA ELEMENT NAME:	PARTNO
DESCRIPTION:	INVENTORY PART NUMBER
OTHER NAMES:	PTNO
VALUE RANGE:	100 TO 5000
DATA TYPE:	NUMERIC
POSITIONS:	4 POSITIONS OR COLUMNS

Figure 3.12: A Typical Data Dictionary Entry

Storing and Retrieving Data

- When an application requests data from the DBMS, the application follows a logical access path
- When the DBMS goes to a storage device to retrieve the requested data, it follows a path to the physical location (physical access path) where the data is stored

Storing and Retrieving Data (continued)

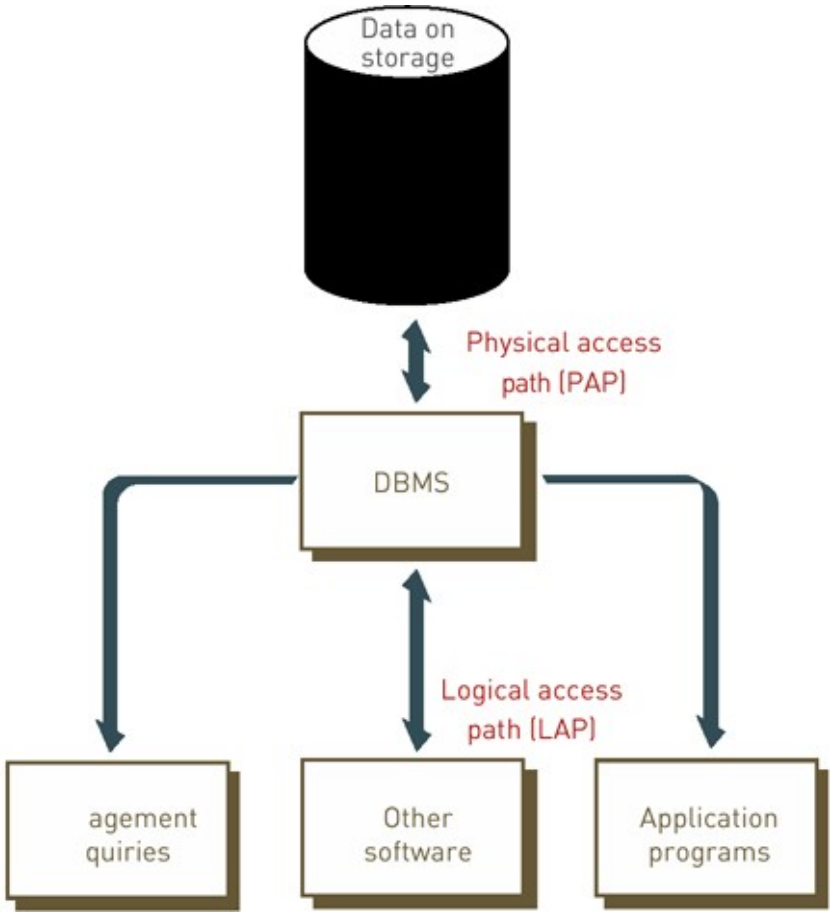


Figure 3.13: Logical and Physical Access Paths

Manipulating Data and Generating Reports

- **Query-By-Example (QBE):** a visual approach to developing database queries or requests
- **Data manipulation language (DML):** commands that manipulate the data in a database
- **Structured Query Language (SQL):** ANSI standard query language for relational databases
- Database programs can produce reports, documents, and other outputs

Manipulating Data and Generating Reports (continued)

