

TUYÊN BỐ BẢN QUYỀN

Tài liệu này thuộc loại sách giáo trình nên các nguồn thông tin có thể được phép dùng nguyên bản hoặc trích dùng cho các mục đích về đào tạo và tham khảo.

Mọi mục đích khác mang tính lệch lạc hoặc sử dụng với mục đích kinh doanh thiếu lành mạnh sẽ bị nghiêm cấm.

LỜI GIỚI THIỆU

Cơ sở dữ liệu là mô đun cơ sở chuyên ngành quan trọng của tin học. Mục đích của giáo trình Cơ sở dữ liệu này là nhằm chuẩn hóa tài liệu học tập cho sinh viên các hệ cao đẳng nghề chuyên ngành Quản Trị Mạng Máy Tính, đồng thời cũng là tài liệu tham khảo đối với các chuyên ngành khác trong lĩnh vực Tin học.

Giáo trình giúp các học viên tiếp cận vấn đề cốt lõi nhất về mặt lý thuyết: các định nghĩa, khái niệm, hệ quả, định lý, giải thuật,... từ đó có thể áp dụng vào Bài toán thực tế thiết kế chuẩn hóa cơ sở dữ liệu nói riêng cũng như thiết kế hệ thống thông tin nói chung.

Giáo trình không đi sâu vào việc chứng minh các định lý mà chú trọng đến việc giải thích ý nghĩa thực tế của các công thức lý thuyết để từ đó hướng dẫn học viên cách tiếp cận tư duy logic, nắm vững kỹ thuật tính toán cũng như các bước triển khai giải quyết các Bài toán thực tế trên khía cạnh công nghệ.

Nội dung giáo trình được chia làm 4 bài:

Bài 1: Tổng quan về cơ sở dữ liệu

Bài 2: Ngôn ngữ truy vấn dữ liệu

Bài 3: Ràng buộc toàn vẹn

Bài 4: Lý thuyết thiết kế cơ sở dữ liệu

Hy vọng cuốn sách còn là tài liệu tham khảo hữu ích cho các đồng nghiệp trong giảng dạy, nghiên cứu trên các lĩnh vực có liên quan.

Giáo trình này ngoài tài liệu tham khảo chính còn có sử dụng các tài liệu tham khảo nội bộ của các đồng nghiệp khác. Rất mong nhận được nhiều ý kiến đóng góp, trao đổi của bạn đọc.

Cần Thơ, ngày 01 tháng 12 năm 2021

Tham gia biên soạn

1. Nguyễn Thị Thúy Lan

MỤC LỤC

TUYÊN BỐ BẢN QUYỀN	1
LỜI GIỚI THIỆU.....	2
MỤC LỤC.....	3
GIÁO TRÌNH MÔ ĐUN	5
Bài 1: TỔNG QUAN VỀ CƠ SỞ DỮ LIỆU	7
1. Một số khái niệm cơ bản.....	7
1.1. Định nghĩa cơ sở dữ liệu.....	7
1.2. Ưu điểm của cơ sở dữ liệu	7
1.3. Các đặc trưng của phương pháp cơ sở dữ liệu.....	7
1.4. Các đối tượng sử dụng CSDL.....	8
1.5. Hệ Quản Trị Cơ Sở Dữ Liệu (Data Base Management System).....	8
1.6. Các Ứng Dụng Của Cơ Sở Dữ Liệu	9
2. Các mô hình dữ liệu	9
3. Mô hình thực thể kết hợp.....	9
3.1. Thực Thể (entity)	9
3.2. Thuộc tính (attribute).....	10
3.3. Loại thực thể (entity type)	10
3.4. Khoá (key)	10
3.5. Mối kết hợp (relationship).....	11
4. Mô hình dữ liệu quan hệ	13
4.1. Các khái niệm cơ bản.....	13
4.2. Chuyển mô hình thực thể kết hợp sang mô hình dữ liệu quan hệ	14
5. Bài tập thực hành.....	15
5.1. Các bước thực hiện	15
5.2. Sinh viên thực hành	16
Bài 2: NGÔN NGỮ TRUY VẤN DỮ LIỆU	20
1. Mở đầu	20
2. Tìm thông tin từ các cột của bảng – Mệnh đề Select.....	22
3. Chọn các dòng của bảng – Mệnh đề Where.....	23
4. Sắp xếp các dòng của bảng – Mệnh đề Order by.....	24
5. Câu lệnh truy vấn lồng nhau	25
6. Gom nhóm dữ liệu – mệnh đề Group by.....	27
7. Bài tập thực hành.....	29
7.1. Các bước thực hiện	29
7.2. Sinh viên thực hành	29
Bài 3. RÀNG BUỘC TOÀN VỆN	31

1. Ràng buộc toàn vẹn	31
2. Phân loại ràng buộc toàn vẹn	33
3. Bài tập thực hành.....	35
3.1. Các bước thực hiện.....	36
3.2. Sinh viên thực hành.....	36
Bài 4. LÝ THUYẾT THIẾT KẾ CƠ SỞ DỮ LIỆU.....	38
1. Các vấn đề gặp phải khi tổ chức dữ liệu:	38
2. Phụ thuộc hàm	38
2.1. Định nghĩa phụ thuộc hàm	39
2.2. Cách xác định phụ thuộc hàm cho lược đồ quan hệ.....	39
2.3. Một số tính chất của phụ thuộc hàm – hệ luật dẫn Armstrong	39
3. Bao đóng của tập phụ thuộc hàm và bao đóng của tập thuộc tính	40
3.1. Bao đóng của tập phụ thuộc hàm F.....	40
3.2. Bao đóng của tập thuộc tính X.....	40
3.3. Bài toán thành viên.....	41
3.4. Thuật toán tìm bao đóng của một tập thuộc tính.....	41
4. Khóa của lược đồ quan hệ - một số thuật toán tìm khóa	42
4.1. Định nghĩa khóa của quan hệ	42
4.2. Thuật toán tìm một khóa của một lược đồ quan hệ.....	42
4.3. Thuật toán tìm tất cả các khóa của một lược đồ quan hệ	42
5. Phủ tối thiểu	43
5.1. Tập phụ thuộc hàm tương đương	43
5.2. Phủ tối thiểu	44
5.3. Thuật toán tìm phủ tối thiểu	45
6. Bài tập thực hành.....	45
6.1. Các bước thực hiện.....	45
6.2. Sinh viên thực hành.....	46
TÀI LIỆU THAM KHẢO	48

GIÁO TRÌNH MÔ ĐUN

Tên mô đun: CƠ SỞ DỮ LIỆU

Mã mô đun: MĐ10

Vị trí, tính chất, ý nghĩa và vai trò của mô đun:

- Vị trí: mô đun này được học sau môn Tin học.
- Tính chất: Cơ sở dữ liệu là mô đun cơ sở nghề bắt buộc của chương trình đào tạo Cao đẳng Quản Trị Mạng Máy Tính.
- Ý nghĩa và vai trò: Đây là mô đun cơ sở của nghề Quản Trị Mạng Máy Tính, cung cấp cho sinh viên các kiến thức tổng quan về cơ sở dữ liệu, ngôn ngữ truy vấn dữ liệu, các thuật toán để thiết kế cơ sở dữ liệu.

Mục tiêu của mô đun:

- Kiến thức:
 - + Trình bày sơ lược các khái niệm về cơ sở dữ liệu, các mô hình dữ liệu.
 - + Trình bày chi tiết mô hình thực thể kết hợp (ERD).
 - + Trình bày cụ thể về mô hình dữ liệu quan hệ.
 - + Trình bày được cách thức truy vấn của dữ liệu quan hệ, diễn hình là ngôn ngữ truy vấn SQL chuẩn.
 - + Trình bày được khái niệm, cách phân loại, các yếu tố ràng buộc toàn vẹn.
 - + Trình bày được khái niệm cơ bản của lý thuyết cơ sở dữ liệu như phụ thuộc hàm, bao đóng, khoá, phủ tối thiểu.
- Kỹ năng:
 - + Phân tích dữ liệu và thiết kế được mô hình thực thể kết hợp.
 - + Áp dụng được các quy tắc chuyển đổi từ mô hình ERD sang mô hình dữ liệu quan hệ.
 - + Thực hiện truy vấn được trên các dữ liệu đã cài đặt.
 - + Xây dựng được các ràng buộc dữ liệu trong một số bài toán cụ thể.
 - + Áp dụng được thuật toán tìm khoá, tìm phủ tối thiểu.
- Năng lực tự chủ và trách nhiệm:
 - + Rèn luyện tính cẩn thận, chính xác, sáng tạo trong quá trình tiếp thu kiến thức và vận dụng vào việc xây dựng các cơ sở dữ liệu cụ thể.
 - + Chủ động, tích cực tìm hiểu các tài liệu và nguồn bài tập liên quan.

Nội dung mô đun:

Số TT	Tên các bài trong mô đun	Thời gian (giờ)			
		Tổng số	Lý thuyết	Thực hành, thí nghiệm, thảo luận, bài tập	Kiểm tra
I	Bài 1: Tổng quan về cơ sở dữ liệu	12	4	7	1
	Một số khái niệm cơ bản	1	1		
	Các mô hình dữ liệu	1	1		
	Mô hình thực thể kết hợp	1	1		

	Mô hình dữ liệu quan hệ	1	1		
	Bài tập thực hành	7		7	
	Kiểm tra	1			1
II	Bài 2: Ngôn ngữ truy vấn dữ liệu	20	9	10	1
	Mở đầu	1	1		
	Tìm thông tin từ các cột của bảng – Mệnh đề Select	1	1		
	Chọn các dòng của bảng – Mệnh đề Where	2	2		
	Sắp xếp các dòng của bảng – Mệnh đề Order by	1	1		
	Câu lệnh truy vấn lồng nhau	2	2		
	Gom nhóm dữ liệu – mệnh đề Group by	2	2		
	Bài tập thực hành	10		10	
	Kiểm tra	1			1
III	Bài 3: Ràng buộc toàn vẹn	4	3	1	
	Ràng buộc toàn vẹn	1	1		
	Phân loại ràng buộc toàn vẹn	2	2		
	Bài tập thực hành	1		1	
IV	Bài 4: Lý thuyết thiết kế cơ sở dữ liệu	24	14	9	1
	Các vấn đề gặp phải khi tổ chức dữ liệu	1	2		
	Các phụ thuộc hàm	3	3		
	Bao đóng của tập phụ thuộc hàm và bao đóng của tập thuộc tính	3	3		
	Khóa của lược đồ quan hệ - Một số thuật toán tìm khóa	3	3		
	Phủ tối thiểu	3	3		
	Bài tập thực hành	9		9	
	Kiểm tra	1			1
	Cộng:	60	30	27	3

Bài 1: TỔNG QUAN VỀ CƠ SỞ DỮ LIỆU

Mã bài: MD10-01

Giới thiệu:

Bài này giới thiệu khái quát về các mô hình dữ liệu cơ bản, các thuật ngữ, khái niệm liên quan trong cơ sở dữ liệu. Thông qua bài học này người đọc sẽ hình dung được những vấn đề cần tiếp cận, khai thác trong mô đun cơ sở dữ liệu.

Mô hình dữ liệu quan hệ là dạng mô hình cơ sở dữ liệu cơ bản, được khai thác về mặt lý thuyết khá triệt để, triển khai về mặt ứng dụng hiệu quả nhất so với các mô hình dữ liệu khác. Trong chương này, chúng ta sẽ tìm hiểu sâu hơn một số khái niệm, thuật ngữ. Đồng thời tìm hiểu một trong những phương pháp xây dựng một mô hình dữ liệu quan hệ dựa trên các nguyên tắc chuyển đổi từ mô hình thực thể mối quan hệ.

Mục tiêu:

- Trình bày sơ lược các khái niệm về cơ sở dữ liệu, các mô hình dữ liệu.
- Trình bày chi tiết mô hình thực thể kết hợp (ERD).
- Phân tích dữ liệu và thiết kế được mô hình thực thể kết hợp.
- Trình bày cụ thể về mô hình dữ liệu quan hệ.
- Áp dụng được các quy tắc chuyển đổi từ mô hình ERD sang mô hình dữ liệu quan hệ.
- Rèn luyện tính cẩn thận, chính xác trong phân tích và giải quyết vấn đề.

Nội dung:

1. Một số khái niệm cơ bản

Mục tiêu: Trình bày sơ lược các khái niệm về cơ sở dữ liệu.

1.1. Định nghĩa cơ sở dữ liệu

Dữ liệu được lưu trữ trên các thiết bị lưu trữ theo một cấu trúc nào đó để phục vụ cho nhiều người dùng với nhiều mục đích khác nhau gọi là cơ sở dữ liệu.

1.2. Ưu điểm của cơ sở dữ liệu

- Giảm sự trùng lặp thông tin xuống mức thấp nhất và do đó bảo đảm được tính nhất quán và toàn vẹn dữ liệu.
- Đảm bảo dữ liệu có thể truy xuất theo nhiều cách khác nhau.
- Khả năng chia sẻ thông tin cho nhiều người sử dụng.

1.3. Các đặc trưng của phương pháp cơ sở dữ liệu

- Tính chia sẻ dữ liệu: dữ liệu được chia sẻ bởi nhiều người dùng hợp pháp.
- Tính giảm thiểu dư thừa dữ liệu: Dữ liệu dùng chung cho nhiều bộ phận được lưu một nơi theo cấu trúc thống nhất.
- Tính tương thích: Việc loại bỏ dư thừa kéo theo hệ quả là sự tương thích.
- Tính toàn vẹn dữ liệu: Đảm bảo một số ràng buộc toàn vẹn. Khi người dùng chèn, xóa, sửa thì ràng buộc phải được kiểm tra chặt chẽ.
- Tính bảo mật dữ liệu: Đảm bảo an toàn dữ liệu và bảo mật thông tin là quan trọng.
- Tính đồng bộ dữ liệu: Thông thường cơ sở dữ liệu được nhiều người dùng truy cập đồng thời. Cần có cơ chế bảo vệ chống sự không tương thích.

- Tính độc lập dữ liệu: Sự tách biệt cấu trúc mô tả dữ liệu khỏi chương trình ứng dụng sử dụng dữ liệu gọi là độc lập dữ liệu. Điều này cho phép phát triển tổ chức dữ liệu mà không sửa đổi chương trình ứng dụng.

1.4. Các đối tượng sử dụng CSDL

Đối tượng sử dụng là người khai thác cơ sở dữ liệu thông qua hệ quản trị CSDL. Có thể phân làm ba loại đối tượng: Người quản trị CSDL, người phát triển và lập trình ứng dụng, người dùng cuối cùng.

Người quản trị CSDL: Là người hàng ngày chịu trách nhiệm quản lý và bảo trì CSDL như:

- + Sự chính xác, toàn vẹn và bảo mật của dữ liệu và ứng dụng trong CSDL.
- + Lưu trữ dự phòng và phục hồi CSDL.
- + Giữ liên lạc với người phát triển và lập trình ứng dụng, người dùng cuối.
- + Bảo đảm sự hoạt động hiệu quả của CSDL và hệ quản trị CSDL

Người phát triển và lập trình ứng dụng: là những người chuyên nghiệp về lĩnh vực tin học có trách nhiệm thiết kế, tạo dựng và bảo trì thông tin cuối cùng cho người dùng.

Người dùng cuối là những người không chuyên trong lĩnh vực tin học, họ là các chuyên gia trong các lĩnh vực khác có trách nhiệm cụ thể trong công việc. Họ khai thác CSDL thông qua chương trình (phần mềm ứng dụng) được xây dựng bởi người phát triển ứng dụng hay công cụ truy vấn của hệ quản trị CSDL.

1.5. Hệ Quản Trị Cơ Sở Dữ Liệu (Data Base Management System)

Để giải quyết tốt những vấn đề mà cách tổ chức CSDL đặt ra như đã nói ở trên, cần thiết phải có những phần mềm chuyên dùng để khai thác chúng. Những phần mềm này được gọi là các hệ quản trị CSDL. Các hệ quản trị CSDL có nhiệm vụ hỗ trợ cho các nhà phân tích thiết kế CSDL cũng như những người khai thác CSDL. Hiện nay trên thị trường phần mềm đã có những hệ quản trị CSDL hỗ trợ được nhiều tiện ích như: MS Access, Visual Foxpro, SQL Server, Oracle, ...

Mỗi hệ quản trị CSDL đều được cài đặt dựa trên một mô hình dữ liệu cụ thể. Dù là dựa trên mô hình dữ liệu nào, một hệ quản trị CSDL cũng phải hội đủ các yếu tố sau:

- **Ngôn ngữ giao tiếp giữa người sử dụng và CSDL**, bao gồm: Ngôn ngữ mô tả dữ liệu: Để cho phép khai báo cấu trúc của CSDL, khai báo các mối liên hệ của dữ liệu và các quy tắc quản lý áp đặt lên các dữ liệu đó. Ngôn ngữ thao tác dữ liệu: Cho phép người sử dụng có thể cập nhật dữ liệu (thêm/sửa/xoá). Ngôn ngữ truy vấn dữ liệu: cho phép người khai thác sử dụng để truy vấn các thông tin cần thiết trong Cơ sở dữ liệu. Ngôn ngữ quản lý dữ liệu: cho phép những người quản trị hệ thống thay đổi cấu trúc của các bảng dữ liệu, khai báo bảo mật thông tin và cấp quyền hạn khai thác CSDL cho người sử dụng,...

- **Từ điển dữ liệu**: Dùng để mô tả các ánh xạ liên kết, ghi nhận các thành phần cấu trúc của CSDL, các chương trình ứng dụng, mật mã, quyền hạn sử dụng,...

- **Cơ chế giải quyết vấn đề tranh chấp dữ liệu**: Mỗi hệ quản trị CSDL cũng có thể cài đặt một cơ chế riêng để giải quyết các vấn đề này. Một số biện pháp sau đây thường được sử dụng: thứ nhất: cấp quyền ưu tiên cho từng người sử dụng; thứ hai: Đánh dấu yêu cầu truy xuất dữ liệu, phân chia thời gian, người nào có yêu cầu trước thì có quyền truy xuất dữ liệu trước,...

- **Hệ quản trị CSDL cũng phải có cơ chế sao lưu (backup) và phục hồi (restore) dữ liệu khi có sự cố xảy ra**. Điều này có thể thực hiện sau một thời gian nhất

định hệ quản trị CSDL sẽ tự động tạo ra một bản sao CSDL, cách này hơi tốn kém, nhất là đối với CSDL lớn.

