



Bài giảng

Cơ sở dữ liệu quan hệ

Bộ môn Tin học Cơ bản, Khoa CNTT, ĐH Mỏ - Địa Chất



MỤC LỤC

Chương 1. TỔNG QUAN VỀ CƠ SỞ DỮ LIỆU.....	4
Bùi 1. Các khái niệm cơ bản về cơ sở dữ liệu (2 tiết).....	4
1. Các hệ thống tệp truyền thống.....	4
2. Khái niệm và phân loại các hệ quản trị cơ sở dữ liệu.....	5
Bùi 2. Kiến trúc một hệ cơ sở dữ liệu (1 tiết).....	9
1. Mô hình kiến trúc 3 mức.....	9
2. Ưu điểm của kiến trúc 3 mức	11
3. Kết luận.....	11
Chương 2. MÔ HÌNH DỮ LIỆU.....	13
Bùi 1. Sơ đồ thực thể liên kết (1 tiết).....	13
1. Các khái niệm chung.....	13
Bùi 2. Các mô hình dữ liệu (2 tiết).....	15
1. Mô hình quan hệ (Relational model).....	15
2. Mô hình mạng (Network model).....	16
3. Mô hình phân cấp (Hierarchical model).....	17
4. Đánh giá và so sánh 3 loại mô hình dữ liệu	17
5. Kết luận và bài tập.....	18
Chương 3. MÔ HÌNH DỮ LIỆU QUAN HỆ.....	20
Bùi 1. Các khái niệm cơ bản (1 tiết).....	20
1. Thuộc tính và miền thuộc tính.....	20
2. Quan hệ.....	21
3. Khoá (Key).....	22
Bùi 2. Các phép toán đại số quan hệ (3 tiết).....	23
1. Định nghĩa.....	23
2. Các phép toán đại số.....	23
3. Các ví dụ về tìm kiếm bằng đại số quan hệ.....	28
4. Kết luận chung và bài tập:.....	28
Chương 4. NGÔN NGỮ VẤN TIN CẤU TRÚC SQL (STRUCTURE QUERY LANGUAGE).....	29
Bùi 1. Ngôn ngữ định nghĩa dữ liệu (2 tiết).....	29
1. Giới thiệu về ngôn ngữ SQL.....	29
2. Các mệnh đề của ngôn ngữ SQL.....	29
Bùi 2. Ngôn ngữ thao tác dữ liệu (3 tiết).....	32
1. Khối SELECT.....	32
2. Các mệnh đề tìm kiếm.....	32
3. Các mệnh đề cập nhật dữ liệu.....	37
4. Kết luận và bài tập.....	39
I - Khái quát về CSDL & CSDL quan hệ.....	40
I.1. Khái niệm CSDL	40
1. Định nghĩa.....	40

2. Các tiêu chuẩn của một CSDL.....	40
3. Hệ QTCSDL.....	40
4. Hệ thống thông tin.....	41
5. Kiến trúc một CSDL.....	41
I.2.Các mô hình CSDL.....	41
6. Mô hình phân cấp.....	41
7. Mô hình mạng.....	41
8. Mô hình quan hệ.....	41
I.3.Các khái niệm cơ bản về CSDL quan hệ.....	41
9. Thuộc tính.....	41
10. Quan hệ.....	41
11. Phụ thuộc hàm.....	41
12. Khóa.....	42
I.4.Các phép toán trên các quan hệ.....	42
13. Phép chiếu.....	42
14. Phép nối.....	42
15. Phép tách.....	42
16. Phép chọn : lấy ra các hàng có các thuộc tính thoả mãn một số tiêu chuẩn cho trước.....	42
II-Thiết kế hệ thống CSDL quan hệ.....	42
II.1.Phân tích tư liệu của XN.....	42
1. Xác định danh sách các thuộc tính.....	42
3. Tìm phụ thuộc hàm giữa các thuộc tính.....	43
II.2.Chuẩn hoá các quan hệ.....	43
5. Đưa quan hệ về dạng chuẩn 1 NF.....	43
7. Đưa quan hệ về dạng chuẩn 2NF.....	44
9. Dạng chuẩn 3NF.....	44
10. Dạng chuẩn BCNF.....	45

Chương 1. TỔNG QUAN VỀ CƠ SỞ DỮ LIỆU

Hệ quản trị cơ sở dữ liệu, đặc biệt là cơ sở dữ liệu quan hệ là một hệ thống phần mềm có vai trò quan trọng trong các hệ thống lập trình. Cũng giống như các loại phần mềm hệ thống chủ yếu khác như: trình biên dịch và hệ điều hành, các nguyên lý của hệ quản trị cơ sở dữ liệu đã được phát triển từ khá lâu. Những khái niệm này rất hữu ích, không những giúp cho việc sử dụng hiệu quả các hệ quản trị cơ sở dữ liệu mà còn hỗ trợ trong việc thiết kế và cài đặt chúng.

Bộ 1. Các khái niệm cơ bản về cơ sở dữ liệu (2 tiết)

1. Các hệ thống tệp truyền thống

1.1. Bài toán:

Giả thiết bài toán quản lý thư viện có hai chức năng như sau:

- Chức năng 1 (CN1): In ra danh mục sách với tệp sách chứa các thông tin: tên sách, mã sách, tên tác giả, nhà xuất bản,...
- Chức năng 2 (CN2): Thống kê mỗi loại sách có trong thư viện với số lượng cụ thể.

Để giải quyết bài toán trên có thể có hai cách:

Cách 1: Viết CN2 độc lập CN1. Với CN2 tạo một tệp sách phân loại (tệp 2) chứa các thông tin: mã sách, tên sách, nhà xuất bản, số lượng,...

