

ỦY BAN NHÂN DÂN THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG CAO ĐẲNG LÝ TỰ TRỌNG THÀNH PHỐ HỒ CHÍ MINH

GIÁO TRÌNH

MÔN HỌC/MÔ ĐUN: CƠ SỞ DỮ LIỆU

NGÀNH/NGHỀ: TIN HỌC ỨNG DỤNG, LẬP TRÌNH MÁY TÍNH, THIẾT KẾ
TRANG WEB, AN NINH MẠNG, QUẢN TRỊ MẠNG MÁY TÍNH, CÔNG NGHỆ
KỸ THUẬT MÁY TÍNH, ĐỒ HỌA ĐA TRUYỀN THÔNG, THƯƠNG MẠI
ĐIỆN TỬ

TRÌNH ĐỘ: TRUNG CẤP

*Ban hành kèm theo Quyết định số: /QĐ-LTT-ĐT ngày tháng năm 2021 của
Hiệu trưởng Trường Cao đẳng Lý Tự Trọng Thành phố Hồ Chí Minh*

TP. HCM năm 2021

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

Giáo trình môn Cơ sở dữ liệu được biên soạn dựa theo chương trình chi tiết học phần Cơ sở dữ liệu của hệ Cao đẳng ngành Công nghệ Thông tin. Đây là môn học cơ sở ngành và là môn học chung của ngành Công nghệ Thông tin.

Giáo trình cung cấp một số kiến thức căn bản về mô hình quan hệ, ràng buộc toàn vẹn, ngôn ngữ truy vấn SQL và chuẩn hóa Cơ sở dữ liệu với mục tiêu là giúp -SV có khả năng thiết kế Cơ sở dữ liệu đạt chuẩn.

Giáo trình là tài liệu giảng dạy và học tập chính thức cho SV khoa Công nghệ Thông tin trường CĐ Lý Tự Trọng TP. HCM.

Thành phố Hồ Chí Minh, ngày 01 tháng 08 năm 2021

Giảng viên biên soạn: Nguyễn Văn Giang

MỤC LỤC

CHƯƠNG 1: MÔ HÌNH QUAN HỆ	4
1. CƠ SỞ DỮ LIỆU VÀ HỆ QUẢN TRỊ CƠ SỞ DỮ LIỆU	4
1.1. Cơ sở dữ liệu (<i>Database, DB</i>).....	4
1.2. Hệ quản trị cơ sở dữ liệu (<i>Database Management System, DBMS</i>)	4
1.3. Người dùng (<i>User</i>)	5
2. MÔ HÌNH QUAN HỆ (<i>RELATIONAL MODEL</i>)	5
2.1. Khái niệm mô hình quan hệ	5
2.2. Thuộc tính (<i>Attribute</i>)	7
2.3. Lược đồ quan hệ (<i>Relation Schema</i>).....	7
2.4. Quan hệ (<i>Relation</i>)	8
2.5. Bộ (<i>Tuple</i>)	9
2.6. Khóa (<i>key</i>).....	9
BÀI TẬP CHƯƠNG 1.....	10
Chương 2: RÀNG BUỘC TOÀN VỆN.....	14
1. RÀNG BUỘC TOÀN VỆN – CÁC YẾU TỐ CỦA RÀNG BUỘC TOÀN VỆN... ..	14
1.1. Ràng buộc toàn vẹn	14
1.2. Các yếu tố của ràng buộc toàn vẹn.....	15
2. PHÂN LOẠI RÀNG BUỘC TOÀN VỆN.....	16
2.1. Ràng buộc toàn vẹn có bối cảnh là một quan hệ.....	16
2.2. Ràng buộc toàn vẹn có bối cảnh là nhiều quan hệ	17
BÀI TẬP CHƯƠNG 2.....	18
CHƯƠNG 3: PHỤ THUỘC HÀM	19
1. PHỤ THUỘC HÀM	19
1.1. Khái niệm phụ thuộc hàm	19
1.1.1. Định nghĩa	20
1.1.2. Phụ thuộc hàm hiển nhiên (<i>Trivial Dependency</i>)	21
1.1.3. Bao đóng của tập phụ thuộc hàm :	21
Hệ luật dẫn Armstrong (<i>Phần đọc thêm</i>)	21

1.2. Bao đóng của tập thuộc tính.....	22
1.2.1. Định nghĩa	22
1.2.2. Thuật toán tìm X^+	22
2. PHỦ TỐI THIỂU CỦA TẬP PHỤ THUỘC HÀM	24
2.1. Phủ của tập phụ thuộc hàm	24
2.2. Phụ thuộc hàm đầy đủ	25
2.3. Tập phụ thuộc hàm không dư thừa.....	25
2.4. Phủ tối thiểu (Minimal Cover)	25
3. KHÓA CỦA LƯỢC ĐỒ QUAN HỆ	27
3.1. Định nghĩa	27
3.2. Thuật toán tìm tất cả các khóa của lược đồ quan hệ	27
BÀI TẬP CHƯƠNG 3.....	29
Chương 4: CHUẨN HÓA CƠ SỞ DỮ LIỆU	33
1 DẠNG CHUẨN CỦA LƯỢC ĐỒ QUAN HỆ	33
1.1. Dạng chuẩn 1 (First Normal Form).....	33
1.2. Dạng chuẩn 2 (Second Normal Form)	34
1.3. Dạng chuẩn 3 (Third Normal Form)	35
1.4. Dạng chuẩn của lược đồ cơ sở dữ liệu	37
2. CHUẨN HÓA LƯỢC ĐỒ BẰNG CÁCH PHÂN RÃ	37
2.1. Thuật toán phân rã lược đồ quan hệ về dạng chuẩn 3 bảo toàn thông tin và phụ thuộc hàm	37
2.2. Ví dụ tổng hợp.....	38
BÀI TẬP CHƯƠNG 4.....	41
CÁC ĐỀ THI MẪU.....	43
ĐỀ SỐ 1	43
ĐỀ SỐ 2	44
ĐỀ SỐ 3	46
ĐỀ SỐ 4	47
ĐỀ SỐ 5	48

GIÁO TRÌNH MÔN HỌC/MÔ ĐUN

Tên môn học/mô đun: CƠ SỞ DỮ LIỆU

Mã môn học/mô đun:

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

- Vị trí: Học song song hoặc sau môn học Tin học.
- Tính chất: Là môn học bắt buộc nằm trong các môn học cơ sở. Môn học này yêu cầu phải có tư duy logic và kiến thức về toán lớp 10 (tập hợp).

Mục tiêu của môn học/mô đun:

- **Về kiến thức:** Sau khi học xong người học có các kiến thức cơ bản về mô hình quan hệ, các ràng buộc toàn vẹn, ngôn ngữ SQL, phụ thuộc hàm, khóa, phủ tối thiểu, các dạng chuẩn,....
- **Về kỹ năng:**
 - + Xác định được khóa chính và khóa ngoại trong lược đồ cơ sở dữ liệu (LĐ CSDL)
 - + Xác định được các ràng buộc toàn vẹn trong LĐ CSDL
 - + Tìm được tất cả các khóa của lược đồ quan hệ (LĐQH)
 - + Tìm được phủ tối thiểu của tập phụ thuộc hàm
 - + Xác định được dạng chuẩn cao nhất của LĐQH, LĐCSDL.
 - + Chuẩn hóa LĐQH về dạng chuẩn 3
 - + Thực hiện được các câu truy vấn đơn giản bằng ngôn ngữ SQL
- **Về năng lực tự chủ và trách nhiệm:**
 - + Có khả năng tự duy độc lập trong việc phân tích một bài toán thực tế để xây dựng CSDL đạt dạng chuẩn 3.
 - + Có khả năng tự tìm tòi, học hỏi, tích lũy kinh nghiệm liên quan đến môn học.

Nội dung của môn học/mô đun:

CHƯƠNG 1: MÔ HÌNH QUAN HỆ

Giới thiệu:

- Mô hình Dữ liệu Quan hệ (Relational Data Model) do TS. E. F. Codd đưa ra năm 1970.
- Đây là mô hình cung cấp một cấu trúc dữ liệu đơn giản và đồng bộ dựa trên khái niệm quan hệ, là khái niệm toán học dựa trên nền tảng lý thuyết tập hợp.
- Là cơ sở của các Hệ quản trị Cơ sở Dữ liệu như Oracle, SQL Server,...

Mục tiêu:

Sau khi học xong người học có khả năng:

- Hiểu được mục đích, vị trí, vai trò của môn học;
- Phát biểu được các khái niệm trong mô hình quan hệ;
- Xác định được khóa chính và khóa ngoại trong LĐ CSDL

Nội dung chính:

1. CƠ SỞ DỮ LIỆU VÀ HỆ QUẢN TRỊ CƠ SỞ DỮ LIỆU

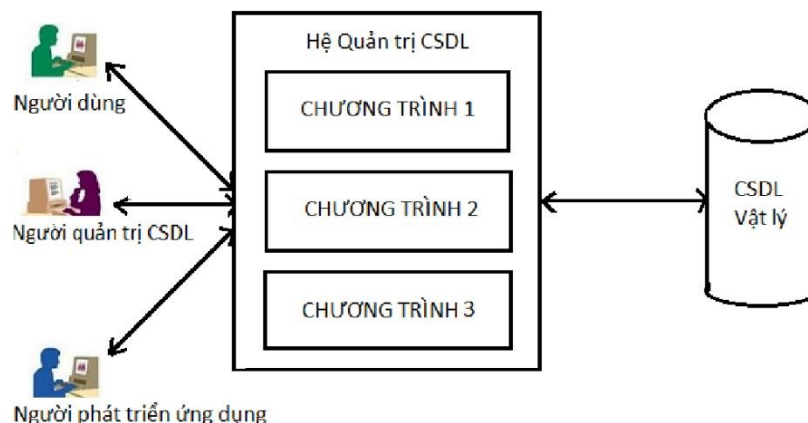
1.1. Cơ sở dữ liệu (Database, DB)

Cơ sở dữ liệu (CSDL) là tập hợp các thông tin được tổ chức một cách riêng biệt, có cấu trúc, phục vụ cho một tổ chức.

1.2. Hệ quản trị cơ sở dữ liệu (Database Management System, DBMS)

Một hệ quản trị cơ sở dữ liệu (HQTCSDL) là một tập hợp các phần mềm quản lý CSDL và cung cấp các dịch vụ xử lý CSDL.

- HQTCSDL cung cấp một giao diện giữa người sử dụng và dữ liệu, biến đổi CSDL vật lý thành CSDL logic.



- Một HQTCSDL phải đáp ứng các yêu cầu sau :
 - + Có khả năng quản lý một tập hợp lớn dữ liệu.
 - + Đảm bảo sự đáng tin cậy (như là cung cấp các chức năng sao lưu dự phòng và khôi phục dữ liệu để đề phòng trường hợp gặp sự cố)
 - + Đảm bảo sự bí mật của dữ liệu (như là sự phân quyền, cho phép truy xuất hay không đối với người dùng).
 - + Đảm bảo sự hiệu quả (tốn một lượng thích hợp về thời gian và không gian cho việc truy xuất dữ liệu, đặc biệt là khi dữ liệu được dùng đồng thời bởi nhiều người).
- Một CSDL là một tập hợp dữ liệu được quản lý bởi một HQTCSDL.
- Các ví dụ về HQTCSDL là MS Access, SQL Server và Oracle Database System,...

1.3. Người dùng (User)

Người dùng khai thác CSDL thông qua HQTCSDL có thể phân thành ba loại: người quản trị CSDL, người phát triển và lập trình ứng dụng, người dùng cuối.

- Người quản trị CSDL: chịu trách nhiệm quản lý và bảo trì CSDL như:
 - + Sự chính xác, toàn vẹn của dữ liệu và bảo mật của CSDL.
 - + Lưu phòng hồ và phục hồi CSDL khi gặp sự cố.
 - + Bảo đảm sự hoạt động trôi chảy và hiệu quả của CSDL và HQTCSDL
- Người phát triển và lập trình ứng dụng: là những người chuyên nghiệp về máy tính có trách nhiệm thiết kế, tạo dựng và bảo trì hệ thống tin cho người dùng cuối.
- Người dùng cuối: là những người có thể không chuyên về máy tính, có trách nhiệm cụ thể trong tổ chức. Họ khai thác CSDL thông qua HQTCSDL.

2. MÔ HÌNH QUAN HỆ (RELATIONAL MODEL)

2.1. Khái niệm mô hình quan hệ

Mô hình Dữ liệu Quan hệ (gọi tắt là Mô hình Quan hệ) bao gồm:

- Một hệ thống các ký hiệu để mô tả dữ liệu dưới dạng dòng và cột như quan hệ, bộ, thuộc tính, khóa chính, khoá ngoại, ...
- Một tập hợp các phép toán thao tác trên dữ liệu như các phép toán đại số quan hệ.
- Ràng buộc toàn vẹn trên các quan hệ.