- **Hệ quản trị CSDL phải cung cấp một giao diện thân thiện, dễ sử dụng.**

1.6. Các Ứng Dụng Của Cơ Sở Dữ Liệu

Hiện nay, hầu như CSDL gắn liền với mọi ứng dụng của tin học; chẳng hạn như việc quản lý hệ thống thông tin trong các cơ quan nhà nước, việc lưu trữ và xử lý thông tin trong các doanh nghiệp, trong các lĩnh vực nghiên cứu khoa học, trong công tác giảng dạy, cũng như trong việc tổ chức thông tin đa phương tiện,...

2. Các mô hình dữ liệu

Mục tiêu: Trình bày giới thiệu sơ lược các mô hình dữ liệu.

Mô hình dữ liệu là sự trừu tượng hoá môi trường thực. Mỗi loại mô hình dữ liệu đặc trưng cho một cách tiếp cận dữ liệu khác nhau của những nhà phân tích thiết kế CSDL. Mỗi loại mô hình dữ liệu đều có những ưu điểm và những mặt hạn chế của nó, nhưng vẫn có những mô hình dữ liệu nổi trội và được nhiều người quan tâm nghiên cứu.

Sau đây chúng ta sẽ đi qua lịch sử phát triển của các mô hình dữ liệu.

- Vào những năm sáu mươi, thế hệ đầu tiên của CSDL ra đời dưới dạng mô hình thực thể kết hợp, mô hình mạng và mô hình phân cấp.

- Vào những năm bảy mươi, thế hệ thứ hai của CSDL ra đời. Đó là mô hình dữ liệu quan hệ do E.F. Codd phát minh. Mô hình này có cấu trúc logic chặt chẽ. Đây là mô hình đã và đang được sử dụng rộng khắp trong công tác quản lý trên phạm vi toàn cầu. Việc nghiên cứu mô hình dữ liệu quan hệ nhằm vào lý thuyết chuẩn hoá các quan hệ và là một công cụ quan trọng trong việc phân tích thiết kế các hệ CSDL hiện nay. Mục đích của nghiên cứu này nhằm bỏ đi các phần tử không bình thường của quan hệ khi thực hiện các phép cập nhật, loại bỏ các phần tử dư thừa.

- Sang thập kỷ tám mươi, mô hình CSDL thứ ba ra đời, đó là mô hình cơ sở dữ liệu hướng đối tượng, mô hình cơ sở dữ liệu phân tán, mô hình cơ sở dữ liệu suy diễn,...

Trong phần tiếp theo sau đây, chúng tôi sẽ trình bày về mô hình dữ liệu tiêu biểu nhất để thiết kế (bước đầu) một ứng dụng tin học, đó là mô hình thực thể kết hợp. Trong các chương còn lại của giáo trình này chúng tôi sẽ trình bày về mô hình dữ liệu quan hệ.

3. Mô hình thực thể kết hợp

Mục tiêu: Trình bày chi tiết mô hình thực thể kết hợp (ERD), có thể phân tích dữ liệu và thiết kế được mô hình thực thể kết hợp.

Hiện nay mô hình dữ liệu quan hệ thường được dùng trong các hệ quản trị CSDL, đây là mô hình dữ liệu ở mức vật lý. Để thành lập được mô hình này, thường là phải dùng mô hình dữ liệu ở mức quan niệm để đặc tả, một trong những mô hình ở dạng đó là mô hình thực thể kết hợp (sau đó mới dùng một số quy tắc để chuyển hệ thống từ mô hình này về mô hình dữ liệu quan hệ – các quy tắc này sẽ được nói đến trong mục 2.2).

Sau đây là các khái niệm của mô hình thực thể kết hợp.

3.1. Thực Thể (entity)

Thực thể là một sự vật tồn tại và phân biệt được, chẳng hạn sinh viên Nguyễn Văn Thành, lớp Cao Đẳng Tin Học 2A, MÔ ĐUN Cơ Sở Dữ Liệu, xe máy có biển

số đăng ký 52-0549,... là các ví dụ về thực thể.

3.2. Thuộc tính (attribute)

Các đặc điểm riêng của thực thể gọi là các thuộc tính.

Chẳng hạn các thuộc tính của sinh viên Nguyễn Văn Thành là: mã số, sinh viên, giới tính, ngày sinh, hộ khẩu thường trú, lớp đang theo học,...

3.3. Loại thực thể (entity type)

Là tập hợp các thực thể có cùng thuộc tính. Mỗi loại thực thể đều phải được đặt tên sao cho có ý nghĩa. Một loại thực thể được biểu diễn bằng một hình chữ nhật.

Ví dụ các sinh viên có mã sinh viên là “02CĐTH019”, “02CĐTH519”, “02TCTH465”,... nhóm lại thành một loại thực thể, được đặt tên là Sinhvien chẳng hạn.

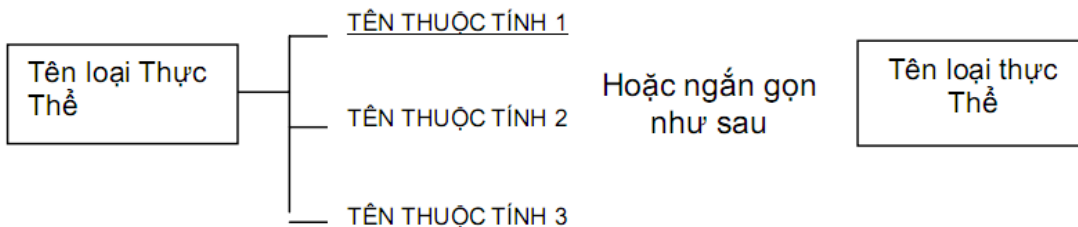
Tương tự trong ứng dụng quản lý điểm của sinh viên (sẽ được trình bày ngay sau đây) ta có các loại thực thể như Monhoc, Lop, Khoa.

3.4. Khoá (key)

Khoá của loại thực thể E là một hay một tập các thuộc tính của E có thể dùng để phân biệt hai thực thể bất kỳ của E.

Ví dụ khoá của loại thực thể Sinhvien là MASV, của Lớp là MALOP, của Khoa là MAKHOA, của Monhoc là MAMH,...

Cần chú ý rằng khi biểu diễn một hệ thống bằng mô hình thực thể kết hợp thì tên của các loại thực thể phải khác nhau. Trong danh sách các thuộc tính của một loại thực thể thì tập thuộc tính khoá thường được gạch dưới liền nét. Nếu một hệ thống có nhiều loại thực thể, để đơn giản hoá mô hình, người ta có thể chỉ nêu tên các loại thực thể; còn các thuộc tính của loại thực thể được liệt kê riêng.



Hình 1.1

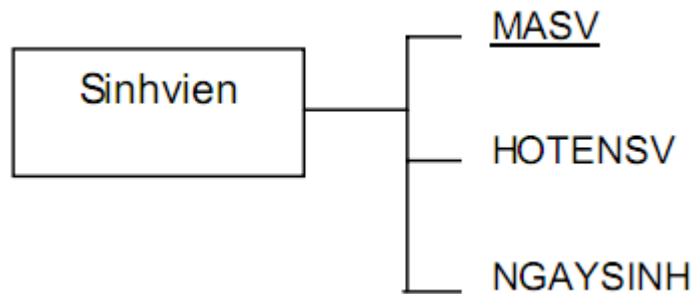
Ví dụ 1.1:

Bài toán quản lý điểm của sinh viên được phát biểu sơ bộ như sau: Mỗi sinh viên cần quản lý các thông tin như: họ và tên (HOTENSV), ngày tháng năm sinh(NGAYSINH), giới tính (NU), nơi sinh(NOISINH), hộ khẩu thường trú (TINH). Mỗi sinh viên được cấp một mã số sinh viên duy nhất (MASV) để phân biệt với mọi sinh viên khác của trường, mỗi sinh viên chỉ thuộc về một lớp nào đó. Mỗi lớp học có một mã số lớp (MALOP) duy nhất để phân biệt với tất cả các lớp học khác trong trường: có một tên gọi (TENLOP) của lớp, mỗi lớp chỉ thuộc về một khoa. Mỗi khoa có một tên gọi (TENKHOA) và một mã số duy nhất (MAKHOA) phân biệt với các khoa khác. Mỗi môn học có một tên gọi (TENMH) cụ thể, được học trong một số đơn vị học trình (DONVIHT))và ứng với môn học là một mã số duy nhất (MAMH) để phân biệt với các môn học khác. Mỗi giảng viên cần quản lý các thông tin: họ và tên(HOTENGV), cấp học vị (HOCVI), thuộc một chuyên ngành (CHUYENNGANH) và được gán cho một mã số duy nhất gọi là mã giảng viên(MAGV) để phân biệt với các giảng viên khác. Mỗi giảng viên có thể dạy nhiều môn ở nhiều khoa, nhưng chỉ thuộc về sự quản lý hành chính của một khoa. Mỗi sinh

viên với một môn học được phép thi tối đa 3 lần, mỗi lần thi (LANTHI), điểm thi (DIEMTHI). Mỗi môn học ở mỗi lớp học chỉ phân công cho một giảng viên dạy (tất nhiên là một giảng viên thì có thể dạy nhiều môn ở một lớp).

Với chương toán trên thì các loại thực thể cần quản lý như: Sinhviên, Môn học, Khoa, Lớp, Giảngviên.

Ví dụ với loại thực thể Sinhviên thì cần quản lý các thuộc tính như: MASV, HOTENSV, NGAYSINH,... và ta có thể biểu diễn như sau:



Hình 1.2

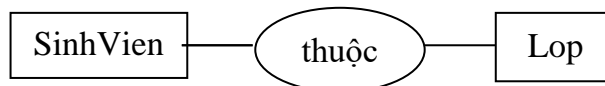
3.5. Môi kết hợp (relationship)

Môi kết hợp diễn tả sự liên hệ giữa các loại thực thể trong một ứng dụng tin học.

Ví dụ môi kết hợp giữa hai loại thực thể Sinhviên và Lop, môi kết hợp giữa Sinhviên với Môn học,...

Môi kết hợp được biểu diễn bằng một hình elip và hai bên là hai nhánh gắn kết với các loại thực thể (hoặc môi kết hợp) liên quan, tên môi kết hợp thường là: *thuộc, gồm, chứa,...*

Chẳng hạn giữa hai loại thực thể Lớp và Khoa có môi kết hợp “thuộc” như sau:



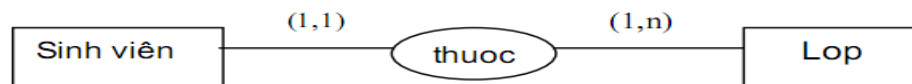
Hình 1.3

Bản số của môi kết hợp:

Bản số của một nhánh R trong môi kết hợp thể hiện số lượng các thực thể thuộc thực thể ở nhánh “bên kia” có liên hệ với một thực thể của nhánh R.

Mỗi bản số là một cặp số (min,max), chỉ số lượng tối thiểu và số lượng tối đa của thực thể khi tham gia vào môi kết hợp đó.

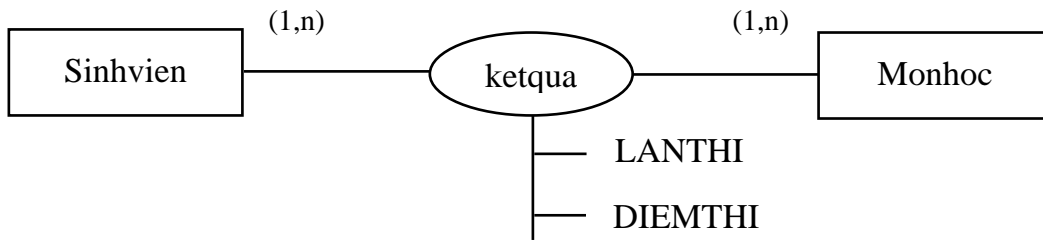
Ví dụ:



Hình 1.4

Có nghĩa là: “mỗi sinh viên thuộc một và chỉ một lớp nên bản số bên nhánh Sinhviên là (1,1), mỗi lớp có 1 đến n sinh viên nên bản số bên nhánh Lop là (1,n)”. Trong một số trường hợp đặc biệt, môi kết hợp có thể có các thuộc tính đi kèm và do đó chúng thường được đặt tên ý với nghĩa đầy đủ hơn.

Ví dụ giữa hai loại thực thể Môn học và Sinhviên có môi kết hợp ketqua với ý nghĩa: “mỗi sinh viên ứng với mỗi lần thi của mỗi môn học có một kết quả điểm thi duy nhất”.



Hình 1.5

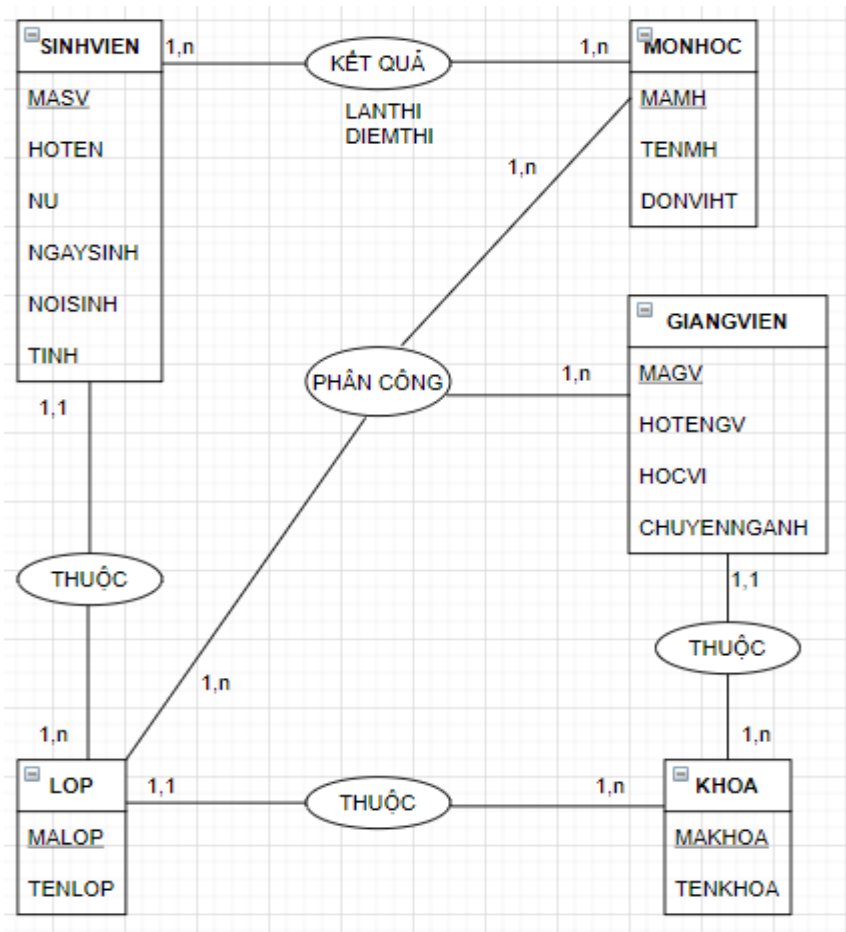
Khoá của mỗi kết hợp: là hợp của các khoá của các loại thực thể liên quan. Chẳng hạn như thuộc tính MAGV là khoá của loại thực thể Giảngvien, MALOP là thuộc tính khoá của loại thực thể Lop, MAMH là thuộc tính khoá của loại thực thể Monhoc, do đó mỗi kết hợp phancong (giữa các loại thực thể Giảngvien,Lop,Monhoc) có khoá là {MAGV,MAMH,MALOP} - phancong là mỗi kết hợp 3 ngôi.

Việc thành lập mô hình thực thể kết hợp cho một ứng dụng tin học có thể tiến hành theo các bước sau:

- b1. Xác định danh sách các loại thực thể
- b2. Xác định các mối kết hợp giữa các loại thực thể để phác thảo mô hình.
- b3. Lập bản số của các mối kết hợp.

Để kết thúc chương này, chúng tôi sẽ lập mô hình thực thể kết hợp cho Bài toán quản lý điểm của sinh viên đã được nêu trong ví dụ 1.1

Ví dụ 1.2:



Hình 1.6

4. Mô hình dữ liệu quan hệ

Mục tiêu: trình bày cụ thể các khái niệm của mô hình dữ liệu quan hệ.

4.1. Các khái niệm cơ bản

1. Thuộc tính

Thuộc tính là các đặc điểm riêng của một đối tượng (đối tượng được hiểu như là một loại thực thể ở mô hình thực thể kết hợp), mỗi thuộc tính có một tên gọi và phải thuộc về một kiểu dữ liệu nhất định.

Kiểu dữ liệu (data type): Các thuộc tính được phân biệt qua tên gọi và phải thuộc một kiểu dữ liệu nhất định (số, chuỗi, ngày tháng, logic, hình ảnh,...). Kiểu dữ liệu ở đây có thể là kiểu vô hướng hoặc là kiểu có cấu trúc. Nếu thuộc tính có kiểu dữ liệu là vô hướng thì nó được gọi là thuộc tính đơn hay thuộc tính nguyên tố, nếu thuộc tính có kiểu dữ liệu có cấu trúc thì ta nói rằng nó không phải là thuộc tính nguyên tố.

Chẳng hạn với sinh viên Nguyễn Văn Thành thì các thuộc tính họ và tên, mã số sinh viên thuộc kiểu chuỗi, thuộc tính ngày sinh thuộc kiểu ngày tháng, họ khẩu thường trú thuộc kiểu chuỗi, thuộc tính hình ảnh thuộc kiểu hình ảnh,...

Miền giá trị (domain of values): Thông thường mỗi thuộc tính chỉ chọn lấy giá trị trong một tập con của kiểu dữ liệu và tập hợp con đó gọi là miền giá trị của thuộc tính đó. Chẳng hạn thuộc tính NỮ có miền giá trị là {nam,nữ}, thuộc tính màu da có miền giá trị là {da trắng, da vàng, da đen, da đỏ}, thuộc tính điểm thi là các số thuộc tập {0; 1; 2;...,10}.

Lưu ý rằng nếu không lưu ý đến ngữ nghĩa thì tên của các thuộc tính thường được ký hiệu bằng các chữ cái in hoa đầu tiên trong bảng chữ cái la tinh: A,B,C,D,... Những chữ cái in hoa X,Y,Z,W,... thường dùng thay cho một nhóm nhiều thuộc tính. Đôi khi còn dùng các ký hiệu chữ cái với các chỉ số A_1, A_2, \dots, A_n để chỉ các thuộc tính trong trường hợp tổng quát hay muốn đề cập đến số lượng các thuộc tính. Tên thuộc tính phải được đặt một cách gọi nhớ, không nên đặt tên thuộc tính quá dài (vì như thế sẽ làm cho việc viết các câu lệnh truy vấn trở nên vất vả hơn), nhưng cũng không nên đặt tên thuộc tính quá ngắn (vì nó sẽ không cho thấy ngữ nghĩa của thuộc tính), đặc biệt không đặt trùng tên hai thuộc tính mang ngữ nghĩa khác nhau thuộc hai đối tượng khác nhau.

