

Mục Lục

Mục Lục	1
GIÁO TRÌNH: CƠ SỞ DỮ LIỆU	4
LỜI GIỚI THIỆU	5
Chương 1: TỔNG QUAN VỀ CƠ SỞ DỮ LIỆU	6
1.1. Tổng quan về Cơ sở dữ liệu	6
1.2. Dữ liệu thông tin	6
1.3. Phương pháp cơ sở dữ liệu.....	7
1.3.1. Cơ sở dữ liệu là gì?	7
1.3.2. Hệ quản trị cơ sở dữ liệu	7
1.3.2.1. Định nghĩa hệ quản trị cơ sở dữ liệu.....	7
1.3.2.2. Các chức năng của một hệ quản trị cơ sở dữ liệu	8
1.3.2.3. Ưu điểm của phương pháp CSDL	10
1.3.2.4. Phân loại cơ sở dữ liệu.....	11
1.4. Phân loại người dùng Cơ sở dữ liệu	12
1.4.1. Người quản trị hệ cơ sở dữ liệu (Database Administrator – DBA).....	13
1.4.2. Người thiết kế cơ sở dữ liệu (Database Designer)	13
1.4.3. Những người sử dụng (End User).....	13
1.4.4. Người phân tích hệ thống và lập trình ứng dụng	13
1.4.5. Người thiết kế và cài đặt hệ quản trị dữ liệu.....	13
1.4.6. Những người phát triển công cụ.....	14
1.4.7. Các thao tác viên và những người bảo trì	14
1.5. Câu hỏi ôn tập.....	14
Chương 2: CÁC MÔ HÌNH DỮ LIỆU.....	15
2.1. Sơ đồ thực thể liên kết	15
2.1.1. Thực thể	15
2.1.2. Sơ đồ thực thể liên kết.....	16
2.1.2.1. Kiểu thực thể, tập thực thể, khóa và tập giá trị.....	16
2.1.2.2. Kiểu liên kết, tập liên kết và các thể hiện.....	17
2.1.2.3. Cấp liên kết, tên vai trò và kiểu liên kết đệ quy	18
2.1.3. Quan hệ	19
2.1.3.1. Các ràng buộc trên các kiểu liên kết.....	19
2.1.3.2. Các kiểu thực thể yếu	20
2.2. Các mô hình dữ liệu	21
2.2.1. Mô hình thực thể quan hệ	21
2.2.1.1. Lớp cha, lớp con và sự thừa kế.....	21

2.2.1.3. Sơ đồ mô hình EER	24
2.3. Câu hỏi ôn tập.....	25
Chương 3: MÔ HÌNH CƠ SỞ DỮ LIỆU QUAN HỆ	26
3.1. Các khái niệm	26
3.1.1. Miền.....	26
3.1.2. Quan hệ	27
3.2. Cơ sở dữ liệu quan hệ.....	27
3.3. Câu hỏi ôn tập.....	30
Chương 4: NGÔN NGỮ THAO TÁC DỮ LIỆU	31
4.1. Đại số quan hệ.....	31
4.1.1. Phép tính tập hợp thông thường.....	31
4.1.1.1. Phép nối (JOIN)	33
4.1.2. Các phép tính đặc biệt.....	36
4.1.2.1. Phép chọn (SELECT).....	36
4.1.2.2. Phép chiếu (PROJECT).....	37
4.1.2.3. Phép đặt lại tên (RENAME)	38
4.2. Các ngôn ngữ cập nhật dữ liệu	39
4.2.1. Phép sửa đổi.....	39
4.2.2. Phép loại bỏ.....	40
4.2.3. Phép thêm	40
4.3. Câu hỏi ôn tập.....	41
Chương 5: NGÔN NGỮ TRUY VẤN DỮ LIỆU SQL	43
5.1. Tổng quan, các qui ước biểu diễn câu lệnh SQL	43
5.1.1. Tổng quan	43
5.1.1.1. Lịch sử phát triển	43
5.1.2. Các qui ước biểu diễn câu lệnh SQL	43
5.1.3. Hiệu chỉnh cấu trúc dữ liệu.....	43
5.1.4. Xoá bảng	44
5.1.4.1. Tạo cơ sở dữ liệu.....	44
5.1.4.2. Xóa cơ sở dữ liệu.....	44
5.1.4.3. Tạo bảng dữ liệu	44
5.1.5. Xóa bỏ bảng dữ liệu.....	45
5.1.6. Sửa đổi cấu trúc bảng dữ liệu	45
5.2. Ngôn ngữ cập nhật dữ liệu	45
5.2.1. Thêm bản ghi	45
5.2.2. Cập nhật bản ghi	45

5.2.3. Xoá bản ghi	45
5.3. Ngôn ngữ truy vấn dữ liệu	45
5.3.1. Truy vấn đơn giản	46
5.3.2. Truy vấn có điều kiện	46
5.3.2.1. Lọc các bản ghi trở về với mệnh đề WHERE	46
5.3.2.2. Truy vấn gộp nhóm với mệnh đề GROUP BY VÀ HAVING	46
5.3.2.3. Sắp xếp các bản ghi kết quả với mệnh đề ORDER BY	47
5.3.2.4. Truy vấn dữ liệu từ nhiều hàng	47
5.4. CÂU HỎI ÔN TẬP.....	48
Chương 6: LÝ THUYẾT THIẾT KẾ CƠ SỞ DỮ LIỆU	49
6.1. Phụ thuộc hàm.....	49
6.1.1. Định nghĩa.....	49
1.2. Hệ tiên đề cho phụ thuộc hàm	51
6.1.3. Khái niệm bao đóng.....	53
6.1.4. Tính toán bao đóng.....	53
6.1.5. Tính tương đương của bao đóng tập phụ thuộc hàm	54
6.1.5.1. Định nghĩa.....	54
6.1.5.2. Cách chứng minh 2 tập phụ thuộc hàm F và G là tương đương.....	54
6.1.6. Phủ tối thiểu.....	54
6.1.6.1. Khái niệm	54
6.1.6.2. Thuật toán tìm phủ tối thiểu của tập phụ thuộc hàm F.....	54
6.2. Phép tách các lược đồ quan hệ.....	55
6.3. Chuẩn hóa lược đồ quan hệ	55
6.3.1. Chuẩn hóa 1NF	55
6.3.2. Chuẩn hóa 2NF	56
6.3.3. Chuẩn hóa 3NF	56
6.3.3.1. Định nghĩa cổ điển	56
6.3.3.2. Định nghĩa tổng quát.....	57
6.4. Chuẩn hóa BCNF	57
6.5. CÂU HỎI ÔN TẬP.....	58
Chương 7: GIẢI BÀI TẬP	60

GIÁO TRÌNH: CƠ SỞ DỮ LIỆU

Tên mô đun: Cơ sở dữ liệu

Mã số mô đun: MH09

I. Vị trí, tính chất của môn học:

- Vị trí: Môn học được bố trí sau khi sinh viên học xong các môn học chung, trước các môn học/mô-đun đào tạo chuyên môn nghề.
- Tính chất: Là môn học chuyên môn nghề bắt buộc, là nền tảng để HSSV có thể thiết kế cơ sở dữ liệu xây dựng ứng dụng và học các môn học mô đun liên quan.

II. Mục tiêu môn học:

- Về kiến thức:

- + Trình bày được công dụng của cơ sở dữ liệu.
- + Phân tích và trình bày được các khái niệm về cơ sở dữ liệu như: khóa, lược đồ quan hệ, phụ thuộc hàm,...
- + Ghi nhớ được các ngôn ngữ thao tác dữ liệu, ngôn ngữ truy vấn dữ liệu SQL.