Cách này có một số nhược điểm như sau:

- Giữa tệp 1 và tệp 2 có một số trường trùng nhau dẫn tới dư thừa dữ liệu
- Dữ liệu lưu lặp ở cả hai tệp, khi dữ liệu thay đổi phải cập nhật ở cả hai tệp. Dữ liệu không nhất quán khi tiến hành các thao tác sửa đổi (*hiện tượng dị thường*)

Cách 2: Tận dụng tệp sách đã có (tệp 1). Sửa tệp 1 thành tệp mới có chứa thêm trường số lượng. Đặc điểm của cách 2:

- Tránh được dư thừa và lặp lại dữ liệu
- Phải sửa lại CN1 để sử dụng tệp mới (tệp 2) -> chương trình phụ thuộc vào dữ liệu

1.2. Phương án đề xuất theo hướng mới

Cả hai cách trên đều có những hạn chế nhất định, do đó cần đưa ra một phương án mới, hiệu quả hơn, khắc phục được những hạn chế trên đây, với mục đích xây dựng một hệ thống thỏa mãn các yêu cầu sau:

- Tránh dư thừa dữ liệu

Bài giảng Cơ sở dữ liệu quan hệ

- Không phụ thuộc dữ liệu
- Các thao tác tra cứu, tìm kiếm, cập nhật nhanh chóng và hiệu quả

2. Khái niệm và phân loại các hệ quản trị cơ sở dữ liệu

2.1. Cơ sở dữ liệu là gì ?

Để dễ dàng cho việc giải thích các khái niệm, trước hết xem xét hệ thống bán vé máy bay bằng máy tính. Dữ liệu lưu trữ trong máy tính bao gồm thông tin về hành khách, chuyến bay, đường bay, v.v... Mọi thông tin về mối quan hệ này được biểu diễn trong máy thông qua việc đặt chỗ của khách hàng. Vậy làm thế nào để biểu diễn được dữ liệu đó và để đảm bảo cho khách hàng đi đúng chuyến.

Dữ liệu trên được lưu trữ trong máy theo một quy định nào đó và được gọi là *cơ sở dữ liệu* (viết tắt CSDL, tiếng Anh là *Database*)

Theo một định nghĩa khác: cơ sở dữ liệu là bộ lưu trữ các dữ liệu tác nghiệp của một xí nghiệp, được lưu trữ để phục vụ cho các ứng dụng.

Ví dụ 2.1: Xí nghiệp là một thư viện, dữ liệu tác nghiệp là: sách, đọc giả, yêu cầu, v.v...

2.2. Hệ quản trị cơ sở dữ liệu

Phần chương trình để có thể xử lý, thay đổi cơ sở dữ liệu gọi là *Hệ quản trị cơ sở dữ liệu* (viết tắt HQTCSDL, tiếng Anh là *Database management system*). Theo định nghĩa này HQTCSDL có nhiệm vụ rất quan trọng như là một bộ *diễn dịch* (*interpreter*) với ngôn ngữ bậc cao nhằm giúp người sử dụng có thể dùng được hệ thống mà ít nhiều không cần quan tâm đến thuật toán chi tiết hoặc biểu diễn dữ liệu trong máy.

Theo một cách hiểu khác:

HQTCSDL là một phần mềm cho phép tạo lập CSDL và điều khiển hoặc truy nhập CSDL đó, đặc biệt HQTCSDL đảm bảo tính độc lập dữ liệu (là sự bất biến của các chương trình ứng dụng đối với các thay đổi về cấu trúc lưu trữ và chiến lược truy nhập).

Ví dụ 2.2: Một số Hệ QTCSDL thông dụng hiện nay: MS Access, SQL Server (của hãng Microsoft), Oracle (của hãng Oracle), DB2, FoxPro, v.v...

a) Hệ quản trị cơ sở dữ liệu hỗ trợ các tính năng sau:

- Định nghĩa dữ liệu (*Database definition*)

Bài giảng Cơ sở dữ liệu quan hệ

- Xây dựng dữ liệu (*Database construction*) : Chức năng định nghĩa và xây dựng dữ liệu hỗ trợ người dùng xây dựng các bộ dữ liệu riêng.
 - Thao tác dữ liệu (*Database manipulation*): các thao tác cập nhật, tìm kiếm, sửa, xoá,...
 - Quản trị dữ liệu (*Database administrator*): phân quyền sử dụng, bảo mật thông tin,...
 - Bảo vệ dữ liệu (*Database protection*): thực hiện các thao tác sao chép, phục hồi, tránh mất mát dữ liệu.
- b) Ngôn ngữ của hệ quản trị cơ sở dữ liệu bao gồm:
- Ngôn ngữ con định nghĩa dữ liệu (*Database Definition Language - DDL*): cung cấp các câu lệnh cho phép mô tả, định nghĩa các đối tượng của CSDL.
 - Ngôn ngữ con thao tác dữ liệu (*Database Manipulation Language - DML*): dùng để thao tác, xử lý trên các đối tượng của CSDL như thêm, xoá, sửa, tìm kiếm, v.v...
 - Ngôn ngữ con kiểm soát dữ liệu (*Database Control Language - DCL*): điều khiển tính đồng thời (tương tranh) đối với dữ liệu.