Trong nhiều hệ quản trị cơ sở dữ liệu, người ta thường đưa thêm vào miền giá trị của các thuộc tính một giá trị đặc biệt gọi là giá trị rỗng (NULL). Tuỳ theo ngữ cảnh mà giá trị này có thể đặc trưng cho một giá trị không thể xác định được hoặc một giá trị chưa được xác định ở vào thời điểm nhập tin nhưng có thể được xác định vào một thời điểm khác.

2. Lược Đồ Quan Hệ (relation schema)

Tập tất cả các thuộc tính cần quản lý của một đối tượng cùng với các mối liên hệ giữa chúng được gọi là lược đồ quan hệ. Lược đồ quan hệ Q với tập thuộc tính $\{A_1, A_2, \dots, A_n\}$ được viết là $Q(A_1, A_2, \dots, A_n)$,

ký hiệu: $Q^+ = \{A_1, A_2, \dots, A_n\}$.

Chẳng hạn lược đồ quan hệ Sinhviên với các thuộc tính như đã được liệt kê trong ví dụ 1.1 được viết như sau:

Sinhvien(MASV,HOTENSV,PHAI,NGAYSINH,NOISINH,TINH,MALOP)

Thường thì khi thành lập một lược đồ quan hệ, người thiết kế gắn cho nó một ý nghĩa nhất định, gọi là tân từ của lược đồ quan hệ. Chẳng hạn tân từ của lược đồ quan hệ Sinhvien là: “Mỗi sinh viên có mỗi MASV duy nhất. Mỗi MASV xác định các thuộc tính còn lại của sinh viên đó như HOTENSV, PHAI, NGAYSINH, NOISINH, TINH, MALOP”.

Khi phát biểu tân từ cho một lược đồ quan hệ, người thiết kế cần phải mô tả đầy đủ ý nghĩa để người khác tránh hiểu nhầm. Dựa vào tân từ này, người ta xác định được tập khoá, siêu khoá của lược đồ quan hệ (sẽ được trình bày trong những mục kế tiếp).

Nhiều lược đồ quan hệ cùng nằm trong một hệ thống thông tin được gọi là một lược đồ cơ sở dữ liệu.

Khái niệm lược đồ quan hệ ứng với khái niệm loại thực thể ở mô hình thực thể kết hợp.

3. Quan Hệ (relation)

Sự thể hiện của lược đồ quan hệ ở một thời điểm nào đó được gọi là quan hệ, rõ ràng là trên một lược đồ quan hệ có thể xác định nhiều quan hệ. Thường ta dùng các ký hiệu như R, S, Q để chỉ các lược đồ quan hệ, còn quan hệ thường được dùng bởi các ký hiệu là r, s, q, ...

Về trục quan thì quan hệ (hay bảng quan hệ) như là một bảng hai chiều gồm các dòng và các cột.

Một quan hệ có n thuộc tính được gọi là quan hệ n ngôi.

Để chỉ quan hệ r xác định trên lược đồ quan hệ Q ta có thể viết r(Q).

4. Bộ (Tuple)

Mỗi bộ là những thông tin về một đối tượng thuộc một quan hệ, bộ cũng còn được gọi là mẫu tin.

Thường người ta dùng các chữ cái thường (như t, ...) để biểu diễn bộ trong quan hệ.

4.2. Chuyển mô hình thực thể kết hợp sang mô hình dữ liệu quan hệ

Mục tiêu: Áp dụng được các quy tắc chuyển đổi mô hình thực thể kết hợp ERD sang mô hình dữ liệu quan hệ (dạng lược đồ).

Quy tắc 1:

Chuyển đổi mỗi loại thực thể thành một lược đồ quan hệ, các thuộc tính của loại thực thể thành các thuộc tính của lược đồ quan hệ, thuộc tính khoá của loại thực thể là thuộc tính khoá của lược đồ quan hệ.

Chẳng hạn loại thực thể Sinhvien ở ví dụ 1.2 khi áp dụng quy tắc 1 thì sẽ được chuyển thành lược đồ quan hệ Sinhvien như sau:

Sinhvien(MASV, HOTENSV, NU, NGAYSINH, TINH, ...)

Quy tắc 2:

Nếu mỗi kết hợp mà cả hai nhánh của nó đều có bản số max là n thì mỗi kết hợp này sẽ được chuyển thành một lược đồ quan hệ K' gồm các thuộc tính của mỗi kết hợp K, cộng thêm các thuộc tính khoá của hai lược đồ quan hệ A, B tương ứng với hai thực thể tham gia vào mỗi kết hợp. Khoá của lược đồ quan hệ K' gồm cả hai

khoá của hai lược đồ quan hệ A và B.

Chẳng hạn mỗi kết hợp Phancong giữa ba loại thực thể Giangvien, Monhoc và Lop được chuyển thành lược đồ quan hệ Phancong và có tập khoá là {MAGV,MAMH,MALOP} như sau:
Phancong(MAGV,MAMH,MALOP)

Quy tắc 3:

Mỗi kết hợp mà một nhánh có bản số là n (nhánh B) và nhánh còn lại có bản số max là 1 (nhánh A) thì loại bỏ mỗi kết hợp này khỏi mô hình thực thể kết hợp và thêm các thuộc tính khoá của lược đồ tương ứng với loại thực thể ở nhánh B vào lược đồ tương ứng với loại thực thể ở nhánh A (khoá của B sẽ thành khoá ngoại của A). Nếu mỗi kết hợp có các thuộc tính thì những thuộc tính này cũng được thêm vào lược đồ quan hệ tương ứng với loại thực thể ở nhánh A.

Chẳng hạn mỗi kết hợp *thuộc* giữa hai loại thực thể Sinhvien và Lop nên lược đồ quan hệ Sinhvien được sửa thành như sau:
Sinhvien(MASV,HOTENSV,PHAI,NGAYSINH, TINH,MALOP)

Quy tắc 4:

Nếu mỗi kết hợp mà cả hai nhánh đều có bản số max là 1 thì áp dụng quy tắc 3 cho một trong hai nhánh tùy chọn.

Ví dụ:

Sau đây là mô hình dữ liệu quan hệ được chuyển từ mô hình thực thể kết hợp ở ví dụ 1.2.

SINHVIENT(MASV, HOTENSV, PHAI, NGAYSINH, NOISINH,TINH, MALOP)

LOP(MALOP, TENLOP, MAKHOA)

KHOA(MAKHOA, TENKHOA)

MONHOC(MAMH, TENMH, DONVIHT)

GIANGVIEN(MAGV, HOTENGV, HOCVI, CHUYENNGANH, MAKHOA)

KETQUA(MASV, MAMH, LANTHI1, LANTHI2, DIEMTHI)

PHANCONG(MALOP,MAMH,MAGV)

5. Bài tập thực hành

Dựa vào các phân tích sơ bộ dưới đây, hãy lập mô hình thực thể kết hợp (gồm loại thực thể, mỗi kết hợp, bản số, thuộc tính của loại thực thể, khoá của loại thực thể) cho Bài toán quản lý sau:

QUẢN LÝ ĐẶT BÁO CỦA KHÁCH HÀNG

Tại một nơi phát hành báo cần phải quản lý thông tin đặt báo của khách hàng. Có nhiều thể loại báo (nhật báo, nguyệt sang, tạp chí, tham luận,...), mỗi thể loại cần lưu lại mã thể loại, tên thể loại. Mỗi tờ báo cần lưu lại mã báo là gì, tên báo (kiến thức ngày nay, tuổi trẻ ngày, tuổi trẻ tuần, Sài Gòn giải phóng,...) hình thức (quý, ngày, tuần), đơn giá kỳ. Mỗi một thể loại báo có nhiều tờ báo. Khách hàng đến đặt báo cần lưu lại thông tin: mã khách hàng, tên khách hàng, ngày đặt. Một khách hàng có thể đặt nhiều tờ báo khác nhau. Cần lưu lại: tờ báo mà khách muốn đặt, số kỳ, thành tiền.

5.1. Các bước thực hiện

❖ Xây dựng mô hình thực thể kết hợp

Bước 1: Xác định danh sách các loại thực thể

Bước 2: Xác định các mối kết hợp giữa các loại thực thể để phác thảo mô hình.

Bước 3: Lập bản số của các mối kết hợp

❖ **Chuyển từ mô hình thực thể kết hợp sang mô hình dữ liệu quan hệ**

Quy tắc 1:

Chuyển đổi mỗi loại thực thể thành một lược đồ quan hệ, các thuộc tính của loại thực thể thành các thuộc tính của lược đồ quan hệ, thuộc tính khoá của loại thực thể là thuộc tính khoá của lược đồ quan hệ.

Quy tắc 2:

Nếu mỗi kết hợp mà cả hai nhánh của nó đều có bản số max là n thì mỗi kết hợp này sẽ được chuyển thành một lược đồ quan hệ K' gồm các thuộc tính của mỗi kết hợp K, cộng thêm các thuộc tính khoá của hai lược đồ quan hệ A, B tương ứng với hai thực thể tham gia vào mỗi kết hợp. Khoá của lược đồ quan hệ K' gồm cả hai khoá của hai lược đồ quan hệ A và B.

Quy tắc 3:

Mỗi kết hợp mà một nhánh có bản số là n (nhánh B) và nhánh còn lại có bản số max là 1 (nhánh A) thì loại bỏ mỗi kết hợp này khỏi mô hình thực thể kết hợp và thêm các thuộc tính khoá của lược đồ tương ứng với loại thực thể ở nhánh B vào lược đồ tương ứng với loại thực thể ở nhánh A (khoá của B sẽ thành khoá ngoại của A). Nếu mỗi kết hợp có các thuộc tính thì những thuộc tính này cũng được thêm vào lược đồ quan hệ tương ứng với loại thực thể ở nhánh A.

Quy tắc 4:

Nếu mỗi kết hợp mà cả hai nhánh đều có bản số max là 1 thì áp dụng quy tắc 3 cho một trong hai nhánh tuỳ chọn.

5.2. Sinh viên thực hành

❖ **Xây dựng mô hình thực thể kết hợp**

Bước 1: Xác định danh sách các loại thực thể

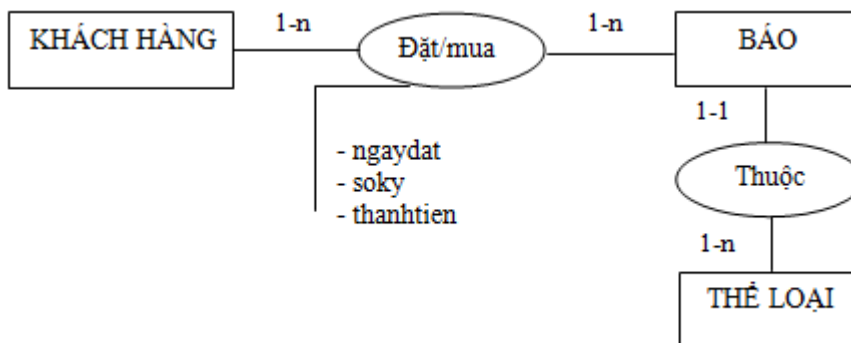
- Thực thể **BÁO**: có các thuộc tính: mã báo (mabao) là thuộc tính định danh; tên báo (tenbao), hình thức (hinhthuc), đơn giá kỳ (dongiak).y).
- Thực thể **THẺ LOẠI**: có các thuộc tính: mã thẻ loại (matheloai) là thuộc tính định danh, tên thẻ loại (tentheloai).
- Thực thể **KHÁCH HÀNG**: có các thuộc tính: mã khách hàng (makhachhang), tên khách hàng (tenkhachhang).

Bước 2: Xác định các mối kết hợp giữa các loại thực thể để phác thảo mô hình.

- **BÁO** thuộc **THẺ LOẠI**.

- **KHÁCH HÀNG** đặt/mua **BÁO**: có các thuộc tính riêng phát sinh trong quá trình đặt.mua là: ngày đặt (ngaydat), số kỳ (soky), thành tiền (thanhtien).

Phác thảo mô hình:



Bước 3: Lập bản số của các mối kết hợp: gắn các quan hệ sau lên mô hình phác thảo,

ta có được mô hình thực thể mối quan hệ sơ bộ cuối cùng.

- Quan hệ Đặt/mua giữa KHÁCH HÀNG và BÁO là quan hệ (1-n)-(1-n)
- Quan hệ Thuộc giữa BÁO và THỂ LOẠI là quan hệ (1-1)-(1-n).

❖ **Chuyển từ mô hình thực thể kết hợp sang mô hình dữ liệu quan hệ**

Áp dụng các quy tắc chuyển đổi từ E-R sang mô hình dữ liệu quan hệ, ta có:

- Chuyển đổi mỗi loại thực thể thành một lược đồ quan hệ (quy tắc 1).

Ta có các lược đồ quan hệ sau:

R₁: KHÁCH HÀNG (makhachhang, tenkhachhang)

R₂: BÁO (mabao, tenbao, dongiak, hinhthuc)

R₃: THỂ LOẠI (matheloai, tentheloai)

- Chuyển các quan hệ và bản số quan hệ:

+ Quan hệ **Đặt/mua** là quan hệ nhiều-nhiều có thuộc tính riêng, áp dụng quy tắc 2, ta có được lược đồ quan hệ sau:

R₄: ĐẶT BÁO (makhachhang, mabao, ngaydat, soky, thanhtien)

+ Quan hệ **Thuộc** là quan hệ một – nhiều (một cho nhánh THỂ LOẠI), áp dụng quy tắc 3, ta có lược đồ quan hệ sau:

R₅: BÁO (mabao, tenbao, dongiak, hinhthuc, matheloai)

Những trọng tâm cần chú ý trong bài

- Khi xây dựng mô hình thực thể kết hợp: áp dụng đúng thứ tự các bước 1, 2, 3
- Khi chuyển từ mô hình thực thể kết hợp sang mô hình dữ liệu quan hệ: áp dụng đúng quy tắc 1, 2, 3, 4

Bài tập mở rộng và nâng cao

Bài 1. QUẢN LÝ VIỆC MƯỢN/TRẢ SÁCH Ở MỘT THƯ VIỆN

Một thư viện tổ chức việc cho mượn sách như sau: Mỗi quyển sách được đánh một mã sách (MASACH) dùng để phân biệt với các quyển sách khác (giả sử nếu một tác phẩm có nhiều bản giống nhau hoặc có nhiều tập thì cũng xem là có mã sách khác nhau), mỗi mã sách xác định các thông tin khác như: tên sách (TENSACH), tên tác giả (TACGIA), nhà xuất bản (NHAXB), năm xuất bản (NAMXB). Mỗi độc giả được thư viện cấp cho một thẻ thư viện, trong đó có ghi rõ mã độc giả (MADG), cùng với các thông tin khác như: họ tên (HOTEN), ngày sinh (NGAYSINH), địa chỉ (DIACHI), nghề nghiệp (NGHENGHIEP).

Cứ mỗi lượt mượn sách, độc giả phải đăng ký các quyển sách cần mượn vào một phiếu mượn, mỗi phiếu mượn có một số phiếu mượn (SOPM) khác nhau, mỗi phiếu mượn xác định các thông tin như: ngày mượn sách (NGAYMUON), mã độc giả. Các các quyển sách trong cùng một phiếu mượn không nhất thiết phải trả trong một lần. Mỗi quyển sách có thể thuộc nhiều phiếu mượn khác nhau (tất nhiên là tại các thời điểm khác nhau).

Bài 2. QUẢN LÝ LỊCH DẠY CỦA GIÁO VIÊN

Để quản lý lịch dạy của các giáo viên và lịch học của các lớp, một trường tổ chức như sau: Mỗi giáo viên có một mã số giáo viên (MAGV) duy nhất, mỗi MAGV xác định các thông tin như: họ và tên giáo viên (HOTEN), số điện thoại (DTGV). Mỗi giáo viên có thể dạy nhiều môn cho nhiều khoa nhưng chỉ thuộc sự quản lý hành chính của một khoa nào đó. Mỗi môn học có một mã số môn học (MAMH) duy nhất, mỗi môn học xác

định tên môn học (TENMH). Ứng với mỗi lớp thì mỗi môn học chỉ được phân cho một giáo viên. Mỗi phòng học có một số phòng học (PHONG) duy nhất, mỗi phòng có một chức năng (CHUCNANG); chẳng hạn như phòng lý thuyết, phòng thực hành máy tính, phòng nghe nhìn, xưởng thực tập cơ khí,... Mỗi khoa có một mã khoa (MAKHOA) duy nhất, mỗi khoa xác định các thông tin như: tên khoa (TENKHOA), điện thoại khoa(DTKHOA). Mỗi lớp có một mã lớp (MALOP) duy nhất, mỗi lớp có một tên lớp (TENLOP), sĩ số lớp (SISO). Mỗi lớp có thể học nhiều môn của nhiều khoa nhưng chỉ thuộc sự quản lý hành chính của một khoa nào đó. Hàng tuần, mỗi giáo viên phải lập lịch báo giảng cho biết giáo viên đó sẽ dạy những lớp nào, ngày nào (NGAYDAY), môn gì?, tại phòng nào, từ tiết nào (TUTIET) đến tiết nào (ĐENTIET), tựa đề chương dạy (BAIDAY), những ghi chú (GHICHU) về các tiết dạy này, đây là giờ dạy lý thuyết (LYTHUYET) hay thực hành - giả sử nếu LYTHUYET=1 thì đó là giờ dạy thực hành và nếu LYTHUYET=2 thì đó là giờ lý thuyết, một ngày có 16 tiết, sáng từ tiết 1 đến tiết 6, chiều từ tiết 7 đến tiết 12, tối từ tiết 13 đến 16. Một số yêu cầu của hệ thống này như:: Lập lịch dạy trong tuần của các giáo viên. Tổng số dạy của các giáo viên theo từng môn cho từng lớp,

Bài 3. QUẢN LÝ HỌC VIÊN Ở MỘT TRUNG TÂM TIN HỌC

Trung tâm tin học KTCT thường xuyên mở các lớp tin học ngắn hạn và dài hạn. Mỗi lớp ngắn hạn có một hoặc nhiều MÔ ĐUN (chẳng hạn như lớp Tin học văn phòng thì có các môn : Word, Power Point, Excel, còn lớp lập trình Pascal thì chỉ học một môn Pascal). Các lớp dài hạn (chẳng hạn như lớp kỹ thuật viên đồ họa đa truyền thông, lớp kỹ thuật viên lập trình, lớp kỹ thuật viên phần cứng và mạng,) thì có thể học nhiều học phần và mỗi học phần có thể có nhiều MÔ ĐUN. Mỗi học viên có một mã học viên(MAHV) duy nhất và chỉ thuộc về một lớp duy nhất (nếu học viên cùng lúc học nhiều lớp thì ứng với mỗi lớp, học viên đó có một MAHV khác nhau). Mỗi học viên xác định họ tên (HOTEN), ngày sinh (NGAYSINH), nơi sinh (NOISINH), phái nam hay nữ (PHAI), nghề nghiệp (NGHENGHIEP) - nghề nghiệp là SINH VIÊN, GIÁO VIÊN, KỸ SU, , BUÔN BÁN,... Trung tâm KTCT có nhiều lớp, mỗi lớp có một mã lớp duy nhất (MALOP), mỗi lớp xác định các thông tin: tên lớp (TENLOP), thời khoá biểu, ngày khai giảng (NGAYKG), học phí (HOCPhi).