- Về kỹ năng:

- + Xây dựng được các mô hình quan hệ.
- + Thiết kế được cơ sở dữ liệu
- + Phân tích được chương trình: xác định nhiệm vụ chương trình (phải làm gì).
- + Chuyển các câu hỏi sau khi tối ưu hoá bằng sơ đồ sang ngôn ngữ SQL.
- + Viết chương trình và thực hiện chương trình trên máy tính có ứng dụng csdl

- Về năng lực tự chủ và trách nhiệm:

- Có năng lực về chuyên môn thích nghi với các môi trường làm việc khác nhau
- Tự học tập, tích lũy kiến thức, kinh nghiệm để nâng cao trình độ chuyên môn nghiệp vụ

LỜI GIỚI THIỆU

Cơ sở dữ liệu là một môn học bắt buộc đối với sinh viên chuyên ngành công nghệ thông tin và điện tử viễn thông. Những năm gần đây trong các kỳ thi chuyển đổi, liên thông, thi tuyển cao học và nghiên cứu sinh đều có mảng kiến thức về cơ sở dữ liệu. Điều đó khẳng định yếu tố quan trọng của Cơ sở dữ liệu đối với các kỹ sư, thạc sĩ công nghệ thông tin và điện tử viễn thông tương lai.

Để hoàn thành cuốn sách, tác giả đã có quá trình ấp ủ và chuẩn bị những tư liệu cần thiết từ nhiều năm nay. Một mặt, đó là quá trình sưu tầm, nghiên cứu, tham khảo các tài liệu trong nước và ngoài nước có liên quan. Mặt khác, đó là quá trình học hỏi từ thực tế, từ kinh nghiệm giảng dạy nhiều năm trong nghề. Cuối cùng, những kiến thức và kinh nghiệm học hỏi đã được thể hiện qua cuốn giáo trình và các chương trình đào tạo cho sinh viên chuyên ngành công nghệ thông tin, tin học ứng dụng. Cuốn giáo trình đã liên tục được bổ sung, gọt giũa, điều chỉnh và trở thành những chất liệu có giá trị, đảm bảo điều kiện chín muồi cho cuốn sách ra đời.

Về bố cục, cuốn sách gồm 7 chương, trình bày các kỹ thuật cơ sở của cơ sở dữ liệu truyền thống, đó là mô hình liên kết thực thể, mô hình cơ sở dữ liệu quan hệ. Bài giảng cũng trình bày cách thiết kế một cơ sở dữ liệu quan hệ, cách sử dụng các phép toán đại số quan hệ để tạo, cập nhật và truy vấn cơ sở dữ liệu và khái niệm phụ thuộc hàm ứng dụng trong lý thuyết thiết kế và chuẩn hóa cơ sở dữ liệu quan hệ. Kèm theo đó là bài tập củng cố cho các chương. Đặc biệt ở chương 7 sẽ là gợi ý hướng dẫn và có lời giải cho một số các bài tập cơ bản ở mỗi chương.

Sau nhiều năm nghiên cứu và giảng dạy, tác giả nhận thấy nhiều giáo trình, nhiều sách về cơ sở dữ liệu quá chú trọng vào lý thuyết mà quên mất bài tập chính là củng cố cho những gì lý thuyết mà chúng ta đang học, hoặc có bài tập cũng chỉ đưa ra mà không có lời giải, như vậy nếu học sinh, sinh viên muốn tự nghiên cứu và học tập cũng rất khó.

Giáo trình này cần thiết cho tất cả các đối tượng muốn tìm hiểu và thiết kế các cơ sở dữ liệu quan hệ ứng dụng trong công tác quản lý.

Hy vọng cuốn sách không chỉ là tài liệu học tập hữu ích với các bạn học sinh, sinh viên mà còn cần thiết với những bạn đọc yêu công nghệ thông tin nói chung và cơ sở dữ liệu nói riêng.

Như đã nêu, mặc dù quá trình chuẩn bị và biên soạn sách được tiến hành khá thận trọng, các nội dung của cuốn sách hầu như đã được trải nghiệm trên thực tế trước khi công bố chính thức, song vì đây là lần biên soạn đầu tiên nên sách khó tránh khỏi những thiếu sót. Rất mong bạn đọc đón nhận tác phẩm này và sẵn lòng đóng góp những ý kiến để sách được hoàn thiện hơn trong những lần xuất bản sau.

Xin chân thành cảm ơn!

Người biên soạn: Trần Thị Thúy Mai

Chương 1: TỔNG QUAN VỀ CƠ SỞ DỮ LIỆU

Đây là chương mở đầu, bởi vậy trong chương này sẽ tập trung trình bày những vấn đề khái quát nhất về cơ sở dữ liệu. Nội dung của chương sẽ giúp cho người đọc:

- Hiểu được các khái niệm về cơ sở dữ liệu.
- Hiểu được kiến trúc của 1 hệ cơ sở dữ liệu.
- Phân tích được các đặc điểm chung của dữ liệu, thông tin và tính độc lập dữ liệu
- Phân loại được người dùng CSDL.

1.1. Tổng quan về Cơ sở dữ liệu

Các cơ sở dữ liệu và các hệ cơ sở dữ liệu đã trở thành một thành phần chủ yếu trong cuộc sống hàng ngày của xã hội hiện đại. Trong vòng một ngày con người có thể có nhiều hoạt động cần có sự giao tiếp với cơ sở dữ liệu như: đến ngân hàng để rút tiền và gửi tiền, đăng ký chỗ trên máy bay hoặc khách sạn, truy cập vào thư viện đã tin học hoá để tìm sách báo, đặt mua tạp chí ở một nhà xuất bản... Tại các ngân hàng, các cửa hàng, người ta cũng cập nhật tự động việc quản lý tiền bạc, hàng hoá.

Tất cả các giao tiếp như trên được gọi là các ứng dụng của cơ sở dữ liệu truyền thống. Trong các cơ sở dữ liệu truyền thống, hầu hết các thông tin được lưu giữ và truy cập là văn bản hoặc số. Những năm gần đây, những tiến bộ về kỹ thuật đã đưa đến những ứng dụng mới của cơ sở dữ liệu. Các cơ sở dữ liệu đa phương tiện bây giờ có thể lưu trữ hình ảnh, phim và tiếng nói. Các hệ thống thông tin địa lý có thể lưu trữ và phân tích các bản đồ, các dữ liệu về thời tiết và các ảnh vệ tinh. Kho dữ liệu và các hệ thống phân tích trực tuyến được sử dụng trong nhiều công ty để lấy ra và phân tích những thông tin có lợi từ các cơ sở dữ liệu rất lớn nhằm đưa ra các quyết định. Các kỹ thuật cơ sở dữ liệu động và thời gian thực được sử dụng trong việc kiểm tra các tiến trình công nghiệp và sản xuất. Các kỹ thuật tìm kiếm cơ sở dữ liệu đang được áp dụng cho World Wide Web để cung cấp việc tìm kiếm các thông tin cần thiết cho người sử dụng bằng cách duyệt qua Internet.

Để hiểu được các cơ sở kỹ thuật của cơ sở dữ liệu chúng ta phải bắt đầu từ các cơ sở kỹ thuật của cơ sở dữ liệu truyền thống. Mục đích của bài giảng này là nghiên cứu các cơ sở kỹ thuật đó.