Hình 1- Hệ CSDL đa người dùng

c) Hệ quản trị cơ sở dữ liệu đa người dùng:

- Cơ sở dữ liệu ở dạng hợp nhất (hay là tập hợp toàn bộ dữ liệu tác nghiệp của xí nghiệp), có hai tính chất cơ bản sau:
 - Không dư thừa dữ liệu: cố gắng tối thiểu và kiểm soát sự dư thừa.
 - Sử dụng chung nguồn dữ liệu: chia sẻ nhiều người sử dụng.
- Người sử dụng hệ QTCSDL bao gồm:
 - Người phân tích hệ thống (*System analyst*)
 - Người thiết kế CSDL (*Database designer*)
 -)
 - Người viết chương trình ứng dụng (*Application programmer*): xây dựng các chương trình ứng dụng dựa trên các cơ sở dữ liệu đã có.
 - Người sử dụng cuối (*end - user*): là người truy nhập vào CSDL từ một thiết bị đầu cuối.

Bài giảng Cơ sở dữ liệu quan hệ

- Người quản trị CSDL (*Database administrator*): thường là một người hoặc một nhóm người có nhiệm vụ điều khiển toàn bộ hệ CSDL.

2.3. Phân loại các hệ quản trị cơ sở dữ liệu

Hệ QTCSDL có thể được phân loại dựa theo nhiều tiêu chí khác nhau, và mỗi cách phân loại có thể có nhiều kiểu hệ QTCSDL. Trong phần này mô tả 3 kiểu hệ QTCSDL khác nhau được phân loại theo quan điểm chung nhất.

a) Hệ QTCSDL loại 1:

Người dùng vừa thiết kế, sử dụng, và quản trị, hay còn gọi là hệ QTCSDL đơn người dùng (hệ CSDL cá nhân, nhỏ). Thường sử dụng để giải quyết những nhiệm vụ đơn lẻ với một người hoặc một vài người.

Hình 2- Mô hình hệ QTCSDL đơn người dùng

b) Hệ QTCSDL loại 2:

Người sử dụng cuối truy nhập CSDL thông qua các thiết bị truy nhập đầu cuối (*terminal*), còn gọi là hệ QTCSDL đa người dùng (hệ QTCSDL trung tâm).

Hình 3- Mô hình hệ QTCSDL đa người dùng

c) Hệ QTCSDL loại 3:

Những yêu cầu của người sử dụng cuối và trình ứng dụng được xử lý tại các máy trạm (*Client*), chỉ những yêu cầu nào cần tới CSDL mới được chuyển tới hệ QTCSDL nằm trên máy chủ (*Server*), đây là mô hình hệ QTCSDL Client / Server.

Hình 4 - Mô hình hệ QTCSDL Client / Server

Nhận xét: Cả 3 loại hệ QTCSDL trên đều đặt CSDL tại một nơi. Do đó hệ QTCSDL có tính chất tập trung.

Ví dụ 2.3:

- Một công ty muốn lưu trữ và duy trì thông tin về các nhân viên cung cấp (*supplier*) và các mặt hàng (*part*).
- Các thông tin lưu trữ cần thiết về nhân viên cung cấp bao gồm: số hiệu nhân viên, họ tên, ngày sinh, mức lương, và địa chỉ thành phố.
- Thông tin về mặt hàng bao gồm: số hiệu mặt hàng, tên mặt hàng, màu sắc và giá.
- Việc chuyển hàng được mô tả thông qua số hiệu nhân viên, tên mặt hàng và số lượng.

NHAN_VIEN

s#	Ho_Ten	Thanh_Ph	Nam_Sinh	Luong
----	--------	----------	----------	-------

Bài giảng Cơ sở dữ liệu quan hệ

10	Lê Văn A	Ha Noi	1960	400
20	Hoàng Thị B	HCM	1970	500
30	Lê Văn Sơn	Hai Phong	1945	600

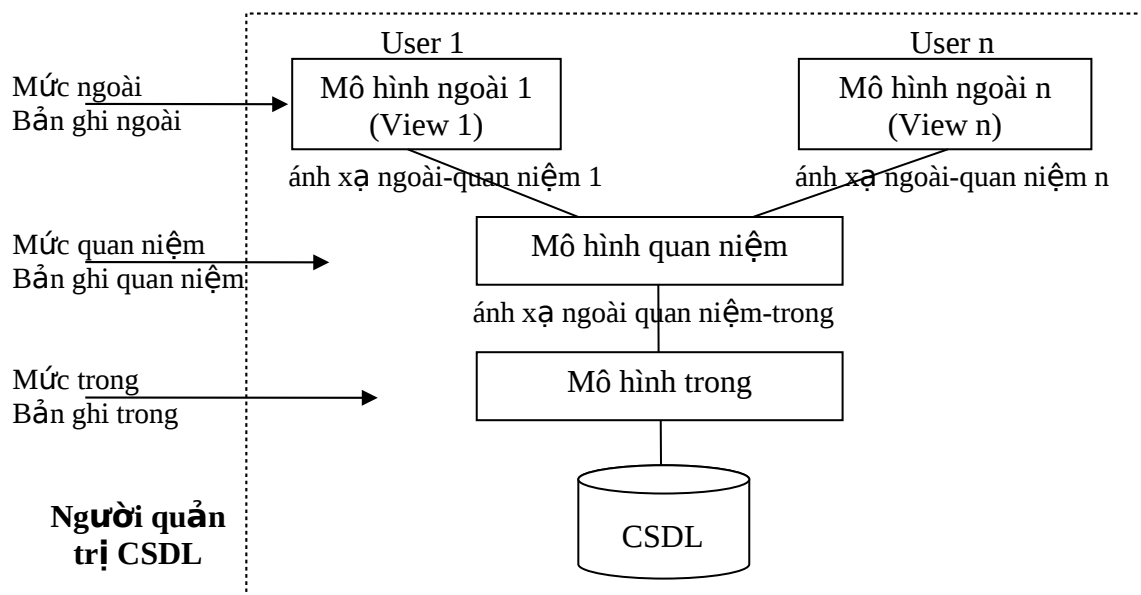
Hình 5 - Quan hệ NHÂN VIÊN

Bộ 2. Kiến trúc một hệ cơ sở dữ liệu (1 tiết)

1. Mô hình kiến trúc 3 mức

Một CSDL được phân thành các mức khác nhau như trên hình. Ở đây có thể xem như chỉ có một CSDL đơn giản và có một hệ phần mềm QTCSDL.

