

Giai đoạn 3 Thiết kế hệ thống

Chương 8 Thiết kế dữ liệu

Nội dung

- ❖ Giải thích các khái niệm thiết kế dữ liệu và các cấu trúc dữ liệu.
- ❖ Mô tả hệ thống xử lý tập tin và các loại tập tin khác nhau.
- ❖ Mô tả hệ thống CSDL, các đặc điểm và các ưu điểm của nó.
- ❖ Định nghĩa các thành phần của hệ quản trị CSDL (DBMS).
- ❖ Giải thích các khái niệm của kho dữ liệu và khai phá dữ liệu.

Nội dung

- ❖ Hiểu các thuật ngữ thiết kế CSDL, gồm các thực thể, vùng tin, vùng tin chung, mẫu tin, tập tin, bảng và vùng khóa.
- ❖ Giải thích các mối liên kết dữ liệu và vẽ sơ đồ liên kết thực thể.
- ❖ Định nghĩa lượng số, ký hiệu lượng số và ký hiệu chân quạ.
- ❖ Giải thích chuẩn hóa, gồm các ví dụ về dạng chuẩn 1, 2 và 3.

Nội dung

- ❖ Mô tả các mô hình dữ liệu phân cấp, mô hình dữ liệu mạng, mô hình dữ liệu quan hệ và mô hình dữ liệu hướng đối tượng.
- ❖ Phân biệt giữa các mẫu tin luận lý và mẫu tin vật lý, bàn luận các dạng lưu trữ dữ liệu và các vùng tin kiểu ngày tháng.
- ❖ Giải thích các tiêu chuẩn điều khiển dữ liệu.

Các khái niệm thiết kế dữ liệu

❖ Các điều cần biết trước khi xây dựng một HTTT.

- ▶ Dữ liệu được tổ chức như thế nào.
- ▶ Các đặc điểm của hệ thống hướng tập tin và hệ thống hướng CSDL.
- ▶ Vấn đề kho dữ liệu.
- ▶ Vấn đề khai phá dữ liệu.
- ▶ Các thành phần của một hệ quản trị CSDL.

Các khái niệm thiết kế dữ liệu

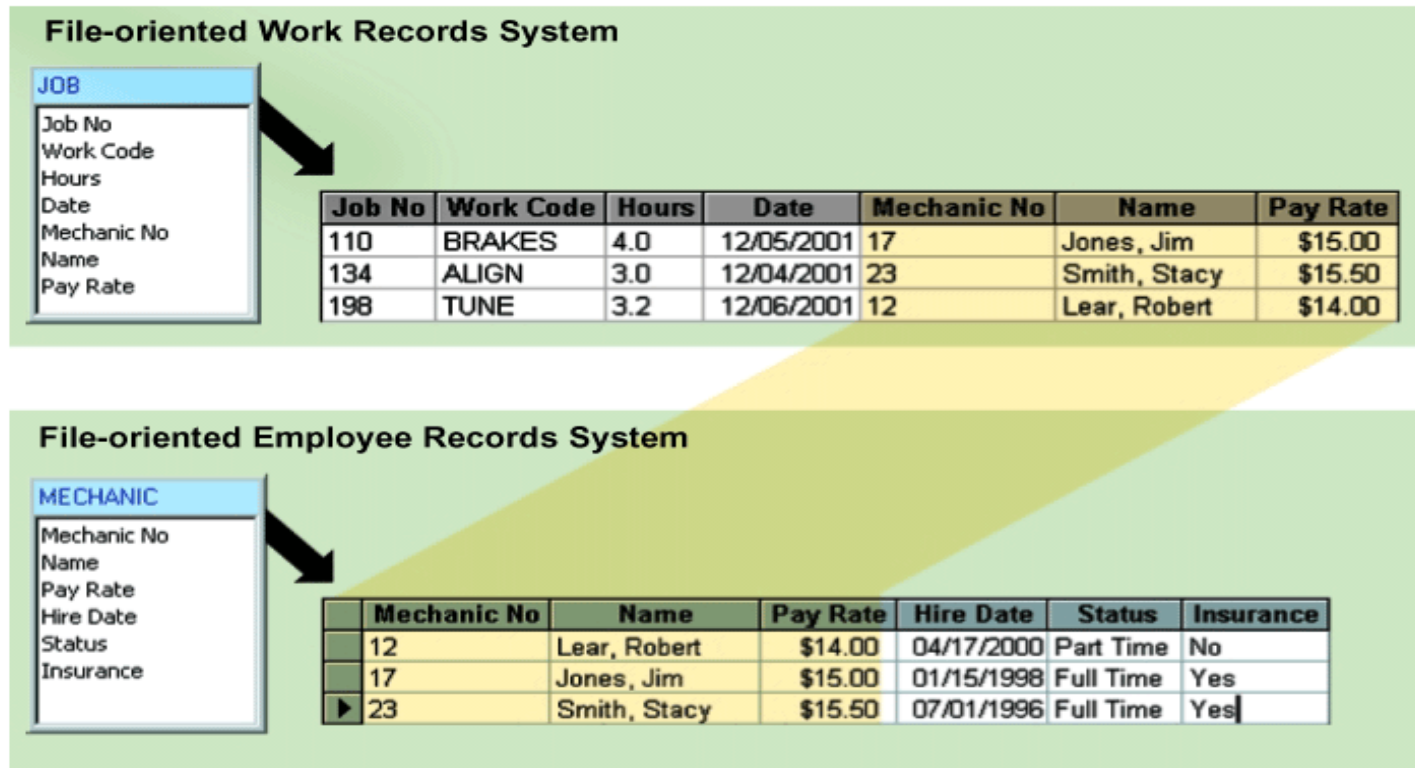
❖ Các cấu trúc dữ liệu

- ▶ Một tập tin chứa dữ liệu về người, vị trí, sự vật hoặc sự kiện tương tác với HTTT.
- ▶ Hệ thống hướng tập tin xử lý các tập tin bằng cách sử dụng việc xử lý tập tin.
- ▶ CSDL bao gồm các tập tin liên kết với nhau, còn được gọi là bảng, tạo thành một cấu trúc dữ liệu tổng thể.

Các khái niệm thiết kế dữ liệu

❖ Hệ thống hướng tập tin

AUTO REPAIR SHOP

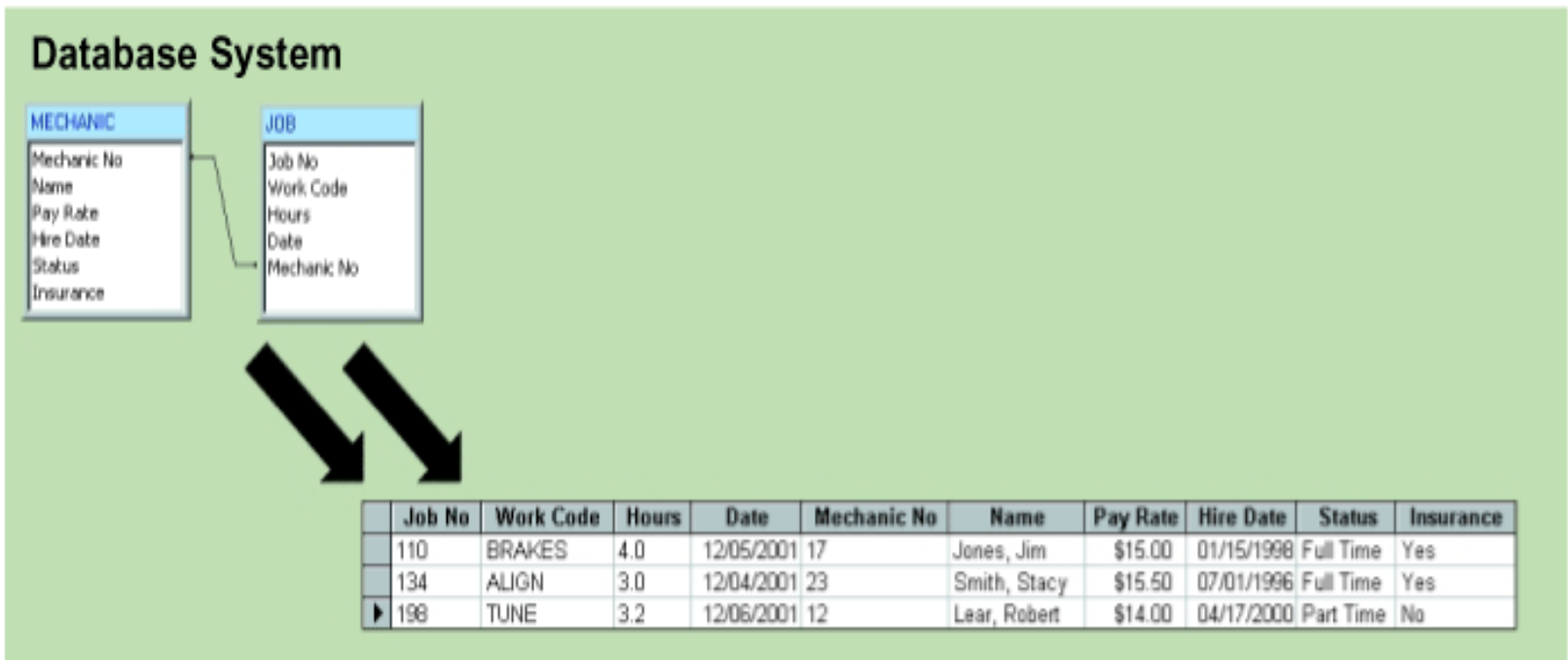


Hình 8.1. Phân xưởng sửa chữa ô tô có tập tin dữ liệu *Job* ghi nhận các công việc và tập tin dữ liệu *Mechanic* ghi nhận các nhân viên. Dữ liệu của một số nhân viên bị trùng lặp trong cả hai tập tin.

Các khái niệm thiết kế dữ liệu

❖ Hệ thống cơ sở dữ liệu

AUTO REPAIR SHOP



Hình 8.2. Thiết kế CSDL của phân xưởng sửa chữa ô tô liên kết hai tập tin dữ liệu với nhau và tránh trùng lặp dữ liệu.



Hệ thống xử lý tập tin

❖ Hệ thống xử lý tập tin

▶ *file processing system*

▶ **Hệ thống xử lý tập tin** là tập hợp các chương trình dùng để lưu trữ, thao tác và truy xuất các tập tin dữ liệu (có kích thước lớn).

- Các tập tin dữ liệu được lưu trữ trong các thư mục (*folder*).

Hệ thống xử lý tập tin

- ❖ **Các thành phần của hệ thống xử lý tập tin**
 - ▶ **Phần cứng:** máy tính.
 - ▶ **Phần mềm:**
 - Hệ điều hành.
 - Các tiện ích.
 - Các tập tin.
 - Các chương trình quản lý tập tin.
 - Các chương trình ứng dụng tạo các báo cáo từ các dữ liệu được lưu trữ trong các tập tin.
 - ▶ **Con người:** người quản lý, chuyên gia, người lập trình, người sử dụng cuối cùng.

Hệ thống xử lý tập tin

- ❖ Các thành phần của hệ thống xử lý tập tin
 - ▶ **Các thủ tục:** các lệnh và các qui tắc chi phối việc thiết kế và sử dụng các thành phần của phần mềm.
 - ▶ **Dữ liệu:** tập hợp các sự kiện.

Hệ thống xử lý tập tin

❖ Quản lý dữ liệu của hệ thống xử lý tập tin

► Ngôn ngữ lập trình:

- Làm gì? Làm như thế nào?
- Các ngôn ngữ lập trình: 3GL (*Third-Generation Language*) - COBOL, FORTRAN, BASIC, ...

► Các chương trình xử lý tập tin: tập hợp các sự kiện.

- Tạo cấu trúc tập tin.
- Thêm dữ liệu vào tập tin.
- Xóa dữ liệu của tập tin.
- Sửa dữ liệu của tập tin.
- Liệt kê dữ liệu của tập tin.

Hệ thống xử lý tập tin

❖ Nhược điểm của hệ thống xử lý tập tin

▶ Phụ thuộc dữ liệu - chương trình

- *program - data dependence*
- Tất cả các chương trình ứng dụng phải duy trì siêu dữ liệu (phần mô tả) của các tập tin mà chúng sử dụng.

▶ Dư thừa dữ liệu / Trùng lặp dữ liệu

- *data redundancy / duplication of data*
- Các hệ thống / chương trình khác nhau có các bản dữ liệu riêng biệt của cùng dữ liệu.

▶ Hạn chế việc dùng chung dữ liệu

- Mỗi ứng dụng có các tập tin riêng biệt, ít sử dụng chung dữ liệu với các hệ thống khác.