1.2. Dữ liệu thông tin

Dữ liệu thông tin có các tính chất sau:

- Một cơ sở dữ liệu biểu thị một khía cạnh nào đó của thế giới thực như hoạt động của một công ty, một nhà trường, một ngân hàng... Những thay đổi của thế giới thực phải được phản ánh một cách trung thực vào trong cơ sở dữ liệu. Những thông tin được đưa vào trong cơ sở dữ liệu tạo thành một không gian cơ sở dữ liệu hoặc là một “thế giới nhỏ” (miniworld).
- Một cơ sở dữ liệu là một tập hợp dữ liệu liên kết với nhau một cách logic và mang một ý nghĩa cố hữu nào đó. Một cơ sở dữ liệu không phải là một tập hợp tùy tiện.
- Một cơ sở dữ liệu được thiết kế và được phổ biến cho một mục đích riêng. Nó có một

nhóm người sử dụng có chủ định và có một số ứng dụng được xác định phù hợp với mối quan tâm của người sử dụng. Nói cách khác, một cơ sở dữ liệu có một nguồn cung cấp dữ liệu, một mức độ tương tác với các sự kiện trong thế giới thực và một nhóm người quan tâm tích cực đến các nội dung của nó

Một cơ sở dữ liệu có thể có cỡ tùy ý và có độ phức tạp thay đổi. Có những cơ sở dữ liệu chỉ gồm vài trăm bản ghi (như cơ sở dữ liệu phục vụ việc quản lý lương ở một cơ quan nhỏ), và có những cơ sở dữ liệu có dung lượng rất lớn (như các cơ sở dữ liệu phục vụ cho việc tính cước điện thoại, quản lý nhân sự trên một phạm vi lớn). Các cơ sở dữ liệu phải được tổ chức quản lý sao cho những người sử dụng có thể tìm kiếm dữ liệu, cập nhật dữ liệu và lấy dữ liệu ra khi cần thiết. Một cơ sở dữ liệu có thể được tạo ra và duy trì một cách thủ công và cũng có thể được tin học hoá. Một cơ sở dữ liệu tin học hoá được tạo ra và duy trì bằng bằng một nhóm chương trình ứng dụng hoặc bằng một *hệ quản trị cơ sở dữ liệu*.

1.3. Phương pháp cơ sở dữ liệu

1.3.1. Cơ sở dữ liệu là gì?

Cơ sở dữ liệu và kỹ thuật cơ sở dữ liệu đã có ảnh hưởng rất lớn đến việc sử dụng máy tính. Có thể nói rằng cơ sở dữ liệu đóng vai trò quan trọng trong mọi lĩnh vực có sử dụng máy tính như giáo dục, thương mại, kỹ nghệ, khoa học, thư viện,.... Thuật ngữ cơ sở dữ liệu trở thành một thuật ngữ phổ dụng.

Một **cơ sở dữ liệu** là một tập hợp các dữ liệu có liên quan với nhau, được lưu trữ trên máy tính, có nhiều người sử dụng và được tổ chức theo một mô hình. **Dữ liệu** là những sự kiện có thể ghi lại được và có ý nghĩa.

Ví dụ, để quản lý việc học tập trong một môi trường đại học, các dữ liệu là các thông tin về sinh viên, về các môn học, điểm thi....Chúng ta tổ chức các dữ liệu đó thành các bảng và lưu giữ chúng vào sổ sách hoặc sử dụng một phần mềm máy tính để lưu giữ chúng trên máy tính. Ta có một tập các dữ liệu có liên quan đến nhau và mang nhiều ý nghĩa, đó là một cơ sở dữ liệu.

1.3.2. Hệ quản trị cơ sở dữ liệu

1.3.2.1. Định nghĩa hệ quản trị cơ sở dữ liệu

Một **hệ quản trị cơ sở dữ liệu** là một tập hợp chương trình giúp cho người sử dụng tạo ra, duy trì và khai thác một cơ sở dữ liệu. Nó là một hệ thống phần mềm phổ dụng, làm dễ quá trình định nghĩa, xây dựng và thao tác cơ sở dữ liệu cho các ứng dụng khác nhau.

Định nghĩa một cơ sở dữ liệu bao gồm việc đặc tả các kiểu dữ liệu, các cấu trúc và các ràng buộc cho các dữ liệu sẽ được lưu trữ trong cơ sở.

Xây dựng một cơ sở dữ liệu là quá trình lưu trữ các dữ liệu trên các phương tiện lưu trữ được hệ quản trị cơ sở dữ liệu kiểm soát.

Thao tác một cơ sở dữ liệu bao gồm các chức năng như truy vấn cơ sở dữ liệu để lấy ra các dữ liệu cụ thể, cập nhật cơ sở dữ liệu để phản ánh các thay đổi trong thế giới

nhỏ và tạo ra các báo cáo từ các dữ liệu.

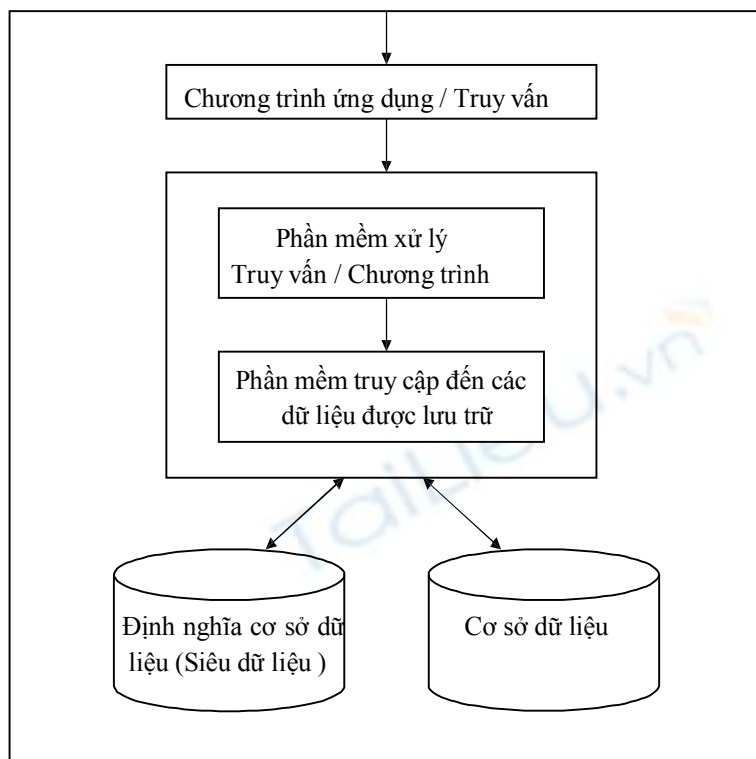
Các hệ quản trị cơ sở dữ liệu dùng để thể hiện một cơ sở dữ liệu tin học hoá có thể là phổ dụng (là một phần mềm đóng gói) hoặc có thể là chuyên dụng (là một tập các phần mềm được tạo ra với một mục đích riêng).

Người ta gọi cơ sở dữ liệu và hệ quản trị cơ sở dữ liệu bằng một thuật ngữ chung là **hệ cơ sở dữ liệu**. Môi trường của một hệ cơ sở dữ liệu được mô tả bằng hình vẽ dưới đây (Hình I-1).

1.3.2.2. Các chức năng của một hệ quản trị cơ sở dữ liệu

1. Lưu trữ các định nghĩa, các mối liên kết dữ liệu (gọi là siêu dữ liệu) vào một từ điển dữ liệu. Các chương trình truy cập đến cơ sở dữ liệu làm việc thông qua hệ quản trị cơ sở dữ liệu. Hệ quản trị cơ sở dữ liệu sử dụng dữ liệu trong từ điển dữ liệu để tìm kiếm các cấu trúc thành phần dữ liệu và các mối liên kết được yêu cầu. Mọi sự thay đổi trong các tệp cơ sở dữ liệu sẽ được tự động ghi lại vào từ điển dữ liệu. Như vậy, hệ quản trị cơ sở dữ liệu giải phóng người sử dụng khỏi việc lập trình cho các mối liên kết phức tạp trong mỗi chương trình, việc sửa đổi các chương trình truy cập đến tệp cơ sở dữ liệu đã bị sửa đổi. Nói cách khác, hệ quản trị cơ sở dữ liệu loại bỏ sự phụ thuộc giữa dữ liệu và cấu trúc ra khỏi hệ thống.

Người sử dụng / Người lập trình



Hình I-1. Môi trường hệ cơ sở dữ liệu

2. Tạo ra các cấu trúc phức tạp theo yêu cầu để lưu trữ dữ liệu. Nó giúp người sử dụng làm nhiệm vụ khó khăn là định nghĩa và lập trình cho các đặc trưng vật lý của dữ liệu.