Tại sao phải dùng Mô hình quan hệ ?

- Giảm dư thừa về dữ liệu (dữ liệu trùng lặp không cần thiết)
- Giảm hay loại trừ các kết quả không mong muốn khi cập nhật CSDL.

Ví dụ 1:

MASV	HOTEN	MONHOC	TENKHOA	DIE MTH I
TH102	TRẦN VĂN BÌNH	CƠ SỞ DỮ LIỆU	CÔNG NGHỆ THÔNG TIN	3.0
TH300	NGUYỄN BÍCH HẠNH	CƠ SỞ DỮ LIỆU	CÔNG NGHỆ THÔNG TIN	8.0
TH102	TRẦN VĂN BÌNH	TIN HỌC VĂN PHÒNG	CÔNG NGHỆ THÔNG TIN	6.0
CK905	LE THANH TRUNG	TIN HỌC VĂN PHÒNG	CƠ KHÍ	5.0

Chẳng hạn với bài toán quản lý điểm thi của sinh viên, nếu lưu trữ dữ liệu theo dạng bảng như trên thì các giá trị của các cột HOTEN, MONHOC, TENKHOA sẽ bị trùng lặp. Sự trùng lặp này gây nên một số vấn đề:

- Ta không thể lưu trữ một sinh viên mới khi sinh viên này chưa có điểm thi
- Khi cần sửa đổi họ tên sinh viên thì ta phải sửa tất cả các dòng có liên quan đến sinh viên này. Điều này dễ gây ra tình trạng dữ liệu thiếu nhất quán.
- Khi có nhu cầu xóa điểm thi của một sinh viên kéo theo khả năng xóa luôn họ tên sinh viên đó.

Để khắc phục các vấn đề trên chúng ta phải phân rã bảng trên thành các bảng con như sau :

MASV	MAMH	DIEM	MASV	TENSV	MAKHOA
TH102	CSDL	3.0	TH102	TRẦN VĂN BÌNH	CNTT
TH300	CSDL	8.0	TH300	NGUYỄN BÍCH HẠNH	CNTT
TH102	THVP	6.0	CK905	LE THANH TRUNG	CK
CK905	THVP	5.0			

MAMH	TENMH	MAKHOA	TENKHOA
CSDL	CƠ SỞ DỮ LIỆU	CNTT	CÔNG NGHỆ THÔNG TIN
THVP	TIN HỌC VĂN PHÒNG	CK	CƠ KHÍ

Vấn đề là làm sao biết một cơ sở dữ liệu có dư thừa và không hiệu quả và phương pháp khắc phục như thế nào. Môn học sẽ cung cấp các kiến thức để trả lời câu hỏi này.

2.2. Thuộc tính (Attribute)

Thuộc tính là đặc trưng, tính chất riêng biệt của đối tượng cần được lưu trữ để phục vụ cho việc khai thác CSDL.

Chẳng hạn với bài toán quản lý điểm thi của sinh viên, với đối tượng sinh viên ta cần phải chú ý đến các đặc trưng riêng như họ tên, ngày sinh, phái, tỉnh thường trú, lớp mà sinh viên theo học,... các đặc trưng này gọi là **thuộc tính**.

Các thuộc tính phải thuộc vào một kiểu dữ liệu nhất định (số, chuỗi, ngày tháng, logic, hình ảnh,...). Trong cùng một đối tượng không được có hai thuộc tính cùng tên.

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. Ví dụ điểm thi của sinh viên chỉ là các số nguyên từ 0 đến 10.

2.3. Lược đồ quan hệ (Relation Schema)

Lược đồ quan hệ (LĐQH) là tập tất cả các thuộc tính cần quản lý của một đối tượng cùng với mối liên hệ giữa chúng.

Lược đồ quan hệ Q với tập thuộc tính $\{A_1, A_2, \dots, A_n\}$ được viết là $Q(A_1, A_2, \dots, A_n)$. Tập các thuộc tính của Q được ký hiệu là Q^+ .

Ví dụ: Cho lược đồ quan hệ sinh viên (Đặt tên là SV) mô tả như sau:

SV(MASV, TENSX, GIOITINH, NGAYSINH, MALOP, DIACHI, SDT, DIEMXT)

$Q^+ = \{ \text{MASV, TENSX, GIOITINH, NGAYSINH, MALOP, DIACHI, SDT, DIEMXT} \}$

Thường khi thành lập một lược đồ, người thiết kế luôn gán cho nó một ý nghĩa nhất định, ý nghĩa đó gọi là **tân từ** của lược đồ quan hệ đó. Dựa vào tân từ người ta xác định được tập thuộc tính khóa của lược đồ quan hệ (khái niệm khoá sẽ được trình bày ở phần sau).

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. Chẳng hạn tân từ của lược đồ quan hệ trên là:

"Mỗi sinh viên có một mã duy nhất (MASV) để phân biệt với các sinh viên khác. Mỗi MASV xác định tất cả các thông tin về sinh viên đó như họ tên (TENSX), giới tính (GIOITINH), ngày sinh (NGAYSINH), lớp mà sinh viên đó theo học (MALOP), địa chỉ (DIACHI), số điện thoại (SDT), điểm xét tuyển đầu vào (DIEMXT)"

Nhiều lược đồ quan hệ cùng nằm trong một hệ thống quản lý được gọi là một **lược đồ cơ sở dữ liệu**.

Ví dụ 1.2: Cho lược đồ cơ sở dữ liệu để quản lý điểm sinh viên, gồm các lược đồ quan hệ sau:

SV(MASV, TENSX, GIOITINH, NGAYSINH, MALOP, DIACHI, SDT, DIEMXT)

Lop(MALOP, TENLOP, MAKHOA)

Khoa(MAKHOA, TENKHOA, DTKHOA)

Mon(MAMH, TENMH, LT, TH)

KQ(MASV, MAMH, DIEM)

Phân giải thích các thuộc tính:

Thuộc tính	Mô tả	Thuộc tính	Mô tả
MASV	Mã sinh viên	MAKHOA	Mã khoa
TENSX	Họ tên sinh viên	DTKHOA	Số điện thoại khoa
GIOITINH	Giới tính (Nữ=Yes)	TENKHOA	Tên khoa
NGAYSINH	Ngày sinh	DTKHOA	Số điện thoại khoa
DIACHI	Địa chỉ	TENMH	Tên môn học
SDT	Số điện thoại	LT	Số tiết lý thuyết
DIEMXT	Điểm xét tuyển	TH	Số tiết thực hành
MALOP	Mã lớp	DIEM	Điểm tổng kết môn học
TENLOP	Tên lớp		

2.4. Quan hệ (Relation)

Quan hệ là sự thể hiện của một lược đồ quan hệ một thời điểm nào đó. Một lược đồ quan hệ có thể có rất nhiều quan hệ. Thường ta dùng các ký hiệu như R, S, Q để chỉ các lược đồ quan hệ, còn quan hệ được định nghĩa trên nó tương ứng được ký hiệu là r, s, q.

Ví dụ 1.3: Xét LĐQH Khoa(MAKHOA, TENKHOA, DTKHOA)

Có 2 quan hệ T_{Khoa_1} và T_{Khoa_2} tại hai thời điểm như sau :

T_{Khoa_1} :

MAKHOA	TENKHOA	DTKHOA
CNTT	Công nghệ Thông tin	8110775
DDT	Điện - Điện Tử	8116819
CK	Cơ khí	4490594
...		

T_{Khoa_2} :

MAKHOA	TENKHOA	DTKHOA
CNTT	Công nghệ Thông tin	3811.0521-456
DDT	Điện - Điện Tử	3811.0521-678
CK	Cơ khí	3811.0521-579
...		

2.5. Bộ (Tuple)

Bộ là tập mỗi giá trị liên quan của tất cả các thuộc tính của một lược đồ quan hệ.

Trong ví dụ 1.3, quan hệ T_{Khoa_1} có các bộ:

(“CNTT”, “Công nghệ Thông Tin”, “8110775”), (“DDT”, “Điện-Điện tử” “8116819”),...

Thường người ta dùng các chữ cái thường (như t, p, q, ...) để biểu diễn các bộ. Chẳng hạn để nói bộ t thuộc quan hệ r ta viết: $t \in r$.

Về trực quan thì mỗi quan hệ xem như một bảng, trong đó mỗi cột là thông tin về một thuộc tính, mỗi dòng là một bộ.

2.6. Khóa (key)

2.6.1. Siêu khóa (Super key)

Cho lược đồ quan hệ Q, $S \subseteq Q^+$. S được gọi là một **siêu khóa** của lược đồ quan hệ Q nếu với hai bộ tùy ý trong mọi quan hệ quan hệ r của Q thì giá trị của các thuộc tính trong S là khác nhau.

Trong ví dụ 1.3, lược đồ quan hệ Khoa có các siêu khóa là :

MAKHOA, (MAKHOA, TENKHOA), (MAKHOA, TENKHOA, DTKHOA)

2.6.2. Khóa dự tuyển (Candidate key)

Một lược đồ quan hệ có thể có nhiều siêu khóa. Siêu khóa không chứa các siêu khóa khác được gọi là **khóa dự tuyển**.

Trong ví dụ 1.2:

- Lược đồ quan hệ SV có khóa dự tuyển là : MASV
- Lược đồ quan hệ Khoa có các khóa dự tuyển là: MAKHOA, TENKHOA

2.6.3. Khóa chính (Primary key)

Trong trường hợp lược đồ quan hệ có nhiều khóa dự tuyển, thì khóa được chọn để cài đặt gọi là **khóa chính**.

Trong ví dụ 1.2:

- Lược đồ quan hệ SV có khóa chính là : MASV
- Lược đồ quan hệ Khoa có khóa chính là: MAKHOA

2.6.4. Khóa ngoại (Foreign key)

Một thuộc tính hay tập thuộc tính được gọi là khóa ngoại nếu nó là thuộc tính của một lược đồ quan hệ này nhưng lại là khóa chính của lược đồ quan hệ khác.

Trong ví dụ 1.2:

- Lược đồ quan hệ Lop có khóa ngoại là : MAKHOA
- Lược đồ quan hệ KQ có 2 khóa ngoại là : MASV, MAMH

BÀI TẬP CHƯƠNG 1

Tìm các khóa chính và khóa ngoại trong các lược đồ cơ sở dữ liệu sau :

1.1. Cho lược đồ cơ sở dữ liệu dùng để quản lý hồ sơ sinh viên bao gồm các quan hệ được mô tả bởi các tên từ như sau:

SV(MASV, TENSX, GIOITINH, NGAYSINH, MALOP, DIACHI, SDT, DIEMXT)

Tên từ: Mỗi sinh viên có một mã duy nhất (MASV) để phân biệt với các sinh viên khác. Mỗi MASV xác định tất cả các thông tin về sinh viên đó như họ tên (TENSX), giới tính (GIOITINH), ngày sinh (NGAYSINH), lớp mà sinh viên đó theo học (MALOP), địa chỉ (DIACHI), số điện thoại (SDT), điểm xét tuyển đầu vào (DIEMXT).

Lop(MALOP, TENLOP, MAKHOA)

Tên từ: Mỗi lớp có một mã lớp (MALOP) duy nhất để phân biệt với các lớp khác. Biết MALOP xác định được tên lớp (TENLOP). Mỗi lớp chỉ thuộc về một khoa nào đó (MAKHOA).

Khoa(MAKHOA, TENKHOA, DTKHOA)

Tên từ: Mỗi khoa có một mã khoa (MAKHOA) duy nhất để phân biệt với các khoa khác. Biết MAKHOA xác định tên khoa (TENKHOA), số điện thoại của khoa (DTKHOA).

Mon (MAMH, TENMH, LT, TH)

Tên từ: Mỗi môn học có một mã môn học (MAMH) duy nhất để phân biệt với các môn học khác. Biết MAMH xác định tên môn học (TENMH), số tiết lý thuyết (LT), số tiết thực hành (TH) của môn học đó.

KQ (MASV, MAMH, DIEM)

Tên từ: Mỗi sinh viên ứng với mỗi môn học sẽ có một điểm tổng kết (DIEM)

1.2. Cho lược đồ CSDL dùng để quản lý lao động bao gồm các lược đồ quan hệ sau:

NhanVien(MANV, HOTEN, NGAYSINH, PHAI, DIACHI, MAPB)

Tên từ: Mỗi nhân viên có một mã số nhân viên (MANV) duy nhất. Một mã số nhân viên xác định các thông tin như họ tên (HOTEN), ngày sinh (NGAYSINH), phái (PHAI), địa chỉ (DIACHI) và phòng ban (MAPB) nơi quản lý nhân viên.