Hệ thống xử lý tập tin

❖ Nhược điểm của hệ thống xử lý tập tin

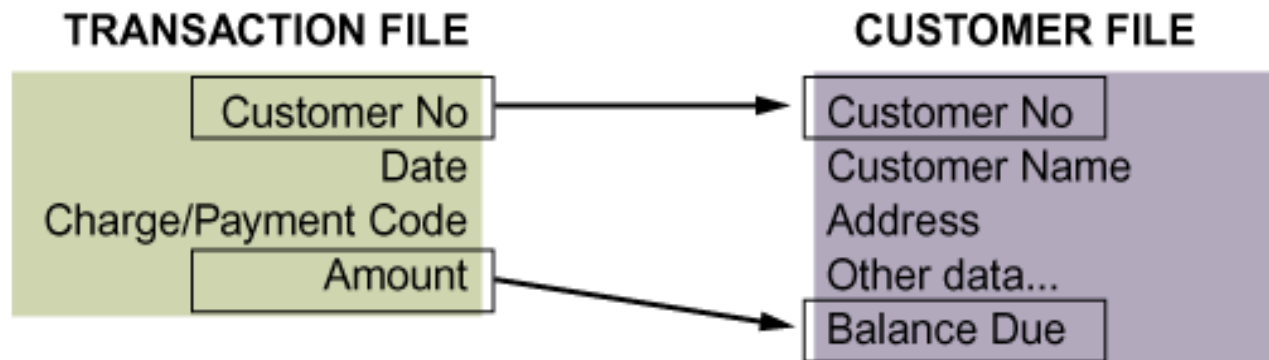
▶ Thời gian phát triển lâu

- Người lập trình phải thiết kế các dạng tập tin dữ liệu riêng và viết cách truy xuất tập tin cho các ứng dụng mới.

▶ Chi phí bảo trì cao

- Các nhược điểm nêu trên làm cho việc bảo trì chương trình gặp nhiều khó khăn, thường chiếm khoảng 80% ngân sách phát triển HTTT.

Hệ thống xử lý tập tin



Hình 8.3. Công ty thẻ tín dụng sử dụng hệ thống xử lý tập tin để thông báo các giao dịch bán hàng ngày từ tập tin Transaction cho tập tin Customer.

Hệ thống xử lý tập tin

- ❖ Các loại tập tin được sử dụng trong hệ thống xử lý tập tin
 - ▶ Tập tin chính
 - ▶ Tập tin bảng
 - ▶ Tập tin giao tác
 - ▶ Tập tin làm việc
 - ▶ Tập tin an toàn
 - ▶ Tập tin lịch sử

Hệ thống xử lý tập tin

❖ Các loại tập tin được sử dụng trong hệ thống xử lý tập tin

► **Tập tin chính (*master file*)**

- Chứa các dữ liệu tương đối lâu dài của một thực thể.
- Có thể chứa dữ liệu của khách hàng, văn phòng đại diện bán hàng, sinh viên, nhân viên hoặc bệnh nhân.

► **Tập tin bảng (*table file*)**

- Chứa các dữ liệu tham chiếu được HTTT sử dụng.
- Tương đối lâu dài.
- Không được HTTT cập nhật.



Hệ thống xử lý tập tin

❖ Các loại tập tin được sử dụng trong hệ thống xử lý tập tin

▶ **Tập tin giao tác (*transaction file*)**

- Chứa các mẫu tin nghiệp vụ hằng ngày và dữ liệu tác nghiệp.
- Tập tin nhập để cập nhật tập tin chính.
- Các tập tin tạm thời.

▶ **Tập tin làm việc (*work file*)**

- Tập tin tạm thời do HTTT tạo ra cho một công việc.
- Còn được gọi là tập tin linh tinh.
- Có thể chứa các bản sao của các mẫu tin của tập tin chính được sử dụng tạm thời.

Hệ thống xử lý tập tin

❖ Các loại tập tin được sử dụng trong hệ thống xử lý tập tin

▶ **Tập tin an toàn (*security file*)**

- Được tạo và lưu trữ để sao lưu và phục hồi.
- Các tập tin an toàn mới phải được tạo thường xuyên để thay thế các tập tin lỗi thời.

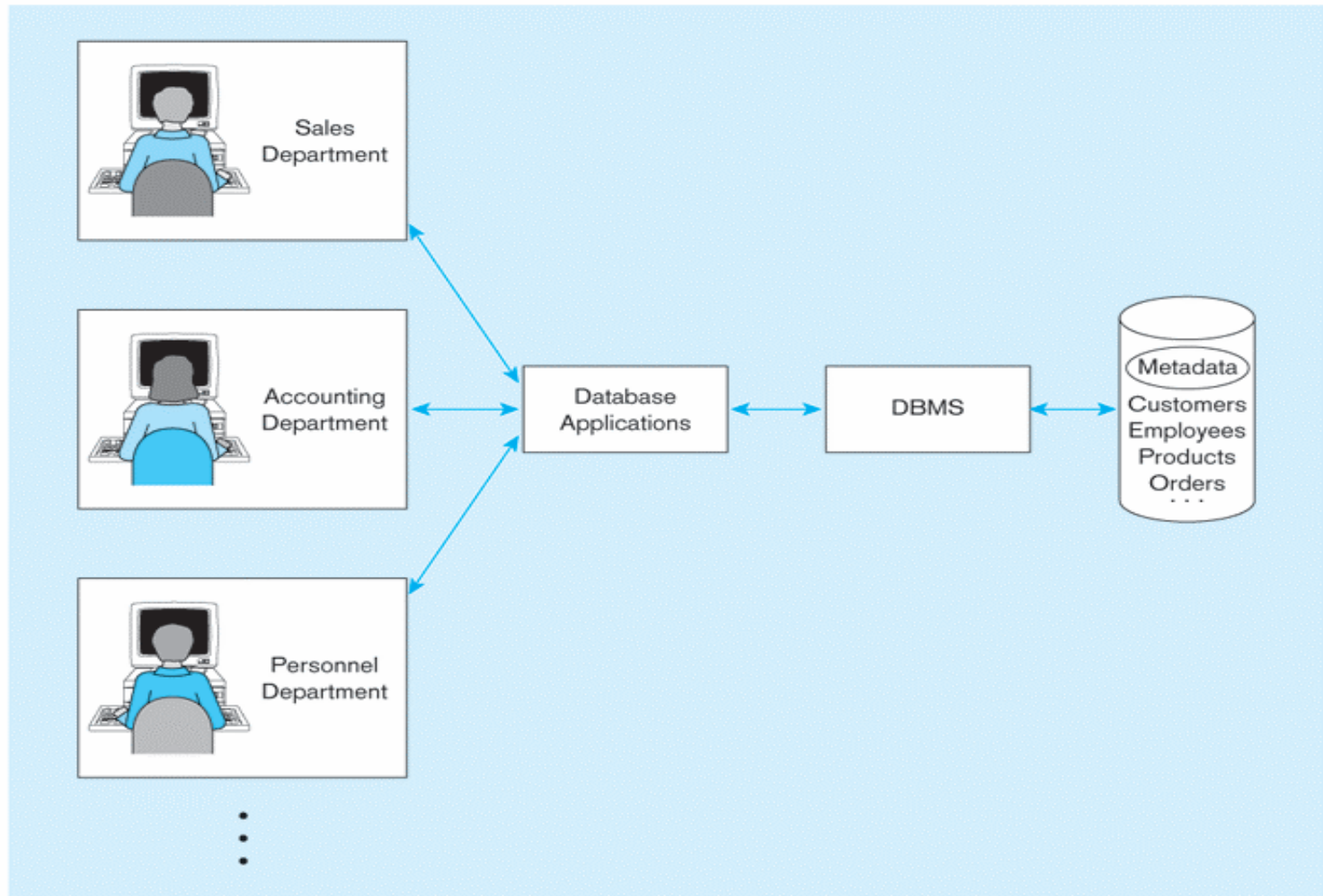
▶ **Tập tin lịch sử (*history file*)**

- Được tạo và lưu trữ để chứa dữ liệu lịch sử.
- Các tập tin lịch sử mới không thay thế các tập tin cũ.

Hệ cơ sở dữ liệu

- ❖ CSDL cung cấp một khung thức tổng quát để tránh dư thừa dữ liệu và hỗ trợ môi trường động thời gian thực.
- ❖ Nhiều hệ thống có thể được xây dựng từ một CSDL.
- ❖ **Các thành phần của hệ CSDL**
 - ▶ Cơ sở dữ liệu.
 - ▶ Hệ quản trị CSDL.
 - ▶ Các chương trình ứng dụng CSDL.

Hệ cơ sở dữ liệu



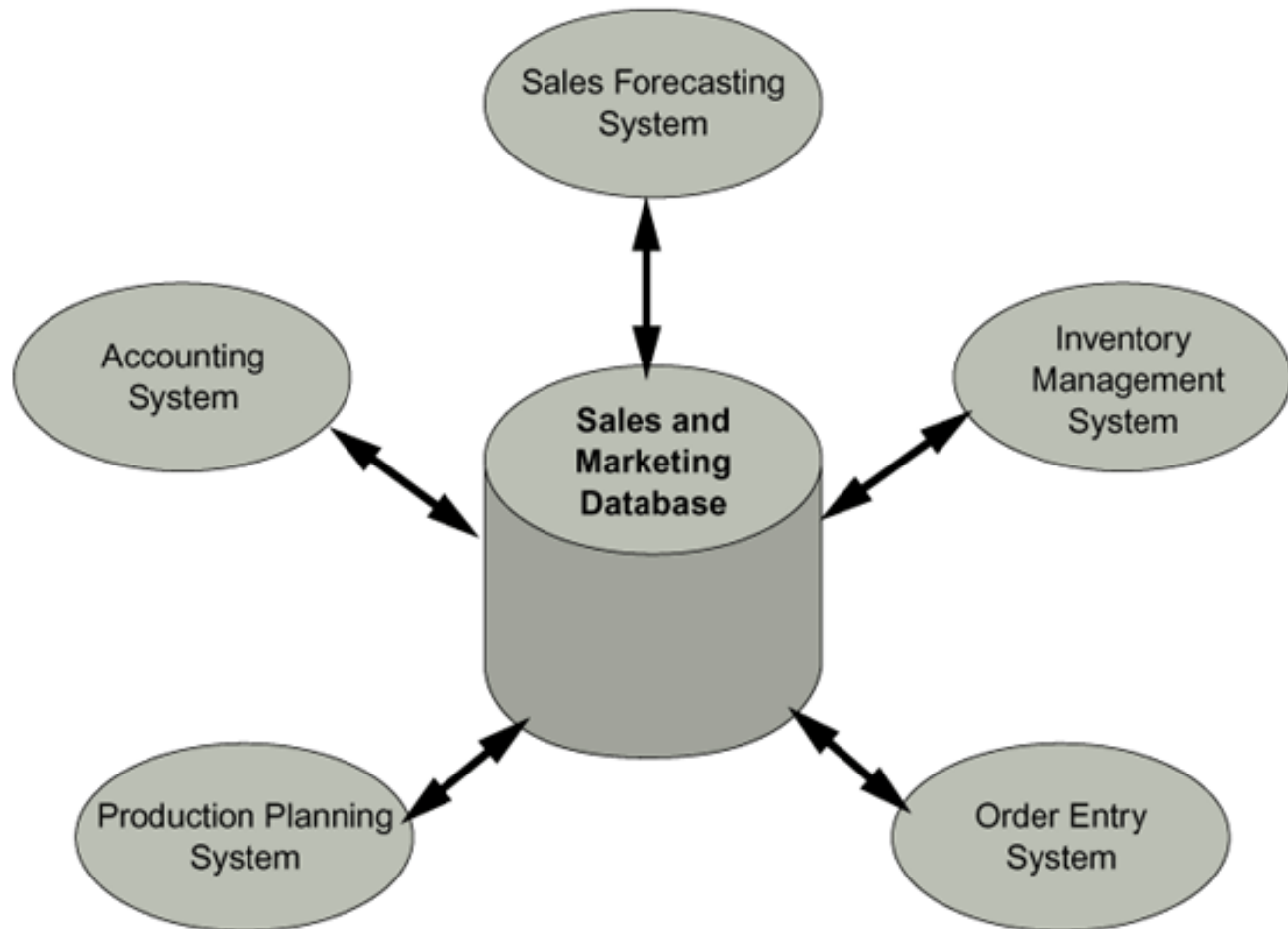
Hình 8.4. Minh họa của một hệ cơ sở dữ liệu.