Chú ý rằng tại một thời điểm, trung tâm có thể mở nhiều lớp cho cùng một chương trình học. Với các lớp dài hạn thì ngày khai giảng được xem là ngày bắt đầu của mỗi học phần và HOC PHÍ là học phí của mỗi học phần, với lớp ngắn hạn thì HOC PHÍ là học phí của toàn khoá học đó.

Trung tâm có nhiều môn học, mỗi môn học có mã môn học (MAMH) duy nhất, mỗi môn học xác định tên môn học(TENMH), số tiết lý thuyết (SOTIETLT), số tiết thực hành (SOTIETTH). Mỗi học viên ứng với mỗi môn học có một điểm thi(DIEMTHI) duy nhất. Mỗi lần đóng học phí, học viên sẽ được trung tâm giao cho một phiếu biên lai thu tiền, mỗi biên lai có một số biên lai duy nhất để quản lý. Một số yêu cầu của hệ thống này như: Lập danh sách những học viên khai giảng khoá ngày nào đó. Lập danh sách các học viên của một lớp ? Cho biết số lượng học viên của mỗi lớp khai giảng khoá ngày nào đó?

Bài 4. QUẢN LÝ COI THI TUYỂN SINH

Một hội đồng coi thi tuyển sinh có nhiều điểm thi, mỗi điểm thi được đặt tại một trường nào đó. Các điểm thi (DIEMTHISO) được đánh số là điểm thi số 1, điểm thi số

2, điểm thi số 3,... Mỗi điểm thi xác định địa chỉ (DIACHIDIEMTHI). Ví dụ: điểm thi số 1, đặt tại trường PTTH Nguyễn Thị Minh Khai, điểm thi số 2 đặt tại trường PTTH Bùi Thị Xuân,... Mỗi thí sinh có một số báo danh (SOBD) duy nhất, mỗi số báo danh xác định các thông tin: họ và tên (HOTEN), ngày sinh (NGAYSINH), phái (PHAI), hộ khẩu thường trú (TINH), đối tượng dự thi (DOITUONG), ngành đăng ký thi, khu vực của thí sinh (KHUVUC), số hiệu phòng thi. Ví dụ: thí sinh Vũ Mạnh Cường, có số báo danh là 02978, sinh ngày 12/12/1984, phái nam, hộ khẩu thường trú tại Chợ Gạo - Tiền Giang, thuộc khu vực 1, đối tượng là 5B, đăng ký dự thi vào ngành có mã ngành là 01, thi tại phòng thi 0178, điểm thi số 1. Mỗi ngành có một mã ngành (MANGANH) duy nhất, mỗi mã ngành xác định tên ngành (TENGANH). Mỗi điểm thi có nhiều phòng thi – mỗi phòng thi (PHONGTHI) được đánh số khác nhau ở tất cả các điểm thi. Trong một phòng thi, danh sách các thí sinh được sắp xếp theo thứ tự alphabet (do đó trong một phòng thi có thể có thí sinh của nhiều ngành khác nhau). Mỗi phòng thi có thêm cột ghi chú (GHICHU) - ghi thêm các thông tin cần thiết như phòng thi đó nằm tại dãy nhà nào. Ví dụ phòng thi 0060 nằm ở dãy nhà H lầu 2 - điểm thi số 1 - trường PTTH Bùi Thị Xuân. Mỗi môn thi có một mã môn thi duy nhất (MAMT), mỗi mã môn thi biết các thông tin như : tên môn thi (TENMT), ngày thi (NGAYTHI), buổi thi (BUOITHI), thời gian làm chương thi được tính bằng phút (PHUT). Thời gian làm chương thi của các môn tối thiểu là 90 phút và tối đa là 180 phút (tuỳ theo kỳ tuyển sinh công nhân, trung cấp, cao đẳng hay đại học). Mỗi ngành có một mã ngành, chẳng hạn ngành Công Nghệ Thông Tin có mã ngành là 1, ngành Công Nghệ Hoá Thực Phẩm có mã ngành là 10,...

Mỗi đơn vị có cán bộ tham gia vào kỳ thi có một mã đơn vị duy nhất (MADONVI), mã đơn vị xác định tên đơn vị (TENDONVI). Nếu là cán bộ, công nhân viên của trường thì đơn vị là khoa/phòng quản lý cán bộ đó, nếu là giáo viên từ các trường khác thì ghi rõ tên đơn vị đó. Chẳng hạn cán bộ Nguyễn Thanh Liêm đơn vị Khoa Công Nghệ Thông Tin, cán bộ coi thi Nguyễn Thị Tuyết Mai, đơn vị trường PTTH Ngôi Sao - Quận 1,... Mỗi cán bộ coi thi chỉ làm việc tại một điểm thi nào đó. Mỗi cán bộ có một mã số duy nhất (MACANBO), mỗi MACANBO xác định các thông tin khác như : họ và tên (HOTENCB), đơn vị công tác, chức vụ (CHUCVU) được phân công tại điểm thi, chẳng hạn chức vụ là điểm trưởng, điểm phó, giám sát, thư ký, cán bộ coi thi, phục vụ,... Ví dụ cán bộ Nguyen Van Thanh đơn vị Khoa Công Nghệ Thông Tin, làm nhiệm vụ thi tại điểm thi số 1, chức vụ là giám sát phòng thi.

Yêu cầu về đánh giá kết quả học tập bài 1

Nội dung:

- + Về kiến thức: Trình bày sơ lược các khái niệm về cơ sở dữ liệu, các mô hình dữ liệu chi tiết mô hình thực thể kết hợp, mô hình dữ liệu quan hệ.
- + Về kỹ năng: Phân tích dữ liệu và thiết kế được mô hình thực thể kết hợp. Áp dụng được các quy tắc chuyển đổi từ mô hình ERD sang mô hình dữ liệu quan hệ.
- + Năng lực tự chủ và trách nhiệm: Rèn luyện tính cẩn thận, chính xác trong phân tích và giải quyết vấn đề.

Phương pháp:

- + Về kiến thức: Được đánh giá bằng hình thức kiểm tra viết, trắc nghiệm, vấn đáp
- + Về kỹ năng: Đánh giá kỹ năng thực hành thiết kế được mô hình thực thể kết hợp. Áp dụng được các quy tắc chuyển đổi từ mô hình ERD sang mô hình dữ liệu quan hệ.
- + Năng lực tự chủ và trách nhiệm: Rèn luyện tính cẩn thận, chính xác trong phân tích và giải quyết vấn đề.

Bài 2: NGÔN NGỮ TRUY VẤN DỮ LIỆU

Mã bài: MD10-02

Giới thiệu:

SQL(STRUCTURE QUERY LANGUAGE) là ngôn ngữ truy vấn dựa trên đại số quan hệ. Ngôn ngữ truy vấn SQL có tập lệnh khá phong phú dùng để thao tác trên cơ sở dữ liệu.

Trong bài này, chúng ta chỉ nghiên cứu về câu lệnh quan trọng nhất của SQL đó là câu lệnh truy vấn dữ liệu SELECT. Kết quả của lệnh select là một quan hệ, quan hệ kết quả này có thể kết xuất ra màn hình, máy in, hoặc là trên các thiết bị lưu trữ thông tin khác.

Mục tiêu:

- Trình bày được cách thức truy vấn của dữ liệu quan hệ, điển hình là ngôn ngữ truy vấn SQL chuẩn.
- Thực hiện truy vấn được trên các dữ liệu đã cài đặt.
- Rèn luyện tính cẩn thận, chính xác trong việc học và làm bài tập.
- Thực hiện các thao tác an toàn với máy tính.

Nội dung:

1. Mở đầu

Mục tiêu:

- Giới thiệu về lịch sử ngôn ngữ SQL, mục đích, lợi ích của ngôn ngữ SQL.
- Trình bày tóm tắt các kiểu dữ liệu cơ bản, các hàm xử lý dữ liệu cơ bản của SQL.

SQL.

- Trình bày cú pháp tổng quát của câu lệnh Select – câu lệnh truy vấn dữ liệu chuẩn của SQL.

Vào những năm 1970, SQL(Structure Query Language) lần đầu được hãng IBM phát triển như một bộ phận của hệ quản trị CSDL mô hình quan hệ có tên là SYSTEM R. Sau đó vào các năm 1980 IBM tiếp tục phát triển SQL cho các hệ quản trị cơ sở dữ liệu nổi tiếng là SQL/DS trên nền hệ điều hành VM, DB2 trên nền hệ điều hành MVS, Hệ quản trị cơ sở dữ liệu mở rộng trên nền hệ điều hành IBM OS/2, Hệ quản trị cơ sở dữ liệu cho hệ thống IBM AS/400. Năm 1986, Viện tiêu chuẩn quốc gia Hoa kỳ (ANSI – American National Standards Institute) và Tổ chức Tiêu chuẩn Quốc tế (ISO – International Standards Organization) đã thừa nhận SQL như là *ngôn ngữ chuẩn xử lý dữ liệu*. Ngôn ngữ chuẩn ANSI SQL tiếp tục được cập nhật vào những năm 1989 và 1992 cho đến hôm nay.

SQL được cài đặt cho hệ thống máy tính lớn (mainframe) cũng như máy tính cá nhân. Bên cạnh các sản phẩm của hãng IBM, cũng cần phải kể đến các hệ quản trị cơ sở dữ liệu nổi tiếng khác như ORACLE của Công ty Oracle, SQL Server của hãng Microsoft, SQLBase của hãng Sybase, Ingres của hãng Relational Technologies, ...

Mục đích của chuẩn SQL là

- Xác định cú pháp và ngữ nghĩa của ngôn ngữ SQL định nghĩa và thao tác dữ liệu.
- Định nghĩa các cấu trúc dữ liệu và phép toán cơ bản để thiết kế, truy cập, lưu trữ, kiểm soát và bảo vệ cơ sở dữ liệu SQL.
- Cung cấp công cụ đảm bảo tính tương thích của cấu trúc dữ liệu và các modul ứng dụng giữa các hệ quản trị cơ sở dữ liệu.

- Xác định chuẩn tối thiểu và chuẩn hoàn chỉnh, cho phép các cấp độ sử dụng SQL khác nhau trong các sản phẩm.

- Cung cấp chuẩn ban đầu, có thể chưa hoàn chỉnh, cho phép mở rộng các chức năng xử lý những vấn đề như sự toàn vẹn tham chiếu, giao thức chuyển đổi, các hàm người dùng, các toán tử nối ngoài phép đẳng nối, và các hệ thống ký tự quốc gia,...

Một ngôn ngữ quan hệ chuẩn như SQL sẽ mang lại các lợi ích sau đây:

- Giảm thiểu chi phí đào tạo. Các hệ quản trị cơ sở dữ liệu có chung ngôn ngữ chuẩn SQL sẽ làm giảm chi phí chuyển đổi từ hệ này sang hệ khác.

- Nâng cao hiệu năng công việc. Các chuyên gia hệ thống thông tin với kiến thức sâu sắc về SQL sẽ nhanh chóng nắm bắt các chương trình ứng dụng của các hệ quản trị cơ sở dữ liệu, vì họ đã quen thuộc với ngôn ngữ của các chương trình này.

- Tính khả chuyển của các ứng dụng. Các ứng dụng có thể dễ dàng sử dụng trên các hệ thống khác nhau, nhưng cùng sử dụng SQL.

- Tăng tuổi thọ của các ứng dụng. Một ngôn ngữ chuẩn có xu hướng tồn tại thời gian dài, điều đó làm giảm áp lực viết lại chương trình.

- Làm giảm sự phụ thuộc vào nhà cung cấp. Vì SQL là ngôn ngữ chung nên người dùng dễ dàng sử dụng nhiều sản phẩm của các nhà cung cấp khác nhau, với giá cả cạnh tranh.

- Khả năng giao tiếp giữa các hệ thống chéo. Các hệ quản trị cơ sở dữ liệu quan hệ và các chương trình ứng dụng khác nhau có thể dễ dàng giao tiếp và hợp tác để xử lý dữ liệu và thực hiện chương trình người dùng.

SQL là ngôn ngữ có cấu trúc

Trong câu lệnh của SQL có một số mệnh đề tuân theo những cú pháp riêng của nó. Có 4 loại lệnh trong SQL:

- Các lệnh truy vấn dữ liệu (Data Query).
- Các lệnh định nghĩa dữ liệu (Data Definition Language).
- Các lệnh truy xuất và cập nhật dữ liệu (Data Manipulation Language).
- Các lệnh điều khiển quyền truy cập dữ liệu (Data Control Language).

Kiểu dữ liệu

SQL có các kiểu dữ liệu sau:

INTEGER : Kiểu số nguyên 2 byte -2 147 483 648 → 2 147 483 64 .

SMALLINT : Kiểu số nguyên 1 byte - 32 768 → 32 767.

DECIMAL(n,d) : Kiểu số thực độ dài n (kể cả dấu), số chữ số thập phân d

FLOAT(n,d) : Kiểu số thực khoa học độ dài n , số chữ số thập phân d

CHAR(n) : Kiểu chuỗi ký tự độ dài n

DATE : Kiểu ngày tháng (ngày/tháng/năm)

LOGICAL : Kiểu lôgic, nhận giá trị *true* hoặc *false*

VARCHAR(n) : Kiểu chuỗi ký tự độ dài thay đổi, tối đa n ký tự

LONGVARCHAR : Kiểu chuỗi ký tự độ dài thay đổi (chứa dữ liệu như *ghi chú*,...)

Và các kiểu khác như: Kiểu tiền tệ, Kiểu hình ảnh, Kiểu âm thanh...

Cú pháp tổng quát:

```
SELECT [DISTINCT] <biểu thức 1> [AS <tên 1>] [...] | *
FROM <bảng 1> [<bí danh 1>] [...]
[INTO <dbf đích>]
[WHERE <điều kiện nối > [AND | OR <điều kiện lọc>]]
[GROUP BY <cột nhóm 1> [...]]
[HAVING <điều kiện nhóm>]]
```

[ORDER BY <biểu thức sắp xếp 1> [ASC | DESC] [...]]
[UNION | INTERSECT | MINUS <câu truy vấn khác>]

Các thành phần cơ bản của câu lệnh:

- Biểu thức sau SELECT có thể bao gồm:
 - + Danh sách các cột, kể cả các biểu thức chứa các cột, của các bảng hoặc khung nhìn khai báo sau FROM. Các biểu thức ngăn cách nhau bằng dấu phẩy (,) và có thể đổi tên bằng tùy chọn AS <tên >.
 - + * là ký tự đại diện tất cả các thuộc tính của bảng sau FROM
 - + Các hàm tính toán : COUNT, SUM, AVG, MIN, MAX
- Biểu thức sau FROM gồm một hoặc danh sách các bảng quan hệ (có thể đặt bí danh).
 - INTO <dbf đích>: Lưu bảng kết quả vào đĩa.
 - Biểu thức sau WHERE bao gồm :
 - + Các thuộc tính của các bảng quan hệ sau FROM
 - + Các toán tử số học : +, -, *, /
 - + Các toán tử so sánh, có thể so sánh với ANY (giá trị nào đó trong 1 tập hợp) hoặc ALL (tất cả giá trị trong 1 tập hợp)
 - + Các toán tử Boolean : AND, OR, NOT
 - + Các toán tử tập hợp : UNION (hợp), INTERSECT (giao), MINUS (hiệu).
 - + Các quan hệ bao hàm : IN (∈), NOT IN (∉), CONTAINS (chứa), DOES NOT CONTAIN (không chứa)
 - + Điều kiện tồn tại : EXISTS (tồn tại), NOT EXIST (không tồn tại)
 - + Câu vấn tin con kiểu SELECT ... FROM ... WHERE ...

2. Tìm thông tin từ các cột của bảng – Mệnh đề Select

Mục tiêu: Trình bày cú pháp, chức năng của mệnh đề Select

Cú pháp:

```
SELECT [DISTINCT] <biểu thức 1> [AS <tên 1>] [...] | *  
FROM <bảng >
```

Công dụng:

Xác định nội dung kết quả gồm các cột liệt kê sau SELECT hoặc tất cả các cột <*> lấy từ <Bảng>.

Từ khoá DISTINCT dùng liệt kê những giá trị duy nhất tránh trùng lặp.

Từ khoá AS dùng đặt bí danh

Ví dụ

Cho lược đồ cơ sở dữ liệu khách hàng

KHACHHANG(MaKH, TenKH, DiaChi, SoDu)

HOPDONG(SoHD, MaKH, MaNCC, TenHang, SoLuong)

NHACC(MaNCC, TenNCC, DiaChiNCC, TenHang, DonGia)

- Trả về các bản ghi gồm tất cả các cột của bảng NHACC.

```
SELECT *
```

```
FROM NHACC
```

- Trả về các bản ghi gồm 2 cột TenHang, DonGia của bảng NHACC.

```
SELECT TenHang, DonGia
```

```
FROM NHACC
```

hoặc

```
SELECT DISTINCT TenHang, DonGia
FROM NHACC
```

- Trả về các bản ghi gồm 1 cột TenHang với giá trị duy nhất của bảng NHACC.

```
SELECT DISTINCT TenHang
FROM NHACC
```

3. Chọn các dòng của bảng – Mệnh đề Where

Mục tiêu: Trình bày cú pháp, chức năng của mệnh đề Where trong câu lệnh truy vấn Select.

Cú pháp:

```
SELECT [DISTINCT] <biểu thức 1> [AS <tên 1>] [... ] | *
FROM <bảng >
WHERE <điều kiện lọc>
```

Công dụng:

Trả về các bản ghi gồm các cột liệt kê sau SELECT hoặc tất cả các cột <*> lấy từ <Bảng> gồm các bộ thỏa mãn <điều kiện lọc>.