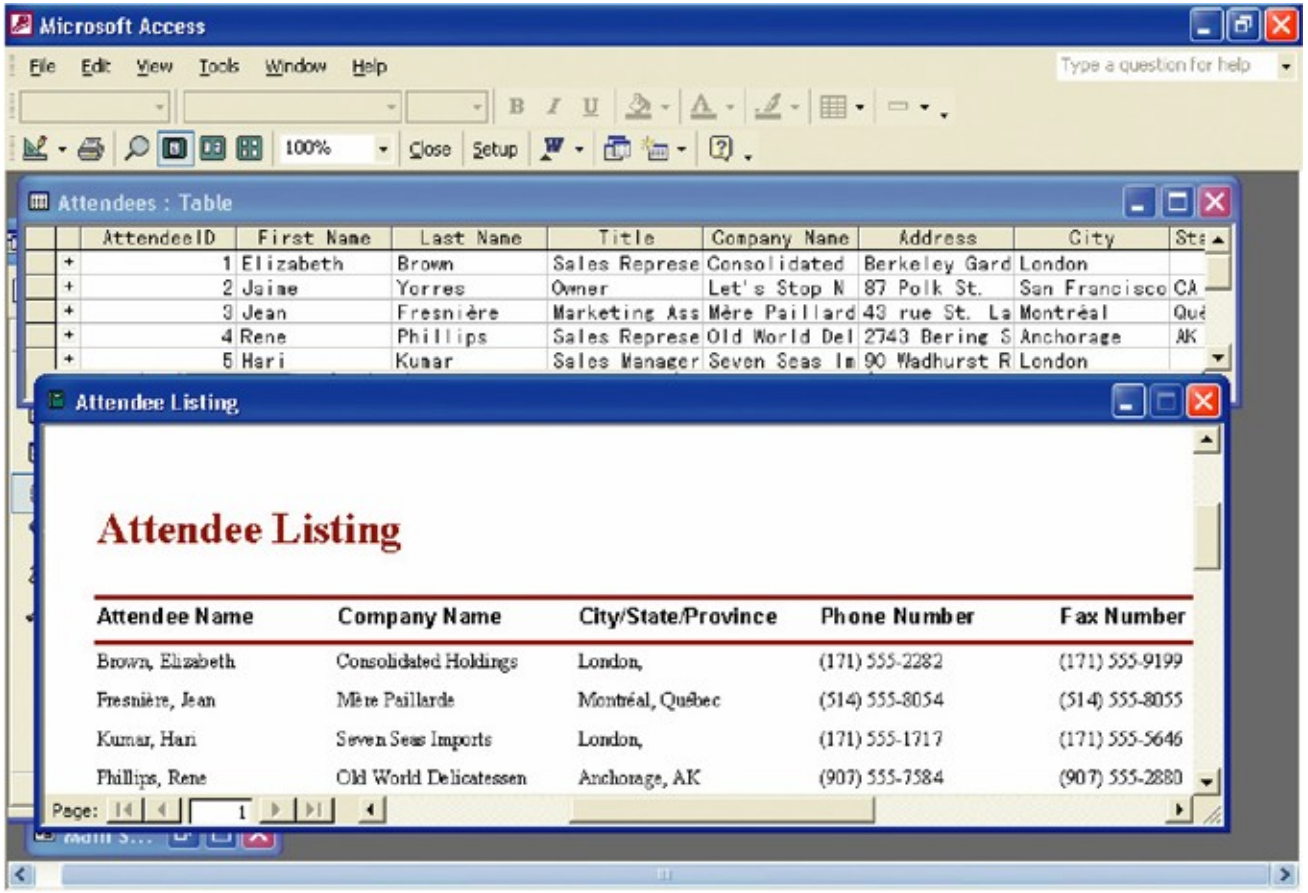


Figure 3.16: Database Output

Quản trị CSDL - Database Administration

- **Database administrator (DBA):** directs or performs all activities to maintain a database environment
  - Designing, implementing, and maintaining the database system and the DBMS
  - Establishing policies and procedures
  - Training employees

Các hệ QTCSDL phổ biến - Popular Database Management Systems

- Popular DBMSs for end users: Microsoft Access and Corel Paradox
- The complete database management software market includes databases by IBM, Oracle, and Microsoft
- Examples of open-source database systems: PostgreSQL and MySQL
- Many traditional database programs are now available on open-source operating systems

Special-Purpose Database Systems

- Summation and Concordance
- CaseMap
- LiveNote
- Scottish Intelligence Database (SID)
- GlobalSpec

Lựa chọn một Hệ QTCSDL - Selecting a Database Management System

- Important characteristics of databases to consider:
  - Size of the database
  - Number of concurrent users
  - Performance
  - Ability to be integrated with other systems
  - Features of the DBMS
  - Vendor considerations
  - Cost of the system

Using Databases with Other Software

- Database management systems are often used with other software packages or the Internet
- A database management system can act as a front-end application or a back-end application
  - Front-end application: interacts with users
  - Back-end application: interacts with applications

Database Applications: Linking the Company Database to the Internet

- Corporate databases can be accessed by customers, suppliers, and employees through:
  - The Internet
  - Intranets
  - Extranets
- **Semantic Web:** Developing a seamless integration of traditional databases with the Internet

Data Warehouses, Data Marts, and Data Mining

- **Data warehouse:** collects business information from many sources in the enterprise
- **Data mart:** a subset of a data warehouse
- **Data mining:** an information-analysis tool for discovering patterns and relationships in a data warehouse or a data mart

Data Warehouses, Data Marts, and Data Mining (continued)