3. Biến đổi các dữ liệu được nhập vào để phù hợp với các cấu trúc dữ liệu ở điểm 2. Như vậy, hệ quản trị cơ sở dữ liệu giúp người sử dụng phân biệt dạng logic và dạng vật lý của dữ liệu. Bằng việc duy trì sự độc lập dữ liệu, hệ quản trị cơ sở dữ liệu chuyển các yêu cầu logic thành các lệnh định vị một cách vật lý và lấy ra các dữ liệu yêu cầu. Điều đó cũng có nghĩa là hệ quản trị cơ sở dữ liệu tạo khuôn dạng cho các dữ liệu được lấy ra để làm cho nó phù hợp với mong muốn logic của người sử dụng.

4. Tạo ra một hệ thống bảo mật và áp đặt tính bảo mật và riêng tư trong cơ sở dữ liệu.

5. Tạo ra các cấu trúc phức tạp cho phép nhiều người sử dụng truy cập đến dữ liệu

6. Cung cấp các thủ tục sao lưu và phục hồi dữ liệu để đảm bảo sự an toàn và toàn vẹn dữ liệu.

7. Xúc tiến và áp đặt các quy tắc an toàn để loại bỏ vấn đề toàn vẹn dữ liệu. Điều đó cho phép ta làm tối thiểu sự dư thừa dữ liệu và làm tối đa tính nhất quán dữ liệu.

8. Cung cấp việc truy cập dữ liệu thông qua một ngôn ngữ truy vấn. Một ngôn ngữ truy vấn là một ngôn ngữ phi thủ tục cho phép người sử dụng chỉ ra cái gì cần phải làm mà không cần phải chỉ ra nó được làm như thế nào. Các hệ quản trị cơ sở dữ liệu cũng cung cấp việc truy cập dữ liệu cho những người lập trình thông qua các ngôn ngữ thủ tục.

1.3.2.3. Ưu điểm của phương pháp CSDL

Trước khi khái niệm cơ sở dữ liệu ra đời, hệ thống tệp (file) là một phương pháp được áp dụng trong việc quản lý. Một tệp có thể được xem là một cặp hồ sơ lưu trữ các thông tin liên quan đến từng công việc riêng biệt. Ví dụ, trong một cơ quan, bộ phận tài vụ sẽ có một cặp hồ sơ liên quan đến lương của các nhân viên, bộ phận tổ chức có cặp hồ sơ liên quan đến vấn đề nhân sự... Việc xử lý để lấy ra các thông tin như là các thống kê về lương, về quá trình công tác... lúc đầu được thực hiện một cách thủ công. Dần dần, khối lượng thông tin ngày càng lớn, việc xử lý thông tin ngày càng phức tạp, người ta sử dụng máy tính vào việc quản lý. Các cặp hồ sơ được chuyển thành các tệp trên máy tính và việc xử lý thông tin được thực hiện bằng cách lập trình (trong một ngôn ngữ lập trình thế hệ 3).

Việc quản lý theo giải pháp hệ thống tệp có rất nhiều nhược điểm. Thứ nhất, đó là sự dư thừa thông tin: cùng một thông tin được lưu trữ nhiều lần (chẳng hạn, danh sách nhân viên có mặt trong tệp lương và cũng có mặt cả trong tệp nhân sự). Điều đó gây ra việc lãng phí bộ nhớ và dễ gây sai sót trong khi cập nhật dữ liệu, dễ sinh ra các dữ liệu không đúng đắn. Thứ 2 đó là việc phụ thuộc giữa chương trình ứng dụng và dữ liệu. Mỗi

khi có sự thay đổi cấu trúc tệp và các dữ liệu trong tệp, chương trình ứng dụng khai thác thông tin trên tệp đó cũng thay đổi theo. Điều đó gây ra khó khăn lớn cho việc bảo trì. Giải pháp cơ sở dữ liệu ra đời đã giải quyết được những nhược điểm đó. Cụ thể, giải pháp cơ sở dữ liệu có những đặc trưng sau:

1. Bản chất tự mô tả của hệ cơ sở dữ liệu.

Một đặc trưng cơ bản của giải pháp cơ sở dữ liệu là hệ thống cơ sở dữ liệu không chỉ gồm có bản thân cơ sở dữ liệu mà còn có cả định nghĩa hoặc mô tả đầy đủ về cấu trúc cơ sở dữ liệu và các ràng buộc. Định nghĩa này được lưu trữ trong từ điển hệ thống, nó chứa các thông tin như là cấu trúc của mỗi tệp, kiểu và dạng lưu trữ của từng mục dữ liệu. Các thông tin được lưu giữ trong từ điển gọi là siêu dữ liệu (meta-data) và chúng mô tả cấu trúc của dữ liệu nguyên thủy (hình I-1). Phần mềm hệ quản trị cơ sở dữ liệu và những người sử dụng cơ sở dữ liệu sử dụng từ điển để lấy thông tin về cấu trúc của cơ sở dữ liệu

2. Sự độc lập giữa chương trình và dữ liệu.

Trong hệ thống tệp, cấu trúc của các tệp cơ sở dữ liệu được nhúng vào trong các chương trình truy cập, vì vậy bất kỳ một thay đổi nào về cấu trúc của một tệp cũng đòi hỏi phải thay đổi tất cả các chương trình truy cập đến tệp đó. Ngược lại, các chương trình truy cập của hệ quản trị cơ sở dữ liệu không đòi hỏi việc thay đổi như thế. Cấu trúc của các tệp dữ liệu được lưu trữ trong từ điển tách rời với các chương trình truy cập. Tính chất này gọi là sự độc lập dữ liệu – chương trình.

3. Hỗ trợ các khung nhìn dữ liệu nhiều thành phần.

Một cơ sở dữ liệu có nhiều người sử dụng, mỗi một người có thể đòi hỏi một phối cảnh hoặc một khung nhìn (view) khác nhau. Một khung nhìn có thể là một tập con của cơ sở dữ liệu hoặc nó có thể chứa các dữ liệu ảo, đó là các dữ liệu được trích ra từ các tệp cơ sở dữ liệu khác nhau nhưng không được lưu trữ một cách rõ ràng. Một hệ quản trị cơ sở dữ liệu nhiều người sử dụng phải cung cấp nhiều công cụ để định nghĩa các khung nhìn nhiều thành phần.

4. Chia sẻ dữ liệu và nhiều người sử dụng.

Một hệ quản trị cơ sở dữ liệu nhiều người sử dụng phải cho phép nhiều người sử dụng truy cập đồng thời đến cơ sở dữ liệu. Hệ quản trị cơ sở dữ liệu phải có phần mềm kiểm tra cạnh tranh để đảm bảo rằng các người sử dụng cập nhật đến cùng một cơ sở dữ liệu phải được thực hiện theo cách được kiểm tra để cho kết quả của các cập nhật là đúng đắn.

1.3.2.4. Phân loại cơ sở dữ liệu

Có rất nhiều mô hình dữ liệu đã được đề nghị. Chúng ta có thể phân loại các mô hình dữ liệu dựa trên các khái niệm mà chúng sử dụng để mô tả các cấu trúc cơ sở dữ liệu.

Các mô hình dữ liệu bậc cao hoặc mô hình dữ liệu mức quan niệm cung cấp các khái niệm gắn liền với cách cảm nhận dữ liệu của nhiều người sử dụng. Các mô hình này

tập trung vào bản chất logic của biểu diễn dữ liệu, nó quan tâm đến *cái* được biểu diễn trong cơ sở dữ liệu chứ không phải *cách* biểu diễn dữ liệu.