PhongBan(MAPB, TENPB)

Tên từ: Mỗi phòng ban có một mã phòng ban (MAPB) duy nhất, mã phòng ban xác định tên phòng ban (TENPB)

CongTrinh(MACT, TENCT, DIADIEM, NGAYCAPGP, NGAYKC, NGAYHT)

Tân từ: Mỗi công trình có một mã số công trình (MACT) duy nhất. Mã số công trình xác định các thông tin như tên gọi công trình (TENCT), địa điểm (DIADIEM), ngày công trình được cấp giấy phép xây dựng (NGAYCAPGP), ngày khởi công (NGAYKC), ngày hoàn thành (NGAYHT).

Cong(MACT, MANV, SLNGAYCONG)

Tân từ: Lược đồ quan hệ Cong ghi nhận số lượng ngày công (SLNGAYCONG) của một nhân viên (MANV) tham gia vào công trình (MACT).

1.3. Cho lược đồ CSDL để quản lý thư viện như sau:

Sach(MASH, TENSACH, TACGIA, NHAXB, NAMXB)

Tân từ: Mỗi quyển sách được đánh một mã sách (MASH) dùng để phân biệt với các quyển sách khác, 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).

DauSach(MADAUSACH, BAN,TAP, MASH)

Tân từ: Mỗi quyển sách có thể có nhiều bản giống nhau và cũng có thể có nhiều tập.

DocGia(MADG, HOTEN, NGAYSINH, DIACHI, NGHENGHIEP)

Tân từ: Mỗi đọc giả được thư viện cấp cho một thẻ thư viện, trong đó có ghi rõ mã đọc giả (MADG), cùng với các thông tin khác như : họ tên (HOTEN), ngày sinh (NGAYSINH), địa chỉ (DIACHI), nghề nghiệp(NGHENGHIEP).

PhieuMuon(SOPM, NGAYMUON, MADG)

Tân từ: Cứ mỗi lượt mượn sách, đọc giả phải ghi các quyển sách cần mượn vào một phiếu mượn, mỗi phiếu mượn có một số phiếu mượn (SOPM) duy nhất, mỗi phiếu mượn xác định các thông tin như: ngày mượn (NGAYMUON), đọc giả mượn.

ChiTietMuon(SOPM, MADAUSACH, NGAYTRA)

Tân từ: Trên phiếu mượn ghi các quyển sách mượn và ngày trả (NGAYTRA). Các quyển sách trong cùng một phiếu mượn không nhất thiết phải trả trong cùng một ngày.

1.4. Cho lược đồ CSDL để quản lý bán hàng như sau:

Khach (MAKH, TENKH, DCKH, DTKH)

Tân từ: Mỗi khách hàng có một mã khách hàng phân biệt (MAKH), mỗi MAKH xác định tên khách hàng (TENKH), địa chỉ (DCKH), số điện thoại (DTKH).

NhanVien(MANV, TENNV, NGAYSINH, DIACHI, DTHOAI)

Tân từ: Mỗi nhân viên có một mã nhân viên phân biệt (MANV), mỗi MANV xác định tên nhân viên (TENNV), ngày sinh (NGAYSINH), địa chỉ (DCNV), số điện thoại (DTNV).

Hang(MAMH, TENMH, DVT, DONGIA,SLTON)

Tân từ: Mỗi mặt hàng có một mã hàng phân biệt (MAMH), mỗi MAMH xác định tên hàng (TENMH), đơn vị tính (DVT), đơn giá (DONGIA), số lượng tồn kho (SLTON).

DonDatHang(SODDH, NGAYDH, MAKH)

Tân từ: Mỗi đơn đặt hàng có một mã số riêng biệt (SODDH). Biết SODDH xác định một ngày đặt hàng (NGAYDH) và mã khách hàng tương ứng (MAKH). Mỗi khách hàng trong một ngày có thể đặt hàng nhiều lần.

ChiTietDatHang(SODH, MAMH, SLDAT)

Tân từ: Mỗi đơn đặt hàng có thể bao gồm nhiều mặt hàng. Biết mã số đơn đặt hàng và mã mặt hàng thì biết được số lượng đặt hàng(SLDAT).

HoaDon(SOHD, NGAYLAP, NGAYXUAT, MANV, MAKH, SODDH)

Tân từ: Mỗi hoá đơn có một mã số phân biệt (SOHD) , mỗi hoá đơn bán hàng có thể gồm nhiều mặt hàng. Mỗi SOHD xác định ngày lập hoá đơn (NGAYLAP), ngày xuất kho giao hàng cho khách (NGAYXUAT), mã nhân viên lập hoá đơn, mã khách hàng và ứng với đơn đặt hàng nào.

ChiTietHoaDon (SOHD, MAMH, SLBAN)

Tân từ: Ứng với mỗi SOHD và MAMH xác định số lượng bán (SLBAN) của một mặt hàng trong một hoá đơn.

Phieuthu (SOPT, SOHD, NGAYTHU, SOTIEN)

Tân từ: Với mỗi hóa đơn khách hàng có thể nộp tiền nhiều lần. Mỗi lần nộp tiền sẽ có một phiếu thu riêng. Mỗi phiếu thu có một số phiếu thu để phân biệt (SOPT), biết SOPT xác định SOHD, ngày thu tiền (NGAYTHU) và số tiền thu (SOTIEN).

Chương 2: RÀNG BUỘC TOÀN VỆN

Giới thiệu:

- Ràng buộc toàn vẹn (Integrity Constraint) là các quy tắc trong một cơ sở dữ liệu nhằm kiểm tra tính đúng đắn và hợp lệ của dữ liệu trước khi lưu trữ.

Mục tiêu:

Sau khi học xong người học có khả năng:

- Hiểu được khái niệm ràng buộc toàn vẹn và các yếu tố của ràng buộc toàn vẹn;
- Xác định và phát biểu được các loại ràng buộc toàn vẹn trong lược đồ cơ sở dữ liệu.

Nội dung chính:

1. RÀNG BUỘC TOÀN VỆN – CÁC YẾU TỐ CỦA RÀNG BUỘC TOÀN VỆN

1.1. Ràng buộc toàn vẹn

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 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 (RBTV)**.

Nhiệm vụ của người phân tích thiết kế là phải phát hiện càng đầy đủ các RBTV càng tốt và mô tả chúng một cách chính xác trong hồ sơ phân tích thiết kế - đó là một việc làm rất quan trọng. RBTV được xem như là một công cụ để diễn đạt ngữ nghĩa của CSDL. Một CSDL được thiết kế công kênh nhưng nó thể hiện được đầy đủ ngữ nghĩa của thực tế vẫn có giá trị cao hơn rất nhiều so với một cách thiết kế gọn nhẹ nhưng nghèo nàn về ngữ nghĩa vì thiếu các RBTV của CSDL.

Công việc kiểm tra RBTV thường được tiến hành vào thời điểm cập nhật dữ liệu (thêm, sửa, xoá). Những RBTV phát sinh phải cần được ghi nhận và xử lý một cách tường minh (thường là bởi một hàm chuẩn hoặc một đoạn chương trình).

RBTV và kiểm tra sự vi phạm RBTV là hai trong số những vấn đề quan trọng trong quá trình phân tích thiết kế cơ sở dữ liệu, nếu không quan tâm đúng mức đến những vấn đề trên, thì có thể dẫn đến những hậu quả nghiêm trọng về tính an toàn và toàn vẹn dữ liệu , đặc biệt là đối với những cơ sở dữ liệu lớn.

Để minh họa, trong chương này ta sẽ xét lược đồ CSDL trong bài tập 1.4 như sau :

Khach (MAKH, TENKH, DCKH, DTKH)

NhanVien (MANV, TENNV, NGAYSINH, DIACHI, DTHOAI)

Hang(MAMH, TENMH, DVT, DONGIA,SLTON)

DonDatHang(SODDH, NGAYDH, MAKH)

ChiTietDatHang(SODH, MAMH, SLDAT)

HoaDon(SOHD, NGAYLAP, NGAYXUAT, MANV, MAKH, SODDH)

ChiTietHoaDon (SOHD, MAMH, SLBAN)

Phieuthu (SOPT, SOHD, NGAYTHU, SOTIEN)

Quy ước: Gọi T_{Khach} , T_{NV} , T_{Hang} , T_{DDH} , T_{CTDH} , T_{HD} , T_{CTHD} , T_{PT} lần lượt là các quan hệ hay bảng thể hiện của các lược đồ Khach, NhanVien, HangHoa, DonDatHang, ChiTietDatHang, HoaDon, ChiTietHoaDon, PhieuThu.

1.2. Các yếu tố của ràng buộc toàn vẹn

Mỗi RBTV có ba yếu tố: điều kiện, bối cảnh, bảng tầm ảnh hưởng

1.2.1. Điều kiện

Điều kiện của RBTV là sự mô tả và biểu diễn nội dung của nó.

Điều kiện của một RBTV 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 RBTV cũng có thể được biểu diễn bằng phụ thuộc hàm (khái niệm phụ thuộc hàm sẽ được đề cập trong chương 4)

Sau đây là một số RBTV trên lược đồ CSDL Quản lý bán hàng.

Ví dụ 2.1: Ta có RBTV R1 sau :

Mỗi khách hàng có một mã số duy nhất không trùng với bất kỳ khách hàng nào khác

RBTV trên có thể diễn tả như sau:

$\forall t_1, t_2 \in T_{Khach}$

$t_1 \neq t_2 \Rightarrow t_1.MAKH \neq t_2.MAKH$

Cuối $\forall$

1.2.2. Bối cảnh

Bối cảnh của một RBTV 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 RBTV này khi tiến hành cập nhật dữ liệu. Bối cảnh của một RBTV có thể là một hoặc nhiều quan hệ.

Chẳng hạn với RBTV R1 ở trên thì bối cảnh của nó là T_{Khach}

1.2.3. Bảng tầm ảnh hưởng

Trong quá trình phân tích thiết kế một CSDL, người phân tích cần lập bảng tầm ảnh hưởng cho một RBTV nhằm xác định thời điểm cần phải tiến hành kiểm tra RBTV.

Thông thường, thời điểm cần phải kiểm tra RBTV là thời điểm cập nhật dữ liệu.

Một bảng tầm ảnh hưởng của một RBTV có dạng sau:

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

Bảng này chứa toàn các ký hiệu +, -

Chẳng hạn + tại dòng r₁, cột Thêm có nghĩa là khi thêm một bộ vào quan hệ r₁ thì cần kiểm tra RBTV.

Dấu - tại ô dòng r₂, cột sửa thì có nghĩa là khi sửa một bộ trên quan hệ r₂ thì không cần kiểm tra RBTV.

Các hệ quản trị cơ sở dữ liệu thường có các cơ chế tự động kiểm tra các RBTV về khóa nội, khóa ngoại, giá trị NOT NULL.

Ví dụ: Ràng buộc R₁ có bảng tầm ảnh hưởng

R ₁	Thêm	Sửa	Xoá
T _{Khach}	+	- (*)	-

(*) : Theo quy ước các HQT CSDL không được phép sửa giá trị của thuộc tính khóa

2. PHÂN LOẠI RÀNG BUỘC TOÀN VỆN

Trong quá trình phân tích thiết kế CSDL, người phân tích phải phát hiện tất cả các RBTV tiềm ẩn trong CSDL đó. Việc phân loại các RBTV là rất có ích, nó nhằm giúp cho người phân tích có được một định hướng để phát hiện các RBTV, tránh bỏ sót. Các RBTV có thể được chia làm hai loại chính như sau:

2.1. Ràng buộc toàn vẹn có bối cảnh là một quan hệ

2.1.1. RBTV liên bộ:

RBTV liên bộ là RBTV giữa các bộ trong cùng một quan hệ.

Thông thường các RBTV liên bộ là RBTV về khóa nội. Đây là loại RBTV rất phổ biến, nó có mặt trong mọi lược đồ quan hệ của CSDL và thường được các hệ quản trị CSDL tự động kiểm tra.

Trong ví dụ 2.1, RBTV R_1 là RBTV liên bộ.

Ví dụ 2.2: Với 2 dòng bất kỳ trong T_{CTHD} không thể trùng nhau ở cả 2 cột MAHD và MAMH.

Ta có RBTV R_2 sau :

$$\forall t_1, t_2 \in T_{CTHD}$$

$$t_1 \neq t_2 \Rightarrow t_1.MAHD \neq t_2.MAHD \text{ OR } t_1.MAMH \neq t_2.MAMH$$

Cuối $\forall$

R_2	Thêm	Sửa	Xoá
T_{CTHD}	+	-(*)	-

2.1.2. RBTV về miền giá trị

RBTV 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 2.3: Số lượng bán của các mặt hàng trong hóa đơn phải là số dương nên ta có RBTV R_3 sau :

$$\forall t \in T_{HD}$$

$$t.SLBAN > 0$$

Cuối $\forall$

R_3	Thêm	Sửa	Xoá
T_{HD}	+	+	-

2.1.3. RBTV liên thuộc tính

RBTV liên thuộc tính (trên một quan hệ) là mối liên hệ giữa các thuộc tính trong một lược đồ quan hệ.