Người dùng truy nhập thông qua mô hình ngoài, qua ánh xạ ngoài quan niệm đi vào mô hình quan niệm, qua ánh xạ quan niệm trong vào mô hình trong, và truy nhập vào CSDL thực hiện các yêu cầu. Sau khi thực hiện các yêu cầu, ánh xạ ngược trở lại phía người dùng.



Hình 6 - Kiến trúc hệ QTCSDL

Đây là kiến trúc chuẩn 3 mức:

- **Mức ngoài:** là mức sát với người dùng. Mức này tương ứng với cách nhìn riêng của người dùng đối với hệ QTCSDL.
- **Mức trong (mức vật lý):** là mức sát với lưu trữ thực sự của các tệp dữ liệu theo một cấu trúc nào đó trên các thiết bị nhớ thứ cấp (như đĩa từ, băng từ...).
- **Mức khái niệm:** là mức trung gian. CSDL mức khái niệm là một sự biểu diễn trừu tượng của CSDL vật lý (hay có thể nói: CSDL mức vật lý là sự cài đặt cụ thể của CSDL mức khái niệm).

Bài giảng Cơ sở dữ liệu quan hệ

- **Các khung nhìn (view):** là cách nhìn hay quan niệm của từng người sử dụng đối với CSDL mức khái niệm. Sự khác nhau giữa khung nhìn và mức khái niệm thực chất là không lớn.

Tương ứng với 3 mức trên là 3 mô hình cụ thể của hệ QTCSDL bao gồm: mô hình ngoài, mô hình quan niệm, mô hình trong. Các mô hình này tương tác thông qua các ánh xạ.

- (1) Mô hình ngoài: của một người dùng là tập hợp tất cả dữ liệu của người dùng được phép nhìn thấy và truy nhập vào, thường chỉ là một bộ phận của toàn bộ CSDL.
- (2) Mô hình quan niệm: bao gồm tập toàn bộ dữ liệu của xí nghiệp và được biểu diễn ít nhiều trừu tượng.
- (3) Mô hình trong: bao gồm tập toàn bộ dữ liệu của xí nghiệp được biểu diễn sát với lưu trữ thật sự trên các thiết bị nhớ.

Mô hình ngoài thường được biểu diễn thông qua bản ghi gọi là bản ghi ngoài, tương ứng với mô hình quan niệm và mô hình trong ta có bản ghi quan niệm và bản ghi trong.

Do có sự khác nhau giữa bản ghi ngoài và bản ghi quan niệm, để ánh xạ tới người dùng cần có ánh xạ ngoài-quan niệm. Tương ứng cần có ánh xạ quan niệm-trong giữa mô hình quan niệm và mô hình trong.

Thực chất các mô hình ngoài, mô hình quan niệm và mô hình trong là các sơ đồ kiểu bản ghi, tương ứng các ánh xạ là ánh xạ từ kiểu bản ghi này sang kiểu bản ghi khác.

HỆ QTCSDL trên được quản lý bởi người quản trị CSDL:

- Là người quyết định nội dung thông tin của CSDL, được hoàn thành thông qua viết sơ đồ quan niệm hay biểu diễn mô hình quan niệm.
- Là người quyết định cấu trúc lưu trữ và số lượng truy nhập đối với toàn bộ dữ liệu thông qua viết sơ đồ trong.
- Với sự giúp đỡ của người sử dụng, viết sơ đồ ngoài tương ứng với người sử dụng.
- Đưa ra các phương án sao lưu, phục hồi cho CSDL.
- Đưa ra các cách kiểm soát thẩm quyền và kiểm tra tính đúng đắn của CSDL.

2. Ưu điểm của kiến trúc 3 mức

Tại sao phân chia hệ QTCSDL thành 3 mức ? Việc phân chia như vậy có đạt được mục tiêu đặt ra không ?

Kiến trúc một hệ CSDL cần bảo đảm các mục tiêu sau:

- (1) Tránh dư thừa dữ liệu
- (2) Tính độc lập giữa dữ liệu và chương trình ứng dụng.
- (3) Tra cứu, tìm kiếm, cập nhật dữ liệu nhanh chóng.
- (4) Đảm bảo tính an toàn và toàn vẹn dữ liệu.

Chứng minh tính đúng đắn:

(1) Tránh dư thừa dữ liệu: Tất cả các mô hình ngoài (thực chất là ảo) đều truy nhập vào cùng CSDL (tính hợp nhất của dữ liệu) do đó đảm bảo không dư thừa dữ liệu.

(2) Tính độc lập:

Ba mức trong hệ QTCSDL (mức ngoài, mức quan niệm và mức trong) độc lập với nhau, khi xoá hay thêm logic diễn ra trên mô hình quan niệm tương ứng là sự thay đổi vật lý diễn ra ở mô hình trong (cấu trúc lưu trữ), do đó đảm bảo tính độc lập logic (tính bất biến của ứng dụng với sự thay đổi cơ sở ở mức logic).

Ba mức đảm bảo tính độc lập: khi mô hình trong thay đổi chỉ cần sửa ánh xạ trong quan niệm, các mô hình quan niệm và ứng dụng của người dùng không cần thay đổi (đảm bảo tính độc lập vật lý). Tương tự, khi mô hình quan niệm thay đổi chỉ cần sửa ánh xạ ngoài quan niệm.