Các *mô hình dữ liệu bậc thấp* hoặc các *mô hình dữ liệu vật lý* cung cấp các khái niệm mô tả chi tiết về việc các dữ liệu được lưu trữ trong máy tính như thế nào. Các khái niệm do mô hình dữ liệu vật lý cung cấp nói chung có ý nghĩa đối với các chuyên gia máy tính chứ không có ý nghĩa mấy đối với người sử dụng thông thường. Ở giữa hai loại mô hình này là một lớp các *mô hình dữ liệu thể hiện*, chúng cung cấp những khái niệm mà người sử dụng có thể hiểu được và không xa với cách tổ chức dữ liệu bên trong máy tính. Người ta còn gọi loại mô hình dữ liệu này là loại *mô hình dữ liệu mức logic*. Các mô hình dữ liệu thể hiện che giấu một số chi tiết về việc lưu trữ dữ liệu nhưng có thể được cài đặt trực tiếp trên hệ thống máy tính. Trong chương II, chúng ta sẽ nghiên cứu một mô hình dữ liệu mức quan niệm,

mô hình thực thể - liên kết, gọi tắt là mô hình ER (Entity – Relationship Model). Mô hình này sử dụng các khái niệm thực thể, thuộc tính, mối liên kết, để diễn đạt các đối tượng của thế giới thực. Một *thực thể* diễn đạt một đối tượng hoặc một khái niệm của thế giới thực. Ví dụ, một thực thể là một nhân viên hoặc một dự án được mô tả trong cơ sở dữ liệu. Một thuộc tính diễn đạt một đặc trưng nào đó của thực thể. Chẳng hạn, họ tên, lương... là các thuộc tính của thực thể nhân viên. Một mối liên kết giữa hai hay nhiều thực thể diễn đạt một mối quan hệ qua lại giữa các thực thể. Ví dụ, giữa thực thể nhân viên và thực thể dự án có mối liên kết một nhân viên *làm việc trên* một dự án. Mô hình dữ liệu hướng đối tượng cũng là một mô hình dữ liệu bậc cao. Nó sử dụng các khái niệm như lớp, phương thức, thông điệp... Bạn đọc có thể tìm hiểu về mô hình này trong các tài liệu [1], [2].

Các mô hình dữ liệu thể hiện là các mô hình được sử dụng thường xuyên nhất trong các hệ cơ sở dữ liệu thương mại. Ba mô hình nổi tiếng thuộc loại này là mô hình quan hệ, mô hình mạng và mô hình phân cấp. Các mô hình mạng và phân cấp ra đời trước và được sử dụng rộng rãi trong quá khứ (trước 1970). Vào đầu những năm 70, mô hình quan hệ ra đời. Do tính ưu việt của nó, mô hình quan hệ dần dần thay thế các mô hình mạng và phân cấp. Chúng ta sẽ nghiên cứu về mô hình quan hệ trong chương III.

Các mô hình dữ liệu vật lý mô tả cách lưu trữ dữ liệu trong máy tính giới thiệu các thông tin như khuôn dạng bản ghi, sắp xếp bản ghi, đường truy cập...

1.4. Phân loại người dùng Cơ sở dữ liệu

Với một cơ sở dữ liệu lớn, rất nhiều người tham gia vào việc thiết kế, sử dụng và duy trì cơ sở dữ liệu. Những người liên quan đến hệ cơ sở dữ liệu được chia thành hai nhóm chính. Nhóm thứ nhất gồm những người mà công việc của họ liên quan hàng ngày đến cơ sở dữ liệu, đó là những người quản trị cơ sở dữ liệu, thiết kế cơ sở dữ liệu, sử dụng cơ sở dữ liệu, phân tích hệ thống và lập trình ứng dụng. Nhóm thứ hai gồm những người làm việc để duy trì môi trường hệ cơ sở dữ liệu nhưng không quan tâm đến bản thân cơ sở dữ liệu, đó là những người thiết kế và cài đặt hệ quản trị cơ sở dữ

liệu, phát triển công cụ, thao tác viên và bảo trì.

1.4.1. Người quản trị hệ cơ sở dữ liệu (Database Administrator – DBA)

Trong một tổ chức có nhiều người cùng sử dụng các tài nguyên, cần phải có một người giám sát và quản lý. Trong môi trường hệ cơ sở dữ liệu, các tài nguyên là cơ sở dữ liệu, hệ quản trị cơ sở dữ liệu và các phần mềm liên quan. Người quản trị hệ cơ sở dữ liệu là người chịu trách nhiệm quản lý các tài nguyên đó. Người này chịu trách nhiệm về việc cho phép truy cập cơ sở dữ liệu, tổ chức và hướng dẫn việc sử dụng cơ sở dữ liệu, cấp các phần mềm và phần cứng theo yêu cầu.

1.4.2. Người thiết kế cơ sở dữ liệu (Database Designer)

Người này chịu trách nhiệm xác định các dữ liệu sẽ được lưu giữ trong cơ sở, chọn các cấu trúc thích hợp để biểu diễn và lưu giữ các dữ liệu đó. Những nhiệm vụ này được thực hiện trước khi cơ sở dữ liệu được cài đặt và phổ biến. Người thiết kế có trách nhiệm giao thiệp với những người sử dụng tương lai để hiểu được các đòi hỏi của họ và đưa ra một thiết kế thoả mãn các yêu cầu đó. Anh ta cũng có nhiệm vụ giao thiệp với các nhóm người sử dụng và có khả năng hỗ trợ các yêu cầu của các nhóm.

1.4.3. Những người sử dụng (End User)

Những người sử dụng là những người mà công việc của họ đòi hỏi truy cập đến cơ sở dữ liệu để truy vấn, cập nhật và sinh ra các thông tin. Có thể chia những người sử dụng thành hai nhóm chính: những *người sử dụng thụ động* (tức là những người sử dụng không có nhiều kiến thức về hệ cơ sở dữ liệu) và những *người sử dụng chủ động* (là những người có hiểu biết tốt về hệ cơ sở dữ liệu).

Chức năng công việc của những người sử dụng thụ động (chiếm phần lớn những người sử dụng) gắn liền với việc truy vấn và cập nhật thường xuyên cơ sở dữ liệu bằng cách sử dụng các câu hỏi và các cập nhật chuẩn (gọi là các giao tác định sẵn) đã được lập trình và kiểm tra cẩn thận. Những người này chỉ cần học một ít về các phương tiện do hệ quản trị cơ sở dữ liệu cung cấp và hiểu các kiểu giao tác chuẩn đã được thiết kế và cài đặt là đủ.

Những người sử dụng chủ động có hiểu biết tốt về hệ cơ sở dữ liệu, họ có thể tự cài đặt các ứng dụng riêng của mình để làm thoả mãn các yêu cầu phức tạp của họ.

1.4.4. Người phân tích hệ thống và lập trình ứng dụng

Người phân tích hệ thống xác định các yêu cầu của những người sử dụng (chủ yếu là những người sử dụng thụ động) để đặc tả các chương trình phù hợp với yêu cầu của họ.

Người viết chương trình ứng dụng thể hiện các đặc tả của những người phân tích thành chương trình, sau đó kiểm thử, sửa lỗi làm tài liệu và bảo trì các giao tác định sẵn.

1.4.5. Người thiết kế và cài đặt hệ quản trị dữ liệu

Đó là những người thiết kế, cài đặt các mô đun, giao diện của hệ quản trị cơ sở dữ liệu thành các phần mềm đóng gói. Một hệ quản trị cơ sở dữ liệu là một hệ thống phần mềm phức tạp bao gồm nhiều thành phần (mô đun). Đó là các mô đun cài đặt từ

điền dữ liệu, ngôn ngữ truy vấn, bộ xử lý giao diện, truy cập dữ liệu, kiểm tra cạnh tranh, phục hồi và an toàn. Hệ quản trị cơ sở dữ liệu phải giao tiếp với các hệ thống phần mềm khác như hệ điều hành và các chương trình dịch cho nhiều ngôn ngữ khác nhau.

1.4.6. Những người phát triển công cụ

Là những người thiết kế và cài đặt các công cụ (tool), đó là các phần mềm đóng gói làm dễ việc thiết kế và sử dụng cơ sở dữ liệu.

1.4.7. Các thao tác viên và những người bảo trì

Là những người chịu trách nhiệm về việc chạy và bảo trì phần cứng và phần mềm của hệ thống.