Ví dụ 2.4: Do yêu cầu phải lập hóa đơn trước khi xuất hàng nên ta có RBTV R_4 sau:

$$\forall t \in T_{HD}$$

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

Cuối $\forall$

R_4	Thêm	Sửa	Xoá
T_{HD}	+	+	-

2.2. Ràng buộc toàn vẹn có bối cảnh là nhiều quan hệ

2.2.1. RBTV phụ thuộc tồn tại:

Là RBTV về khoá ngoại.

Ví dụ 2.5:

Các khách hàng trong bảng T_{DDH} phải có trong bảng T_{Khach} nên ta có RBTV R_5 như sau:

$$R_5.T_{DDH}[MAKH] \subseteq T_{Khach}[MAKH]$$

R_5	Thêm	Sửa	Xoá
T_{Khach}	-	- (*)	+
T_{DDH}	+	+	-

2.2.2. RBTV 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 2.6: Do yêu cầu phải đặt hàng trước khi xuất hàng nên ta có RBTV R_6 như sau:

$$\forall t_1 \in T_{DDH}, t_2 \in T_{HD}$$

Nếu $t_1.SODH=t_2.SODH$ thì

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

Cuối $\forall$:

R_6	Thêm	Sửa	Xoá
T_{DDH}	-	+	-
T_{HD}	+	+	-

BÀI TẬP CHƯƠNG 2

Hãy xác định tất cả các RBTV của các LDCSDL trong các bài tập chương 1.

CHƯƠNG 3: PHỤ THUỘC HÀM

Giới thiệu:

Phụ thuộc hàm (Functional dependency) là một công cụ dùng để biểu diễn một cách hình thức các RBTV (gọi tắt là ràng buộc). Phương pháp biểu diễn này có rất nhiều ưu điểm, và đây là một công cụ cực kỳ quan trọng, gắn chặt với lý thuyết thiết kế cơ sở dữ liệu.

Phụ thuộc hàm được ứng dụng trong việc giải quyết các bài toán tìm khóa, tìm phủ tối thiểu và chuẩn hóa cơ sở dữ liệu.

Mục tiêu:

Sau khi học xong người học có khả năng:

- Hiểu các khái niệm phụ thuộc hàm, bao đóng của tập phụ thuộc hàm, bao đóng của tập thuộc tính, phủ tối thiểu của phụ thuộc hàm, khóa của LDQH;
- Phát biểu được phương pháp và thực hiện được thao tác tìm bao đóng của tập thuộc tính, phủ tối thiểu của phụ thuộc hàm, tìm tất cả các khóa của LDQH.

Nội dung chính:

1. PHỤ THUỘC HÀM

1.1. Khái niệm phụ thuộc hàm

Ví dụ 4.1: Cho quan hệ PhanCong như sau:

PhanCong (PHICONG, MAYBAY, NGAYKH, GIOKH)

Nguyễn Văn Hoàng	83	9/8/2021	10:15AM
Nguyễn Văn Hoàng	116	10/8/2021	1:25PM
Trần Thành Nam	281	8/8/2021	5:50AM
Trần Thành Nam	301	12/8/2021	6:35PM
Trần Thành Nam	83	11/8/2021	10:15AM
Hoàng Ngọc Châu	83	13/8/2021	10:15AM
Hoàng Ngọc Châu	116	12/8/2021	1:25PM
Lê Quyết Thắng	281	9/8/2021	5:50AM

Lê Quyết Thắng	281	13/8/2021	5:50AM
Lê Quyết Thắng	412	15/8/2021	1:25PM

Quan hệ PhanCong diễn tả phi công nào lái máy bay nào và máy bay khởi hành vào thời gian nào. Không phải sự phối hợp bất kỳ nào giữa phi công, máy bay và ngày giờ khởi hành cũng đều được chấp nhận mà chúng có các điều kiện ràng buộc qui định sau:

- + Mỗi máy bay có một giờ khởi hành duy nhất.
- + Nếu biết phi công, biết ngày giờ khởi hành thì biết được máy bay do phi công ấy lái.
- + Nếu biết máy bay, biết ngày khởi hành thì biết phi công lái chuyến bay ấy.

Các ràng buộc này là các ví dụ về phụ thuộc hàm và được phát biểu lại như sau:

- + MAYBAY xác định GIOKH
- + {PHICONG,NGAYKH,GIOKH} xác định MABAY
- + {MAYBAY,NGAYKH} xác định PHICONG

hay

- + GIOKH phụ thuộc hàm vào MAYBAY
- + MABAY phụ thuộc hàm vào {PHICONG,NGAYKH,GIOKH}
- + PHICONG phụ thuộc hàm vào {MAYBAY,NGAYKH}

và được ký hiệu như sau:

- + MAYBAY $\rightarrow$ GIOKH
- + PHICONG,NGAYKH,GIOKH $\rightarrow$ MABAY
- + MAYBAY,NGAYKH $\rightarrow$ PHICONG

Trong cách ghi trên ta đã ký hiệu **MAYBAY** thay cho {**MAYBAY**}.

1.1.1. Định nghĩa

Cho lược đồ quan hệ Q và X, Y là hai tập con khác rỗng của Q^+ .

Ta nói X xác định Y hay Y phụ thuộc hàm vào X , ký hiệu $X \rightarrow Y$ nếu với mọi quan hệ r trên Q , ta có : $\forall t_1, t_2 \in r: t_1.X = t_2.X \Rightarrow t_1.Y = t_2.Y$

Ví dụ 4.2: Cho lược đồ quan hệ Q gồm 2 quan hệ r_1 và r_2 :

r_1 :

A	B	C	D
x	u	x	y
y	x	w	x
z	u	u	y
y	z	w	z

r_2 :

A	B	C	D
x	u	w	z
x	t	w	y
z	t	y	y
y	z	w	z

Các phụ thuộc hàm được thỏa là : $A \rightarrow C, B \rightarrow D$

1.1.2. Phụ thuộc hàm hiển nhiên (Trivial Dependency)

Dễ thấy : Nếu $X \supseteq Y$ thì $X \rightarrow Y$.

Trong trường hợp này $X \rightarrow Y$ được gọi là **phụ thuộc hàm hiển nhiên**.

Ví dụ : Phụ thuộc hàm $A \rightarrow A, AB \rightarrow A$ là các phụ thuộc hàm hiển nhiên.

1.1.3. Bao đóng của tập phụ thuộc hàm :

Người ta thường dùng ký hiệu F để chỉ tập các phụ thuộc hàm của lược đồ quan hệ Q . Ta có thể đánh số các phụ thuộc hàm của F là $f_1, f_2, \dots, f_m$. Quy ước rằng chỉ cần mô tả các phụ thuộc hàm không hiển nhiên trong tập F (các phụ thuộc hàm hiển nhiên được ngầm hiểu là đã có trong F).

Ta nói phụ thuộc hàm $X \rightarrow Y$ được suy diễn logic từ F nếu một quan hệ r thỏa mãn tất cả các phụ thuộc hàm của F thì cũng thỏa phụ thuộc hàm $X \rightarrow Y$.

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

Các tính chất của tập F^+ :

1. Tính phản xạ: Với mọi tập phụ thuộc hàm F ta luôn luôn có $F \subseteq F^+$
2. Tính đơn điệu: Nếu $F \subseteq G$ thì $F^+ \subseteq G^+$
3. Tính lũy đẳng: Với mọi tập phụ thuộc hàm F ta luôn luôn có $(F^+)^+ = F^+$.

Hệ luật dẫn Armstrong (Phần đọc thêm)

Cho X, Y, Z, W là tập con của Q^+ , hệ luật dẫn Armstrong gồm 6 luật sau :

- A1. Luật phản xạ (reflexive rule): $X \rightarrow X$
- A2. Luật thêm vào (augmentation rule): Cho $X \rightarrow Y \Rightarrow XZ \rightarrow Y$
- A3. Luật bắc cầu (transitive rule): Cho $X \rightarrow Y, Y \rightarrow Z \Rightarrow X \rightarrow Z$
- A4. Luật phân rã (decomposition rule): Cho $X \rightarrow YZ \Rightarrow X \rightarrow Y$

A5. Luật bắc cầu giả (pseudo transitive rule): Cho $X \rightarrow Y, YZ \rightarrow W \Rightarrow XZ \rightarrow W$

A6. Luật hợp (union rule): Cho $X \rightarrow Y, X \rightarrow Z \Rightarrow X \rightarrow YZ$

Các luật A4, A5 và A6 được suy dẫn từ các luật A1, A2 và A3. Hệ tiên đề Armstrong (Armstrong's Axioms) gồm 3 luật: A1, A2 và A3.

1.2. Bao đóng của tập thuộc tính

1.2.1. Định nghĩa

Cho lược đồ quan hệ Q , tập các phụ thuộc hàm F , $X \subseteq Q^+$. Bao đóng (closure) của X dựa trên F , ký hiệu X^+ hay X^+_F được định nghĩa như sau :

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

Nhận xét :

1. $X \subseteq X^+$
2. Nếu $X \subseteq Y$ thì $X^+ \subseteq Y^+$
3. $X \rightarrow Y \in F^+ \Leftrightarrow Y \subseteq X^+$

Từ nhận xét 3 ta thấy để kiểm tra $X \rightarrow Y \in F^+$ hay không ta chỉ cần kiểm tra xem $Y \subseteq X^+$ hay không ?

1.2.2. Thuật toán tìm X^+

Dữ liệu vào: $Q^+, F^+, X \subseteq Q^+$

Dữ liệu ra: X^+

Phương pháp:

Bước 1: $X^+ = X$

Bước 2: $Temp = X^+$

$\forall U \rightarrow V \in F : \text{nếu } U \subseteq X^+ \text{ thì } X^+ = X^+ \cup V$

Bước 3: Nếu $X^+ = Temp$ thì X^+ là kết quả cần tìm và kết thúc thuật toán.

Ngược lại quay lại bước 2.

Ví dụ 4.3: Cho lược đồ quan hệ $Q(A,B,C,D,E,G,H)$ và tập phụ thuộc hàm F

$$F = \{ \begin{array}{lll} f_1: & B & \rightarrow A, \\ f_2: & DA & \rightarrow CE, \\ f_3: & D & \rightarrow H, \\ f_4: & GH & \rightarrow C, \\ f_5: & AC & \rightarrow D \end{array} \}$$

1. Tính $(AC)^+$ và $(BD)^+$.

2. Chứng minh $AC \rightarrow EH, BD \rightarrow CE \in F^+$

Giải :

1. Tính $(AC)^+$:

Bước 1: $X^+ = AC$

Bước 2: $Temp = AC$

$f_5: AC \rightarrow D \in F$ có $AC \subseteq X^+$ nên $X^+ = ACD$

Bước 3: $X^+ \neq Temp$ quay lại bước 2.

Bước 2: $Temp = ACD$

$f_2: DA \rightarrow CE \in F$ có $AD \subseteq X^+$ nên $X^+ = ACDE$

$f_3: D \rightarrow H \in F$ có $D \subseteq X^+$ nên $X^+ = ACDEH$

Bước 3: $X^+ \neq Temp$ quay lại bước 2.

Bước 2: $Temp = ACDEH$

Không có phụ thuộc hàm nào làm thay đổi X^+ được nữa.

Bước 3: $X^+ = Temp$. Thuật toán dừng, kết quả $X^+ = ACDEH$

Vậy $(AC)^+ = ACDEH$

Tính $(BD)^+$: Cách thực hiện tương tự như trên, nhưng có thể trình bày ngắn gọn như sau:

X^+	Phụ thuộc hàm áp dụng
BD	$f_1: B \rightarrow A$
BDA	$f_2: DA \rightarrow CE$
BDACE	$f_3: D \rightarrow H$
BDACEH	

Vậy $(BD)^+ = BDACEH$

2. Chứng minh $AC \rightarrow EH, BD \rightarrow CE \in F^+$

Vì $EH \in (AC)^+ = ACDEH$ nên $AC \rightarrow EH \in F^+$

Vì $CE \in (BD)^+ = BDACEH$ nên $BD \rightarrow CE \in F^+$

2. PHỦ TỐI THIỂU CỦA TẬP PHỤ THUỘC HÀM

2.1. Phủ của tập phụ thuộc hàm

2.1.1. Phủ của tập phụ thuộc hàm

Cho hai tập phụ thuộc hàm F và G . Ta nói F phủ G nếu $F^+ \supseteq G^+$

2.1.2. Hai tập phụ thuộc hàm tương đương

Hai tập phụ thuộc hàm F và G được gọi là tương đương (equivalent) nếu:

$F^+ = G^+$, ký hiệu $F \equiv G$.