(3) Tra cứu, tìm kiếm và cập nhật dữ liệu nhanh chóng: Sử dụng các thuật toán (ví dụ: đánh chỉ số cho CSDL,...) cho phép tra cứu, tìm kiếm và cập nhật nhanh dữ liệu.

(4) Tính an toàn và toàn vẹn dữ liệu: Mô hình ngoài chính là kỹ thuật đảm bảo tính an toàn, toàn vẹn dữ liệu. Mô hình quan niệm là mô hình Ổn định vì mô hình ngoài phụ thuộc người dùng, mô hình trong phụ thuộc thiết bị thứ cấp.

3. Kết luận

Trong chương này chúng ta đã làm quen và tìm hiểu các khái niệm cơ bản liên quan đến CSDL, kiến trúc và các ưu điểm của hệ QTCSDL so với cách lưu trữ trước đây. Vậy khi nào cần và khi nào không cần tới hệ QTCSDL ?

3.1. Tại sao cần hệ QTCSDL

- Để đảm bảo tính tiêu chuẩn hoá: các hệ CSDL khác nhau, dữ liệu của các chương trình ứng khác nhau dựa trên một tiêu chuẩn chung.
- Cung cấp các công cụ định nghĩa và thao tác dữ liệu linh hoạt
- Tích hợp với nhiều trình ứng dụng khác nhau: các ngôn ngữ lập trình, các ứng dụng hỗ trợ phân tích thiết kế, v.v...

3.2. Khi nào không cần hệ QTCSDL

- Khi chúng ta giải quyết các vấn đề đơn giản mà các chương trình ứng dụng có thể thực hiện tốt không cần tới hệ CSDL.
- Khi hệ thống CSDL không đáp ứng được yêu cầu về hiệu năng như: tốc độ, tính bảo mật, định dạng dữ liệu cần lưu trữ, v.v..
- Khi không cần thiết đa người dùng cùng truy nhập vào một CSDL chung.

3.3. Tài liệu tham khảo

- Lê Tiến Vương, *Nhập môn Cơ sở dữ liệu quan hệ*, NXB Thống kê, Chương 1
- P. O' Neil, *Database - Principles, Programming, Performance*, Chương 1.1, 1.2
- R. Elmasri, S.B. Navathe, *Fundamentals of Database Systems*, Chương 1

Chương 2. MÔ HÌNH DỮ LIỆU

Trong chương này, chúng ta sẽ xem xét các mô hình chính được sử dụng trong các hệ thống cơ sở dữ liệu. Trong đó đặc biệt nhấn mạnh về mô hình thực thể liên kết, mô hình này được dùng chủ yếu làm công cụ thiết kế CSDL.

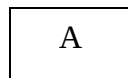
Bộ 1. Sơ đồ thực thể liên kết (1 tiết)

1. Các khái niệm chung

1.1. Thực thể

Thực thể được định nghĩa là một đối tượng có thực hay trừu tượng mà ta muốn lưu trữ thông tin về nó.

Để biểu diễn (hay kí hiệu) thực thể: sử dụng hình chữ nhật bao quanh tên thực thể (tên thực thể biểu diễn là danh từ).



A: tên thực thể, là danh từ

1.2. Thuộc tính

Một thực thể được xây dựng bởi một tập thuộc tính đặc trưng cho thực thể.

Biểu diễn:

- a₁ ai : thuộc tính thứ i của thực thể A
- (là danh từ)
- ai
-
- an

Ví dụ 1.2: Thực thể sách gồm các thuộc tính: mã sách, tên sách, tác giả, nhà xuất bản

1.3. Liên kết thực thể

Liên kết thực thể chỉ mối quan hệ ràng buộc giữa các thực thể. Liên kết thực thể chia thành 3 loại như sau:

a) Kiểu liên kết một thực thể:

Là liên kết của thực thể đó với chính nó (hay còn gọi là liên kết đệ quy)

Biểu diễn:

b) Kiểu liên kết hai thực thể:

Liên kết hai thực thể là mối liên kết giữa hai thực thể khác nhau. Tên của các liên kết được biểu diễn bằng các động từ. Có 3 loại liên kết hai thực thể:

Bài giảng Cơ sở dữ liệu quan hệ

- Liên kết 1 -1: là liên kết thoả mãn điều kiện nếu xuất hiện một thực thể A thì xuất hiện một thực thể B hoặc ngược lại. Biểu diễn như sau:
- Liên kết 1 - nhiều: là liên kết thoả mãn điều kiện nếu xuất hiện một thực thể A thì xuất hiện nhiều thực thể B hoặc ngược lại, nếu xuất hiện một thực thể B thì xuất hiện nhiều thực thể A. Biểu diễn như sau:

Ví dụ 1.3.1: Xét quan hệ giữa nhân viên và phòng trong một công ty. Đây là quan hệ một nhiều (một nhân viên thuộc chỉ một phòng còn một phòng có thể có nhiều nhân viên).

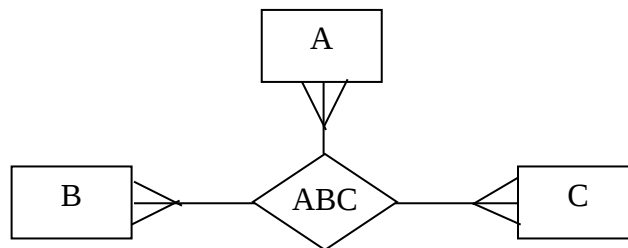
- Liên kết nhiều -nhiều: là liên kết thoả mãn điều kiện xuất hiện nhiều thực thể A thì xuất hiện nhiều thực thể B hoặc ngược lại. Biểu diễn như sau:

Ví dụ 1.3.2: Quan hệ giữa thực thể sách và thực thể độc giả là quan hệ nhiều - nhiều (giả thiết một độc giả có thể mượn nhiều loại sách cùng một lúc và một loại sách có nhiều bản sao có thể cho mượn).

c) Liên kết nhiều thực thể:

Là mối liên kết trong đó có nhiều hơn hai thực thể. Để biểu diễn liên kết nhiều thực thể và đơn giản hoá khi biểu diễn ta quy các liên kết nhiều thực thể này về các liên kết hai thực thể bằng cách đưa thêm vào thực thể trung gian (kí hiệu là: TG).