Hệ cơ sở dữ liệu

❖ Hệ quản trị CSDL

- ▶ DBMS – *DataBase Management System*
- ▶ **Hệ quản trị CSDL** là tập hợp các chương trình dùng để quản lý cấu trúc và dữ liệu của CSDL và điều khiển truy xuất dữ liệu trong CSDL.
- ▶ Cho phép người sử dụng định nghĩa, tạo lập và bảo trì CSDL và cung cấp các truy xuất CSDL.

Hệ cơ sở dữ liệu



Hình 8.5. Môi trường CSDL tiêu biểu gồm một CSDL phục vụ cho năm hệ thống nghiệp vụ riêng biệt.

Hệ cơ sở dữ liệu

❖ Các chức năng của hệ quản trị CSDL

- ▶ Lưu trữ, truy xuất và cập nhật dữ liệu: DDL (*Data Definition Language*), DML (*Data Manipulation Language*).
- ▶ Quản lý giao tác (*transaction management*).
- ▶ Điều khiển tương tranh (*concurrency control*).
- ▶ Chép lưu và phục hồi dữ liệu.
- ▶ Bảo mật dữ liệu: DCL (*Data Control Language*).
- ▶ Hỗ trợ truyền thông dữ liệu.
- ▶ Duy trì tính toàn vẹn / nhất quán dữ liệu.
- ▶ Cung cấp các tiện ích.

Hệ cơ sở dữ liệu

❖ Các thành phần của môi trường hệ quản trị CSDL

▶ Phần cứng

- Máy tính.

▶ Phần mềm

- Hệ điều hành.
- Hệ quản trị CSDL.
- Các chương trình ứng dụng CSDL.

▶ Con người

- Người quản trị dữ liệu và người quản trị CSDL.
- Người thiết kế CSDL.
- Người phát triển ứng dụng.
- Người sử dụng cuối cùng.

Hệ cơ sở dữ liệu

❖ Các thành phần của môi trường hệ quản trị CSDL

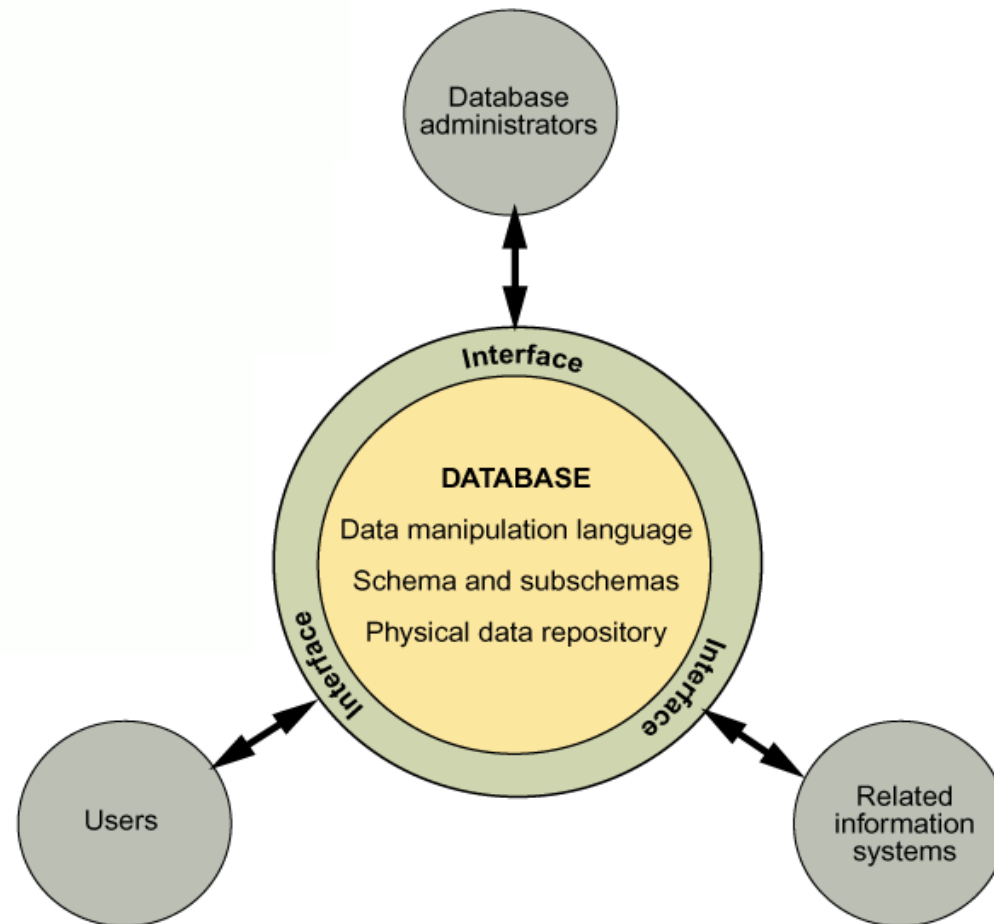
▶ Các thủ tục

- Đăng nhập vào hệ quản trị CSDL.
- Sử dụng các chức năng của hệ quản trị CSDL, các chương trình ứng dụng.
- Khởi động và thoát khỏi hệ quản trị CSDL.
- Chép lưu và phục hồi CSDL.
- Thay đổi cấu trúc và lưu trữ CSDL.
- Nâng cao hiệu suất của chương trình ứng dụng.

▶ Dữ liệu

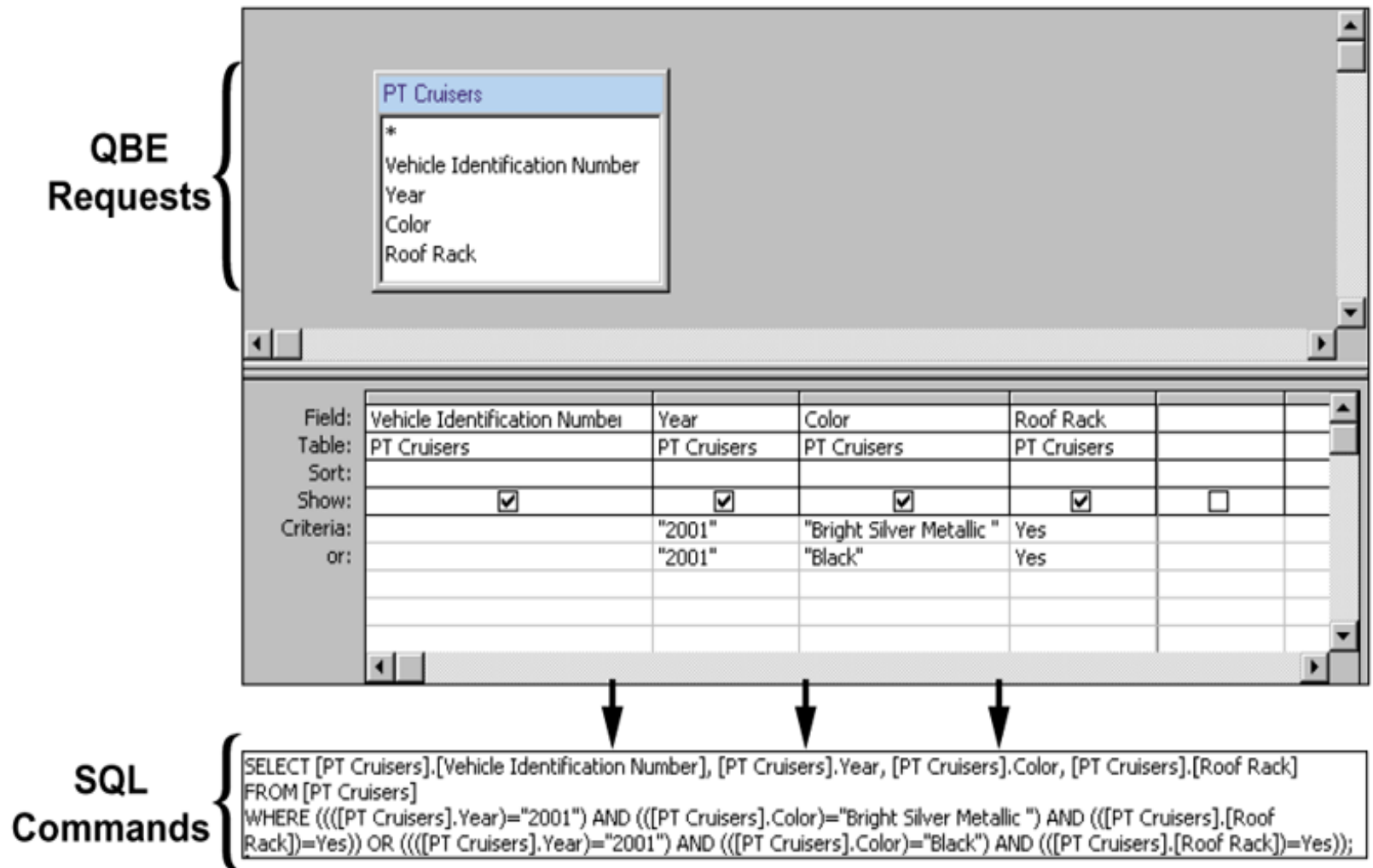
- Siêu dữ liệu.
- Các dữ liệu tác nghiệp.

Hệ cơ sở dữ liệu



Hình 8.6. Hệ quản trị CSDL có ngôn ngữ thao tác dữ liệu để hỗ trợ giao tiếp cho người sử dụng, người quản trị CSDL và các HTTT liên quan.

Hệ cơ sở dữ liệu



Hình 8.7. Yêu cầu QBE tạo ra lệnh SQL.

Hệ cơ sở dữ liệu

Microsoft Access

File Edit View Insert Format Records Tools Window Help

Type a question for help

ORDER_t : Table

Order ID	Order Date	Customer ID
1001	10/21/2004	1
1002	10/21/2004	8
1003	10/22/2004	15
1004	10/22/2004	5
1005	10/24/2004	3
1006	10/24/2004	2
1007	10/27/2004	11
1008	10/30/2004	12
1009	11/5/2004	4
1010	11/5/2004	1
0		0

Record: 1 of 10

Order_line_t : Table

Order_ID	Product_ID	Ordered_Quantity
1001	1	2
1001	2	2
1001	4	1
1002	3	5
1003	3	3
1004	6	2
1004	8	2
1005	4	4
1006	4	1
1006	5	2
1006	7	2
1007	1	3
1007	2	2
1008	3	3
1008	8	3
1009	4	2
1009	7	3
1010	8	10
0	0	0

Record: 1 of 18

Datasheet View

Hình 8.8. Cơ sở dữ liệu quan hệ.

Hệ cơ sở dữ liệu

❖ Ưu điểm của cách tiếp cận CSDL

- ▶ Độc lập dữ liệu - chương trình (*data-program independence*)
 - Hệ quản trị CSDL chứa các siêu dữ liệu, do đó các ứng dụng không cần quan tâm đến các dạng thức của dữ liệu.
 - Hệ quản trị CSDL quản lý các truy vấn và cập nhật dữ liệu, do đó các ứng dụng không cần xử lý việc truy xuất dữ liệu.
- ▶ Giảm tối thiểu sự dư thừa dữ liệu.
- ▶ Nâng cao tính nhất quán / toàn vẹn dữ liệu.

Hệ cơ sở dữ liệu

❖ Ưu điểm của cách tiếp cận CSDL

- ▶ **Nâng cao việc dùng chung dữ liệu.**
 - Những người sử dụng khác nhau có những cái nhìn khác nhau về dữ liệu.
- ▶ **Tăng hiệu suất phát triển ứng dụng.**
- ▶ **Tuân thủ các tiêu chuẩn.**
 - Tất cả các truy xuất đều được thực hiện theo cùng một cách.
- ▶ **Nâng cao chất lượng của dữ liệu.**
 - Các ràng buộc, các qui tắc hợp lệ của dữ liệu.