1.5. Câu hỏi ôn tập

- 1.5.1. Định nghĩa các thuật ngữ : cơ sở dữ liệu, hệ quản trị cơ sở dữ liệu, hệ cơ sở dữ liệu, từ điển cơ sở dữ liệu, mô hình cơ sở dữ liệu.
- 1.5.2. Nêu các tính chất của một cơ sở dữ liệu
- 1.5.3. Nêu các chức năng của một hệ quản trị cơ sở dữ liệu
- 1.5.4. Giải thích các đặc trưng của giải pháp cơ sở dữ liệu
- 1.5.5. Định nghĩa mô hình cơ sở dữ liệu và phân loại

Chương 2: CÁC MÔ HÌNH DỮ LIỆU

Trong chương này chúng ta sẽ làm quen với mô hình thực thể - liên kết, gọi tắt là mô hình ER (Entity-Relationship Model). Đó là một mô hình dữ liệu mức quan niệm phổ biến, tập trung vào các cấu trúc dữ liệu và các ràng buộc. Mô hình này thường được sử dụng để thiết kế các ứng dụng cơ sở dữ liệu và nhiều công cụ thiết kế cơ sở dữ liệu sử dụng các khái niệm của nó.

Sau khi học xong chương 2, học sinh có khả năng:

- Phân loại được các mô hình dữ liệu, các sơ đồ quan hệ .
- Nắm được tầm quan trọng của các mô hình dữ liệu quan hệ.
- Vận dụng để giải quyết các bài toán về mô hình dữ liệu quan hệ .

2.1. Sơ đồ thực thể liên kết

2.1.1. Thực thể

Thực thể là một “vật” trong thế giới thực, có sự tồn tại độc lập. Một thực thể có thể là cụ thể, tức là chúng ta có thể cảm nhận được bằng các giác quan, hoặc có thể là trừu tượng, tức là cái mà chúng ta không cảm nhận được bằng các giác quan nhưng có thể nhận biết được bằng nhận thức. Một cái ô tô, một nhân viên,... là những thực thể cụ thể. Một đơn vị công tác, một trường học... là những thực thể trừu tượng.

Mỗi một thực thể có các *thuộc tính*, đó là các đặc trưng cụ thể mô tả thực thể đó. Ví dụ, một thực thể *Nhân viên* được mô tả bằng *Họ tên*, *Tuổi*, *Địa chỉ*, *Lương*... của nhân viên đó. Một thực thể cụ thể sẽ có một giá trị cho mỗi thuộc tính của nó. Ví dụ, nhân viên nv1 có các giá trị cho các thuộc tính *Họ tên*, *Tuổi*, *Địa chỉ*, *Lương* của nó là “ Lê Văn”, 32, “Hà nội”, 500000. Các giá trị thuộc tính mô tả mỗi thực thể sẽ trở thành một phần chính của các dữ liệu sẽ được lưu giữ trong cơ sở dữ liệu. Trong mô hình ER có mặt nhiều kiểu thuộc tính: thuộc tính đơn, thuộc tính phức hợp, thuộc tính đơn trị, thuộc tính đa trị, thuộc tính được lưu trữ, thuộc tính suy diễn được, thuộc tính có giá trị không xác định, thuộc tính phức tạp.

Thuộc tính đơn là thuộc tính không thể phân chia ra được thành các thành phần nhỏ hơn. Ví dụ, thuộc tính *Tuổi* của một nhân viên là một thuộc tính đơn. *Thuộc tính phức hợp* là thuộc tính có thể phân chia được thành các thành phần nhỏ hơn, biểu diễn các thuộc tính cơ bản hơn với các ý nghĩa độc lập. Ví dụ, thuộc tính *Họ tên* của thực thể nhân viên có thể phân chia thành các tính *Họ đệm* và *Tên*. Giá trị của một thuộc tính là sự kết hợp kết hợp các giá trị của các thuộc tính thành phần tạo nên nó. Việc phân chia một thuộc tính phức hợp thành các thuộc tính đơn tùy thuộc vào hoàn cảnh cụ thể.

Những thuộc tính có giá trị duy nhất cho một thực thể cụ thể gọi là các *thuộc tính đơn trị*. Ví dụ, *Họ tên* là một thuộc tính đơn trị của thực thể nhân viên, mỗi nhân viên có một họ tên duy nhất. Trong một số trường hợp, một thuộc tính có thể có một tập giá trị cho cùng một thực thể. Những thuộc tính như vậy gọi là *thuộc tính đa trị*. Ví dụ, thuộc tính *Bằng cấp* của một người. Một người có thể không có bằng cấp nào, người

khác có thể có một bằng, người khác nữa có thể có nhiều bằng. Như vậy, các người khác nhau có thể có một số giá trị khác nhau cho thuộc tính Bằng cấp. Thuộc tính Bằng cấp là một thuộc tính đa trị.

Thuộc tính được lưu trữ là các thuộc tính mà giá trị của nó được nhập vào khi cài đặt cơ sở dữ liệu. Trong một số trường hợp, hai hay nhiều thuộc tính có giá trị liên quan đến nhau. Ví dụ, thuộc tính Tuổi và thuộc tính Ngày sinh của một người. Với một người cụ thể, ta có thể tính tuổi của anh ta bằng cách lấy năm hiện tại trừ đi năm của Ngày sinh. Thuộc tính mà giá trị của nó có thể tính được thông qua giá trị của các thuộc tính khác gọi là *thuộc tính suy diễn được*.

Các giá trị không xác định (null values): Trong một số trường hợp, một thực thể cụ thể có thể không có các giá trị áp dụng được cho một thuộc tính. Ví dụ, thuộc tính Số điện thoại của thực thể nhân viên sẽ không có giá trị đối với các nhân viên không có số điện thoại. Trong trường hợp như vậy, ta phải tạo ra một giá trị đặc biệt gọi là giá trị không xác định (null). Giá trị không xác định được tạo ra khi một thuộc tính có giá trị không áp dụng được hoặc khi không biết.

Các thuộc tính phức tạp: Là sự kết hợp của các thuộc tính phức hợp và đa trị.

2.1.2. Sơ đồ thực thể liên kết

2.1.2.1. Kiểu thực thể, tập thực thể, khóa và tập giá trị

Các kiểu thực thể và các tập thực thể: Một cơ sở dữ liệu thường chứa những nhóm thực thể như nhau. Ví dụ, một công ty thuê hàng trăm nhân viên và lưu giữ những thông tin tương tự liên quan đến mỗi nhân viên. Các thực thể nhân viên này chia sẻ các thuộc tính giống nhau nhưng mỗi thực thể có các giá trị riêng cho các thuộc tính đó. Một *kiểu thực thể* là một tập hợp các thực thể có các thuộc tính như nhau. Một kiểu thực thể trong cơ sở dữ liệu được mô tả bằng tên và các thuộc tính. Ví dụ: NHÂNVIÊN (Họ tên, Tuổi, Lương), CÔNGTY (Tên, Địa điểm, Giám đốc). Một tập hợp các thực thể của một kiểu thực thể cụ thể trong cơ sở dữ liệu tại một thời điểm được gọi là một tập thực thể, nó thường được tham chiếu đến bằng cách sử dụng tên của kiểu thực thể. Ví dụ, NHÂNVIÊN vừa dùng để chỉ một kiểu thực thể, vừa để chỉ tập hợp hiện tại của tất cả các thực thể nhân viên trong cơ sở dữ liệu.

Một kiểu thực thể được biểu diễn trong lược đồ ER như là một hộp hình chữ nhật có chứa tên kiểu thực thể. Các thuộc tính được đặt trong các hình ô van và được nối với các kiểu thực thể bằng các đường thẳng. Các thuộc tính phức hợp cũng được nối với các thuộc tính thành phần của nó bằng đường thẳng. Các thuộc tính đa trị được hiển thị trong các hình ô van đúp (hình II-2).