Ghi chú : Sau WHERE có thể dùng các mệnh đề :

- <biểu thức> BETWEEN <giá trị 1> AND <giá trị 2> : biểu diễn điều kiện <giá trị 1> ≤ <biểu thức> ≤ <giá trị 2>

- <biểu thức> NOT BETWEEN <giá trị 1> AND <giá trị 2> : biểu diễn điều kiện <biểu thức> < <giá trị 1> hoặc <biểu thức> > <giá trị 2>

- <biểu thức> IN (<giá trị 1>, <giá trị 2>, ...) : biểu diễn điều kiện <biểu thức> bằng một trong các giá trị <giá trị 1>, <giá trị 2>, ...

- <biểu thức> NOT IN (<giá trị 1>, <giá trị 2>, ...) : biểu diễn điều kiện <biểu thức> khác các giá trị <giá trị 1>, <giá trị 2>, ...

- <biểu thức ký tự> [NOT] LIKE <chuỗi ký tự> : biểu diễn điều kiện <biểu thức ký tự> [không] giống <chuỗi ký tự>

Trong <chuỗi ký tự> có thể dùng dấu % (dấu phần trăm) đại diện cho chuỗi ký tự bất kỳ

- _ (dấu nối) đại diện cho một ký tự bất kỳ

- <trường> IS [NOT] NULL: biểu diễn điều kiện.

giá trị <trường> là [khác] NULL

Ví dụ

- Trả về các bản ghi gồm các cột TenNCC, DiaChiNCC có giá trị duy nhất từ bảng NHACC của các nhà cung cấp hoặc Đường hoặc Sữa.

```
SELECT DISTINCT TenNCC, DiaChiNCC
FROM NHACC
WHERE (TenHang = 'Đường') OR (TenHang = 'Sữa')
```

hoặc

```
SELECT DISTINCT TenNCC, DiaChiNCC
FROM NHACC
WHERE TenHang IN ('Đường', 'Sữa')
```

- Trả về các bản ghi gồm các TenKH, DiaChi của các khách hàng có SoDu không âm từ bảng KHACHHANG

```
SELECT TenKH, DiaChi
FROM KHACHHANG
WHERE SoDu >= 0
```

- Trả về các bản ghi gồm các hợp đồng có SoLuong từ 100 đến 200 từ bảng HOPDONG

```
SELECT *
FROM HOPDONG
WHERE SoLuong BETWEEN 100 AND 200
```
- Trả về các bản ghi gồm các nhà cung cấp có tên bắt đầu bằng 'Cty' từ bảng NHACC

```
SELECT *
FROM NHACC
WHERE TenNCC LIKE 'Cty %'
```
- Trả về các bản ghi gồm các nhà cung cấp có mặt hàng 'Bia' từ bảng NHACC

```
SELECT *
FROM NHACC
WHERE TenHang LIKE 'Bia %'
```
- Chọn những khách hàng chưa có địa chỉ

```
SELECT *
FROM KHACHHANG
WHERE DiaChi IS NULL
```
- Chọn những khách hàng có số dư xác định :

```
SELECT *
FROM KHACHHANG
WHERE SoDu IS NOT NULL
```

4. Sắp xếp các dòng của bảng – Mệnh đề Order by

Mục tiêu: Trình bày cú pháp, chức năng của mệnh đề Order by trong câu lệnh truy vấn Select

Cú pháp:

```
SELECT [DISTINCT] <biểu thức 1> [AS <tên 1>] [...] | *
FROM <bảng >
ORDER BY <biểu thức sắp xếp 1> [ASC | DESC] [...]
```

Công dụng:

Trả về các bản ghi gồm các cột liệt kê sau SELECT hoặc tất cả các cột <*> lấy từ <Bảng> gồm các bộ được sắp xếp theo các biểu thức sau ORDER BY.

Ý nghĩa các tham số :

ASC : tăng dần (ngầm định)
 DESC : giảm dần .

Ví dụ :

- Cho biết danh sách khách hàng sắp xếp theo TENKH tăng dần :

```
SELECT *
FROM KHACHHANG
ORDER BY TenKH
```
- Cho biết danh sách các nhà cung cấp 'Sữa' sắp xếp theo thứ tự DonGia giảm dần:

```
SELECT *
FROM NHACC
WHERE TenHang = 'Sữa'
ORDER BY DonGia DESC
```
- Cho biết danh sách các nhà cung cấp sắp xếp theo thứ tự TenNCC và TenHang :


```
SELECT *
FROM NHACC
ORDER BY TenNCC, TenHang
```

5. Câu lệnh truy vấn lồng nhau

Mục tiêu: Trình bày cú pháp, chức năng của câu lệnh truy vấn Select lồng nhau.

Cú pháp:

```
SELECT [DISTINCT] <biểu thức 1> [AS <tên 1>] [,...] / *
FROM <bảng 1> [<bí danh 1>] [,...]
[WHERE <điều kiện nối , lọc có chứa câu truy vấn con >]
[ORDER BY <biểu thức sắp xếp 1> [ASC / DESC] [,...]]
```

Công dụng:

Cho phép tạo ra các điều kiện chứa quan hệ tập hợp và số liệu tổng hợp.
Trong quan hệ tập hợp, câu truy vấn con đặt trong ngoặc đơn .

Sử dụng các toán tử so sánh : <, <=, =, >, >=, <>

- Liệt kê các khách hàng có số dư lớn nhất:

```
SELECT TenKH
FROM KHACHHANG
WHERE Sodu =
      (SELECT MAX(Sodu)
       FROM KHACHHANG)
```

- Liệt kê nhà cung cấp gạo có DonGia <= DonGia trung bình :

```
SELECT TenNCC
FROM NHACC a
WHERE (a.TenHang = 'Gạo') AND DonGia <=
      (SELECT AVG(b.DonGia)
       FROM NHACC b
       WHERE b.TenHang = 'Gạo')
```

Sử dụng các toán tử so sánh : <, <=, =, >, >=, <> với ANY/ALL

- Liệt kê nhân viên từ bảng NHANVIEN có NgaySinh trùng với người khác

```
SELECT a.Manv, a.HoLot, a.Ten
FROM NHANVIEN a
WHERE a.NgaySinh = ANY
      (SELECT b.NgaySinh
       FROM NHANVIEN b
       WHERE b.Manv <> a.Manv)
```

- Liệt kê nhân viên từ bảng NHANVIEN có NgaySinh không trùng với người khác

```
SELECT a.Manv, a.HoLot, a.Ten
FROM NHANVIEN a
WHERE a.NgaySinh <> ALL
      (SELECT b.NgaySinh
```

```

FROM          NHANVIEN b
WHERE        b.Manv <> a.Manv)

```

Sử dụng toán tử IN (NOT IN) biểu diễn quan hệ ∈

- Liệt kê TenNCC, TenHang và DonGia của các nhà cung cấp có ít nhất 1 mặt hàng trong HOPDONG :

```

SELECT DISTINCT TenNCC, TenHang, DonGia
FROM NHACC
WHERE TenHang IN
      (SELECT TenHang
       FROM HOPDONG )

```

- Liệt kê TenNCC, TenHang và DonGia của các nhà cung cấp có ít nhất 1 mặt hàng mà khách hàng 'KS. Bạch Đằng' đặt hàng :

```

SELECT DISTINCT TenNCC, TenHang, DonGia
FROM NHACC
WHERE TenHang IN
      (SELECT TenHang
       FROM HOPDONG
       WHERE TenKH = 'KS. Bạch Đằng' )

```

- Liệt kê Ten, DiaChi các nhà cung cấp không bán Gạo :

```

SELECT DISTINCT TenNCC, DiaChiNCC
FROM NHACC
WHERE TenNCC NOT IN
      (SELECT TenNCC
       FROM NHACC
       WHERE TenHang = 'Gạo')

```

- Liệt kê hãng có ít nhất 1 mặt hàng giống Cty Lương thực :

```

SELECT DISTINCT a.TenNCC, a.DiaChiNCC
FROM NHACC a
WHERE a.TenHang IN
      (SELECT b.TenHang
       FROM NHACC b
       WHERE b.TenNCC = 'Cty lương thực')

```

- Liệt kê hợp đồng có TenKH và TenHang giống hợp đồng số 3 :

```

SELECT *
FROM HOPDONG a
WHERE (a.TenKH, a.TenHang) IN
      (SELECT b.TenKH, b.TenHang
       FROM HOPDONG
       WHERE b.SoHD = 3)

```

- Liệt kê nhà cung cấp không có mặt hàng trong HOPDONG :

```

SELECT DISTINCT TenNCC

```

```

FROM NHACC
MINUS
SELECT DISTINCT TenNCC
FROM NHACC
WHERE TenHang IN
      (SELECT TenHang
       FROM HOPDONG )

```

- Liệt kê nhà cung cấp có mặt hàng không có trong HOPDONG :

```

SELECT DISTINCT TenNCC
FROM NHACC
WHERE TenHang NOT IN
      (SELECT TenHang
       FROM HOPDONG )

```

- Liệt kê nhà cung cấp, mà tất cả mặt hàng của nó đều có trong HOPDONG

```

SELECT DISTINCT TenNCC
FROM NHACC
GROUP BY TenNCC
HAVING SET(TenHang) IN
      (SELECT TenHang
       FROM HOPDONG )

```

Sử dụng [NOT] EXISTS <bảng> biểu diễn sự tồn tại

- Liệt kê TenKH, DiaChi khách hàng đã đặt hàng :

```

SELECT      MaKH, TenKH, DiaChi
FROM        KHACHHANG K
WHERE       EXISTS
            (SELECT      MaKH
             FROM        HOPDONG H
             WHERE       K.MaKH = H.MaKH)

```

- Liệt kê TenKH, DiaChi khách hàng chưa có hợp đồng:

```

SELECT      MaKH, TenKH, DiaChi
FROM        KHACHHANG K
WHERE       NOT EXISTS
            (SELECT      MaKH
             FROM        HOPDONG H
             WHERE       K.MaKH = H.MaKH)

```

6. Gom nhóm dữ liệu – mệnh đề Group by

Mục tiêu: Trình bày cú pháp, chức năng của mệnh đề Group by trong câu lệnh truy vấn Select.

Cú pháp:

```

SELECT [DISTINCT] <biểu thức 1> [AS <tên 1>] [,...] | *
FROM <bảng 1> [<bí danh 1>] [,...]
[WHERE <điều kiện nối > [AND / OR <điều kiện lọc>]]
GROUP BY <thuộc tính 1> [,...]
      [HAVING <điều kiện nhóm>]

```

Công dụng:

Nhóm các bản ghi (thỏa mãn <điều kiện nối> hoặc <điều kiện lọc>, nếu có) theo danh sách thuộc tính sau GROUP BY , sau đó kết xuất kết quả tổng hợp (thỏa <điều kiện nhóm>).

Các thuộc tính sau GROUP BY phải có trong danh sách thuộc tính sau SELECT.

Ghi chú :

- Hàm tập hợp SET(...) (tập hợp các phần tử) chỉ sử dụng trong điều kiện nhóm.
- Các hàm tính toán có thể sử dụng sau SELECT hoặc trong điều kiện nhóm.

Ví dụ

- Liệt kê các nhà cung cấp cùng tổng số mặt hàng mà nó cung ứng :


```

SELECT TenNCC, COUNT(TenHang) AS SoTenHang
FROM NHACC
GROUP BY TenNCC

```
- Liệt kê các mặt hàng cùng tổng SoLuong trong HOPDONG


```

SELECT TenHang, SUM(SoLuong) AS TongCong
FROM HOPDONG
GROUP BY TenHang

```
- Liệt kê các mặt hàng cùng giá bình quân trong NHACC :


```

SELECT TenHang, AVG(DonGia) AS DonGiaBQ
FROM NHACC
GROUP BY TenHang

```
- Liệt kê các mặt hàng cùng tổng số hãng cung ứng mặt hàng đó :


```

SELECT TenHang, COUNT(TenNCC) AS SoCty
FROM NHACC
GROUP BY TenHang

```
- Liệt kê các mặt hàng có nhiều hãng cung ứng mặt hàng đó :


```

SELECT TenHang, COUNT(TenNCC) AS SoCty
FROM NHACC
GROUP BY TenHang
      HAVING COUNT(TenNCC) > 1

```
- Liệt kê các hãng cung ứng cả Đường và Sữa :


```

SELECT DISTINCT TenNCC
FROM NHACC
GROUP BY TenNCC
      HAVING SET(TenHang) CONTAINS ('Đường', 'Sữa')

```

7. Bài tập thực hành

Cho lược đồ cơ sở dữ liệu quản lý nhân viên của một công ty như sau:

Nhanvien(MANV,HOTEN, PHAI,NGAYSINH,LUONG,MAPB, MACV)

Mỗi nhân viên có một mã nhân viên (MANV) duy nhất, mỗi mã nhân viên xác định họ và tên nhân viên (HOTEN), phái (PHAI), lương (LUONG), mã phòng ban (MAPB), mã chức vụ (MACV).

Phongban(MAPB,TENPB,TRUSO,MANVPHUTRACH,KINHPhi, DOANHThu)

Mỗi phòng ban có tên gọi phòng ban(TENPB), địa điểm đặt trụ sở (TRUSO), mã nhân viên phụ trách(MANVPHUTRACH), kinh phí hoạt động (KINHPhi), và doanhthu(DOANHThu)

Chucvu(MACV,TENCV,LUONGTHAPNHAT,LUONGCAONHAT)

Mỗi chức vụ có tên gọi chức vụ (TENCV), mức lương tối thiểu(LUONGTHAPNHAT), mức lương tối đa (LUONGCAONHAT).

Hãy trả lời câu hỏi sau bằng SQL

Lập danh sách gồm các thông tin về các phòng ban trong công ty như: mã số phòng ban, tên phòng ban, địa điểm trụ sở, mã số người phụ trách, kinh phí hoạt động, doanh thu.

7.1. Các bước thực hiện

Bước 1: Phân tích yêu cầu

Bước 2: Xác định Table

Bước 3: Xác định các cột cần hiển thị

Bước 4: Viết đúng các điều kiện

Bước 5: Viết đúng lệnh

7.2. Sinh viên thực hành

```
SELECT *
```

```
FROM PHONGBAN
```

Những trọng tâm cần chú ý trong bài

- Phân tích yêu cầu chính xác
- Xác định đúng các Table
- Xác định các cột cần hiển thị
- Viết đúng các điều kiện
- Viết đúng lệnh

Bài tập mở rộng và nâng cao

Dựa vào lược đồ cơ sở dữ liệu quản lý nhân viên ở trên. Hãy trả lời câu hỏi sau bằng SQL

2. Lập danh sách những nhân viên sinh nhật trong tháng 10
3. Lập danh sách gồm các thông tin mã số nhân viên, họ và tên và lương cả năm của các nhân viên (giả sử rằng lương cả năm =12*lương)
4. Lập những phòng ban có kinh phí hoạt động cao nhất.

5. Lập danh sách nhân viên của phòng ban có mã số phòng ban là 40.
6. Lập danh sách nhân viên của phòng có mã số phòng ban 10,30,50.
7. Lập danh sách các nhân viên có lương tháng từ 2.500.000 đến 4.000.000
8. Tìm những nhân viên có tuổi cao nhất thuộc phòng ban có MAPB là 10
9. Lập danh sách các nhân viên của phòng 10,30,50. kết quả in ra theo thứ tự tăng dần của mã phòng nếu trùng mã phòng thì sắp xếp giảm dần theo mức lương.
10. Lập danh sách các nhân viên phòng 10,30,50, chỉ in ra những người là lãnh đạo của mỗi phòng ban này.
11. Lập danh sách gồm MAPB, LUONGCAONHAT mà người có mức lương cao nhất của từng phòng ban lớn hơn hoặc bằng 4.000.000
12. Lập mã phòng ban, tên phòng ban, họ và tên của lãnh đạo phòng tương ứng.
13. Lập danh sách những người làm việc cùng phòng với ông Nguyen Van Thanh
14. Lập biết mã số nhân viên, họ và tên, mức lương của người lãnh đạo ông Nguyen Van Thanh.
15. Lập danh sách nhân viên có mức lương lớn hơn hay bằng mức lương cao nhất của phòng ông Nguyen Van Thanh.
16. Cho biết MAPB, TENPB, tổng số nhân viên, mức lương cao nhất, mức lương thấp nhất, mức lương trung bình của từng phòng ban.
17. Cho biết các nhân viên có mức lương cao nhất của các phòng ban.

Yêu cầu về đánh giá kết quả học tập bài 2

Nội dung:

- + Về kiến thức: Trình bày được cách thức truy vấn của ngôn ngữ truy vấn dữ liệu SQL.
- + Về kỹ năng: Thực hiện truy vấn được trên các dữ liệu đã cài đặt.
- + Năng lực tự chủ và trách nhiệm: Rèn luyện tính cẩn thận, tỉ mỉ trong việc học và làm Bài tập. Thực hiện các thao tác an toàn với máy tính.

Phương pháp:

- + Về kiến thức: Được đánh giá bằng hình thức kiểm tra viết, trắc nghiệm, vấn đáp
- + Về kỹ năng: Đánh giá kỹ năng thực hành viết đúng lệnh truy vấn dữ liệu.
- + Năng lực tự chủ và trách nhiệm: Rèn luyện tính cẩn thận, chính xác khi thực hiện lệnh.

Bài 3. RÀNG BUỘC TOÀN VỆN

Mã bài: MD10-03

Giới thiệu:

Ràng buộc toàn vẹn là các quy định, điều kiện từ ứng dụng thực tế. Các điều kiện này là bất biến. Vì vậy, luôn phải đảm bảo cơ sở dữ liệu thỏa ràng buộc toàn vẹn sau mỗi thao tác làm thay đổi tình trạng của cơ sở dữ liệu.

Mục tiêu:

- Trình bày được khái niệm, cách phân loại, các yếu tố ràng buộc toàn vẹn.
- Xây dựng được các ràng buộc dữ liệu trong một số bài toán cụ thể.
- Rèn luyện tính cẩn thận, tỉ mỉ trong việc học và làm bài tập.

Nội dung:

1. Ràng buộc toàn vẹn

1.1 Ràng buộc toàn vẹn

Khái niệm

Trong mỗi CSDL luôn tồn tại nhiều mối liên hệ giữa các thuộc tính, giữa các bộ; sự liên hệ này có thể xảy ra trong cùng một quan hệ hoặc trong các quan hệ của một lược đồ CSDL. Các mối liên hệ này là những điều kiện bất biến mà tất cả các bộ của những quan hệ có liên quan trong CSDL đều phải thỏa mãn ở mọi thời điểm. Những điều kiện bất biến đó được gọi là ràng buộc toàn vẹn.