Biểu diễn liên kết nhiều - nhiều - nhiều như sau:



Khi thêm thực thể trung gian (TG), liên kết nhiều - nhiều - nhiều sẽ chuyển thành 3 liên kết 1 - nhiều (đây là liên kết hai thực thể). Biểu diễn như sau:

Theo định nghĩa liên kết thực thể ta có thể coi liên kết thực thể là một dạng thực thể đặc biệt và cũng được lưu trữ. Do đó CSDL phải lưu trữ cả hai đối tượng: thực thể và liên kết thực thể, đồng thời CSDL phải có khả năng biểu diễn hai đối tượng này.

Bài 2. Các mô hình dữ liệu (2 tiết)

Hiện nay có nhiều loại mô hình dữ liệu khác nhau. Trong bài này chúng ta sẽ chỉ nghiên cứu ba loại mô hình dữ liệu cơ bản thường được sử dụng bao gồm: mô hình quan hệ, mô hình mạng, và mô hình phân cấp.

1. Mô hình quan hệ (Relational model)

Mô hình này dựa trên cơ sở khái niệm lý thuyết tập hợp của các quan hệ, tức là tập các k - bộ với k cố định.

Hay nói một cách đơn giản hơn, trong mô hình quan hệ dữ liệu được biểu diễn dưới dạng các bảng, đặc biệt cả hai đối tượng: thực thể và liên kết thực thể được biểu diễn dưới dạng duy nhất là dạng bảng.

Như vậy sơ đồ thực thể liên kết được ánh xạ thành mô hình quan hệ:

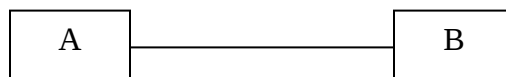
- Tên của thực thể được ánh xạ thành tên bảng, mỗi bảng biểu diễn một thực thể tương ứng.
- Thuộc tính của thực thể tương ứng với các cột trong bảng.

Ví dụ 1.1: Thực thể Nhân viên trong một công ty được biểu diễn dưới dạng bảng như sau:

SS#	HoHo_Tenen	Thanhnh_Phoho	Namam_Sinhnh	Luongng
11	Bùi Ngọc Anh	Hà Nội	1960	600
	Nguyễn Văn Huyền	Hà Nội	1965	500
12		Hải Phòng	1970	400
13	Nguyễn Văn Quang	HCM	1975	400
	Trần Văn Trung			

Hình 7 - Biểu diễn thực thể dưới dạng bảng

- Biểu diễn liên kết thực thể:
 - Trước tiên phải xác định khoá của thực thể, sau đó chuyển khoá này thành cột khoá của bảng tương ứng (khái niệm khoá sẽ được giải thích rõ trong phần sau).
 - Bước tiếp theo là biến đổi liên kết thực thể:
 - Đối với liên kết 1 - 1:** Lấy khoá của thực thể này đặt vào thực thể kia.

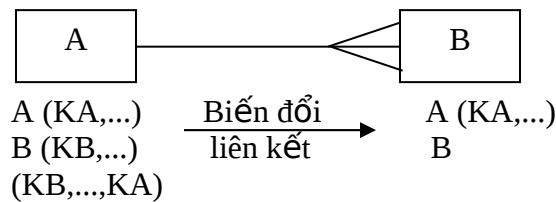


Bộ môn Tin học Cơ bản, Khoa CNTT, ĐHQG Hà Nội

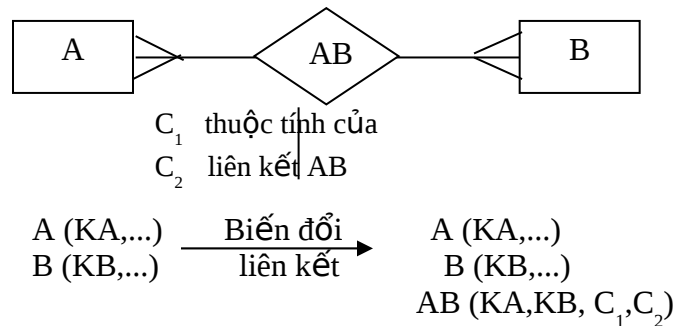
Điểm đến Địa Chấ

(KA,...,KB)
B (KB,...) liên kết B
(KB,...,KA)

Đối với liên kết 1 - nhiều: Lấy khoá của bảng đầu liên kết một đặt vào bảng đầu liên kết nhiều.



Đối với liên kết nhiều - nhiều: Thêm một bảng mới (bảng trung gian) gồm



khoá của các bảng liên kết và thuộc tính liên kết

Ví dụ 1.2: Xét quan hệ giữa nhân viên và phòng trong một công ty. Đây là quan hệ một nhiều (một nhân viên thuộc chỉ một phòng còn một phòng có thể có nhiều nhân viên).

Mô hình quan hệ:

2. Mô hình mạng (Network model)

Dữ liệu được biểu diễn dưới dạng một đồ thị có hướng phức tạp (dạng đa danh sách).