Hệ cơ sở dữ liệu

❖ Ưu điểm của cách tiếp cận CSDL

- ▶ **Nâng cao tính truy xuất và đáp ứng của dữ liệu.**
 - Sử dụng ngôn ngữ truy vấn dữ liệu chuẩn SQL.
- ▶ **Giảm chi phí bảo trì chương trình.**
- ▶ **Bảo mật dữ liệu.**
- ▶ **Chép lưu và phục hồi dữ liệu.**
- ▶ **Điều khiển tương tranh.**

Hệ cơ sở dữ liệu

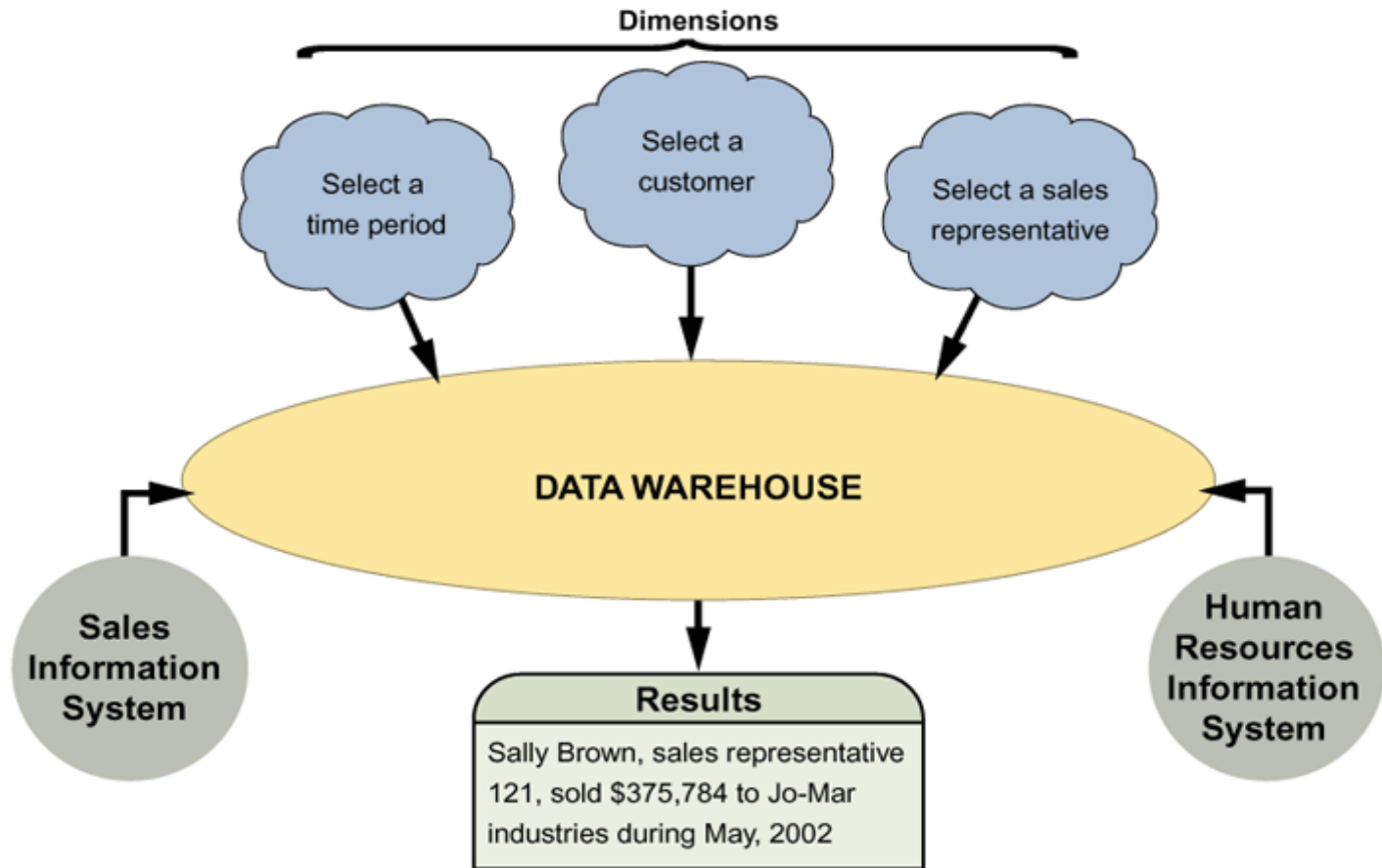
❖ Chi phí và rủi ro của cách tiếp cận CSDL

- ▶ Chi phí ban đầu.
 - Chi phí cài đặt và quản lý.
 - Chi phí chuyển đổi.
- ▶ Chi phí vận hành.
 - Cần nhân viên mới có chuyên môn.
 - Cần phải chép lưu và phục hồi dữ liệu.
- ▶ Mâu thuẫn về mặt tổ chức.
 - Rất khó thay đổi các thói quen cũ.

Kho dữ liệu

- ❖ *data warehousing*
- ❖ **Kho dữ liệu** là sự tập hợp các dữ liệu hướng chủ thể (*subject-oriented*), tích hợp (*integrated*), thay đổi theo thời gian (*time-variant*), bất biến (*nonvolatile*) trong quá trình hỗ trợ ra quyết định.
- ❖ Kho dữ liệu dùng để hỗ trợ phân tích quản lý và hỗ trợ ra quyết định.
- ❖ Chứa dữ liệu của giao tác theo dạng thức cho phép người sử dụng truy xuất, kết hợp và phân tích dữ liệu.

Kho dữ liệu



Hình 8.9. Kho dữ liệu chứa dữ liệu của nhiều HTTT. Bằng cách chọn các giá trị, người sử dụng có thể có thông tin đa chiều từ các dữ liệu được lưu trữ.

Khai phá dữ liệu

- ❖ *data mining*
- ❖ Phần mềm khai phá dữ liệu tìm kiếm các mẫu (*pattern*) có ý nghĩa và các mối liên kết giữa các dữ liệu.
- ❖ Thương mại điện tử sử dụng việc khai phá dữ liệu như là công cụ phân tích hành vi của người sử dụng *Web* và đưa ra các phương hướng.



Các thuật ngữ thiết kế dữ liệu

❖ Thiết kế dữ liệu

- ▶ Các thực thể
- ▶ Các thuộc tính (vùng tin)
- ▶ Các mẫu tin
- ▶ Các tập tin (bảng)

Các thuật ngữ thiết kế dữ liệu

❖ Thực thể

- ▶ *entity*
- ▶ Là người, vị trí, sự vật hoặc sự kiện mà dữ liệu được thu thập và lưu trữ.

❖ Thuộc tính (vùng tin)

- ▶ *attribute*
- ▶ Là đặc tính của một thực thể.

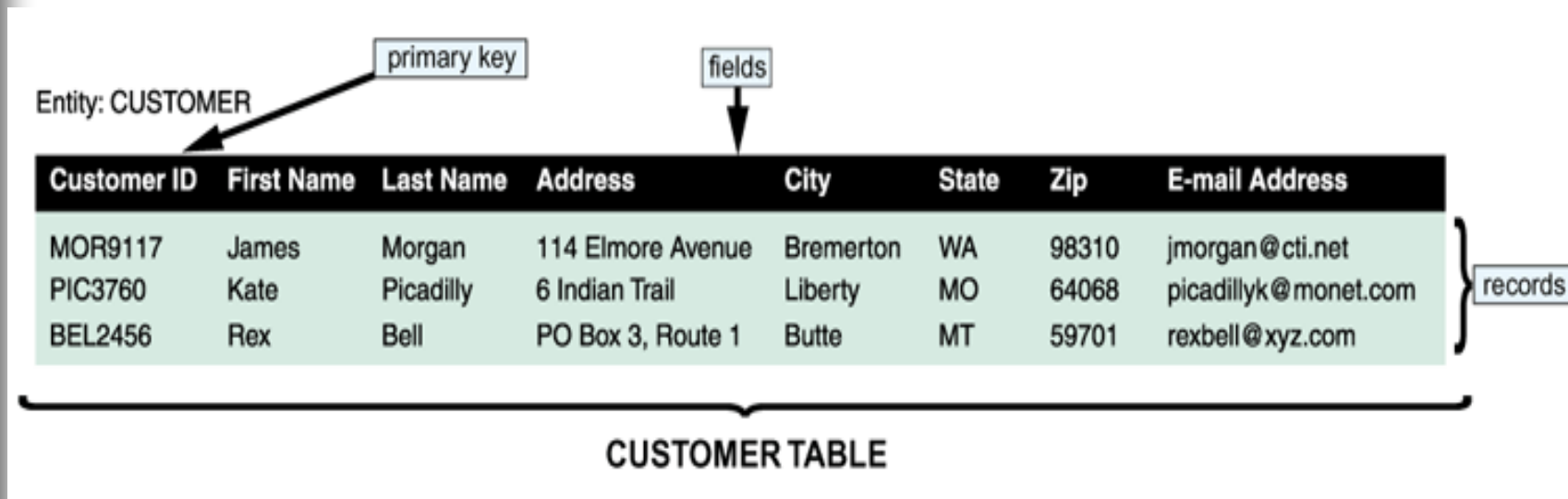
❖ Mẫu tin

- ▶ *record*
- ▶ Là tập hợp các thuộc tính để mô tả một thể hiện của một thực thể.

❖ Tập tin (bảng)

- ▶ Là tập hợp các mẫu tin chứa dữ liệu của thực thể cụ thể.

Các thuật ngữ thiết kế dữ liệu



Hình 8.10. Bảng của thực thể Customer gồm các mẫu tin, các vùng tin và khóa chính.

Các thuật ngữ thiết kế dữ liệu

❖ Khóa

- ▶ **key**
- ▶ Được dùng để tổ chức, truy xuất và duy trì các cấu trúc dữ liệu.
- ▶ Các loại khóa
 - Khóa chính (*primary key*)
 - Khóa dự tuyển (*candidate key*)
 - Khóa ngoại (*foreign key*)
 - Khóa thứ cấp (*secondary key*)

Các thuật ngữ thiết kế dữ liệu

❖ Khóa

▶ Khóa dự tuyển (*candidate key*)

- Một thuộc tính hoặc tổ hợp các thuộc tính *duy nhất* và *tối thiểu* dùng để xác định một thể hiện của một thực thể.
- Khóa bao gồm nhiều thuộc tính được gọi là khóa phức hợp (*composite key*).
- Một thuộc tính là thành phần của khóa được gọi là thuộc tính khóa.
- Một thuộc tính không là thành phần của khóa được gọi là thuộc tính không khóa.

▶ Khóa chính (*primary key*)

- Là khóa dự tuyển được chọn để xác định duy nhất một thể hiện của một thực thể.

Các thuật ngữ thiết kế dữ liệu

❖ Khóa

▶ Khóa ngoại (*foreign key*)

- Vùng tin trong một tập tin trùng tương ứng với khóa chính trong một tập tin khác.
- Ví dụ: mã giảng viên cố vấn là khóa ngoại trong tập tin STUDENT, tương ứng với khóa chính trong tập tin ADVISOR.
- Khóa ngoại không cần phải duy nhất và khác rỗng.
- Khóa chính có chứa khóa ngoại được gọi là *khóa mượn (borrowed key)*.

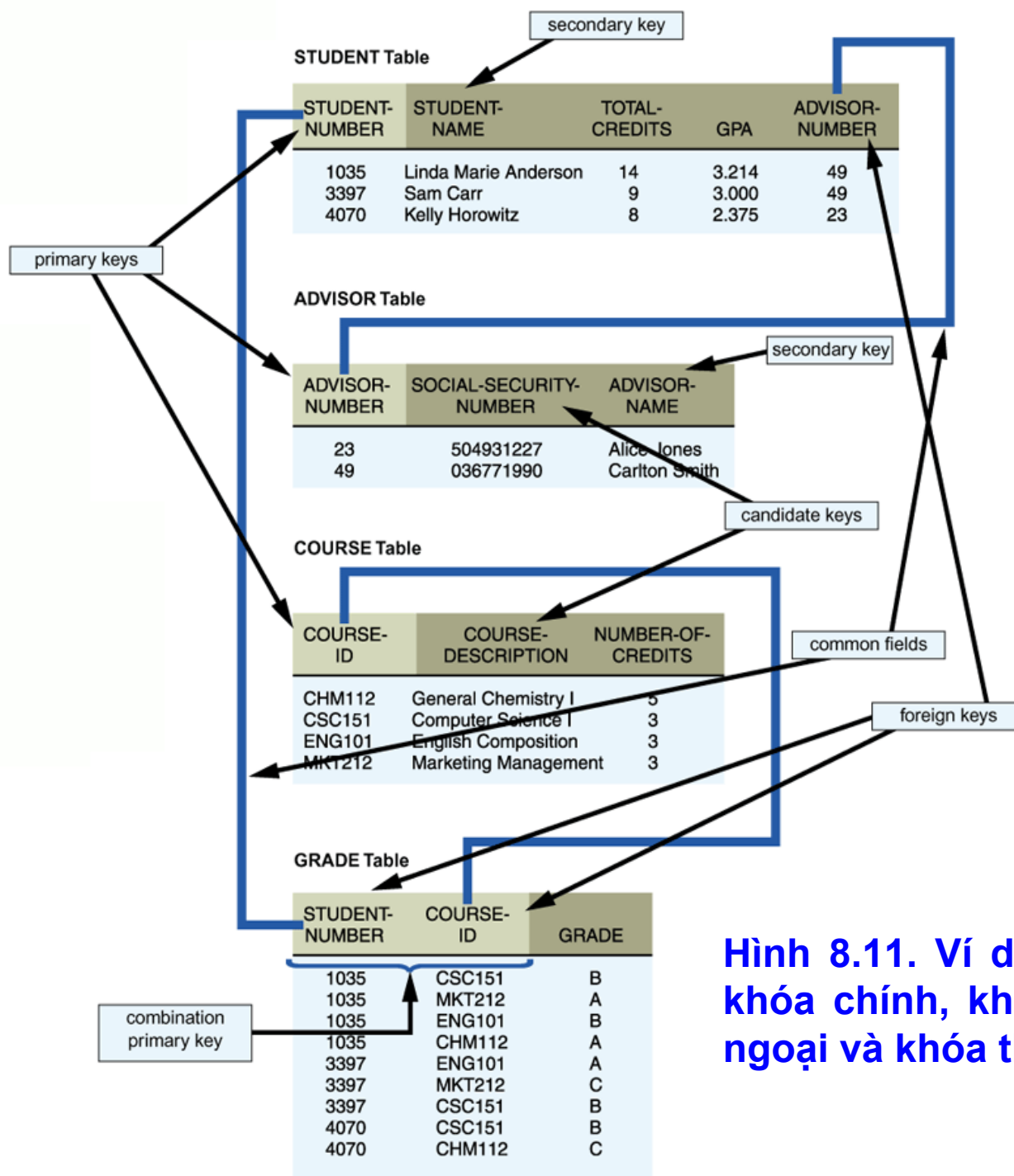


Các thuật ngữ thiết kế dữ liệu

❖ Khóa

▶ Khóa thứ cấp (*secondary key*)

- Một vùng tin hoặc tổ hợp các vùng tin được dùng để truy xuất các mẫu tin.
- Khóa thứ cấp không cần phải duy nhất.