Trong thực tế ràng buộc toàn vẹn là các quy tắc quản lý được áp đặt trên các đối tượng của thế giới thực. Chẳng hạn mỗi sinh viên phải có một mã sinh viên duy nhất, hai thí sinh dự thi vào một trường phải có số báo danh khác nhau, một sinh viên dự thi một MÔ ĐUN không quá 2 lần,...

Ràng buộc toàn vẹn là một điều kiện bất biến không được vi phạm trong một CSDL.

1.2. Các yếu tố của ràng buộc toàn vẹn

Mỗi ràng buộc toàn vẹn có bốn yếu tố: điều kiện, bối cảnh, bảng tầm ảnh hưởng và hành động phải cần thực hiện khi phát hiện có ràng buộc toàn vẹn bị vi phạm:

Điều kiện:

Điều kiện của ràng buộc toàn vẹn là sự mô tả, và biểu diễn hình thức nội dung của nó.

Điều kiện của một ràng buộc toàn vẹn R có thể được biểu diễn bằng ngôn ngữ tự nhiên, ngôn ngữ đại số quan hệ, ngôn ngữ mã giả, ngôn ngữ truy vấn SQL,... ngoài ra điều kiện của ràng buộc toàn vẹn cũng có thể được biểu diễn bằng phụ thuộc hàm.

Ví dụ sau đây chỉ ra một số ràng buộc toàn vẹn trên lược đồ CSDL quản lý sinh viên.

- Mỗi lớp học phải có một mã số duy nhất để phân biệt với các lớp học khác trong trường.

- Mỗi lớp học phải thuộc về một khoa của trường.

- Mỗi sinh viên có một mã số sinh viên duy nhất, không trùng với bất cứ sinh viên nào trong trường.

- Mỗi học viên phải đăng ký vào một lớp học trong trường. Mỗi học viên chỉ được thi tối đa 2 lần cho mỗi MÔ ĐUN.

- Tổng số học viên của một lớp phải lớn hơn hoặc bằng số lượng đếm được của một lớp tại một thời điểm nào đó.

Bối cảnh:

Bối cảnh của ràng buộc toàn vẹn là những quan hệ mà ràng buộc đó có hiệu lực hay nói một cách khác, đó là những quan hệ cần phải được kiểm tra khi tiến hành cập nhật dữ liệu. Bối cảnh của một ràng buộc toàn vẹn có thể là một hoặc nhiều quan hệ.

Chẳng hạn với ràng buộc toàn vẹn R trên thì bối cảnh của nó là quan hệ Sinhvien

Bảng tầm ảnh hưởng

Trong quá trình phân tích thiết kế một CSDL, người phân tích cần lập bảng tầm ảnh hưởng cho một ràng buộc toàn vẹn nhằm xác định thời điểm cần phải tiến hành kiểm tra khi tiến hành cập nhật dữ liệu.

Thời điểm cần phải kiểm tra ràng buộc toàn vẹn chính là thời điểm cập nhật dữ liệu.

Một bảng tầm ảnh hưởng của một ràng buộc toàn vẹn có dạng sau:

Tên RBTV	Thêm(T)	Sửa(S)	Xoá(X)
r ₁	+		
r ₂		-	
r ₃			-(*)
r _n			

Bảng này chứa toàn các ký hiệu + , - hoặc -(*)

- Dấu + tại (dòng r₁, cột Thêm) thì có nghĩa là khi thêm một bộ vào quan hệ r₁ thì RBTV bị vi phạm.

- Dấu - tại ô (dòng r₂, cột Sửa) thì có nghĩa là khi sửa một bộ trên quan hệ r₂ thì RBTV không bị vi phạm.

Quy ước:

- Không được sửa thuộc tính khoá.

- Nếu không bị vi phạm do không được phép sửa đổi thì ký hiệu là -(*)

Hành động cần phải có khi phát hiện có RBTV bị vi phạm:

Khi một ràng buộc toàn vẹn bị vi phạm, cần có những hành động thích hợp. Thông thường có 2 giải pháp:

- Thứ nhất: Đưa ra thông báo và yêu cầu sửa chữa dữ liệu của các thuộc tính cho phù hợp với quy tắc đảm bảo tính nhất quán dữ liệu. Thông báo phải đầy đủ và phải thân thiện với người sử dụng. Giải pháp này là phù hợp cho việc xử lý thời gian thực.

- Thứ hai: Từ chối thao tác cập nhật. Giải pháp này là phù hợp đối với việc xử lý theo lô. Việc từ chối cũng phải được lưu lại bằng những thông báo đầy đủ, rõ ràng vì sao thao tác bị từ chối và cần phải sửa lại những dữ liệu nào?

Khóa nội, khoá ngoại, giá trị NOT NULL là những ràng buộc toàn vẹn miền giá trị của các thuộc tính. Những ràng buộc toàn vẹn này là những ràng buộc toàn vẹn đơn giản trong CSDL.

Các hệ quản trị cơ sở dữ liệu thường có các cơ chế tự động kiểm tra các ràng buộc toàn vẹn về miền giá trị của khóa nội, khoá ngoại, giá trị NOT NULL.

Việc kiểm tra ràng buộc toàn vẹn có thể tiến hành vào những thời điểm sau đây:

- Thứ nhất: Kiểm tra ngay sau khi thực hiện một thao tác cập nhật CSDL. Thao tác cập nhật chỉ được xem là hợp lệ nếu như nó không vi phạm bất cứ một ràng buộc toàn vẹn nào, nghĩa là nó không làm mất tính toàn vẹn của CSDL. Nếu vi phạm ràng buộc toàn vẹn, thao tác cập nhật bị coi là không hợp lệ và sẽ bị hệ thống huỷ bỏ (hoặc có một xử lý thích hợp nào đó).

- Thứ hai: Kiểm tra định kỳ hay đột xuất, nghĩa là việc kiểm tra ràng buộc toàn vẹn được tiến hành độc lập với thao tác cập nhật dữ liệu. Đối với những trường hợp vi phạm ràng buộc toàn vẹn, hệ thống có những xử lý ngầm định hoặc yêu cầu người sử dụng xử lý những sai sót một cách tường minh.

2. Phân loại ràng buộc toàn vẹn

Xét ví dụ sau đây: Cho một CSDL QL BH dùng để quản lý việc đặt hàng và giao hàng của một công ty. Lược đồ CSDL QL BH gồm các lược đồ quan hệ như sau:

Q1: KHACHHANG (MAKH, TENKH, DIACHIKH, DIENTHOAI)

Tên từ:

Mỗi khách hàng có một mã khách hàng (MAKH) duy nhất, mỗi MAKH xác định tên khách hàng (TENKH), địa chỉ (DIACHIKH), số điện thoại (DIENTHOAI).

Q2: MATHANG (MAHANG, TENHANG, QUYCACH, DVTINH)

Tên từ:

Mỗi mặt hàng có một mã hàng (MAHANG) duy nhất, mỗi MAHANG xác định tên hàng (TENHANG), quy cách hàng (QUYCACH), đơn vị tính (DVTINH).

Q3: DATHANG (SODH, MAHANG, SLDAT, NGAYDH, MAKH)

Tên từ:

Mỗi mã số đặt hàng (SODH) xác định một ngày đặt hàng (NGAYDH) và mã khách hàng tương ứng (MAKH). Biết mã số đặt hàng và mã mặt hàng thì biết được số lượng đặt hàng (SLDAT). Mỗi khách hàng trong một ngày có thể có nhiều lần đặt hàng.

Q4: Hoadon (SOHD, NGAYLAP, SODH, TRIGIAHD, NGAYXUAT)

Tên từ:

Mỗi hoá đơn tổng hợp có một mã số duy nhất là SOHD, mỗi hoá đơn bán hàng có thể gồm nhiều mặt hàng. Mỗi hoá đơn xác định ngày lập hoá đơn (NGAYLAP), ứng với số đặt hàng nào (SODH). Giả sử rằng hoá đơn bán hàng theo yêu cầu của chỉ một đơn đặt hàng có mã số là SODH và ngược lại, mỗi đơn đặt hàng chỉ được giải quyết chỉ trong một hoá đơn. Do điều kiện khách quan có thể công ty không giao đầy đủ các mặt hàng cũng như số lượng từng mặt hàng như yêu cầu trong đơn đặt hàng nhưng không bao giờ giao vượt ngoài yêu cầu. Mỗi hoá đơn xác định một trị giá của nhưng các mặt hàng trong hoá đơn (TRIGIAHD) và một ngày xuất kho giao hàng cho khách (NGAYXUAT)

Q5: Chitiethd (SOHD, MAHANG, GIABAN, SLBAN)

Tên từ:

Mỗi SOHD, MAHANG xác định giá bán (GIABAN) và số lượng bán (SLBAN) của một mặt hàng trong một hoá đơn.

Q6: Phiethu (SOPT, NGAYTHU, MAKH, SOTIEN)

Tên từ:

Mỗi phiếu thu có một số phiếu thu (SOPT) duy nhất, mỗi SOPT xác định một ngày thu (NGAYTHU) của một khách hàng có mã khách hàng là MAKH và số tiền thu là SOTIEN. Mỗi khách hàng trong một ngày có thể có nhiều số phiếu thu.

2.1 Ràng buộc toàn vẹn liên bộ

Ràng buộc toàn vẹn về khoá chính:

Đây là một trường hợp đặc biệt của Ràng Buộc toàn Vẹn liên bộ, RBTV này rất phổ biến và thường được các hệ quản trị CSDL tự động kiểm tra.

Ví dụ:

Với r là một quan hệ trên lược đồ quan hệ KHACHHANG ta có RBTV sau:

$R_1: \forall t_1, t_2 \in r$

$t_1.MAKH \neq t_2.MAKH$

Cuối \forall

R_1	Thêm	Sửa	Xoá
Khach	+	-	-

Ràng buộc toàn vẹn về tính duy nhất

Ví dụ: Mỗi phòng ban phải có một tên gọi duy nhất. Ngoài ra nhiều khi ta còn gặp những RBTV khác chẳng hạn như các RBTV trong quan hệ sau đây.

Ví dụ: KETQUA(MASV,MAMH,LAN THI,DIEM)

Mỗi sinh viên chỉ được đăng thi mỗi môn tối đa là 2 lần.

2.2. Ràng buộc toàn vẹn về phụ thuộc tồn tại

Ràng buộc toàn vẹn về khoá ngoại còn được gọi là ràng buộc toàn vẹn phụ thuộc tồn tại. Cũng giống như ràng buộc toàn vẹn về khoá nội, loại ràng buộc toàn vẹn này rất phổ biến trong các CSDL.

Ví dụ:

$R_2: \text{dathang}[MAKH] \subseteq \text{khach}[MAKH]$

R_2	Thêm	Sửa	Xoá
dathang	+	+	-
Khach	-	-	+

2.3. Ràng buộc toàn vẹn về miền giá trị

Ràng buộc toàn vẹn có liên quan đến miền giá trị của các thuộc tính trong một quan hệ. Ràng buộc này thường gặp. Thông thường các hệ quản trị CSDL đã tự động kiểm tra (một số) ràng buộc loại này.

Ví dụ: Với r là một quan hệ của Hoadon ta có ràng buộc toàn vẹn sau

$R_3: \forall t \in r$

$t.TRIGIAHD > 0$

Cuối \forall

R_3	Thêm	Sửa	Xoá
Hoadon	+	+	-

2.4. Ràng buộc toàn vẹn liên thuộc tính

Ràng buộc toàn vẹn liên thuộc tính (một quan hệ) là mối liên hệ giữa các thuộc tính trong một lược đồ quan hệ.

Ví dụ: Với r là một quan hệ của Hoadon ta có ràng buộc toàn vẹn sau:

$$R_4: \forall t \in r$$

$$t.NGAYLAP \leq t.NGAYXUAT$$

Cuối \forall

R ₄	Thêm	Sửa	Xoá
Hoadon	+	+	-

2.5. Ràng buộc toàn vẹn liên thuộc tính liên quan hệ

Ràng buộc loại này là mối liên hệ giữa các thuộc tính trong nhiều lược đồ quan hệ.

Ví dụ: Với r, s lần lượt là quan hệ của DatTenHang và Hoadon. Ta có RBTV R₅ như sau:

$$R_5: \forall t_1 \in r, t_2 \in s$$

Nếu $t_1.SODH=t_2.SODH$ thì

$$t_1.NGAYDH \leq t_2.NGAYXUAT$$

Cuối \forall

R ₅	Thêm	Sửa	Xoá
Dathang	+	-	-
Hoandon	+	+	-

2.6. Ràng buộc toàn vẹn về thuộc tính tổng hợp

Ràng buộc loại này là mối liên hệ giữa các bộ trong một lược đồ cơ sở dữ liệu. Chẳng hạn như tổng số tiền phải trả trong mỗi hoá đơn (chitiethd) phải bằng TRỊ GIÁ HOÁ ĐƠN của hoá đơn đó trong quan hệ Hoadon.

Hoặc số lượng học viên trong một lớp phải bằng SOHOCVIEN của lớp đó.

Ngoài ra còn có một số loại RBTV khác như: RBTV về thuộc tính tổng hợp, RBTV do tồn tại chu trình, RBTV về giá trị thuộc tính theo thời gian.

3. Bài tập thực hành

Việc tổ chức kỳ thi tốt nghiệp của một khoa như sau:

Mỗi thí sinh có một Mã số sinh viên duy nhất (MASV), mỗi MASV xác định được các thông tin: họ và tên (HOTEN), ngày sinh (NGAYSINH), nơi sinh, phái, dân tộc.

Mỗi lớp có một mã lớp (MALOP) duy nhất, mỗi mã lớp xác định các thông tin: tên lớp (TENLOP), mỗi lớp chỉ thuộc sự quản lý của một khoa nào đó. Mỗi khoa có một mã khoa duy nhất (MAKHOA), mỗi mã khoa xác định tên khoa (TENKHOA).

Mỗi thí sinh đều phải dự thi tốt nghiệp ba môn. Mỗi môn thi có một mã môn thi (MAMT) duy nhất, mỗi mã môn thi xác định các thông tin: tên môn thi (TENMT), thời gian làm chương – được tính bằng phút (PHUT), ngày thi (NGAYTHI), buổi thi (BUOITHI), môn thi này là môn lý thuyết hay thực hành (LYTHUYET). Chú ý rằng, nếu một MÔ ĐUN được cho thi ở nhiều hệ thì được đặt MAMT khác nhau (chẳng hạn cả trung cấp và cao đẳng ngành công nghệ thông tin đều thi môn Cơ Sở Dữ Liệu), để diễn tả điều này, mỗi mã MÔ ĐUN cần phải được ghi chú (GHICHU) để cho biết môn thi đó dành cho khối nào trung cấp, hay cao đẳng). Mỗi thí sinh ứng với một môn thi có một điểm thi (DIEMTHI) duy nhất, điểm thi được chấm theo tTenHang điểm 10 và có lấy điểm lẻ đến 0.5. Một thí sinh được coi là đậu tốt nghiệp nếu điểm thi của tất cả các môn của thí sinh đó đều lớn hơn hoặc bằng 5.

Trong một phòng thi có thể có thí sinh của nhiều lớp. Trong một kỳ thi, mỗi thí sinh có thể thi tại những phòng thi (PHONGTHI) khác nhau, chẳng hạn một thí sinh thi tốt nghiệp ba môn là Cơ sở dữ liệu, Lập trình C và Visual Basic thì môn Cơ Sở Dữ Liệu và Lập Trình C thi tại phòng A3.4, còn môn thực hành Visual Basic thi tại phòng máy H6.1

Qua phân tích sơ bộ trên, ta có thể lập một lược đồ cơ sở dữ liệu như sau:

THISINH (MASV, HOTEN, NGAYSINH, NOISINH, PHAI, DANTOC, MALOP)

LOP (MALOP, TENLOP, MAKHOA)

MONTHI (MAMT, TENMT, LYTHUYET, PHUT, NGAYTHI, BUOITHI, GHICHU)

KETQUA (MASV, MAMT, DIEMTHI, PHONGTHI)

KHOA (MAKHOA, TENKHOA)

1. Tìm khoá cho mỗi lược đồ quan hệ trên.

2. Hãy phát biểu các ràng buộc toàn vẹn có trong cơ sở dữ liệu trên.

3.1. Các bước thực hiện

Bước 1: Xác định điều kiện ràng buộc

Bước 2: Xác định bối cảnh

Bước 3: Viết bảng tầm ảnh hưởng

Bước 4: Các hành động cần phải có khi phát hiện có RBTV bị vi phạm

3.2. Sinh viên thực hành

1. Tìm khoá

THISINH (MASV, HOTEN, NGAYSINH, NOISINH, PHAI, DANTOC, MALOP)

LOP (MALOP, TENLOP, MAKHOA)

MONTHI (MAMT, TENMT, LYTHUYET, PHUT, NGAYTHI, BUOITHI, GHICHU)

KETQUA (MASV, MAMT, DIEMTHI, PHONGTHI)

KHOA (MAKHOA, TENKHOA)

2. Ràng buộc toàn vẹn

R1: $\forall t_1, t_2 \in r$

$t_1.MASV \neq t_2.MASV$

R1	Thêm	Sửa	Xoá
THISINH	+	-	-

Những trọng tâm cần chú ý trong bài

- Xác định điều kiện ràng buộc
- Xác định bối cảnh
- Viết bảng tầm ảnh hưởng
- Các hành động cần phải có khi phát hiện có RBTV bị vi phạm

Bài tập mở rộng và nâng cao

Quản lý đăng ký chuyên đề

Phòng giáo vụ tại một trường đại học muốn tin học hóa việc quản lý học các chuyên đề của sinh viên. Sau đây là kết quả của việc phân tích thiết kế ứng dụng trên.

Mỗi sinh viên có một mã số duy nhất, một họ tên, thuộc một phái, có một ngày sinh, một địa chỉ và học một ngành duy nhất.

Mỗi ngành có một mã ngành duy nhất, có một tên ngành duy nhất. Ngoài ra cũng cần lưu lại một con số cho biết số chuyên đề mà một sinh viên theo học một ngành cụ thể phải học, và cũng cần lưu lại tổng số sinh viên đã từng theo học ngành này.

Sinh viên phải học các chuyên đề khác nhau. Mỗi chuyên đề có một mã duy nhất và có một tên duy nhất. Cần lưu lại tên về số sinh viên tối đa có thể chấp nhận được mỗi khi có một lớp mở cho chuyên đề cụ thể.

Mỗi chuyên đề có thể được học bởi sinh viên thuộc nhiều ngành và sinh viên thuộc mỗi ngành phải học nhiều chuyên đề. Mỗi ngành học tối đa là 8 chuyên đề.