Xây dựng một ánh xạ biến đổi sơ đồ thực thể liên kết sang mô hình mạng. Nó chỉ có khả năng biểu diễn trực tiếp đối với các liên kết 1 - 1 hay 1 - nhiều. Do đó để biểu diễn được các liên kết khác như liên kết nhiều - nhiều, cần quy các liên kết này về dạng 1- nhiều bằng cách đưa thêm một thực thể trung gian (TG) vào liên kết.

Chuyển thành liên kết 1- nhiều:

Với mô hình mạng:

- Thực thể chuyển thành kiểu bản ghi.
- Thuộc tính chuyển thành các trường của bản ghi.
- Liên kết giữa các thực thể: một liên kết 1 - nhiều tương ứng tập gán nhãn.
 - Kiểu bản ghi ở đầu 1 gọi là kiểu bản ghi chủ
 - Kiểu bản ghi ở đầu nhiều gọi là kiểu bản ghi thành viên

Bài giảng Cơ sở dữ liệu quan hệ

Biểu diễn:

Chiều mũi tên hướng tới đầu nhiều

Ví dụ 2: Xét quan hệ giữa nhân viên và phòng trong một công ty. Đây là quan hệ một nhiều (một nhân viên thuộc chỉ một phòng còn một phòng có thể có nhiều nhân viên).

3. Mô hình phân cấp (Hierarchical model)

Mô hình dữ liệu là một cây, trong đó các nút biểu diễn các tập thực thể, giữa các nút con và nút cha được liên hệ theo một mối quan hệ xác định.

Mô hình phân cấp là trường hợp đặc biệt của mô hình mạng với một số điều kiện hạn chế sau:

- Không có chu trình.
- Mỗi một kiểu bản ghi chỉ là kiểu bản ghi thành viên nhiều nhất của một tập.

Với mô hình phân cấp khi biểu diễn không cần kí hiệu mũi tên. Trong cấu trúc cây, đầu một tương đương với nút gốc của cây, và đầu nhiều tương ứng với các nút lá của cây.

Ví dụ 3: Xét quan hệ giữa nhân viên và phòng trong một công ty như ví dụ trên.

4. Đánh giá và so sánh 3 loại mô hình dữ liệu

4.1. Cách biểu diễn dữ liệu

- Mô hình quan hệ: cả hai đối tượng thực thể và liên kết thực thể đều biểu diễn dưới dạng bảng
- Mô hình mạng: các đối tượng thực thể biểu diễn dưới dạng bản ghi, liên kết thực thể tương ứng là liên kết móc nối giữa các bản ghi dạng đồ thị có hướng
- Mô hình phân cấp: các đối tượng thực thể biểu diễn dưới dạng bản ghi, liên kết thực thể tương ứng là liên kết giữa các nút của cây

4.2. Cách thực hiện yêu cầu

Ví dụ 4: thực hiện các yêu cầu sau:

- Bổ sung thêm một quyển sách ('S03', 'Mạng máy tính', 'NT Hải') vào CSDL
- Xoá độc giả với mã độc giả là 'ĐG1'

Mô hình quan hệ:

Bài giảng Cơ sở dữ liệu quan hệ

- Với yêu cầu thứ nhất, hệ thống sẽ thêm một hàng trong bảng *Sách*.
- Với yêu cầu thứ hai để xoá độc giả có mã 'ĐG1', trước tiên sẽ xoá độc giả này trong bảng *Mượn* sau đó mới xoá trong bảng *Độc giả*.

Mô hình mạng và mô hình phân cấp:

- Các phép toán thao tác đối mô hình mạng và phân cấp phức tạp hơn đối với mô hình quan hệ (chỉ thao tác trên các bảng).

5. Kết luận và bài tập

Chương này đã trình bày các khái niệm và sơ đồ thực thể liên kết, đặc biệt là ba loại mô hình dữ liệu: mô hình quan hệ, mô hình mạng, và mô hình phân cấp. Đây là các loại mô hình cơ bản để xây dựng các hệ QTCSDL hiện nay.

5.1. Tài liệu tham khảo

- Lê Tiến Vương, *Nhập môn Cơ sở dữ liệu quan hệ*, NXB Giáo dục, Chương 1
- P. O' Neil, *Database - Principles, Programming, Performance*, Chương 1.3, 1.4
- R. Elmasri, S.B. Navathe, *Fundamentals of Database Systems*, Chương 3,4

5.2. Bài tập

Bài 1) Cho sơ đồ thực thể liên kết giữa công việc, nhân viên và phòng như dưới đây, hãy biến đổi sang 3 mô hình dữ liệu:

Bài 2) Cho sơ đồ thực thể liên kết giữa sách và độc giả như sau, hãy hoàn chỉnh sơ đồ sau đó biến đổi sang 3 mô hình dữ liệu.

5.3. Chữa bài tập

Bài 1) biến đổi sang 3 mô hình dữ liệu:

a) Mô hình quan hệ:

Công việc (Mã CV, Tên CV, Định mức)

Nhân Viên (Mã NV, HỌ và tên, Ngày sinh, Mã CV, Mã phòng)

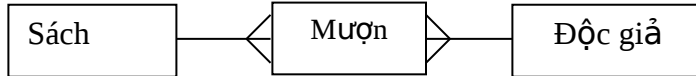
Phòng (Mã phòng, Tên phòng, Số phòng)

b) Mô hình mạng:

c) Mô hình phân cấp:

Bài giảng Cơ sở dữ liệu quan hệ

Bài 2) Chuyển sơ đồ quan hệ về dạng dưới đây, sau đó thực hiện tương tự:



Chương 3. MÔ HÌNH DỮ LIỆU QUAN HỆ

Mô hình dữ liệu quan hệ là mô hình phổ biến và quan trọng, được E.Codd đưa ra vào năm 1970. Một trong những ưu điểm của mô hình dữ liệu quan hệ là có hỗ trợ các ngôn ngữ khai báo, khá đơn giản nhưng hiệu quả và các phép toán trên dữ liệu. Các phép toán này có thể tổ hợp và phân tách dễ dàng nhờ vào một hệ thống kí hiệu đại số gọi là đại số quan hệ (relational algebra).