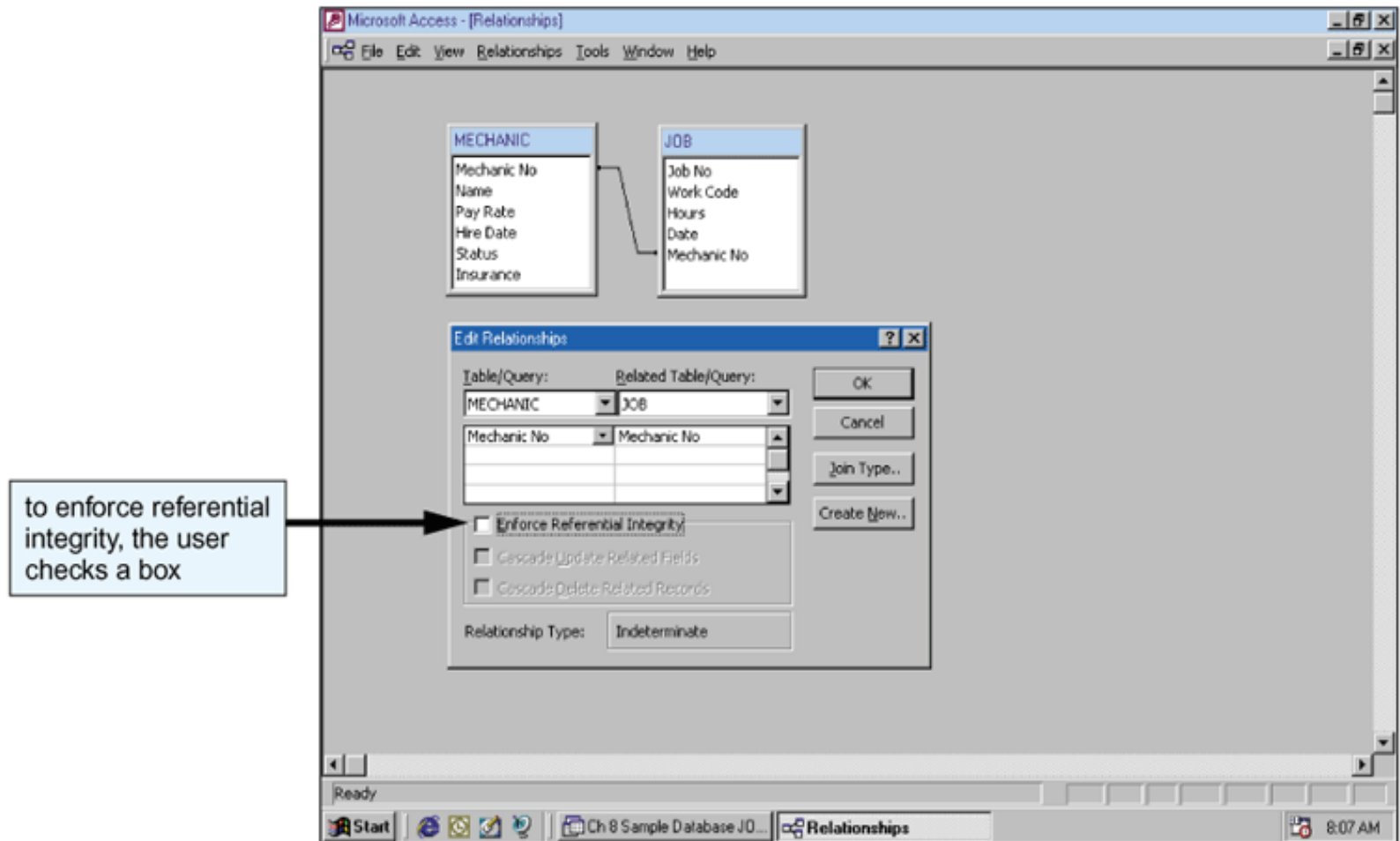
Hình 8.11. Ví dụ về các vùng tin, khóa chính, khóa dự tuyển, khóa ngoại và khóa thứ cấp.

Các thuật ngữ thiết kế dữ liệu

❖ Tính toàn vẹn tham chiếu

- ▶ *referential integrity*
- ▶ Một tập các qui tắc tránh các vấn đề về chất lượng và không nhất quán dữ liệu.
- ▶ Tính toàn vẹn tham chiếu bảo đảm giá trị khóa ngoại không thể được đưa vào trừ khi nó trùng khớp với một giá trị khóa chính trong một tập tin khác.

Các thuật ngữ thiết kế dữ liệu



Hình 8.12. Microsoft Access cho phép người sử dụng đặc tả các qui tắc ràng buộc tham chiếu trong thiết kế CSDL quan hệ.



Các mối liên kết dữ liệu

❖ Mối liên kết

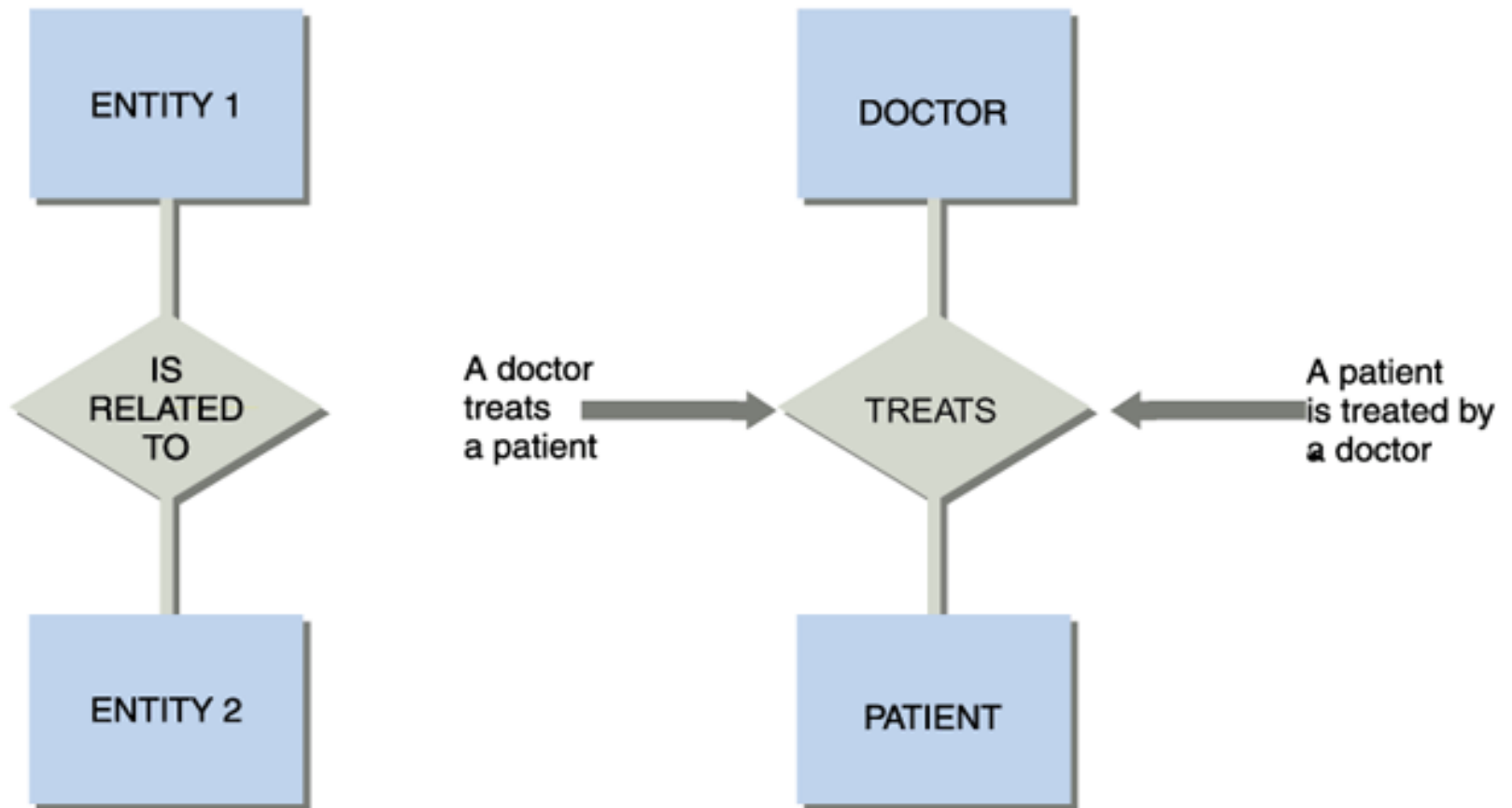
- ▶ *relationship*
- ▶ Là một liên kết luận lý giữa các thực thể dựa vào việc chúng tương tác với nhau như thế nào.

Các mối liên kết dữ liệu

❖ Sơ đồ liên kết thực thể

- ▶ ERD - *Entity-Relationship Diagram*
- ▶ ERD là một sơ đồ đồ họa cho thấy các mối liên kết giữa các thực thể của hệ thống.
- ▶ Mỗi thực thể là một hình chữ nhật, có nhãn là một danh từ.
- ▶ Mỗi mối liên kết là một hình thoi, có nhãn là một động từ.
- ▶ Các loại mối liên kết
 - Một – một (1:1)
 - Một – nhiều (1:M)
 - Nhiều – nhiều (M:N)
- ▶ Một ERD đầy đủ cho thấy tất cả các mối liên kết trong hệ thống.