Vào mỗi học kỳ của mỗi năm học, ta cần lưu lại các chuyên đề nào được mở ra cho học kỳ của năm đó để sinh viên có thể đăng ký. Sinh viên chỉ được đăng ký những chuyên đề có mở.

Khi sinh viên đăng ký học, lưu lại việc đăng ký học một chuyên đề của một sinh viên vào một năm của một học kỳ nào đó.

Một sinh viên chỉ được đăng ký vào các chuyên đề thuộc ngành học của sinh viên đó mà thôi. Mỗi năm có 2 học kỳ. Sinh viên chỉ được đăng ký tối đa là 3 chuyên đề trong một học kỳ mà thôi.

1. Hãy thiết kế mô hình ER cho ứng dụng trên.

2. Chuyển mô hình ER sang mô hình quan hệ. Xác định khóa chính, khóa ngoại và liệt kê có phân loại tất cả ràng buộc toàn vẹn nhận diện được.

Yêu cầu về đánh giá kết quả học tập bài 3

Nội dung:

+ Về kiến thức: Trình bày được khái niệm, cách phân loại, các yếu tố ràng buộc toàn vẹn.

+ Về kỹ năng: Xây dựng được các ràng buộc dữ liệu trong một số Bài toán cụ thể.

+ Năng lực tự chủ và trách nhiệm: Rèn luyện tính cẩn thận, tỉ mỉ trong việc học và làm Bài tập.

Phương pháp:

+ Về kiến thức: Được đánh giá bằng hình thức kiểm tra viết, trắc nghiệm, vấn đáp

+ Về kỹ năng: Đánh giá kỹ năng thực hành viết đúng các ràng buộc toàn vẹn dữ liệu.

+ Năng lực tự chủ và trách nhiệm: Rèn luyện tính cẩn thận, tỉ mỉ trong việc học và làm Bài tập.

Bài 4. LÝ THUYẾT THIẾT KẾ CƠ SỞ DỮ LIỆU

Mã bài: MD10-04

Giới thiệu:

Bài này trình bày những khái niệm cơ bản nhất về mô hình dữ liệu quan hệ của E.F.Codd, gồm các khái niệm về quan hệ, phụ thuộc hàm, hệ tiên đề Armstrong, bao đóng, khoá. Chúng đóng vai trò rất quan trọng trong mô hình dữ liệu quan hệ và được dùng nhiều trong việc thiết kế các hệ quản trị cơ sở dữ liệu (CSDL) hiện nay.

Mục tiêu:

- Trình bày được khái niệm cơ bản của lý thuyết cơ sở dữ liệu như phụ thuộc hàm, bao đóng, khoá, phủ tối thiểu.
- Áp dụng được thuật toán tìm khoá, tìm phủ tối thiểu.
- Rèn luyện tính cẩn thận, chính xác trong phân tích và giải quyết vấn đề.

Nội dung:

1. Các vấn đề gặp phải khi tổ chức dữ liệu:

Mục tiêu: Trình bày được các vấn đề dị thường dữ liệu mắc phải khi thực hiện tổ chức và thiết kế cơ sở dữ liệu.

Khi thiết kế, tổ chức cơ sở dữ liệu quan hệ ta thường đứng trước vấn đề lựa chọn các lược đồ quan hệ: lược đồ nào tốt hơn? Tại sao? Mục này sẽ nghiên cứu một số tiêu chuẩn đánh giá lược đồ quan hệ và các thuật toán giúp chúng ta xây dựng được lược đồ cơ sở dữ liệu quan hệ có cấu trúc tốt.

Có thể nói tổng quát, một lược đồ quan hệ có cấu trúc tốt là lược đồ không chứa sự dư thừa dữ liệu và các dị thường dữ liệu.

- Dư thừa dữ liệu là sự trùng lặp thông tin trong cơ sở dữ liệu.
- Dị thường dữ liệu là các sự cố xảy ra khi cập nhật dữ liệu (lập, dị thường chèn bộ, dị thường xóa bộ, dị thường sửa bộ) làm cho dữ liệu không tương thích, bất định hoặc mất mát.

+ Dị thường do dữ liệu lặp: một số thông tin có thể bị lặp lại một cách vô ích.
+ Dị thường chèn bộ: không thể chèn bộ mới vào quan hệ, nếu không có đầy đủ dữ liệu.

+ Dị thường xóa bộ: ngược lại với dị thường chèn bộ, việc xóa bộ có thể dẫn đến mất thông tin.

+ Dị thường sửa bộ: việc sửa đổi dữ liệu dư thừa có thể dẫn đến sự không tương thích dữ liệu.

Cơ sở lý thuyết của việc thiết kế lược đồ cơ sở dữ liệu quan hệ tốt là khái niệm phụ thuộc dữ liệu. Phụ thuộc dữ liệu biểu diễn các quan hệ nhân quả giữa các thuộc tính trong quan hệ. Cũng dựa trên khái niệm phụ thuộc dữ liệu người ta định nghĩa các dạng chuẩn của lược đồ quan hệ. Còn quá trình biến đổi lược đồ thành lược đồ tương đương thỏa mãn dạng chuẩn gọi là quá trình chuẩn hóa lược đồ quan hệ.

2. Phụ thuộc hàm

Mục tiêu: Trình bày được định nghĩa về phụ thuộc hàm, các tính chất của phụ thuộc hàm (hệ tiên đề Armstrong).

2.1. Định nghĩa phụ thuộc hàm

Cho lược đồ quan hệ $R=(A_1, A_2, \dots, A_n)$ và X, Y là các tập con của $R^+ = \{A_1, A_2, \dots, A_n\}$. Ta nói rằng X xác định hàm Y hay Y phụ thuộc hàm X , ký hiệu $X \rightarrow Y$, nếu mọi quan hệ bất kỳ r của lược đồ R thoả mãn:

$$\forall u, v \in r : u(X) = v(X) \Rightarrow u(Y) = v(Y)$$

Phụ thuộc hàm $X \rightarrow Y$ gọi là phụ thuộc hàm *tâm thường* nếu $Y \subset X$ (hiển nhiên là nếu $Y \subset X$ thì theo định nghĩa ta có $X \rightarrow Y$).

Phụ thuộc hàm $X \rightarrow Y$ gọi là phụ thuộc hàm *nguyên tố* nếu không có tập con thực sự $Z \subset X$ thoả $Z \rightarrow Y$.

Tập thuộc tính $K \subset R$ gọi là *khoá* nếu nó xác định hàm tất cả các thuộc tính và $K \rightarrow R$ là phụ thuộc hàm nguyên tố.

2.2. Cách xác định phụ thuộc hàm cho lược đồ quan hệ

Cách duy nhất để xác định đúng các phụ thuộc thích hợp cho một lược đồ quan hệ là xem xét nội dung tâm từ của lược đồ quan hệ đó.

Ví dụ một số phụ thuộc hàm ứng với từng lược đồ quan hệ được xác định như sau:

MASV \rightarrow HOTENSV, NGAYSINH, MALOP, PHAI
MALOP \rightarrow TENLOP, MAKHOA

2.3. Một số tính chất của phụ thuộc hàm – hệ luật dẫn Armstrong

Để có thể xác định được các phụ thuộc hàm khác từ tập phụ thuộc hàm đã có, ta sử dụng các quy tắc suy diễn đơn giản để kiểm tra xem một phụ thuộc hàm có được suy diễn logic từ F hay không.

Một trong các quy tắc suy diễn đó gọi là hệ tiên đề Armstrong(1974), gồm các luật sau:

- | | |
|-----------------------------------|--|
| 1. Luật phản xạ (reflexivity) | Nếu $Y \subseteq X$ thì $X \rightarrow Y$ |
| 2. Luật tăng trưởng(augmentation) | $X \rightarrow Y \Rightarrow XZ \rightarrow YZ$ |
| 3. Luật bắc cầu(transitivity) | $X \rightarrow Y, Y \rightarrow Z \Rightarrow X \rightarrow Z$ |

Các quy tắc suy rộng:

- | | |
|---|---|
| 4. Luật hợp (the union rule) | Cho $X \rightarrow Y, X \rightarrow Z \Rightarrow X \rightarrow YZ$ |
| 5. Luật bắc cầu giả (the pseudotransitivity rule) | |

$$\text{Cho } X \rightarrow Y, WY \rightarrow Z \Rightarrow XW \rightarrow Z, \text{ Với } X, Y, Z, W \in R^+$$

- | | |
|--|--|
| 6. Luật phân rã (the decomposition rule) | |
|--|--|

$$\text{Cho } X \rightarrow YZ \Rightarrow X \rightarrow Y, X \rightarrow Z$$

Ví dụ:

Cho lược đồ $R(ABC)$ và $F=\{AB \rightarrow C, C \rightarrow A\}$. Dùng các quy tắc Armstrong ta chứng minh rằng $(B,C) \rightarrow (A,B,C)$.

Thật vậy, ta có

	C	\rightarrow	A	(theo giả thiết)
	BC	\rightarrow	AB	(theo luật tăng trưởng)
	C	\rightarrow	C	(theo luật phản xạ)
\Rightarrow	BC	\rightarrow	ABC (đpcm)	(theo luật hợp)

3. Bao đóng của tập phụ thuộc hàm và bao đóng của tập thuộc tính

Mục tiêu: Trình bày khái niệm về bao đóng của tập phụ thuộc hàm và bao đóng tập thuộc tính, các giải thuật xác định bao đóng tương ứng với tập phụ thuộc hàm và tập thuộc tính đã được xác định.

3.1. Bao đóng của tập phụ thuộc hàm F

Bao đóng của tập phụ thuộc hàm F , ký hiệu là F^+ , là tập hợp tất cả các phụ thuộc hàm suy diễn logic từ F :

$$F^+ = \{X \rightarrow Y \mid F \models X \rightarrow Y\}$$

Hay nói cách khác: *Bao đóng (closure) của tập phụ thuộc hàm F* (ký hiệu là F^+) là tập hợp tất cả các phụ thuộc hàm có thể suy ra từ F dựa vào các tiên đề Armstrong.

Rõ ràng $F \subseteq F^+$

Ví dụ: Cho $R=(A,B,C)$ và $F = \{A \rightarrow B, B \rightarrow C\}$. Khi đó bao đóng F^+ gồm các phụ thuộc hàm $X \rightarrow Y$ thoả

(i) X chứa A , Y bất kỳ:

$A,B,C \rightarrow A,B,C;$	$A,B,C \rightarrow A,B;$	$A,B,C \rightarrow A,C;$	$A,B,C \rightarrow B,C;$
$A,B,C \rightarrow A;$	$A,B,C \rightarrow B;$	$A,B,C \rightarrow B;$	$A,B,C \rightarrow C;$
$A,B \rightarrow A,B,C;$	$A,B \rightarrow A,B;$	$A,B \rightarrow A,C;$	$A,B \rightarrow B,C;$
$A,B \rightarrow A;$	$A,B \rightarrow B;$	$A,B \rightarrow B;$	$A,B \rightarrow C;$
$A,C \rightarrow A,B,C;$	$A,C \rightarrow A,B;$	$A,C \rightarrow A,C;$	$A,C \rightarrow B,C;$
$A,C \rightarrow A;$	$A,C \rightarrow B;$	$A,C \rightarrow B;$	$A,C \rightarrow C;$
$A \rightarrow A,B,C;$	$A \rightarrow A,B;$	$A \rightarrow A,C;$	$A \rightarrow B,C;$
$A \rightarrow A;$	$A \rightarrow B;$	$A \rightarrow B;$	$A \rightarrow C;$

(ii) X chứa B nhưng không chứa A , Y không chứa A :

$BC \rightarrow BC;$	$BC \rightarrow B;$	$BC \rightarrow C$
$B \rightarrow BC;$	$B \rightarrow B;$	$B \rightarrow C$

(iii) $C \rightarrow C$

Về mặt lý thuyết ta hoàn toàn có thể xây dựng thủ tục tính bao đóng F^+ của tập phụ thuộc hàm F , nhưng trên thực tế Bài toán xác định F^+ là không khả thi vì với số thuộc tính và phụ thuộc hàm lớn sẽ dẫn đến bùng nổ tổ hợp.

Thay vào đó chúng ta sẽ xem xét một Bài toán khác: "Kiểm tra xem một phụ thuộc hàm có thuộc bao đóng F^+ hay không?". Bài toán này gọi là **Bài toán thành viên**. Bài toán thành viên thiết thực hơn Bài toán tính bao đóng vì trong thực tế rất hiếm khi phải tìm tất cả các phụ thuộc hàm suy diễn lô-gic từ F .

Bài toán thành viên liên quan mật thiết với khái niệm bao đóng của tập thuộc tính.

3.2. Bao đóng của tập thuộc tính X

Bao đóng của tập thuộc tính $X \subset R$ (đối với tập phụ thuộc hàm F), ký hiệu là X_F^+ (X^+), là tập hợp tất cả các thuộc tính phụ thuộc hàm vào X :

$$X^+ = \{A \mid X \rightarrow A \in F^+\}$$

Từ định nghĩa dễ dàng suy ra: $X \subset X^+$ và $X \rightarrow Y \Leftrightarrow Y \subset X^+$. Nghĩa là X^+ là tập thuộc tính lớn nhất phụ thuộc hàm vào X .

Ví dụ: Cho $R(ABC)$ và $F = \{A \rightarrow B, B \rightarrow C\}$. Khi đó ta dễ dàng thấy bao đóng của thuộc tính B là $B^+ = \{B,C\}$ vì $B \rightarrow \{B,C\}$ và B không xác định A .

3.3. Bài toán thành viên

Qua phần trên ta nhận thấy X^+ được định nghĩa thông qua F^+ . Vấn đề nảy sinh khi nghiên cứu lý thuyết CSDL là: Cho trước tập các phụ thuộc hàm F và một phụ thuộc hàm f , Bài toán kiểm tra có hay không $f \in F^+$ gọi là *Bài toán thành viên*.

Để giải quyết Bài toán Bài toán thành viên thật sự không đơn giản; vì mặc dù F là rất nhỏ nhưng F^+ thì có thể rất lớn. Tuy nhiên ta có thể giải bằng cách tính X^+ và so sánh X^+ với tập Y . Dựa vào tính chất $X \rightarrow Y \in F^+ \Leftrightarrow Y \subseteq X^+$, ta có ngay câu trả lời $X \rightarrow Y \in F^+$ hay không? Như vậy thay vì giải Bài toán thành viên ta đưa về giải Bài toán tìm bao đóng của tập thuộc tính.

3.4. Thuật toán tìm bao đóng của một tập thuộc tính

Thuật toán tìm bao đóng với độ phức tạp $O(N^2)$, với N là số lượng thuộc tính của lược đồ quan hệ Q .

Dữ Liệu Vào $Q, F, X \subseteq Q^+$
 Dữ Liệu Ra X^+

```

    Bước 1:      Đặt  $X^+ = X$ 
    Bước 2:      Temp =  $X^+$ 
                  $\forall f \quad U \rightarrow V \in F$ 
                 if ( $U \subseteq X^+$ )
                      $X^+ = X^+ \cup V.$ 
                  $F = F - f;$ 
    Bước 3:      if ( $X^+ = \text{Temp}$ )
    
```

Ví dụ:	" X^+ chính là kết quả cần tìm "	
Cho	Dừng thuật toán	lược đồ quan hệ
tập phụ	else	$Q(\text{ABCDEFGH})$ và
$F = \{B \rightarrow$	trở lại Bước 2:	thuộc hàm
$H, GH \rightarrow C,$		$A, DA \rightarrow CE, D \rightarrow$
Tìm bao đóng của các tập $X = \{AC\}$ dựa trên F .		$AC \rightarrow D\}$.

Giải:

- $X^+ = AC$
- Đặt $\text{Temp} = X^+$
 - + Xét $AC \rightarrow D$, có $AC \subseteq X^+$: $X^+ = X^+ \cup D = ACD.$
 Loại $AC \rightarrow D$ khỏi F . Lặp bước 2
 - + Xét $DA \rightarrow CE$, có $DA \subseteq X^+$: $X^+ = X^+ \cup CE = ACDE.$
 Loại $DA \rightarrow CE$ khỏi F . Lặp bước 2
 - + Xét $D \rightarrow H$, có $D \subseteq X^+$: $X^+ = X^+ \cup H = ACDEH.$
 Loại $D \rightarrow H$ khỏi F Lặp bước 2

Vì các phụ thuộc hàm $U \rightarrow V$ còn lại không thỏa điều kiện $U \subseteq X^+$ nên $X^+ = \text{Temp}$.
 Thuật toán dừng.

Vậy $X^+ = \{ACDEH\}$

4. Khóa của lược đồ quan hệ - một số thuật toán tìm khóa

Mục tiêu: Trình bày được định nghĩa khóa của một lược đồ quan hệ và giải thuật xác định một khóa, xác định tập tất cả các khóa của một lược đồ quan hệ đã cho.

4.1. Định nghĩa khóa của quan hệ

Cho quan hệ $R(A_1, A_2, \dots, A_n)$ được xác định bởi tập thuộc tính R^+ và tập phụ thuộc hàm F định nghĩa trên R , cho $K \subseteq R^+$.

K là một khóa của R nếu thỏa đồng thời cả hai điều kiện sau:

$$1. K \subseteq R^+ \in F^+ \text{ (hay } K^+_{F^+} = R^+)$$

(K chỉ thỏa điều kiện 1 thì được gọi là siêu khóa)

$$2. \text{ Không tồn tại } K' \subset K \text{ sao cho } K'^+ = R^+$$

Tập $S \subseteq \{A_1, \dots, A_n\}$ là *siêu khóa* của R nếu S chứa khóa. Một lược đồ quan hệ có thể có nhiều siêu khóa, nhiều khóa.

4.2. Thuật toán tìm một khóa của một lược đồ quan hệ

$$K = Q^+;$$

While $A \in K$ do

$$\text{if } (K - A)^+ = Q^+ \text{ then } K = K - A$$

K còn lại chính là một khóa cần tìm.

Nếu muốn tìm các khóa khác (nếu có) của lược đồ quan hệ, ta có thể thay đổi thứ tự loại bỏ các phần tử của K .

Ví dụ:

Cho lược đồ quan hệ $R(ABC)$ và tập phụ thuộc hàm

$$F = \{A \rightarrow B; A \rightarrow C; B \rightarrow A\}$$

Hãy tìm một khóa của R .

Giải:

$$K = \{A, B, C\}$$

Loại thuộc tính A , do $(K - A)^+ = R^+$ nên $K = \{B, C\}$

thuộc tính B không loại được do $(K - B)^+ \neq R^+$ nên $K = \{B, C\}$

Loại thuộc tính C , do $(K - C)^+ = R^+$ nên $K = \{B\}$. Vậy một khóa của R là B .