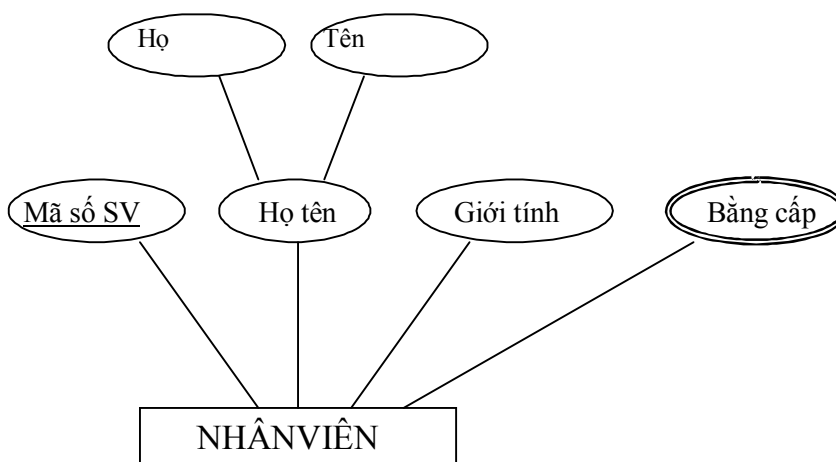
Một kiểu thực thể mô tả một *lược đồ* (hoặc một mục đích) cho một tập các thực thể chia sẻ cùng một cấu trúc. Tập hợp các thực thể của một kiểu thực thể cụ thể được nhóm vào một tập thực thể và được gọi là một *thể hiện* của một kiểu thực thể.

Thuộc tính khóa của một kiểu thực thể: Một ràng buộc quan trọng trên các thực thể của một kiểu thực thể là *khóa*. Một kiểu thực thể thường có một thuộc tính mà các giá

trị của nó là khác nhau đối với mỗi thực thể riêng biệt trong một tập thực thể. Thuộc tính như vậy gọi là *thuộc tính khóa* và các giá trị của nó có thể dùng để xác định từng thực thể một cách duy nhất. Ví dụ, thuộc tính Tên của kiểu thực thể CÔNGTY là khóa của kiểu thực thể đó vì mỗi thực thể công ty có một tên duy nhất. Đôi khi, nhiều thuộc tính kết hợp với nhau tạo thành một khóa, nghĩa là tổ hợp các giá trị của các thuộc tính này phải khác nhau đối với mỗi thực thể. Trong trường hợp như vậy ta có một thuộc tính *khóa phức hợp*. Chú ý rằng khóa phức hợp phải tối thiểu, nghĩa là tất cả các thuộc tính thành phần phải có mặt trong thuộc tính phức hợp để thỏa mãn tính chất duy nhất. Trong biểu đồ đồ họa của mô hình ER, thuộc tính khóa được biểu diễn bằng cách gạch ngang dưới tên của nó (hình II-3).

Khi chỉ ra rằng một thuộc tính là khóa của một kiểu thực thể nghĩa là tính chất duy nhất nêu trên phải được thỏa mãn đối với đối với mỗi mở rộng của kiểu thực thể. Như vậy, ràng buộc khóa cấm hai thực thể bất kỳ có giá trị cho thuộc tính khóa như nhau tại cùng một thời điểm. Đó là một ràng buộc trên tất cả các thể hiện của thực thể. Ràng buộc khóa cũng như các ràng buộc sẽ được giới thiệu về sau được lấy ra từ các ràng buộc của “thế giới nhỏ” của cơ sở dữ liệu.

Một kiểu thực thể có thể có nhiều hơn một thuộc tính khóa. Ví dụ, nếu một công ty có một mã số duy nhất và một tên duy nhất thì các thuộc tính Mã số công ty và Tên công ty đều là các thuộc tính khóa. Một kiểu thực thể cũng có thể không có khóa. Một thực thể không có khóa được gọi là *kiểu thực thể yếu*.



Hình II-2. Biểu diễn kiểu thực thể và các thuộc tính

2.1.2.2. Kiểu liên kết, tập liên kết và các thể hiện

Một *kiểu liên kết* R giữa n kiểu thực thể $E_1, E_2, \dots, E_n$ xác định một *tập liên kết* giữa các thực thể của các kiểu đó. Cũng như các kiểu thực thể và tập thực thể, một kiểu liên kết và tập liên kết tương ứng với nó cũng có tên chung là R. Một cách toán học, tập liên kết R là một tập hợp các thể hiện liên kết r_i , $i = 1, 2, \dots$ trong đó mỗi r_i liên kết

n thực thể riêng biệt $e_1, e_2, \dots, e_n$ và mỗi một thực thể e_j trong r_i là một thành phần của kiểu thực thể E_j , $1 \leq j \leq n$. Như vậy, một kiểu liên kết R là một quan hệ toán học trên $E_1, E_2, \dots, E_n$ hoặc có thể định nghĩa như là một tập con của tích Đề các $E_1 \times E_2 \times \dots \times E_n$. Mỗi kiểu thực thể $E_1, E_2, \dots, E_n$ được gọi là tham gia vào kiểu liên kết R , và tương tự, mỗi thực thể riêng biệt $e_1, e_2, \dots, e_n$ được gọi là tham gia vào thể hiện liên kết $r_i = (e_1, e_2, \dots, e_n)$.

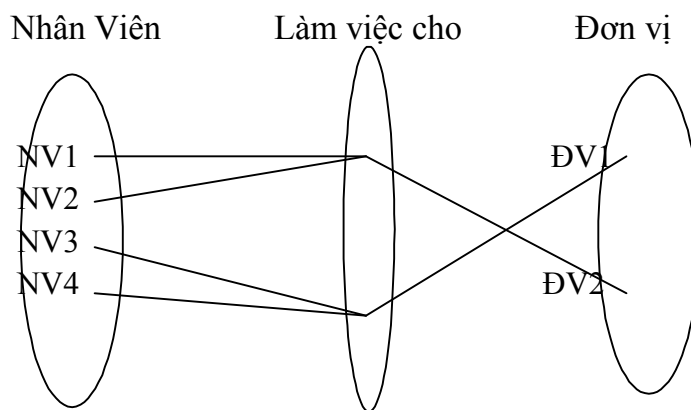
Một cách không hình thức, mỗi thể hiện liên kết r_i trong R là một sự kết hợp của các thực thể, mỗi thực thể thuộc về một kiểu thực thể tham gia vào liên kết. Mỗi liên kết r_i như vậy diễn đạt một sự kiện rằng các thực thể tham gia trong r_i có mối quan hệ với nhau theo một cách nào đó ở trong thế giới thực. Ví dụ, trong thực tế, các nhân viên làm việc cho các đơn vị, như vậy, có một kiểu liên kết liên kết *làm việc cho*, liên kết giữa kiểu thực thể NHÂNVIÊN và kiểu thực thể ĐƠN VỊ.

Trong sơ đồ ER, kiểu liên kết được biểu diễn bằng một hình thoi nối trực tiếp với các hình chữ nhật biểu diễn các kiểu thực thể tham gia vào liên kết. Hình II-4 minh họa kiểu liên kết và thể hiện liên kết.

a) Kiểu liên kết:



b) Thể hiện liên kết:



Hình II-4. Kiểu liên kết và thể hiện liên kết

2.1.2.3. Cấp liên kết, tên vai trò và kiểu liên kết đệ quy

Cấp của một kiểu liên kết là số các kiểu thực thể tham gia vào kiểu liên kết đó. Một kiểu liên kết có thể có cấp 1, cấp 2, cấp 3, Ví dụ, kiểu liên kết *<làm việc cho>* giữa kiểu thực thể NHÂNVIÊN và kiểu thực thể ĐƠN VỊ là một kiểu liên kết cấp 2. Kiểu liên kết *<biết trước>* giữa kiểu thực thể MÔN HỌC với chính nó là một kiểu liên kết cấp 1...

Đôi khi chúng ta có thể coi một kiểu liên kết như một thuộc tính của một kiểu thực thể. Ví dụ, nếu kiểu thực thể NHÂNVIÊN có thuộc tính Đơn vị để chỉ ra tên đơn vị mà nhân viên làm việc cho, thì thuộc tính Đơn vị biểu thị một kiểu liên kết. Nói cách khác, một thuộc tính của một kiểu thực thể hoặc có chức năng biểu thị một đặc trưng của kiểu thực thể, hoặc có chức năng biểu thị một kiểu liên kết giữa kiểu thực thể đó với các kiểu thực thể khác. Các thuộc tính biểu thị một kiểu liên kết có thể đơn trị hoặc đa trị tùy theo bản chất của mối liên kết.