Các mối liên kết dữ liệu



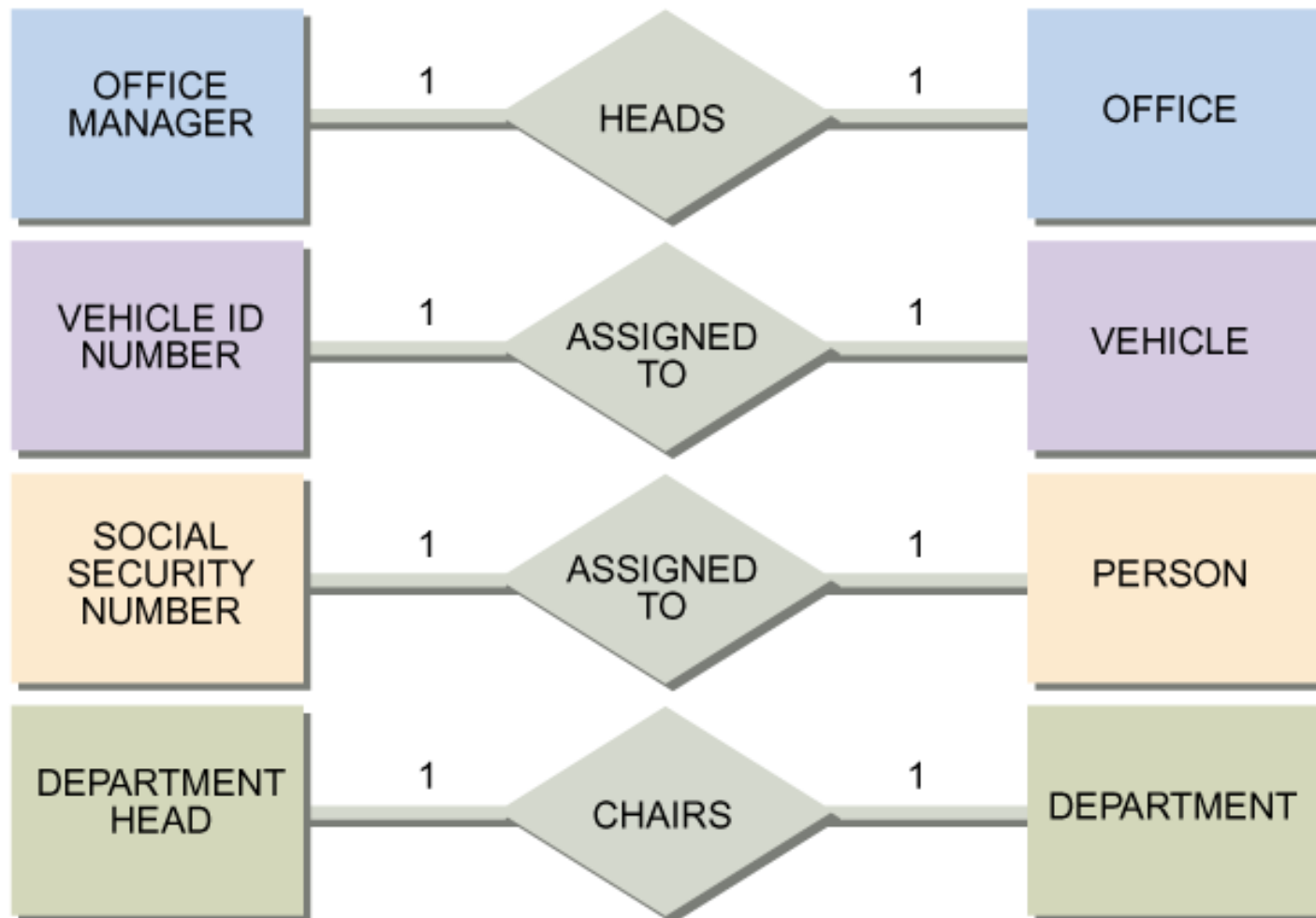
Hình 8.13. Trong sơ đồ liên kết thực thể, các thực thể có tên là một danh từ số ít, các mối liên kết có tên là một động từ.

Các mối liên kết dữ liệu

❖ Mối liên kết một – một (1:1)

- ▶ Tồn tại khi một thể hiện của thực thể *A* chỉ liên kết với một thể hiện của thực thể *B* và ngược lại.
- ▶ Ví dụ
 - Một người quản lý lãnh đạo một cơ quan.
 - Một biển số xe gán cho một xe.
 - Một tài xế lái một xe.
 - Một giảng viên của khoa là chủ nhiệm của một bộ môn.

Các mối liên kết dữ liệu



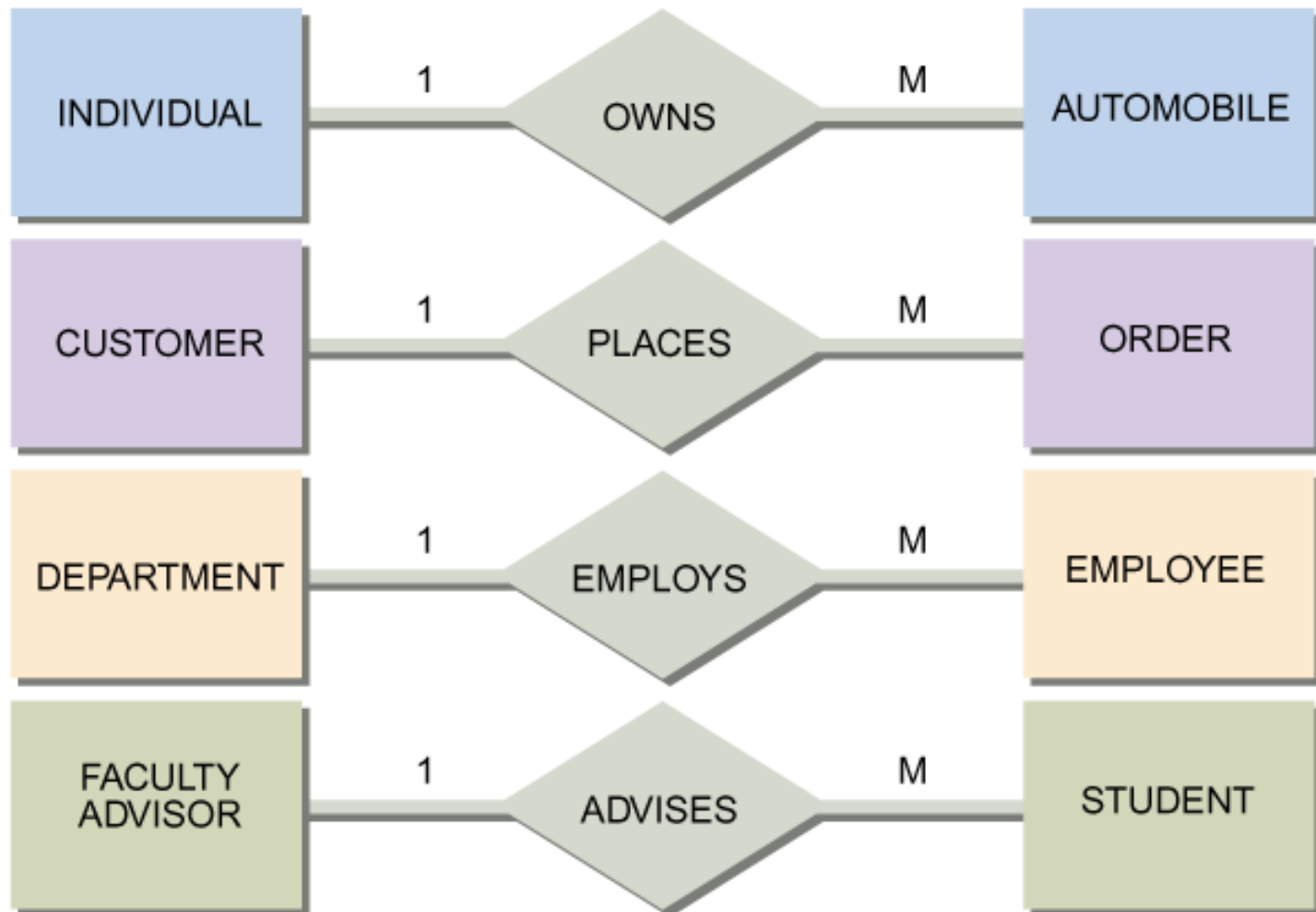
Hình 8.14. Các mối liên kết một-một

Các mối liên kết dữ liệu

❖ Mối liên kết một – nhiều (1:M)

- ▶ Tồn tại khi một thể hiện của thực thể *A* liên kết với nhiều thể hiện của thực thể *B*, nhưng một thể hiện của thực thể *B* chỉ liên kết với một thể hiện của thực thể *A*.
- ▶ Ví dụ
 - Một người sở hữu nhiều xe.
 - Một khách hàng có nhiều đơn đặt hàng.
 - Một phòng có nhiều nhân viên.
 - Một giảng viên hướng dẫn nhiều sinh viên.

Các mối liên kết dữ liệu



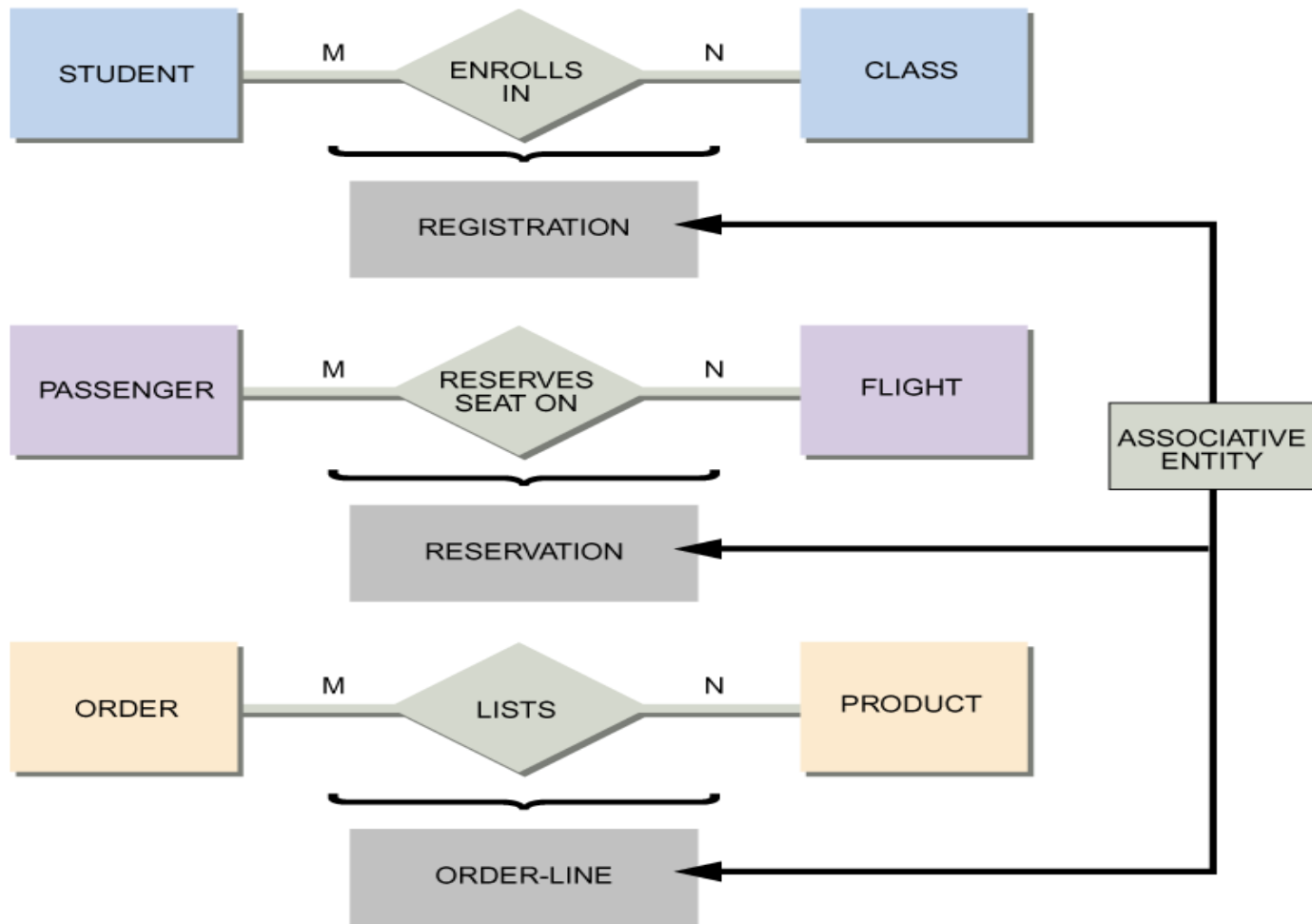
Hình 8.15. Các mối liên kết một-nhiều.

Các mối liên kết dữ liệu

❖ Mối liên kết nhiều – nhiều (M:N)

- ▶ Tồn tại khi một thể hiện của thực thể *A* liên kết với nhiều thể hiện của thực thể *B* và ngược lại.
- ▶ Ví dụ
 - Một sinh viên đăng ký học một hoặc nhiều lớp, mỗi lớp có một hoặc nhiều sinh viên đăng ký.
 - Một hành khách đi một hoặc nhiều chuyến bay, một chuyến bay có nhiều hành khách đi.
 - Một đơn đặt hàng có nhiều sản phẩm, một sản phẩm có trong nhiều đơn đặt hàng.

Các mối liên kết dữ liệu



Hình 8.16. Các mối liên kết nhiều-nhiều. Sự kiện hoặc giao dịch liên kết hai thực thể là một thực thể kết hợp có các thuộc tính riêng.