4.3. Thuật toán tìm tất cả các khóa của một lược đồ quan hệ

Một số khái niệm hỗ trợ cho thuật toán tìm tất cả các khóa sau đây:

- Tập nguồn (TN): chứa tất cả thuộc tính chỉ xuất hiện ở vế trái mà không xuất hiện ở vế phải của tập phụ thuộc hàm và tập các thuộc tính không tham gia vào tập phụ thuộc hàm F .

- Tập đích (TD): chứa tất cả các thuộc tính chỉ xuất hiện ở vế phải mà không xuất hiện ở vế trái của tập phụ thuộc hàm.

- Tập trung gian (TG): chứa tất cả các thuộc tính tham gia vào cả 2 vế của tập phụ thuộc hàm.

Dữ liệu vào: Lược đồ quan hệ R và tập phụ thuộc hàm F .

Dữ liệu ra: Tất cả các khóa K của quan hệ.

Thuật toán:

Bước 0: Tìm tập thuộc tính nguồn (TN), tập thuộc tính trung gian (TG).

Tìm tất cả các tập con của tập trung gian gọi là X_i (bằng phương pháp duyệt nhị phân)

if $TG = \theta$ then
 $K = TN$; kết thúc.
 Ngược lại
 Qua bước 1

Bước 1 Tìm tất cả các tập con của TG: X_i

$S = \phi$

$\forall X_i \in TG$

if $(TN \cup X_i)^+ = R^+$ then

$S = S \cup \{TN \cup X_i\}$

{S là tập các siêu khoá cần tìm}

Bước 2: Tính $TN \cup X_i$

Bước 3: Tính $(TN \cup X_i)^+$

Bước 4: Nếu $X_i^+ = R^+$ thì X_i là siêu khoá

Nếu một tập con $TN \cup X_i$ có bao đóng đúng bằng R^+ thì $TN \cup X_i$ là một siêu khoá của R.

Giả sử sau bước này có m siêu khoá: $S = \{S_1, S_2, \dots, S_m\}$

Bước 5 : Xây dựng tập chứa tất cả các khoá của R từ tập S

Xét mọi S_i, S_j con của S ($i \neq j$), nếu $S_i \subset S_j$ thì ta loại S_j ($i, j = 1..m$), kết quả còn lại chính là tập tất cả các khoá cần tìm.

Ví dụ: Cho lược đồ quan hệ R(ABC) và tập phụ thuộc hàm

$F = \{A \rightarrow B; A \rightarrow C; B \rightarrow A\}$

Hãy tìm tất cả các khóa của R.

Giải: Áp dụng thuật tìm tất cả các khóa đã cho ở trên ta có:

$TN = \{\phi\}$; $TG = \{A, B\}$

Gọi X_i là tập con của tập trung gian. Ta lập bảng như sau:

X_i	$TN \cup X_i$	$(TN \cup X_i)^+$	Siêu khoá	Khóa
ϕ	ϕ	ϕ	-	-
A	A	ABC	A	A
B	B	ABC	B	B
AB	AB	ABC	AB	-

Vậy lược đồ quan hệ R có hai khóa $K_1 = \{A\}$, $K_2 = \{B\}$

5. Phủ tối thiểu

Mục tiêu: Trình bày giải thuật xác định một phủ tối thiểu của tập phụ thuộc hàm đã có sẵn, qua đó trình bày các khái niệm và cách xác định tập phụ thuộc hàm có vẻ phải một thuộc tính, tập phụ thuộc hàm có vẻ trái không dư thừa và tập phụ hàm đầy đủ.

5.1. Tập phụ thuộc hàm tương đương

Cho F và G là hai tập phụ thuộc hàm, ta nói F và G tương đương (hay F phủ G hoặc G phủ F) và ký hiệu là $F^+ = G^+$ nếu và chỉ nếu mỗi phụ thuộc hàm thuộc F đều thuộc G^+ và mỗi phụ thuộc hàm thuộc G đều thuộc F^+ .

Ta nói F phủ G nếu $G^+ \subseteq F^+$

Chẳng hạn cho lược đồ quan hệ $Q(ABCDEFGH)$, thì hai tập phụ thuộc hàm F và G (xác định trên Q) là tương đương.

$F = \{B \rightarrow A; DA \rightarrow CE; D \rightarrow H; GH \rightarrow C; AC \rightarrow D; DG \rightarrow C\}$

$G = \{B \rightarrow A; DA \rightarrow CE; D \rightarrow H; GH \rightarrow C; AC \rightarrow D; BC \rightarrow AC; BC \rightarrow D; DA \rightarrow AH; AC \rightarrow DEH\}$

(Việc kiểm tra các phụ thuộc hàm trong G có được suy diễn từ F và ngược lại xem như Bài tập dành cho bạn đọc).

5.2. Phủ tối thiểu

F_{tt} được gọi là tập phụ thuộc hàm tối thiểu (hay phủ tối thiểu) nếu F thỏa đồng thời ba điều kiện sau:

1. F là tập phụ thuộc hàm có vế trái không dư thừa.
2. F là tập phụ thuộc hàm có vế phải một thuộc tính.
3. F là tập phụ thuộc hàm không dư thừa.

5.2.1. Phụ thuộc hàm có vế trái dư thừa:

F là tập phụ thuộc hàm trên lược đồ quan hệ Q , Z là tập thuộc tính, $Z \rightarrow Y \in F$. Nói rằng phụ thuộc hàm $Z \rightarrow Y$ có vế trái dư thừa (phụ thuộc không đầy đủ) nếu có một $A \in Z$ sao cho:

$$F \equiv F - \{Z \rightarrow Y\} \cup \{(Z-A) \rightarrow Y\}$$

Ngược lại $Z \rightarrow Y$ là phụ thuộc hàm có vế trái không dư thừa hay Y phụ thuộc hàm đầy đủ vào Z (phụ thuộc hàm đầy đủ).

Ta nói F là tập phụ thuộc hàm có vế trái không dư thừa nếu F không chứa phụ thuộc hàm có vế trái dư thừa.

Thuật toán loại khỏi F các phụ thuộc hàm có vế trái dư thừa:

Bước 1: - Xét lần lượt các phụ thuộc hàm $X \rightarrow Y$ của F .

Bước 2: - Với mọi tập con thực sự $X' \neq \emptyset$ của X .

- Nếu $X' \rightarrow Y \in F^+$ thì thay $X \rightarrow Y$ trong F bằng $X' \rightarrow Y$.

- Lặp lại bước 2.

5.2.2. Tập phụ thuộc hàm có vế phải một thuộc tính:

Mỗi tập phụ thuộc hàm F đều tương đương với tập phụ thuộc hàm G mà vế phải của các phụ thuộc hàm trong G chỉ gồm một thuộc tính.

G được gọi là tập phụ thuộc hàm có vế phải một thuộc tính.

Ví dụ:

$F = \{A \rightarrow BC, B \rightarrow C, AB \rightarrow D\}$ ta suy ra

$$F \equiv \{A \rightarrow B, A \rightarrow C, B \rightarrow C, AB \rightarrow D\} = G$$

5.2.3. Tập phụ thuộc hàm không dư thừa:

Nói rằng F là tập phụ thuộc hàm không dư thừa nếu không tồn tại $F' \subset F$ sao cho $F' \equiv F$. Ngược lại F là tập phụ thuộc hàm dư thừa.

Thuật toán loại khỏi F các phụ thuộc hàm dư thừa:

Bước 1: - Lần lượt xét các phụ thuộc hàm $X \rightarrow Y$ của F

Bước 2: - Nếu $X \rightarrow Y$ là thành viên của $F - \{X \rightarrow Y\}$ thì loại $X \rightarrow Y$ khỏi F .

Bước 3: - Lặp lại bước 2 cho các phụ thuộc hàm tiếp theo của F .

5.3. Thuật toán tìm phủ tối thiểu

Từ điều kiện xác định phủ tối thiểu, ta có thuật toán tìm phủ tối thiểu như sau:

Thuật toán:

Bước 1: - Loại khỏi F các phụ thuộc hàm có vế trái dư thừa.

Bước 2: - Tách các phụ thuộc hàm có vế phải trên một thuộc tính thành các phụ thuộc hàm có vế phải một thuộc tính.

Bước 3: - Loại khỏi F các phụ thuộc hàm dư thừa.

Chú ý: Theo thuật toán trên, có thể tìm được nhiều hơn một phủ tối thiểu F_{tt} để $F \equiv F_{tt}$ và nếu thứ tự loại các phụ thuộc hàm khác nhau sẽ thu được các phủ tối thiểu khác nhau.

Ví dụ: cho $R(\text{MSCD}, \text{MSSV}, \text{CD}, \text{HG})$ và tập phụ thuộc hàm F:

$F = \{\text{MSCD} \rightarrow \text{CD}; \text{CD} \rightarrow \text{MSCD}; \text{CD}, \text{MSSV} \rightarrow \text{HG}; \text{MSCD}, \text{HG} \rightarrow \text{MSSV}; \text{CD}, \text{HG} \rightarrow \text{MSSV}; \text{MSCD}, \text{MSSV} \rightarrow \text{HG}\}$

Hãy tìm một F_{tt} của F?

Kết quả ta có được một phủ tối thiểu sau:

$F_{tt} = \{\text{MSCD} \rightarrow \text{CD}; \text{CD} \rightarrow \text{MSCD}; \text{CD}, \text{HG} \rightarrow \text{MSSV}; \text{MSCD}, \text{MSSV} \rightarrow \text{HG}\}$

6. Bài tập thực hành

Cho lược đồ quan hệ $R = \text{ABCDEFGHI}$ và tập phụ thuộc hàm F như sau:

$F = \{ \text{AB} \rightarrow \text{CD}, \text{C} \rightarrow \text{EH}, \text{G} \rightarrow \text{A}, \text{G} \rightarrow \text{H}, \text{CE} \rightarrow \text{H} \}$

1. Tìm khoá của F.

2. Tìm phủ tối thiểu (S) của tập phụ thuộc hàm F.

6.1. Các bước thực hiện

❖ **Tìm khoá:**

Bước 0: Tìm tập thuộc tính nguồn (TN), tập thuộc tính trung gian (TG).

Tìm tất cả các tập con của tập trung gian gọi là X_i (bằng phương pháp duyệt nhị phân)

if $TG = \emptyset$ then

$K = TN$; kết thúc.

Ngược lại

Qua bước 1

Bước 1: Tìm tất cả các tập con của TG: X_i

$S = \emptyset$

$\forall X_i \in TG$

if $(TN \cup X_i)^+ = R^+$ then

$S = S \cup \{TN \cup X_i\}$

{S là tập các siêu khoá cần tìm}

Bước 2: Tính $TN \cup X_i$

Bước 3: Tính $(TN \cup X_i)^+$

Bước 4: Nếu $X_i^+ = R^+$ thì X_i là siêu khoá

Nếu một tập con $TN \cup X_i$ có bao đóng đúng bằng R^+ thì $TN \cup X_i$ là một siêu khoá của R.

Giả sử sau bước này có m siêu khoá: $S = \{S_1, S_2, \dots, S_m\}$

Bước 5: Xây dựng tập chứa tất cả các khoá của R từ tập S

Xét mọi S_i, S_j con của S ($i \neq j$), nếu $S_i \subset S_j$ thì ta loại S_j ($i, j = 1..m$), kết quả còn lại chính là tập tất cả các khoá cần tìm.

❖ **Tìm phủ tối thiểu**

Bước 1: Loại khỏi F các phụ thuộc hàm có vế trái dư thừa.

Bước 2: Tách các phụ thuộc hàm có vế phải trên một thuộc tính thành các phụ thuộc hàm có vế phải một thuộc tính.

Bước 3: Loại khỏi F các phụ thuộc hàm dư thừa.

6.2. Sinh viên thực hành

❖ **Tìm khoá:**

Kết quả: Khoá là BG

❖ **Tìm phủ tối thiểu**

Kết quả: Phủ tối thiểu

$S = \{ AB \rightarrow C, AB \rightarrow D, C \rightarrow E, C \rightarrow E, G \rightarrow A, G \rightarrow H \}$

Những trọng tâm cần chú ý trong bài

- Thực hiện đúng các bước tìm khoá
- Thực hiện đúng các bước tìm phủ tối thiểu

Bài tập mở rộng và nâng cao

Bài 1:

a. Cho lược đồ quan hệ Q(ABCD) và tập phụ thuộc hàm $F = \{A \rightarrow B; BC \rightarrow D\}$.

Những phụ thuộc hàm nào sau đây thuộc F^+ ?

$C \rightarrow D; A \rightarrow D; AD \rightarrow C; AC \rightarrow D; BC \rightarrow A; B \rightarrow CD.$

b. Cho lược đồ quan hệ Q(ABCDEFGH) và tập phụ thuộc hàm

$F = \{ AB \rightarrow C; B \rightarrow D; CD \rightarrow E; CE \rightarrow GH; G \rightarrow A \}$

Những phụ thuộc hàm nào sau đây không thuộc vào F^+ ?

$AB \rightarrow E; AB \rightarrow GH; CGH \rightarrow E; CB \rightarrow E; GB \rightarrow E.$

c. Cho lược đồ quan hệ Q,F như sau: với Q(ABCD) $F = \{A \rightarrow B; A \rightarrow C\}$.

Trong các phụ thuộc hàm sau, những phụ thuộc hàm được suy ra từ F ?

$A \rightarrow D; C \rightarrow D; AB \rightarrow B; BC \rightarrow A; A \rightarrow BC$

Bài 2:

Cho lược đồ quan hệ Q(ABCD) và tập phụ thuộc hàm

$F = \{ A \rightarrow D; D \rightarrow A; AB \rightarrow C \}$

a. Tính AC^+

b. Chứng minh $BD \rightarrow C$

Bài 3:

a. Q(ABCDEG) Cho $F = \{ AB \rightarrow C; C \rightarrow A; BC \rightarrow D; ACD \rightarrow B; D \rightarrow EG; BE \rightarrow C; CG \rightarrow BD; CE \rightarrow AG \}$

$X = [BD], X^+ = ?$

$Y = [CG], Y^+ = ?$

b. Cho lược đồ quan hệ Q và tập phụ thuộc hàm $F = \{ AB \rightarrow E; AG \rightarrow I; BE \rightarrow I; E \rightarrow G; GI \rightarrow H \}$. Chứng minh rằng $AB \rightarrow GH$.

c. Tương tự cho tập phụ thuộc hàm

$F = \{ AB \rightarrow C; B \rightarrow D; CD \rightarrow E; CE \rightarrow GH; G \rightarrow A \}$

Chứng minh rằng $AB \rightarrow E; AB \rightarrow G$

d. Q(ABCDEFGH)

$F = \{ B \rightarrow A; DA \rightarrow CE; D \rightarrow H; GH \rightarrow C; AC \rightarrow D \}$ Hãy tìm một khoá của Q ?

Bài 4:

Hãy tìm tất cả các khoá cho lược đồ quan hệ sau:

Q(BROKER,OFFICE,STOCK,QUANTITY,INVESTOR,DIVIDENT)

$F = \{ STOCK \rightarrow DIVIDENT, INVESTOR \rightarrow BROKER, INVESTOR, STOCK \rightarrow QUANTITY, BROKER \rightarrow OFFICE \}$

Bài 5:

Cho $Q(A,B,C,D)$, $F = \{AB \rightarrow C; D \rightarrow B; C \rightarrow ABD\}$. Hãy tìm tất cả các khoá của Q

Bài 6:

Cho lược đồ quan hệ $Q(MSCD, MSSV, CD, HG)$ và tập phụ thuộc F như sau:

$F = \{MSCD \rightarrow CD; CD \rightarrow MSCD; CD, MSSV \rightarrow HG; MSCD, HG \rightarrow MSSV; CD, HG \rightarrow MSSV; MSCD, MSSV \rightarrow HG\}$

Hãy tìm phủ tối thiểu của F .

Bài 7:

Xác định phủ tối thiểu của tập phụ thuộc hàm sau: $Q(ABCDEG)$

$F = \{AB \rightarrow C; C \rightarrow A; BC \rightarrow D; ACD \rightarrow B; D \rightarrow EG; BE \rightarrow C; CG \rightarrow BD; CE \rightarrow AG\}$

Bài 8:

a. Cho $Q(ABCD)$ và $F = \{AB \rightarrow C; D \rightarrow B; C \rightarrow ABD\}$.

Hãy kiểm tra xem $AB \rightarrow D$ có thuộc F^+ hay không ?

Hãy tìm tất cả các khoá của lược đồ quan hệ Q .

b. Cho $Q(A,B,C,D)$ và $F = \{C \rightarrow A; A \rightarrow C; AD \rightarrow B; BC \rightarrow D; AB \rightarrow D; CD \rightarrow B\}$ Hãy tìm phủ tối thiểu của F .

Trình bày được khái niệm cơ bản của lý thuyết cơ sở dữ liệu như phụ thuộc hàm, bao đóng, khoá, phủ tối thiểu.

Áp dụng được thuật toán tìm khoá, tìm phủ tối thiểu.

Rèn luyện tính cẩn thận, chính xác trong phân tích và giải quyết vấn đề.

Yêu cầu về đánh giá kết quả học tập bài 4

Nội dung:

+ Về kiến thức: Trình bày được khái niệm cơ bản của lý thuyết cơ sở dữ liệu như phụ thuộc hàm, bao đóng, khoá, phủ tối thiểu.

+ Về kỹ năng: Áp dụng được thuật toán tìm khoá, tìm phủ tối thiểu.

+ Năng lực tự chủ và trách nhiệm: Rèn luyện tính cẩn thận, chính xác trong việc học và làm Bài tập.

Phương pháp:

+ Về kiến thức: Được đánh giá bằng hình thức kiểm tra viết, trắc nghiệm, vấn đáp

+ Về kỹ năng: Đánh giá kỹ năng thực hành giải các Bài toán tìm khoá, tìm phủ tối thiểu.

+ Năng lực tự chủ và trách nhiệm: Rèn luyện tính cẩn thận, chính xác trong việc học và làm Bài tập.

TÀI LIỆU THAM KHẢO

1. Lê Tiên Vương, Nhập môn cơ sở dữ liệu quan hệ, Nhà xuất bản Giáo dục.
2. Vũ Đức Thi, Cơ sở dữ liệu kiến thức và thực hành, Nhà xuất bản thống kê.
3. Nguyễn An Tế, Giáo trình nhập môn cơ sở dữ liệu, ĐHKHTN- ĐHQGTPHCM.
4. Đỗ Trung Tuấn, Cơ sở dữ liệu, Nhà xuất bản giáo dục.
5. Giáo trình Cơ sở dữ liệu, ĐHCN TPHCM.