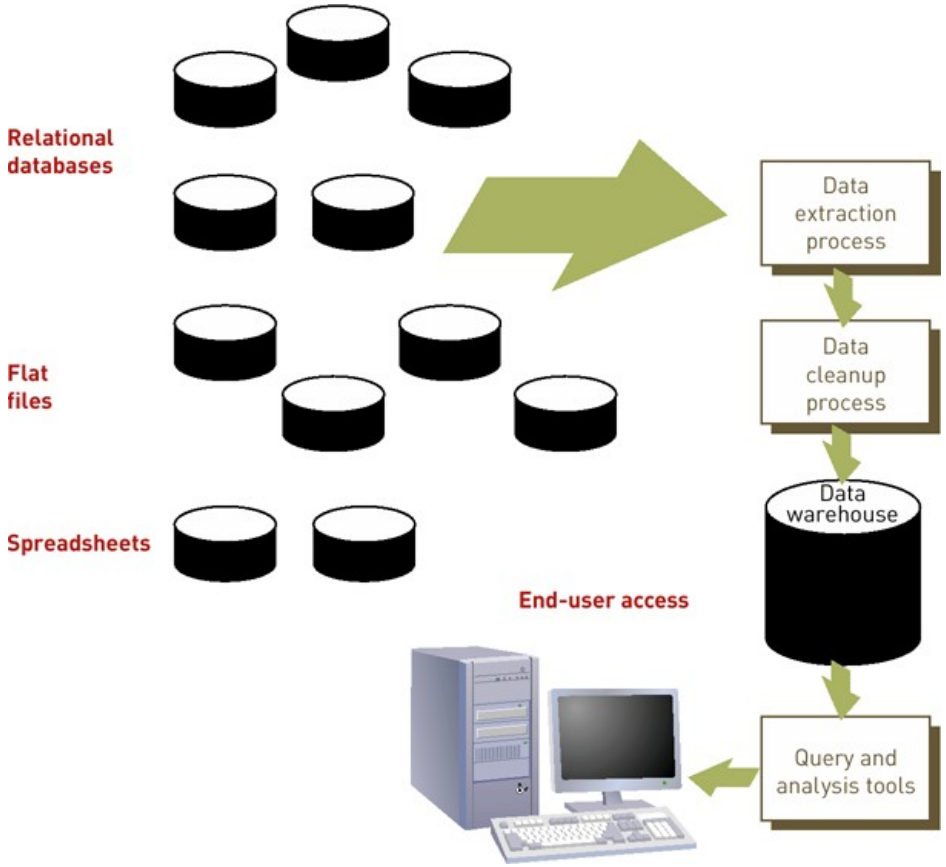


Figure 3.17: Elements of a Data Warehouse

Data Warehouses, Data Marts, and Data Mining (continued)

Application	Description
Branding and positioning of products and services	Enable the strategist to visualize the different positions of competitors in a given market using performance (or importance) data on dozens of key features of the product in question and then to condense all that data into a perceptual map of just two or three dimensions
Customer churn	Predict current customers who are likely to go to a competitor
Direct marketing	Identify prospects most likely to respond to a direct marketing campaign (such as a direct mailing)
Fraud detection	Highlight transactions most likely to be deceptive or illegal
Market basket analysis	Identify products and services that are most commonly purchased at the same time (e.g., nail polish and lipstick)
Market segmentation	Group customers based on who they are or on what they prefer
Trend analysis	Analyze how key variables (e.g., sales, spending, promotions) vary over time

Table 3.3: Common Data-Mining Applications

Business Intelligence

- **Business intelligence (BI):** gathering the right information in a timely manner and usable form and analyzing it to have a positive impact on business
- **Knowledge management:** capturing a company's collective expertise and distributing it wherever it can help produce the biggest payoff

Distributed Databases

- **Distributed database**
  - Data may be spread across several smaller databases connected via telecommunications devices
  - Corporations get more flexibility in how databases are organized and used
- **Replicated database**
  - Holds a duplicate set of frequently used data

Online Analytical Processing (OLAP)

- Software that allows users to explore data from a number of different perspectives

Characteristic	OLAP	Data Mining
Purpose	Supports data analysis and decision making	Supports data analysis and decision making
Type of analysis supported	Top-down, query-driven data analysis	Bottom-up, discovery-driven data analysis
Skills required of user	Must be very knowledgeable of the data and its business context	Must trust in data-mining tools to uncover valid and worthwhile hypothesis

Table 3.4: Comparison of OLAP and Data Mining

Object-Oriented and Object-Relational Database Management Systems

- **Object-oriented database**
  - Stores both data and its processing instructions
  - Method: a procedure or action
  - Message: a request to execute or run a method

Object-Oriented and Object-Relational Database Management Systems (continued)

- **Object-oriented database management system (OODBMS)**
  - Programs that manipulate an object-oriented database and provide a user interface and connections to other application programs
- **Object-relational database management system (ORDBMS)**
  - A DBMS capable of manipulating audio, video, and graphical data

Visual, Audio, and Other Database Systems

- Visual database systems
- Audio database systems
- Virtual database systems
- Spatial data technology

Summary

- Cấu trúc của dữ liệu: bits, characters, fields, records, files, and databases
- Một thực thể là một lớp tổng quát của các đối tượng (objects) mà dữ liệu về chúng được thu thập, lưu trữ và xử lý
- Attribute: các thuộc tính của một thực thể
- Data model: Lược đồ của các thực thể và quan hệ giữa chúng
- Relational model: miêu tả dữ liệu trong một bảng 2 chiều

Summary (continued)

- Selecting eliminates rows according to criteria
- Projecting eliminates columns in a table
- A database management system (DBMS) is a group of programs used as an interface between:
  - The database and application programs
  - The database and the user
- Data dictionary: detailed description of all the data used in the database

Summary (continued)

- Data warehouse: database that collects business information from all aspects of a company's processes, products, and customers
- Data mining: an information-analysis tool for discovering patterns and relationships in a data warehouse
- An object-oriented database stores both data and its processing instructions