Ví dụ 4.4: Cho lược đồ quan hệ $Q(A,B,C,D,E)$ và hai tập phụ thuộc hàm:

$F = \{A \rightarrow BC, A \rightarrow D, CD \rightarrow E\}$ và $G = \{A \rightarrow BCE, A \rightarrow ABD, CD \rightarrow E\}$

a. F có tương đương với G không?

b. F có tương đương với $G' = \{A \rightarrow BCDE\}$ không?

Giải:

a. Ta có $A_G^+ = ABCDE \Rightarrow$ trong G^+ có $A \rightarrow BC$ và $A \rightarrow D \Rightarrow F \subseteq G^+ \Rightarrow F^+ \subseteq G^+ (1)$.

$A_F^+ = ABCDE \Rightarrow$ trong F^+ có $A \rightarrow BCE$ và $A \rightarrow ABD \Rightarrow F^+ \supseteq G \Rightarrow F^+ \supseteq G^+ (2)$

(1) và (2) $\Rightarrow F^+ = G^+ \Rightarrow F \equiv G$.

b. Do $(CD)_{G'}^+ = CD \Rightarrow G'^+$ không chứa phụ thuộc hàm $CD \rightarrow E \Rightarrow F$ không tương đương với G'

2.2. Phụ thuộc hàm đầy đủ

Cho lược đồ quan hệ Q và tập phụ thuộc hàm F , $X \subseteq Q^+$. Thuộc tính $A \in Q^+$ được gọi là *phụ thuộc đầy đủ* vào tập thuộc tính X nếu :

1. $X \rightarrow A \in F^+$
2. Không tồn tại $X' \subset X$ sao cho $X' \rightarrow A \in F^+$

Khi đó phụ thuộc hàm $X \rightarrow A$ được gọi là *phụ thuộc hàm đầy đủ* hay phụ thuộc hàm có vẻ trái không dư thừa.

Ví dụ 4.5: Cho lược đồ quan hệ $Q(A,B,C,D,E)$ và tập phụ thuộc hàm $F=\{A \rightarrow BCD, AD \rightarrow E\}$

Phụ thuộc hàm $AD \rightarrow E$ không là phụ thuộc hàm đầy đủ vì có phụ thuộc hàm $A \rightarrow E \in F^+$.

Phụ thuộc hàm $A \rightarrow BCD$ là phụ thuộc hàm đầy đủ.

Chú ý : Phụ thuộc hàm mà vẻ trái chỉ có một thuộc tính luôn là phụ thuộc hàm đầy đủ.

2.3. Tập phụ thuộc hàm không dư thừa

Tập phụ thuộc hàm F được gọi 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.

Ví dụ 4.6: Cho $F = \{A \rightarrow BC, B \rightarrow D, AB \rightarrow D\}$ thì F dư thừa vì $F \equiv F' = \{A \rightarrow BC, B \rightarrow D\}$

2.4. Phủ tối thiểu (Minimal Cover)

2.4.1. Định nghĩa

F được gọi là một *tập phụ thuộc hàm tối thiểu* nếu F thỏa đồng thời ba điều kiện sau:

1. Vế phải của các phụ thuộc hàm trong F chỉ có một thuộc tính.
2. Mọi phụ thuộc hàm trong F đều là phụ thuộc hàm đầy đủ.
3. F là tập phụ thuộc hàm không dư thừa

Cho tập các phụ thuộc hàm F , nếu F' là tập các phụ thuộc hàm tối thiểu và tương đương với F thì F' được gọi là *phủ tối thiểu* của F .

2.4.2. Thuật toán tìm phủ tối thiểu của một tập phụ thuộc hàm

Dữ liệu vào: Q^+, F

Dữ liệu ra: Phủ tối thiểu F' của F

Phương pháp:

Bước 1: $F' = F$.

Trong F' , tách các phụ thuộc hàm có vẻ phải trên một thuộc tính thành các phụ thuộc hàm có vẻ phải một thuộc tính.

Bước 2: Loại các thuộc tính dư thừa ở vế trái các phụ thuộc hàm trong F'

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

Chú ý: Theo thuật toán trên, từ một tập phụ thuộc hàm F luôn tìm được ít nhất một phủ tối thiểu và nếu thứ tự loại các phụ thuộc hàm trong tập F là khác nhau thì có thể sẽ thu được những phủ tối thiểu khác nhau.

Ví dụ 4.7: Cho lược đồ quan hệ Q(A,B,C,D) và tập phụ thuộc F như sau:

$F = \{AB \rightarrow CD, B \rightarrow C, C \rightarrow D\}$. Hãy tìm phủ tối thiểu của F.

Giải:

Bước 1: $F' = \{AB \rightarrow C, AB \rightarrow D, B \rightarrow C, C \rightarrow D\}$

Bước 2:

$AB \rightarrow C$ là phụ thuộc hàm có vế trái dư thừa thuộc tính A vì $B \rightarrow C \in F'$

$F' = \{B \rightarrow C, AB \rightarrow D, B \rightarrow C, C \rightarrow D\}$

$AB \rightarrow D$ là phụ thuộc hàm có vế trái dư thừa thuộc tính A vì $B \rightarrow D \in F'^+$

(do $B \rightarrow C, C \rightarrow D \in F'$)

$F' = \{B \rightarrow C, B \rightarrow D, B \rightarrow C, C \rightarrow D\}$

Bước 3: Loại một phụ thuộc hàm $B \rightarrow C$ dư thừa do trùng lặp,

$F' = \{B \rightarrow D, B \rightarrow C, C \rightarrow D\}$

Kiểm tra $B \rightarrow D$ có dư thừa không (nghĩa là $B \rightarrow D \in (F' - \{B \rightarrow D\})^+$)

$B^+_{F' - \{B \rightarrow D\}} = BCD$ chứa D nên $B \rightarrow D \in (F' - \{B \rightarrow D\})^+$

Vậy $B \rightarrow D$ dư thừa

$F' = \{B \rightarrow C, C \rightarrow D\}$

Kiểm tra $B \rightarrow C$ có dư thừa không (nghĩa là $B \rightarrow C \in (F' - \{B \rightarrow C\})^+$)

$B^+_{F' - \{B \rightarrow C\}} = B$ không chứa C nên $B \rightarrow C \notin (F' - \{B \rightarrow C\})^+$

Vậy $B \rightarrow C$ không dư thừa

Kiểm tra $C \rightarrow D$ có dư thừa không (nghĩa là $C \rightarrow D \in (F' - \{C \rightarrow D\})^+$)

$C^+_{F' - \{C \rightarrow D\}} = C$ không chứa D nên $C \rightarrow D \notin (F' - \{C \rightarrow D\})^+$

Vậy $C \rightarrow D$ không dư thừa

Kết luận : phủ tối thiểu của F là $F' = \{B \rightarrow C, C \rightarrow D\}$

Ví dụ 4.8: 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

Kết quả: phủ tối thiểu của F là

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

3. KHÓA CỦA LƯỢC ĐỒ QUAN HỆ

3.1. Định nghĩa

Cho lược đồ quan hệ Q và tập phụ thuộc hàm F, $K \subseteq Q^+$.

K được gọi là một *khóa* (key) của Q nếu thỏa đồng thời cả hai điều kiện sau :

1. $K^+ = Q^+$ và
2. Không tồn tại $K' \subset K : K'^+ = Q^+$

Một *siêu khóa* (super key) là một tập hợp chứa khóa (chỉ thỏa điều kiện i)

Thuộc tính A được gọi là *thuộc tính khóa* nếu $A \in K$ với K là khóa bất kỳ của Q. Ngược lại A được gọi là *thuộc tính không khóa*.

Nhận xét :

- Một lược đồ quan hệ có thể có nhiều khóa.
- Q^+ cũng là một siêu khóa.

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

3.2.1. Một số khái niệm và nhận xét

Cho lược đồ quan hệ Q và tập các phụ thuộc hàm trong F.

- **Tập thuộc tính nguồn (TN)** : là tập tất cả các thuộc tính của Q mà không xuất hiện ở vế phải của bất kỳ phụ thuộc hàm nào trong F.
- **Tập thuộc tính đích (TĐ)** : là tập tất cả các thuộc tính chỉ xuất hiện ở vế phải mà không xuất hiện ở vế trái của các phụ thuộc hàm trong F.
- **Tập thuộc tính trung gian (TG)**: là tập tất cả các thuộc tính của Q mà không thuộc tập thuộc tính nguồn hay tập thuộc tính đích.

Nhận xét : Gọi K là một khóa bất kỳ của Q, ta luôn có :

- i. $TN \subseteq K$
- ii. $TĐ \cap K = \emptyset$
- iii. Nếu $TG = \emptyset$ hay $(TN)^+ = Q^+$ thì khóa duy nhất là TN

3.3.2. Thuật toán

Dữ liệu vào: Q^+, F

Dữ liệu ra: Tập tất cả các khóa TK

Phương pháp:

Bước 1: Tìm tập thuộc tính nguồn (TN), tập thuộc tính trung gian (TG)

Bước 2: Nếu $TG = \emptyset$ hay $(TN)^+ = Q^+$ thì khóa duy nhất là TN, kết thúc thuật toán.

Ngược lại thực hiện bước 3.

Bước 3: Tìm tất cả tập con X_i của TK

Bước 4: $TK = \emptyset$

$\forall X_i \subseteq TG : \text{nếu } (TN \cup X_i)^+ = Q^+ \text{ thì } TK = TK \cup \{TN \cup X_i\}$

Ví dụ 4.9: Cho lược đồ quan hệ $Q(C, S, Z)$ và tập phụ thuộc hàm F như sau:

$F = \{f_1: CS \rightarrow Z; f_2: Z \rightarrow C\}$. Tìm tất cả các khóa của Q ?

Giải:

$TN = \{S\}; TG = \{C, Z\}$

Gọi X_i là các tập con của tập TG:

X_i	$(TN \cup X_i)$	$(TN \cup X_i)^+$	Siêu khóa	Khó a

ϕ	S	S		
C	SC	Q^+	SC	SC
Z	SZ	Q^+	SZ	SZ
CZ	SCZ	Q^+	SCZ	

Kết quả quan hệ trên có 2 khóa là : {S,C} và {S,Z}

BÀI TẬP CHƯƠNG 3

4.1. Cho quan hệ sau:

r:

A	B	C	D	E
a ₁	b ₁	c ₁	d ₁	e ₁
a ₁	b ₂	c ₂	d ₂	d ₁
a ₂	b ₁	c ₃	d ₃	e ₁
a ₂	b ₁	c ₄	d ₃	e ₁
a ₃	b ₂	c ₅	d ₁	e ₁

Phụ thuộc hàm nào sau đây thỏa r:

$A \rightarrow D$, $AB \rightarrow D$, $C \rightarrow BDE$, $E \rightarrow A$, $A \rightarrow E$

4.2. Cho lược đồ quan hệ $Q(A,B,C,D,E,G)$, tập các phụ thuộc hàm

$F = \{f_1: A \rightarrow C;$

$f_2: A \rightarrow EG;$

$f_3: B \rightarrow D;$

$f_4: G \rightarrow E\}$

Tính $(AB)^+, (CGD)^+$

4.3. Cho $F = \{AB \rightarrow C, B \rightarrow D, CD \rightarrow E, CE \rightarrow GH, G \rightarrow A\}$

Hãy chứng tỏ các phụ thuộc hàm $AB \rightarrow E, AB \rightarrow G \in F^+$

(Ghi chú : với bài toán không nói gì về lược đồ quan hệ Q ta ngầm hiểu Q^+ là tập thuộc tính có trong F nghĩa là $Q^+ = \{AB, BC, CD, DE, EG, GH, HA, AC, BD, CE, FH, AG\}$)

4.4. Cho $F = \{AB \rightarrow E, AG \rightarrow I, BE \rightarrow I, E \rightarrow G, GI \rightarrow H\}$.

Hãy chứng tỏ phụ thuộc hàm $AB \rightarrow GH \in F^+$

4.5. Cho $G = \{AB \rightarrow C, A \rightarrow B, B \rightarrow C, A \rightarrow C\}$, $F = \{AB \rightarrow C, A \rightarrow B, B \rightarrow C\}$

F có tương đương với G không?

4.6. Với mỗi lược đồ CSDL sau, hãy :

- Xác định phủ tối thiểu của các tập phụ thuộc hàm F .