Các mối liên kết dữ liệu

❖ Một ERD đầy đủ cho thấy tất cả các mối liên kết trong hệ thống

▶ Ví dụ

- Một văn phòng đại diện bán hàng phục vụ một hoặc nhiều khách hàng, mỗi khách hàng chỉ có một văn phòng đại diện bán hàng.
- Một khách hàng có một hoặc nhiều đơn đặt hàng, mỗi đơn đặt hàng chỉ của một khách hàng
- Một đơn đặt hàng có một hoặc nhiều sản phẩm, mỗi sản phẩm có thể có trong một hoặc nhiều đơn đặt hàng.
- Một kho chứa một hoặc nhiều sản phẩm, mỗi sản phẩm có thể được chứa trong một hoặc nhiều kho.

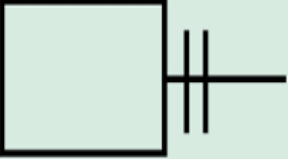
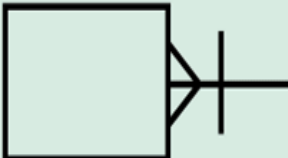




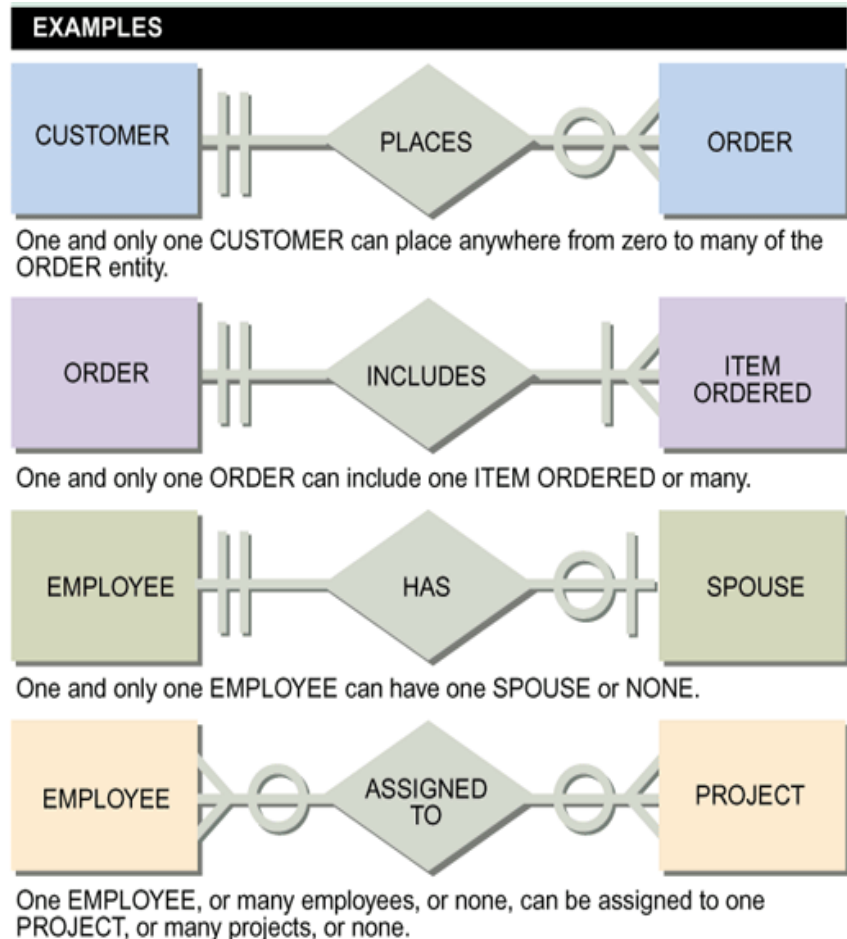
Các mối liên kết dữ liệu

❖ Lượng số

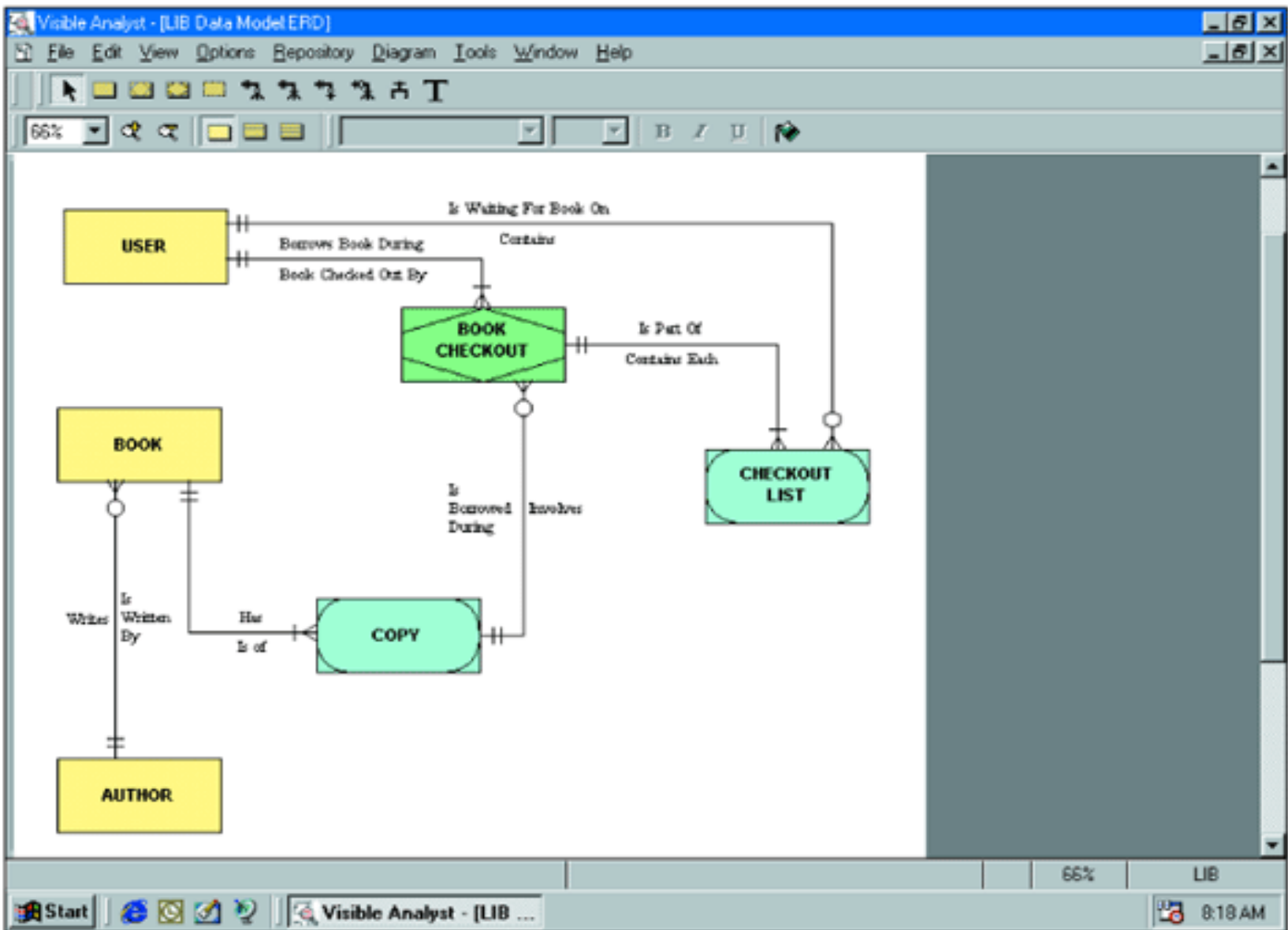
- ▶ *cardinality*
- ▶ Mô tả các thể hiện của một thực thể liên kết như thế nào với các thực thể khác.
- ▶ Mối liên kết bắt buộc và mối liên kết tùy chọn
- ▶ Ký hiệu chân quạ là một cách thức biểu diễn lượng số.
- ▶ Hầu hết các công cụ CASE hỗ trợ việc vẽ ERD.

Các mối liên kết dữ liệu

SYMBOL	MEANING	UML REPRESENTATION
	One and only one	1
	One or many	1..*
	Zero, or one, or many	0..*
	Zero, or one	0..1



Hình 8.18. Ký hiệu chân quạ là phương pháp chung biểu diễn lượng số. Các ví dụ cho thấy sử dụng các ký hiệu khác nhau để mô tả mối liên kết giữa hai thực thể.



Hình 8.19. Sơ đồ liên kết thực thể của Hệ thống thư viện được vẽ trong *Visible Analyst*. Các mối liên kết được mô tả theo cả hai chiều.

Các mối liên kết dữ liệu

❖ Tạo một ERD

1. Xác định các thực thể
2. Xác định tất cả các sự kiện và các hoạt động của các thực thể.
3. Phân tích bản chất của sự tương tác.
4. Vẽ ERD.

Chuẩn hóa dữ liệu

- ❖ **Chuẩn hoá dữ liệu** (*data normalization*) là một quá trình thuận nghịch từng bước để thay thế tập hợp các quan hệ cho trước thành các quan hệ có cấu trúc đơn giản hơn và chuẩn hơn.
- ❖ Chuẩn hóa dữ liệu nhằm để cải tiến một thiết kế CSDL luận lý thỏa mãn các ràng buộc toàn vẹn và tránh dữ liệu bị lặp lại không cần thiết.

Chuẩn hóa dữ liệu

- ❖ Mục đích của chuẩn hoá dữ liệu là loại bỏ các bất thường của một quan hệ để có được các quan hệ có cấu trúc tốt hơn, nhỏ hơn.
 - ▶ Quan hệ có cấu trúc tốt (*well-structured relation*).
 - Có sự dư thừa dữ liệu là tối thiểu.
 - Cho phép người sử dụng thêm, sửa, xóa dữ liệu mà không gây ra sự mâu thuẫn dữ liệu.



Chuẩn hóa dữ liệu

❖ Các dạng chuẩn

- ▶ Dạng chuẩn 1 (1NF - *first normal form*)
- ▶ Dạng chuẩn 2 (2NF - *second normal form*)
- ▶ Dạng chuẩn 3 (3NF - *third normal form*)
- ▶ Dạng chuẩn BCNF (*Boyce-Codd normal form*)

Các mô hình dữ liệu

❖ CSDL phân cấp và CSDL mạng

▶ CSDL phân cấp (*hierarchical database*)

- Dữ liệu được tổ chức giống như cây gia phả hoặc sơ đồ tổ chức.
- Mẫu tin cha có thể có nhiều mẫu tin con – các mẫu tin con chỉ có một cha.
- Con trỏ liên kết mỗi mẫu tin cha với mỗi mẫu tin con.
- Việc duy trì phức tạp và khó khăn.

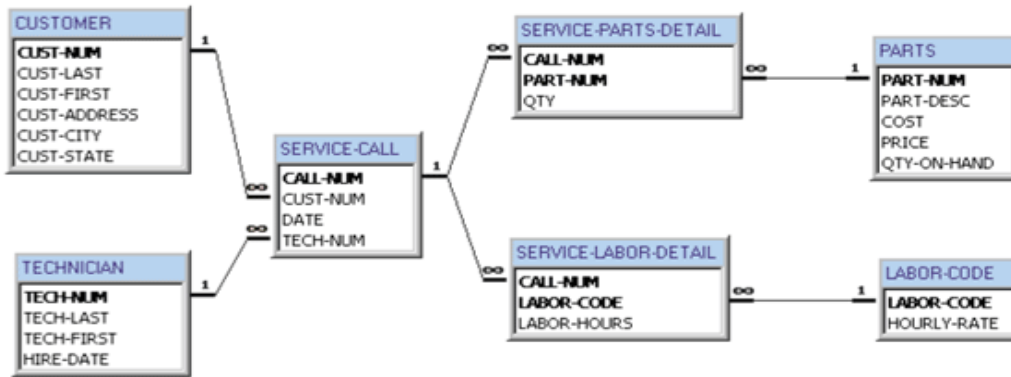
▶ CSDL mạng (*network database*)

- Phức tạp hơn CSDL phân cấp.
- Một mẫu tin con có thể liên kết với nhiều mẫu tin cha.
- Phức tạp cùng với nhiều nhược điểm giống như mô hình phân cấp.

Các mô hình dữ liệu

❖ CSDL quan hệ

- ▶ *relational database*
- ▶ Dữ liệu được tổ chức thành các quan hệ hoặc các bảng hai chiều.
- ▶ Mỗi hàng là một bộ (mẫu tin) và mỗi cột là một thuộc tính (vùng tin).
- ▶ Một vùng tin chung xuất hiện trong nhiều bảng và được sử dụng để tạo các mối liên kết.
- ▶ Mô hình dữ liệu quan hệ là cách tiếp cận thiết kế chiếm ưu thế.



Hình 8.20. CSDL quan hệ sử dụng các vùng tin chung để tạo thành cấu trúc dữ liệu tổng thể.