Các tên vai trò và các kiểu liên kết đệ quy: Mỗi một kiểu thực thể tham gia vào một kiểu liên kết có một vai trò cụ thể trong liên kết. *Tên vai trò* dùng để chỉ rõ vai trò của các thực thể của kiểu thực thể tham gia liên kết, nó giúp đỡ việc giải thích ý nghĩa của liên kết. Ví dụ, trong kiểu liên kết NHÂNVIÊN <làm việc cho> ĐƠN VỊ, vai trò của các thực thể của kiểu thực thể NHÂNVIÊN là nhân viên hoặc công nhân còn vai trò của các thực thể của kiểu thực thể ĐƠN VỊ là đơn vị hoặc nơi thuê công nhân. Nếu các kiểu thực thể tham gia vào kiểu liên kết là khác nhau thì tên vai trò là hoàn toàn không cần thiết bởi vì có thể sử dụng tên các kiểu thực thể làm tên vai trò. Tuy nhiên, trong một số trường hợp, một kiểu thực thể có thể tham gia vào một kiểu liên kết với các vai trò khác nhau. Trong những trường hợp như vậy, tên vai trò trở nên cần thiết để phân biệt ý nghĩa của mỗi sự tham gia. Các kiểu liên kết như vậy gọi là *kiểu liên kết đệ quy*. Ví dụ, trong số các nhân viên làm việc cho một đơn vị, có các nhân viên được phân công giám sát các nhân viên khác. Như vậy sẽ có một kiểu liên kết giữa các thực thể của kiểu thực thể NHÂNVIÊN: NHÂNVIÊN <giám sát> NHÂNVIÊN. Kiểu thực thể NHÂNVIÊN tham gia hai lần vào kiểu liên kết <giám sát>, một lần với vai trò người giám sát, một lần với vai trò người bị giám sát.

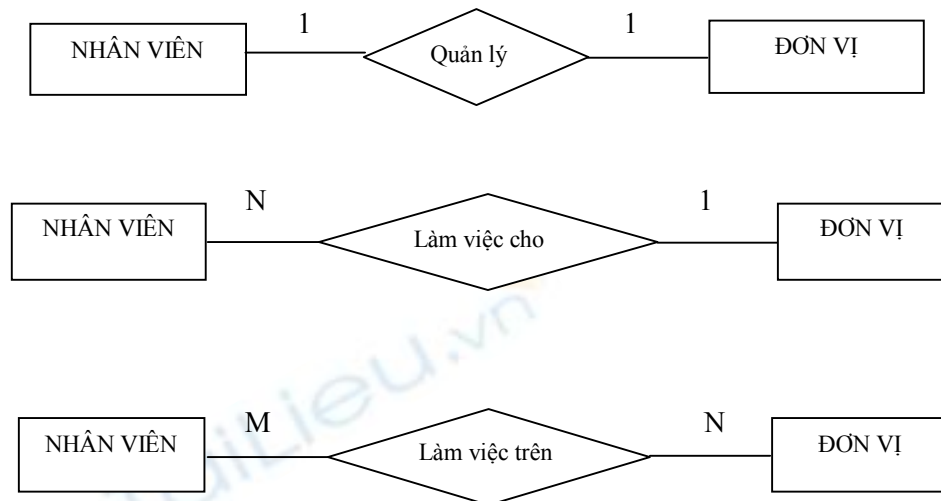
2.1.3. Quan hệ

2.1.3.1. Các ràng buộc trên các kiểu liên kết

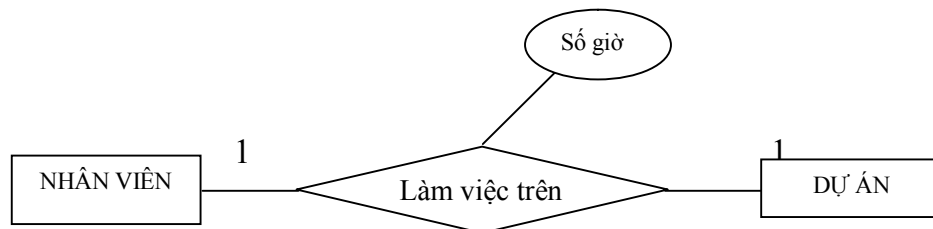
Các kiểu liên kết thường có một số ràng buộc để hạn chế số các tổ hợp có thể của các thực thể có thể tham gia trong tập hợp liên kết tương ứng. Các ràng buộc này được xác định từ tình trạng của thế giới thực mà kiểu liên kết biểu diễn. Ví dụ, nếu công ty có quy chế là một nhân viên chỉ làm việc cho một đơn vị thì chúng ta phải mô tả ràng buộc này trong lược đồ. Có hai loại ràng buộc chính: tỷ số lực lượng và sự tham gia.

Tỷ số lực lượng: Tỷ số lực lượng cho một kiểu liên kết chỉ ra số các thể hiện liên kết mà một thực thể có thể tham gia. Với các kiểu liên kết cấp 2, có thể có các tỷ số lực lượng 1:1, 1:N, và M:N. Một kiểu liên kết có tỷ số lực lượng 1:1 giữa hai kiểu thực thể A và B có nghĩa là trong kiểu liên kết đó, một thực thể của kiểu A chỉ liên kết với một thực thể của kiểu B và ngược lại, một thực thể của kiểu B chỉ liên kết với một thực thể của kiểu A. Tỷ số lực lượng 1:N có nghĩa là một thực thể của kiểu A có thể liên kết với nhiều thực thể của kiểu B nhưng một thực thể của kiểu B chỉ liên kết với một thực thể của kiểu A. Trong kiểu liên kết có tỷ số lực lượng M:N, mỗi thực thể của kiểu A có thể liên kết với nhiều thực thể của kiểu B và ngược lại, mỗi thực thể của kiểu B có

thể liên kết với nhiều thực thể của kiểu A. Trong biểu diễn của lược đồ ER, các tỷ số lực lượng được biểu diễn bằng cách ghi 1, N, M trên các hình thoi biểu diễn kiểu liên kết (hình II-5) .



Hình II-5. Tỷ số lực lượng của các kiểu liên kết



2.1.3.2. Các kiểu thực thể yếu

Các kiểu thực thể không có các thuộc tính khoá cho chính mình được gọi là các kiểu thực thể yếu. Ngược lại, các kiểu thực thể thông thường (nghĩa là có thuộc tính khoá) được gọi là kiểu thực thể chủ của một kiểu thực thể yếu được xác định bằng cách liên kết 1p.với các thực thể cụ thể của một kiểu thực thể khác phối hợp với một số giá trị thuộc tính của nó. Ta gọi kiểu thực thể khác đó là kiểu thực thể xác định hoặc *kiểu thực thể chủ*. Ta gọi kiểu liên kết giữa kiểu thực thể yếu và kiểu thực thể chủ của nó là liên kết xác định của thực thể yếu. Một kiểu thực thể yếu luôn luôn có một ràng buộc tham gia toàn bộ (tồn tại phụ thuộc) vào liên kết xác định của nó bởi vì một kiểu thực thể yếu không thể được xác định mà không có kiểu thực thể chủ. Ví dụ, trong một công ty, con của nhân viên và nhân viên có thể hưởng chế độ bảo hiểm theo nhân viên. Như vậy, sẽ có một kiểu liên kết NHÂNVIÊN <có> <CON>. Đây là một kiểu liên kết có tỷ số lực lượng 1:N. Các thuộc tính của kiểu thực thể CON là Họ tên, Ngaysinh, Giớitính. Hai người con của hai nhân viên khác nhau có thể có cùng giá trị cho các thuộc tính nhưng nó là hai thực thể khác nhau. Chúng chỉ được xác định như hai thực thể khác nhau sau khi xác định một thực thể nhân viên cụ thể có liên quan đến từng người phụ