- Tìm tất cả các khóa của Q

a. $Q(A, B, C, D)$ $F = \{CA \rightarrow D; A \rightarrow B\}$

b. $Q(S, D, I, M)$ $F = \{SI \rightarrow D; SD \rightarrow M\}$

c. $Q(A, B, C, D, E, I)$, $F = \{ACD \rightarrow EBI, CE \rightarrow AD\}$

d. $Q(S, N, D, T, X)$ $F = \{S \rightarrow N; S \rightarrow D; S \rightarrow T; S \rightarrow X\}$

e. $Q(A, B, C, D, E, G)$, $F = \{A \rightarrow BC, C \rightarrow DE, E \rightarrow G\}$

f. $Q(A, B, C, D, E, G, H)$, $F = \{C \rightarrow AB, D \rightarrow E, B \rightarrow G\}$

g. $Q(A, B, C, D, E, G, H)$, $F = \{A \rightarrow BC, D \rightarrow E, H \rightarrow G\}$

h. $Q(A, B, C, D, E, G)$, $F = \{AB \rightarrow C, C \rightarrow B, ABD \rightarrow E, G \rightarrow A\}$

i. $Q(ABCSXYZ)$, $F = \{S \rightarrow ABX, AX \rightarrow B, BY \rightarrow C, CZ \rightarrow X\}$

j. $Q(A, B, C, D, E, H)$, $F = \{A \rightarrow E, C \rightarrow D, E \rightarrow DH\}$

k. $Q(ABCDEFGH)$, $F = \{A \rightarrow H, AB \rightarrow CH, BC \rightarrow D, G \rightarrow B\}$

l. $Q(A, B, C)$, $F = \{A \rightarrow BC, B \rightarrow AC, C \rightarrow A\}$

m. $Q(A,B,C,D)$, $F = \{ AB \rightarrow C, D \rightarrow B, C \rightarrow ABD \}$

n. $Q(C,T,H,R,S,G)$, $F = \{ C \rightarrow T, HR \rightarrow C, HCT \rightarrow R, CS \rightarrow G, HS \rightarrow RC \}$

o. $Q(NGAY, GIO, PHONG, MONHOC, GIAOVIEN)$

$F = \{$
 $NGAY, GIO, PHONG \rightarrow MONHOC$
 $MONHOC, NGAY \rightarrow GIAOVIEN$
 $NGAY, GIO, PHONG \rightarrow GIAOVIEN$
 $MONHOC \rightarrow GIAOVIEN \}$

p. $Q(TENTAU, LOAITAU, MACHUYEN, LUONGHANG, BENCANG, NGAY)$

$F = \{$
 $TENTAU \rightarrow LOAITAU$
 $MACHUYEN \rightarrow TENTAU, LUONGHANG$
 $TENTAU, NGAY \rightarrow BENCANG, MACHUYEN \}$

q. $Q(BROKER, OFFICE, STOCK, QUANTITY, INVESTOR, DIVIDENT)$

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

r. $Q(MAHV, HOTEN, NGAYSINH, MALOP, TENLOP, NGAYKG, MAMH, TENMH, SOTIET, DIEMTHI, SOBL SOTIEN)$

$F = \{$
 $MAHV \rightarrow HOTEN, NGAYSINH, MALOP,$
 $MALOP \rightarrow NGAYKG, TENLOP,$
 $MAMH \rightarrow TENMH, SOTIET,$
 $MAHV, MAMH \rightarrow DIEMTHI,$
 $SOBL \rightarrow MAHV, SOTIEN \}$

Chương 4: CHUẨN HÓA CƠ SỞ DỮ LIỆU

Giới thiệu:

Chuẩn hóa CSDL là quá trình phân rã các LĐQH thành các lược đồ con đạt dạng chuẩn nào đó để làm giảm dư thừa và phụ thuộc dữ liệu.

Mục tiêu: Sau khi học xong người học có khả năng.

- Phát biểu được định nghĩa các dạng chuẩn của LĐQH 1,2,3;
- Phát biểu được các bước để kiểm tra LĐQH đạt các dạng chuẩn 1,2,3.
- Xác định được dạng chuẩn cao nhất của lược đồ quan hệ.
- Phân rã được lược đồ quan hệ về dạng chuẩn 3 bảo toàn thông tin và phụ thuộc hàm.

Nội dung chính:

1 DẠNG CHUẨN CỦA LƯỢC ĐỒ QUAN HỆ

Trong thực tế, một ứng dụng cụ thể có thể được thiết kế theo nhiều cách thành nhiều lược đồ cơ sở dữ liệu khác nhau. Khi đánh giá, lựa chọn một lược đồ CSDL người ta thường dựa trên hai yếu tố là sự trùng lặp thông tin và chi phí kiểm tra các RBTV. Để có cơ sở định mức sự trùng lặp thông tin của các lược đồ CSDL, người ta xây dựng một hệ thống các dạng chuẩn (normal form).

1.1. Dạng chuẩn 1 (First Normal Form)

Một lược đồ quan hệ Q ở dạng chuẩn 1 nếu toàn bộ các thuộc tính của mọi bộ đều mang giá trị đơn.

Ví dụ 5.1: Xét quan hệ

MASV	HOVATEN	KHOA	TENMONHOC	DIEM
21001023	NGUYEN THI THU	CONG NGHE THONG TIN	KY THUAT LAP TRINH	6
			TOAN ROI RAC	8
			CO SO DU LIEU	4
21001030	LE VAN THANH	DIEN TU	VI XU LY	4

Quan hệ này không đạt chuẩn 1 vì các thuộc tính TENMONHOC, DIEM của bộ thứ nhất không mang giá trị đơn (chẳng hạn sinh viên NGUYEN THI THU có thuộc tính TENMONHOC là KY THUAT LAP TRINH, TOAN ROI RAC, CO SO DU LIEU).

Ta có thể đưa quan hệ trên về dạng chuẩn 1 như sau:

Khoa Công nghệ Thông tin- Trường CD Lý Tự Trọng TP. HCM

MASV	HOVATEN	KHOA	TENMONHOC	DIEM
21001023	NGUYEN THI THU	CONG NGHE THONG TIN	KY THUAT LAP TRINH	6
21001023	NGUYEN THI THU	CONG NGHE THONG TIN	TOAN ROI RAC	8
21001023	NGUYEN THI THU	CONG NGHE THONG TIN	CO SO DU LIEU	4
21001030	LE VAN THANH	DIEN TU	VI XULY	4

Từ phần này trở về sau ta quy ước mọi lược đồ quan hệ đều đạt dạng chuẩn 1.

1.2. Dạng chuẩn 2 (Second Normal Form)

1.2.1. Định nghĩa

Một lược đồ quan hệ Q ở dạng chuẩn 2 nếu Q đạt chuẩn 1 và mọi thuộc tính không khóa của Q đều phụ thuộc đầy đủ vào khóa.

1.2.2. Định lý

- Nếu tập thuộc tính không khóa của Q bằng rỗng thì Q đạt dạng chuẩn 2.
- Nếu tất cả khóa của Q chỉ gồm một thuộc tính thì Q đạt dạng chuẩn 2.

1.2.3. Thuật toán kiểm tra lược đồ quan hệ đạt dạng chuẩn 2

Dữ liệu vào: Q^+ , F, TK (tập tất cả các khóa của Q)

Dữ liệu ra: Kết luận Q đạt dạng chuẩn 2 không

Phương pháp:

Bước 1: Tìm tập thuộc tính không khóa N

Bước 2: Với mỗi thuộc tính không khóa $A \in N$

 Với mỗi khóa $K \in TK$

 Nếu $X \rightarrow A$ là phụ thuộc hàm không đầy đủ thì kết thúc thuật toán và kết luận Q không đạt dạng chuẩn 2.

Bước 3: Kết luận Q đạt dạng chuẩn 2.

Ví dụ 5.2: Cho lược đồ quan hệ $Q(A,B,C)$ và tập phụ thuộc hàm $F=\{AB\rightarrow C, B\rightarrow C\}$

Q có đạt dạng chuẩn 3 không ?

Giải:

Q có một khóa duy nhất là AB nên C là thuộc tính không khóa.

$AB\rightarrow C$ là phụ thuộc hàm đầy đủ vì $B\rightarrow C\in F$, nên C không phụ thuộc đầy đủ vào khóa AB.

Kết luận Q chỉ đạt dạng chuẩn 1 mà không đạt dạng chuẩn 2.

Ví dụ 5.3: Cho lược đồ quan hệ $Q(S,I,D,M)$ và tập phụ thuộc hàm $F=\{SI\rightarrow D, SD\rightarrow M\}$

Q có đạt dạng chuẩn 3 không ?

Giải:

Q có một khóa duy nhất là SI.

Thuộc tính không khóa : D,M

$SI\rightarrow D$ là phụ thuộc hàm đầy đủ vì $S\rightarrow D, I\rightarrow D \notin F^+$

$SI\rightarrow M$ là phụ thuộc hàm đầy đủ vì $S\rightarrow M, I\rightarrow M \notin F^+$

Vậy Q đạt dạng chuẩn 2.

1.3. Dạng chuẩn 3 (Third Normal Form)

1.3.1. Định nghĩa

Một lược đồ quan hệ Q ở dạng chuẩn 3 nếu với mọi phụ thuộc hàm $X\rightarrow A\in F^+$ ($A\in Q^+$, $A\notin X$) thì X chứa khóa hay A là thuộc tính khóa.

Nhận xét :

- Nếu tập thuộc tính không khóa của Q bằng rỗng thì Q đạt dạng chuẩn 3.

Ví dụ 5.4: Cho lược đồ quan hệ $Q(C,S,Z)$ và tập phụ thuộc hàm $F=\{CS\rightarrow Z, Z\rightarrow C\}$

Q có đạt dạng chuẩn 3 không ?

Giải:

Q có 2 khóa là CS, SZ.

Q có không có thuộc tính không khóa nên đạt dạng chuẩn 3

1.3.2. Các định lý

Định lý 1:

Cho lược đồ quan hệ Q và tập phụ thuộc hàm F gồm các phụ thuộc hàm về phải có một thuộc tính. Q đạt dạng chuẩn 3 nếu và chỉ nếu với mọi phụ thuộc hàm $X \rightarrow A \in F$ ($A \in Q^+$, $A \notin X$) thì X chứa khóa hay A là thuộc tính khóa.

Định lý 2:

Nếu Q đạt dạng chuẩn 3 thì cũng đạt dạng chuẩn 2.

1.3.3. Thuật toán kiểm tra lược đồ quan hệ đạt dạng chuẩn 3

Dữ liệu vào: Q^+ , F , TK (tập tất cả các khóa của Q)

Dữ liệu ra: Kết luận Q đạt dạng chuẩn 3 không

Phương pháp:

Bước 1: Tìm tập phụ thuộc hàm $F' \equiv F$, F' gồm các phụ thuộc hàm mà về phải có một thuộc tính.

Bước 2: Nếu $\forall X \rightarrow A \in F' : X$ chứa khóa hay A là thuộc tính khóa thì Q đạt dạng chuẩn 3, ngược lại Q không đạt dạng chuẩn 3.

Ví dụ 5. 5: Cho lược đồ quan hệ $Q(S,I,D,M)$ và tập phụ thuộc hàm $F=\{SI \rightarrow D, SD \rightarrow M\}$

Q có đạt dạng chuẩn 3 không ?

Giải:

Q có một khóa duy nhất là SI .

$F=\{SI \rightarrow D, SD \rightarrow M\}$ gồm các phụ thuộc hàm mà về phải có một thuộc tính.

Xét $SD \rightarrow M \in F$: có về trái không chứa khóa và về phải không là thuộc tính khóa.

Vậy Q không đạt dạng chuẩn 3 mà chỉ đạt dạng chuẩn 2.

1.4. Dạng chuẩn của lược đồ cơ sở dữ liệu

Dạng chuẩn của một lược đồ cơ sở dữ liệu là dạng chuẩn thấp nhất trong các dạng chuẩn của các lược đồ quan hệ con.

2. CHUẨN HÓA LƯỢC ĐỒ BẰNG CÁCH PHÂN RÃ

2.1. Thuật toán phân rã lược đồ quan hệ về dạng chuẩn 3 bảo toàn thông tin và phụ thuộc hàm

Trong trường hợp lược đồ quan hệ chưa đạt dạng chuẩn 3, ta có thể phân rã lược đồ này thành các lược đồ con đạt dạng chuẩn 3 bằng thuật toán sau :

Dữ liệu vào: Lược đồ quan hệ Q và tập các phụ thuộc hàm F là phủ tối thiểu, K là tập tất cả các khóa.

Dữ liệu ra: Một phân rã bảo toàn thông tin và phụ thuộc hàm của Q sao cho mỗi lược đồ quan hệ con đều đạt dạng chuẩn 3.

Phương pháp:

Bước 1: Tách ra khỏi Q các thuộc tính không nằm trong một phụ thuộc hàm nào của F .

Bước 2: Nếu có một phụ thuộc hàm nào của F mà liên quan đến tất cả các thuộc tính của Q thì Q không thể phân rã được nữa.

Bước 3: Gom các phụ thuộc hàm có cùng vế trái

Bước 4: Với mỗi phụ thuộc hàm $X \rightarrow A \in F$, phân rã Q thành các lược đồ con XA (Nếu có 2 lược đồ Q_i, Q_j mà $Q_i \subseteq Q_j$ thì nhập Q_i vào Q_j)

Bước 5: Nếu chưa có lược đồ con nào chứa một khóa của Q thì bổ sung một khóa Y vào phân rã.