CUSTOMER

CUST-NUM	CUST-LAST	CUST-FIRST	CUST-ADDRESS	CUST-CITY	CUST-STATE
1456	Jones	Albert	43 Oak Drive	Denver	CO
1933	Valdez	Trini	1011 Hilltop Road	Charlotte	NC
2836	Gonzalez	Juan	678 8th Street	Seattle	WA
6321	Ellis	Mary	101 Main Street	San Jacinto	CA
8760	Sassone	Jackie	9 Blueberry Court	Boston	MA
8848	McMorley	John	321 Smith Lane	Hudson	FL

Primary Key
Common Field

SERVICE-CALL

CALL-NUM	CUST-NUM	DATE	TECH-NUM
10796	1456	12/10/2001	04
10797	2836	12/10/2001	17
10798	8848	12/12/2002	21
10799	1456	12/17/2001	04
10800	6321	12/17/2001	21
10801	1456	12/19/2001	09

TECH-NUM	TECH-LAST	TECH-FIRST	HIRE-DATE
04	Robison	Gerry	6/1/1995
09	Whitfield	David	8/1/1997
17	Kaplan	Carl	1/1/1990
20	Chou	Leland	12/1/2000
21	Johnson	Marie	2/1/2001

SERVICE PARTS-DETAIL

CALL-NUM	PART-NUM	QTY
10796	DD-9192	4
10796	SS-1111	2
10797	AB-6784	6
10799	AB-6784	1
10799	YY-9651	15
10800	SS-1024	8
10801	XC-3133	12

PART-NUM	PART-DESC	COST	PRICE	QTY-ON-HAND
AB-6784	METER	\$98.50	\$125.00	33
DD-9192	DISC	\$1.30	\$1.69	220
SS-1024	COVER	\$16.00	\$19.79	78
SS-1099	EMBLEM	\$38.00	\$50.00	41
SS-1111	JUMPER	\$12.50	\$15.00	3
XC-3133	CABLE	\$6.75	\$7.99	77
YY-9651	PIN	\$0.99	\$1.25	1056

SERVICE LABOR-DETAIL

CALL-NUM	LABOR-CODE	LABOR-HOURS
10796	A	1.5
10796	C	2.0
10798	B	5.0
10799	C	1.5
10799	D	1.5
10801	A	2.5

LABOR-CODE

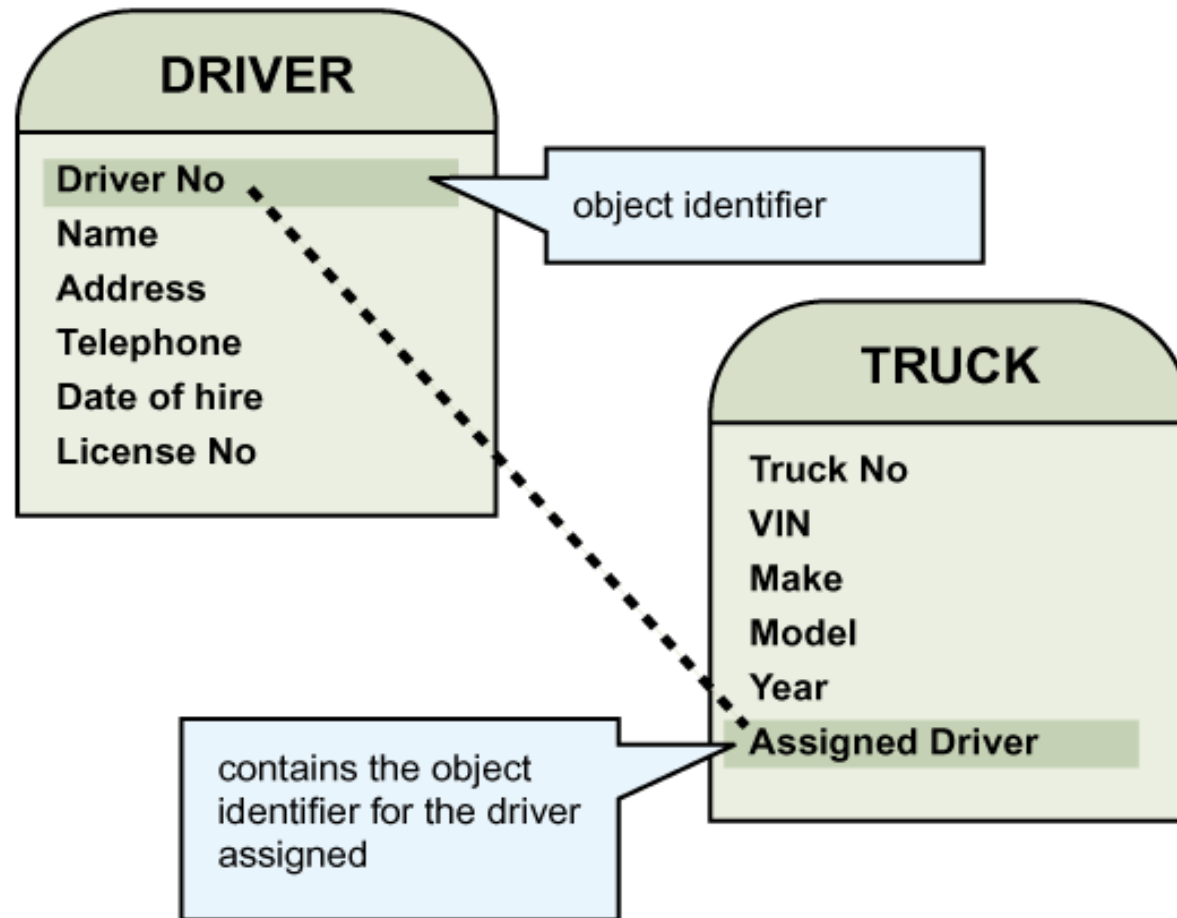
LABOR-CODE	HOURLY-RATE
A	\$35.00
B	\$45.00
C	\$55.00
D	\$65.00

Các mô hình dữ liệu

❖ CSDL hướng đối tượng

- ▶ *object-oriented database*
- ▶ Là sự tiếp nối của quá trình phân tích hướng đối tượng.
- ▶ Ngôn ngữ định nghĩa đối tượng (ODL) là tập chuẩn thiết kế được đề ra bởi *Object Database Management Group* (ODMG).
- ▶ Mỗi đối tượng có một thuộc tính xác định (*object identifier*) duy nhất.
- ▶ Thuộc tính xác định cho phép đối tượng tương tác với các đối tượng khác và tạo thành các mối quan hệ.

Các mô hình dữ liệu



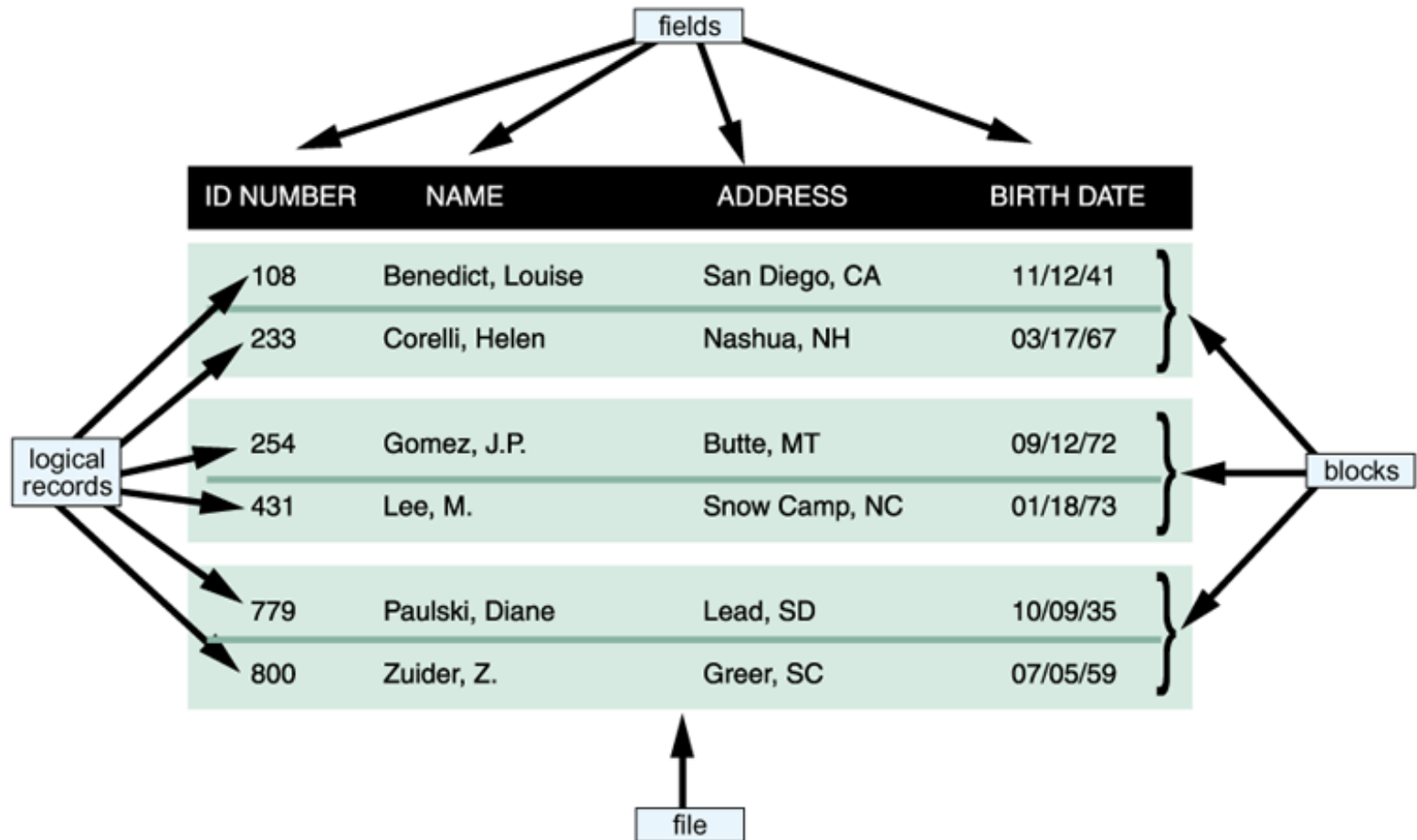
Hình 8.21. Các đối tượng có thuộc tính xác định cho phép chúng liên lạc và tương tác với nhau.

Lưu trữ dữ liệu

❖ Mẫu tin luận lý và mẫu tin vật lý

- ▶ **Bit**
- ▶ **Byte (ký tự)**
- ▶ **Vùng tin (phần tử dữ liệu / mục dữ liệu)**
- ▶ **Mẫu tin luận lý (*logical record*)**
 - Chứa các giá trị của vùng tin để mô tả người, vị trí, sự vật, sự kiện, ...
- ▶ **Mẫu tin vật lý (khối) (*physical record, block*)**
 - Đơn vị nhỏ nhất của dữ liệu có thể được truy xuất .
 - Chứa một hoặc nhiều mẫu tin luận lý, dựa vào hệ số khối.

Lưu trữ dữ liệu



Hình 8.22. Tập tin dữ liệu bao gồm các khối vật lý chứa một hoặc nhiều mẫu tin luận lý. Mỗi mẫu tin luận lý là một tập các giá trị của các vùng tin có liên quan với nhau.



Lưu trữ dữ liệu

❖ Các dạng lưu trữ dữ liệu

- ▶ EBCDIC
- ▶ ASCII
- ▶ *Unicode*
- ▶ Nhị phân

Lưu trữ dữ liệu

❖ Các vùng tin dữ liệu

- ▶ *International Organization for Standardization* (ISO) qui định dạng 4 ký số cho năm, 2 ký số cho tháng và 2 ký số cho ngày (YYYYMMDD).
- ▶ Ngày *Julian* là số có 5 ký số với 2 ký số đầu tiên biểu diễn năm và 3 ký số cuối cùng biểu diễn số thứ tự của ngày trong năm.
- ▶ Ngày tuyệt đối là tổng số ngày kể từ một ngày cơ bản cụ thể nào đó.
- ▶ Ngày *Julian* và ngày tuyệt đối được tính toán dễ dàng.

Điều khiển dữ liệu

- ❖ Điều khiển phải bao gồm các tiêu chuẩn để bảo đảm dữ liệu là đúng, đầy đủ và an toàn.
 - ▶ Tên người sử dụng và mật mã.
 - ▶ Mã hóa (*encryption*).
 - ▶ Các thủ tục chép lưu và phục hồi.
 - ▶ Các tập tin nhật ký kiểm tra (*audit log file*).
 - ▶ Các vùng tin kiểm tra (*audit field*).