Ghi chú : Sau bước 4, ta được phân rã bảo toàn các phụ thuộc hàm.

Sau bước 5, ta được phân rã bảo toàn thông tin (dữ liệu).

Ví dụ 5.6: Cho lược đồ quan hệ $Q(C,T,H,R,S,G)$ và tập phụ thuộc hàm F (là phủ tối thiểu) như sau :

$$F=\{C \rightarrow T, HR \rightarrow C, HT \rightarrow R, CS \rightarrow G, HS \rightarrow R\}$$

Hãy phân rã Q thành tước đồ CSDL bảo toàn thông tin và phụ thuộc hàm.

Giải:

Áp dụng thuật toán:

(Bước 1, 2, 3 không cần làm)

Bước 4: Q được phân rã thành 5 lược đồ con đạt dạng chuẩn 3:

$Q_1(C,T)$ có tập phụ thuộc hàm $F_1=\{ C \rightarrow T \}$

$Q_2(H,R,C)$ có tập phụ thuộc hàm $F_2=\{ HR \rightarrow C \}$

$Q_3(H,T,R)$ có tập phụ thuộc hàm $F_3=\{ HT \rightarrow R \}$

$Q_4(C,S,G)$ có tập phụ thuộc hàm $F_4=\{ CS \rightarrow G \}$

$Q_5(H,S,R)$ có tập phụ thuộc hàm $F_5=\{ HS \rightarrow R \}$

Bước 5: Q có khóa duy nhất là HS. Vì $HS \subseteq Q_5$ nên phân rã bảo toàn thông tin và phụ thuộc hàm

2.2. Ví dụ tổng hợp

Ví dụ 5.7: Cho lược đồ quan hệ Q(S,I,D,M) và tập các phụ thuộc hàm:

$$F=\{SI \rightarrow DM, SD \rightarrow M, D \rightarrow M\}$$

1. Tính D^+ , $(SD)^+$, $(SI)^+$

2. Tìm tất cả các khóa của Q

3. Tìm phủ tối thiểu của F

4. Xác định dạng chuẩn cao nhất của Q

5. Nếu Q chưa đạt dạng chuẩn 3, hãy phân rã Q thành lược đồ cơ sở dữ liệu đạt dạng chuẩn 3 bảo toàn phụ thuộc hàm và thông tin.

Giải :

1. Tính bao đóng

D^+	Phụ thuộc hàm	$(SD)^+$	Phụ thuộc hàm	$(SI)^+$	Phụ thuộc hàm
D	$D \rightarrow M$	SD SDM	$SD \rightarrow M$	SI SIDM	$SI \rightarrow DM$
DM					

Vậy $D^+=DM$, $(SD)^+=SDM$, $(SI)^+=SIDM$

2. Tìm tất cả các khóa của Q

$TN=\{S,I\}$ $TD=\{M\}$ $TG=\{D\}$

$(SI)^+=SIDM=Q^+$ nên Q có một khóa duy nhất là SI

3. Tìm phủ tối thiểu của F' của F

Bước 1: Tách vế phải

$F'=\{f_1: SI \rightarrow D, f_2: SI \rightarrow M, f_3: D \rightarrow M, f_4: D \rightarrow M\}$

Bước 2: Loại bỏ các thuộc tính dư thừa ở vế trái

Hiển nhiên các phụ thuộc hàm f_3, f_4 không có thuộc tính dư thừa ở vế trái.

Phụ thuộc hàm $f_2: SI \rightarrow M$ không có thuộc tính dư thừa ở vế trái vì :

$S^+=S$ nên $S \rightarrow D \notin F'^+$

$I^+=I$ nên $S \rightarrow D \notin F'^+$

Phụ thuộc hàm $f_3: SD \rightarrow M$, dư thuộc tính S ở vế trái vì có phụ thuộc hàm $D \rightarrow M \in F'$

$F'=\{f_1: SI \rightarrow D, f_2: SI \rightarrow M, f_3: D \rightarrow M, f_4: D \rightarrow M\}$

Bước 3: Loại bỏ các phụ thuộc hàm dư thừa

Phụ thuộc hàm $f_1: SI \rightarrow D$ không bỏ được vì:

$(SI)^+_{F'} - \{SI \rightarrow D\} = SIM$, không chứa D.

Phụ thuộc hàm $f_2: SI \rightarrow M$ bỏ được vì:

$(SI)^+_{F'} - \{SI \rightarrow M\} = SIDM$, chứa M

$f_3=f_4$ nên f_4 bỏ được

Xét phụ thuộc hàm $f3: D \rightarrow M$ không bỏ được vì:

$(D)+F'-\{D \rightarrow M\} = D$, không chứa M

Vậy phủ tối thiểu của F là $F' = \{SI \rightarrow D, D \rightarrow M\}$

4. Xác định dạng chuẩn cao nhất của Q

Khóa duy nhất của Q là: SI

Thuộc tính khóa : S, I

Thuộc tính không khóa là : D, M

Kiểm tra dạng chuẩn 2:

Ta có $SI \rightarrow D$ là phụ thuộc hàm đầy đủ vì $S \rightarrow D \notin F'+, I \rightarrow D \notin F'+$

Ta có $SI \rightarrow M$ là phụ thuộc hàm đầy đủ vì $S \rightarrow M \notin F'+, I \rightarrow M \notin F'+$

Vậy các thuộc tính không khóa đều phụ thuộc đầy đủ vào khóa nên Q đạt dạng chuẩn 2.

Kiểm tra dạng chuẩn 3:

Phụ thuộc hàm $D \rightarrow M$ có vế trái D không chứa khóa và vế phải M cũng không là thuộc tính khóa nên Q không đạt dạng chuẩn 3. Dạng chuẩn cao nhất của Q là dạng chuẩn 2

5. Phân rã Q thành lược đồ cơ sở dữ liệu đạt dạng chuẩn 3 bảo toàn phụ thuộc hàm và thông tin.

Do phủ tối thiểu $F' = \{SI \rightarrow D, D \rightarrow M\}$ nên Q có thể phân rã thành 2 lược đồ con ở dạng chuẩn 3:

$Q_1 (S, I, D)$, với tập phụ thuộc hàm $F_1 = \{SI \rightarrow D\}$

$Q_2 (D, M)$, với tập phụ thuộc hàm $F_2 = \{D \rightarrow M\}$

Vì Q_1 chứa khóa SI nên phân rã này bảo toàn thông tin và phụ thuộc hàm.

BÀI TẬP CHƯƠNG 4

5.1. Hãy xác định dạng chuẩn cao nhất của các LĐQH sau. Nếu chưa đạt dạng chuẩn 3, hãy phân rã các LĐQH này thành các lược đồ con đạt 3NF (phân rã bảo toàn thông tin và phụ thuộc hàm)

- a. $Q(A,B,C,D) F=\{ CA \rightarrow D; A \rightarrow B \}$
- b. $Q(S,D,I,M) F=\{ SI \rightarrow D; SD \rightarrow M \}$
- c. $Q(A, B, C, D, E, I), F=\{ ACD \rightarrow EBI, CE \rightarrow AD \}$
- d. $Q(S,N,D,T,X) F=\{ S \rightarrow N; S \rightarrow D; S \rightarrow T; S \rightarrow X \}$
- e. $Q(A,B,C,D,E,G), F=\{ A \rightarrow BC, C \rightarrow DE, E \rightarrow G \}$
- f. $Q(A,B,C,D,E,G,H), F=\{ C \rightarrow AB, D \rightarrow E, B \rightarrow G \}$
- g. $Q(A,B,C,D,E,G,H), F=\{ A \rightarrow BC, D \rightarrow E, H \rightarrow G \}$
- h. $Q(A,B,C,D,E,G), F=\{ AB \rightarrow C, C \rightarrow B, ABD \rightarrow E, G \rightarrow A \}$
- i. $Q(ABCSXYZ), F=\{ S \rightarrow ABX, AX \rightarrow B, BY \rightarrow C, CZ \rightarrow X \}$
- j. $Q(A,B,C,D,E,H), F=\{ A \rightarrow E, C \rightarrow D, E \rightarrow DH \}$
- k. $Q(ABCDEFGH), F=\{ A \rightarrow H, AB \rightarrow CH, BC \rightarrow D, G \rightarrow B \}$
- l. $Q(A,B,C), F=\{ A \rightarrow BC, B \rightarrow AC, C \rightarrow A \}$
- m. $Q(A,B,C,D), F=\{ AB \rightarrow C, D \rightarrow B, C \rightarrow ABD \}$
- n. $Q(C,T,H,R,S,G), F=\{ C \rightarrow T, HR \rightarrow C, HCT \rightarrow R, CS \rightarrow G, HS \rightarrow RC \}$
- o. $Q(NGAY, GIO, PHONG, MONHOC, GIAOVIEN)$
 $F=\{ \quad \quad \quad NGAY, GIO, PHONG \quad \quad \rightarrow MONHOC$

MONHOC, NGAY → GIAOVIEN

NGAY, GIO, PHONG → GIAOVIEN

MONHOC → GIAOVIEN}

p. Q(TENTAU, LOAITAU, MACHUYEN, LUONGHANG, BENCANG, NGAY)

F={ TENTAU → LOAITAU

MACHUYEN → TENTAU, LUONGHANG

TENTAU,NGAY → BENCANG, MACHUYEN}

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

F={ STOCK → DIVIDENT

INVESTOR → BROKER

INVESTOR, STOCK → QUANTITY

BROKER → OFFICE }

r. Q(MAHV, HOTEN, NGAYSINH, MALOP, TENLOP, NGAYKG, MAMH, TENMH, SOTIET, DIEMTHI, SOBL SOTIEN)

F={ MAHV → HOTEN, NGAYSINH, MALOP,

MALOP → NGAYKG, TENLOP,

MAMH → TENMH, SOTIET,

MAHV, MAMH → DIEMTHI,

SOBL → MAHV, SOTIEN}

5.2. Cho lược đồ quan hệ Q và tập phụ thuộc hàm F được cho như sau:

Q (A, B, C, D, E, G, H, K, L, M, N),

F={ C → D,E; G → H,K; A,G → L; M → A,N; A → B,C}

CÁC ĐỀ THI MẪU

ĐỀ SỐ 1

Câu I:

Cho lược đồ cơ sở dữ liệu **TUYỂN SINH** bao gồm các lược đồ quan hệ sau:

DiemThi(MADT, DCHI)

Một hội đồng coi thi tuyển sinh có nhiều địa điểm thi, mỗi điểm thi có một mã số điểm thi riêng biệt (MADT). Mỗi mã số điểm thi xác định địa chỉ điểm thi (DCHI).

ThiSinh(SOBD, HOTEN, NGSINH, MANGANH, SOPHONG)

Mỗi thí sinh đăng ký dự thi sẽ được cấp một số báo danh (SOBD) phân biệt. Mỗi số báo danh xác định các thông tin: họ và tên (HOTEN), ngày sinh (NGSINH), mã ngành đăng ký dự thi (MANGANH) và số hiệu phòng thi (SOPHONG).

Nganh(MANGANH, TENNGANH)

Mỗi ngành có một mã ngành (MANGANH) duy nhất phân biệt, mỗi mã ngành xác định tên ngành (TENNGANH).

Phon(SOPHONG, MADT)

Mỗi phòng thi có một số hiệu phòng thi (SOPHONG) duy nhất phân biệt, trong một phòng thi có thể có các thí sinh của nhiều ngành thi khác nhau. Mỗi phòng thi thuộc về một điểm thi duy nhất.

1. Tìm các khóa chính và khóa ngoại cho mỗi lược đồ quan hệ trên.
2. Xác định và lập bảng tầm ảnh hưởng cho các RBTV có trong lược đồ CSDL trên.

3. Viết các câu truy vấn sau bằng ngôn ngữ SQL

- Lập danh sách thí sinh thi tại điểm thi có mã là 'HVT'. Thông tin gồm tất cả các thuộc tính của LĐQH ThiSinh và được sắp tăng dần theo từng phòng
- Đếm số lượng thí sinh của từng điểm thi. Thông tin gồm MADT, DCDT, SL (số lượng thí sinh).
- Tìm điểm thi có nhiều thí sinh nhất. Thông tin gồm MADT, DCDT, SL (số lượng thí sinh).

Câu II:

Cho lược đồ quan hệ $Q(A, B, C, D, E, G, H)$ và tập phụ thuộc hàm:

$$F = \{ E \rightarrow C, H \rightarrow E, A \rightarrow D, AE \rightarrow H, DG \rightarrow B, DG \rightarrow C \}$$

- Cho biết phụ thuộc hàm $AH \rightarrow CD$ có được suy dẫn logic từ F không ?
- Hãy xác định tất cả các khóa của Q
- Hãy xác định dạng chuẩn cao nhất của Q ?
- Tìm phủ tối thiểu của F .
- Phân rã Q về dạng chuẩn 3, yêu cầu phân rã bảo toàn thông tin và phụ thuộc hàm.

ĐỀ SỐ 2

Câu I:

Cho lược đồ cơ sở dữ liệu quản lý điểm thi của các sinh viên bao gồm các lược đồ quan hệ sau:

SinhVien(MASV, HOTEN, MALOP)

Mỗi sinh viên có một mã số sinh viên (MASV) duy nhất để phân biệt với các sinh viên khác. Biết MASV xác định họ và tên sinh viên (HOTEN). Mỗi sinh viên phải thuộc về một lớp học nào đó (MALOP).

LopHoc(MALOP, TENLOP, MASV)

Mỗi lớp học có một mã lớp (MALOP) duy nhất để phân biệt với các lớp học khác. Biết MALOP xác định tên lớp (TENLOP) và sinh viên làm lớp trưởng (MASV).

MonHoc (MAMH, TENMH, ĐVHT)

Mỗi môn học có một mã môn học (MAMH) duy nhất để phân biệt với các môn học khác. Mỗi MAMH xác định tên gọi của môn học (TENMH) và ứng với một số đơn vị học trình nhất định(ĐVHT).

Ketqua(MASV, MAMH, DIEMTHI)

Mỗi sinh viên (MASV) ứng với một môn học (MAMH) có một điểm thi (DIEMTHI) duy nhất. Điểm thi hợp lệ là các số nguyên từ 0 đến 10.

1. Tìm các khóa chính và khóa ngoại cho mỗi lược đồ quan hệ trên và vẽ mối quan hệ (relationships) giữa các lược đồ quan hệ trên.
2. Xác định và lập bảng tầm ảnh hưởng cho các RBTV có trong lược đồ CSDL trên.
3. Viết các câu truy vấn sau bằng ngôn ngữ SQL
 - a. Lập danh sách các SV có điểm thi < 5. Thông tin gồm tất cả thuộc tính của LĐQH SinhVien.
 - b. Đếm số môn có điểm thi <5 từng SV. Thông tin gồm MASV, HOTEN, MALOP, SO_MON (số môn nợ).
 - c. In danh sách các SV không bị nợ môn. Thông tin gồm tất cả thuộc tính của LĐQH SinhVien.

Câu II

Cho lược đồ quan hệ $Q(A, B, C, D)$ và tập phụ thuộc hàm

$$F=\{AB \rightarrow CD, BC \rightarrow D, C \rightarrow D\}$$

1. Cho biết phụ thuộc hàm $BD \rightarrow C$ có được suy dẫn logic từ F không ?
2. Hãy xác định tất cả các khóa của Q
3. Hãy xác định dạng chuẩn cao nhất của Q ?
4. Tìm phủ tối thiểu của F .
5. Phân rã Q về dạng chuẩn 3, yêu cầu phân rã bảo toàn thông tin và phụ thuộc hàm.

ĐỀ SỐ 3

Câu I :

Cho lược đồ cơ sở dữ liệu quản lý hàng hóa như sau:

Hang (MAHH, TENHH, DVT, GIA, MANSX, MALOAI)

Tên từ: Mỗi mặt hàng có một mã hàng phân biệt duy nhất (MAMH). Biết MAMH xác định được tên hàng (TENMH), đơn vị tính (DVT), đơn giá (GIA), nhà sản xuất (MANSX), loại hàng (MALOAI) của mặt hàng đó.

Loai (MALOAI, TENLOAI)

Tên từ: Mỗi loại hàng có một mã loại hàng phân biệt duy nhất (MALOAI). Biết (MALOAI) xác định được tên loại (TENLOAI).

NSX (MANSX, TENNSX, DCNSX, DTHOAI)

Tên từ: Mỗi nhà sản xuất có một mã phân biệt duy nhất (MANSX). Biết (MANSX) xác định được tên (TENNSX), địa chỉ (DCNSX) và số điện thoại (DTHOAI) của nhà sản xuất đó.

DonDH (MADH, NGAYDH, NGAYGH, TENKH, DCKH)

Tên từ: Mỗi đơn đặt hàng có một mã đơn đặt hàng (MADDH). Biết MADDH xác định ngày đặt hàng (NGAYDH), ngày giao hàng (NGAYGH), họ tên khách hàng (TENKH), địa chỉ khách hàng (DCKH).

CT_DD (MADH, MAHH, SL)

Tên từ: Mỗi đơn đặt hàng có thể có nhiều mặt hàng, ứng với mỗi mặt hàng sẽ xác định số lượng đặt hàng (SL).

1. Tìm các khóa chính và khóa ngoại cho mỗi lược đồ quan hệ trên.
2. Xác định và lập bảng tầm ảnh hưởng cho các RBTV có trong lược đồ CSDL trên.
3. Viết các câu truy vấn sau bằng ngôn ngữ SQL
 - a. Lập danh sách các đơn đặt hàng có đặt hàng của nhà SX có mã là 'HSG'. Thông tin gồm tất cả thuộc tính của LĐQH DonDH.
 - b. Lập danh sách số lượng mặt hàng của từng loại hàng. Thông tin gồm MALOAI, TENLOAI, SLMH (Số lượng mặt hàng).

c. Tìm nhà SX cung cấp nhiều loại hàng nhất. Thông tin gồm MANSX, TENNSX, SO_LOAI_MH (Số loại mặt hàng).

Câu II :

Cho lược đồ quan hệ $Q(A, B, C, D, E, G)$ và tập các phụ thuộc hàm

$$F = \{AB \rightarrow C, AC \rightarrow DB, D \rightarrow EG\}$$

1. Cho biết phụ thuộc hàm $AD \rightarrow CG$ có được suy dẫn logic từ F không ?
2. Hãy xác định tất cả các khóa của Q
3. Hãy xác định dạng chuẩn cao nhất của Q ?
4. Tìm phủ tối thiểu của F .
5. Phân rã Q về dạng chuẩn 3, yêu cầu phân rã bảo toàn thông tin và phụ thuộc hàm.

ĐỀ SỐ 4

Câu I:

Cho lược đồ cơ sở dữ liệu quản lý nhân sự như sau:

PhongBan (MAPB, TENPB, MATP)

Tên từ: Mỗi phòng ban có một mã phòng duy nhất phân biệt (MAPB). Biết (MAPB) xác định tên phòng (TENPB) và nhân viên làm trưởng đó (MATP).

NhanVien (MANV, HOTEN, LUONG, MAPB, PHAI)

Tên từ: Mỗi nhân viên có một mã số riêng duy nhất phân biệt (MANV). Biết MANV xác định được họ tên (HOTEN), mức lương (LUONG), phòng công tác (MAPB), giới tính (PHAI) của nhân viên đó.

DeAn (MADA, TENDA, DDIEM, TGBD, TGKT, MAPB)

Tên từ: Mỗi đề án có một mã đề án duy nhất phân biệt (MADA). Biết (MADA) xác định được tên đề án (TENDA), địa điểm thực hiện (DDIEM), thời gian bắt đầu thực hiện (TGBD) và thời gian kết thúc (TGKT), phòng thực hiện (MAPB) đề án đó.

PhanCong (MANV, MADA, TGBD, TGKT)

Tân từ: Mỗi nhân viên thực hiện một đề án và từ trong một khoảng thời gian xác định với ngày bắt đầu (NGBD) và ngày kết thúc (NGKT)

1. Tìm các khóa chính và khóa ngoại cho mỗi lược đồ quan hệ trên..
2. Xác định và lập bảng tầm ảnh hưởng cho các RBTV có trong lược đồ CSDL trên.
3. Viết các câu truy vấn sau bằng ngôn ngữ SQL
 - a. Lập danh sách các nhân viên tham gia đề án được bắt đầu từ năm 2021. Thông tin gồm tất cả các thuộc tính của LĐQH NhanVien
 - b. Đếm số lượng đề án đã hoàn thành của từng phòng ban. Thông tin gồm MAPB, TENPB, SL_DA (Số lượng đề án)
 - c. Tìm 3 phòng ban đã hoàn thành nhiều đề án trong năm 2021 nhất. Thông tin gồm MAPB, TENPB, SL_DA (Số lượng đề án)

Câu II :

Cho quan hệ $Q(G, H, I, K, L, M)$ và tập các phụ thuộc hàm

$$F = \{GH \rightarrow L, I \rightarrow M, L \rightarrow K, HM \rightarrow G, GK \rightarrow I, H \rightarrow L\}$$

1. Cho biết phụ thuộc hàm $GH \rightarrow KM$ có được suy dẫn logic từ F không ?
2. Hãy xác định tất cả các khóa của Q
3. Hãy xác định dạng chuẩn cao nhất của Q ?
4. Tìm phủ tối thiểu của F .
5. Phân rã Q về dạng chuẩn 3, yêu cầu phân rã bảo toàn thông tin và phụ thuộc hàm.

ĐỀ SỐ 5

CÂU I:

Cho lược đồ cơ sở dữ liệu dùng để quản lý việc khám bệnh tại một bệnh viện tư nhân cho các bệnh nhân có mua bảo hiểm y tế của bệnh viện, gồm các lược đồ quan hệ sau:

BacSi(MABS, TENBS, SĐT, CK)

Tên từ : Mỗi bác sĩ có một mã số riêng (MABS) để phân biệt với các bác sĩ khác, mỗi bác sĩ có một tên (TENBS), số điện thoại (SDT) và có chuyên môn thuộc một chuyên khoa (CK).

BenhNhan(MABN, TENBN, NAMSINH, DCHI)

Tên từ : Mỗi bệnh nhân có một mã số riêng (MABN) để phân biệt với các bệnh nhân khác, mỗi bệnh nhân có một tên (TENBN), năm sinh (NAMSINH), địa chỉ (DCHI).

TheBH(MATHE, MABN, NGAYBD, NGAYKT)

Tên từ : Mỗi bệnh nhân thuộc diện có bảo hiểm y tế khi tới khám bệnh đều phải xuất trình thẻ bảo hiểm y tế. Mỗi thẻ có một mã số riêng (MATHE), biết MATHE xác định được MABN, ngày bắt đầu được bảo hiểm (NGAYBD), ngày kết thúc bảo hiểm (NGAYKT). Ở lần mua bảo hiểm tiếp theo, bệnh nhân sẽ được cấp thẻ khác với mã số khác nhưng mã bệnh nhân không đổi.

PhieuKham(MASO, NGAYKHAM, MATHE, MABS)

Tên từ : Mỗi lần khám bệnh, bệnh nhân sẽ được phát một mã số riêng (MASO) không trùng với mã số nào khác trong ngày. Thông qua mã số này bệnh nhân biết được phòng khám và số thứ tự của mình. Biết MASO và ngày khám (NGAYKHAM) thì xác định được mã thẻ của bệnh nhân, bác sĩ khám (MABS). Trong một ngày bệnh nhân có thể khám nhiều lần.

Câu hỏi:

1. Tìm khóa chính và khóa ngoại cho mỗi lược đồ quan hệ trên.
2. Xác định và lập bảng tầm ảnh hưởng cho các RBTV có trong lược đồ CSDL trên.
2. Xác định và lập bảng tầm ảnh hưởng cho các RBTV có trong lược đồ CSDL trên.
3. Viết các câu truy vấn sau bằng ngôn ngữ SQL
 - a. Lập danh sách các bệnh nhân có khám bệnh vào ngày 31/12/2020. Thông tin gồm tất cả thuộc tính của LĐQH BenhNhan.
 - b. Lập danh sách các bệnh nhân đã từng khám bác sĩ ‘Tran Dong A’. Thông tin gồm tất cả thuộc tính của LĐQH BenhNhan.
 - c. Lập danh sách các bệnh nhân có khám bệnh cùng bác sĩ với bệnh nhân có mã thẻ BHYT là ‘CH479123456789’ trong ngày 31/12/2020. Thông tin gồm tất cả thuộc tính của LĐQH BenhNhan.

CÂU II :

Cho lược đồ quan hệ HoaDon và tập các phụ thuộc hàm F như sau:

HoaDon (SOHD, KHACH, NGAYLAP, MATHANG, DONGIA, SOLUONG)

$F = \{ \text{SOHD} \rightarrow \text{KHACH}, \text{NGAYLAP},$
 $\text{SOHD}, \text{MATHANG} \rightarrow \text{DONGIA}, \text{SOLUONG} \}$

1. Cho biết phụ thuộc hàm KHACH, SOHD $\rightarrow$ DONGIA có được suy dẫn logic từ F không ?
2. Tìm tất cả các khóa của LĐQH Hoadon
3. Hãy cho biết LĐQH HoaDon có đạt dạng chuẩn nào ? Tại sao?
4. Nếu lược đồ chưa đạt dạng chuẩn 3 hãy phân rã thành các lược đồ con đạt dạng chuẩn 3, xác định khóa chính cho các lược đồ con này.