Bộ 1. Các khái niệm cơ bản (1 tiết)

1. Thuộc tính và miền thuộc tính

□ Theo khái niệm toán học, mô hình quan hệ được hiểu là quan hệ với ý nghĩa của lý thuyết tập hợp, tức là tập con tích Đề - Các của các miền. Với miền (domain) là một tập các giá trị. Ví dụ: tập các số nguyên là một miền, tập các âm kí tự tạo thành tên người trong tiếng Anh có độ dài không quá 30 kí tự là một miền; tập 2 số $\{0,1\}$ cũng là một miền v.v...

Gọi D_1, D_2, \dots, D_n , là n miền. Tích Đề - Các của n miền là $D_1 \times D_2 \times \dots \times D_n$ là tập tất cả n - bộ (n - tuples) (v_1, v_2, \dots, v_n) sao cho $v_i \in D_i$, với $i = 1..n$

Ví dụ 1. 1: $n = 2, D_1 = \{0,1\}, D_2 = \{a,b,c\}$, khi đó

$$D_1 \times D_2 = \{(0,a), (0,b), (0,c), (1,a), (1,b), (1,c)\}$$

□ Theo cách hiểu khác, miền thuộc tính là một tập các giá trị mà từ đó có thể rút ra những giá trị thực sự xuất hiện trong cột biểu diễn thuộc tính đó.

Ví dụ 1. 2: miền mã sách, kí hiệu DOM (Mã Sách)

$$\text{DOM (Mã Sách)} = \{ 'S01', 'S02', \dots, 'S99' \}$$

miền tên tác giả, kí hiệu DOM (Tên TG)

$$\text{DOM (Tên TG)} = \{ \text{char}(40) \}$$

2. Quan hệ

2.1. Định nghĩa 1

Cho n tập D_1, D_2, \dots, D_n , không nhất thiết phân biệt nhau, khi đó R được định nghĩa là 1 quan hệ trên n tập đó nếu R là một tập nào đó các bộ được sắp (d_1, d_2, \dots, d_n) sao cho $d_i \in D_i$

Ví dụ 2.1: quan hệ $r = \{ (d_1^i, d_2^i, \dots, d_n^i) \mid i = 1..m \}$ là n bộ được sắp thứ i.

Đặt tên các cột là A_1, A_2, \dots, A_n :

A_1	..., A_i , ...	A_n
d_1^1	..., d_i^1 , ...	d_n^1
d_1^2	..., d_i^2 , ...	d_n^2
\vdots	\vdots	\vdots
d_1^m	..., d_i^m , ...	d_n^m

Mỗi quan hệ có thể biểu diễn dưới dạng bảng, khi đó một dòng tương ứng với một bộ của quan hệ, một cột tương ứng với các giá trị của một thuộc tính.

Quan hệ	Bảng	Tập
Bộ	Dòng	Bản ghi
Thuộc tính	Cột	Trường

Có hai đại lượng đo kích thước quan hệ:

- Số thuộc tính của quan hệ: gọi là bậc quan hệ
- Số bộ: gọi là lực lượng của quan hệ

2.2. Định nghĩa 2

Gọi $R = \{A_1, A_2, \dots, A_n\}$ là tập hữu hạn của các thuộc tính, mỗi thuộc tính A_i với $i = 1, \dots, n$ có miền giá trị tương ứng là $DOM(A_i)$. Khi đó r là quan hệ xác định trên n thuộc tính nếu r là tập con hoặc trùng với tích Đề Các của D_i ($i = 1..n$)

$$r \subseteq D_1 \times D_2 \times \dots \times D_n$$

Khi đó kí hiệu là $r(R)$ hoặc $r(A_1, A_2, \dots, A_n)$.

Ví dụ 2.2: Hình cho thấy quan hệ NHAN_VIEN bao gồm các thuộc tính Ho_Ten, Thanh_Pho, Nam_Sinh và Luong là một quan hệ 4 ngôi.

NHAN_VIEN (Ho_Ten, Thanh_Pho, Nam_Sinh, Luong)

	Ho_Ten	Thanh_Pho	Nam_Sinh	Luong
t_1	Lê Văn A	Ha Noi	1960	400
t_2	Hoàng Thị B	HCM	1970	500

t ₃	Lê Văn Sơn	Hai Phong	1945	600
----------------	------------	-----------	------	-----

Hình 8 - Quan hệ NHÂN VIÊN

t₁ = (Lê Văn A, Ha Noi, 1960, 400) là một bộ của quan hệ NHAN_VIEN

3. Khoá (Key)

3.1. Định nghĩa

Khoá của một quan hệ r xác định trên tập thuộc tính $R = \{A_1, A_2, \dots, A_n\}$ là tập con $K \subseteq \{A_1, A_2, \dots, A_n\}$ nếu với mọi $t_1, t_2 \in r$, $t_1 \neq t_2$ đều tồn tại một thuộc tính $A \in K$ sao cho giá trị của thuộc tính t_1 tại A khác giá trị t_2 tại A ($t_1[A] \neq t_2[A]$).

Nói một cách khác, không tồn tại hai bộ mà có giá trị bằng nhau trên mọi thuộc tính của K . Điều kiện này có thể viết ($t_1[K] \neq t_2[K]$). Do vậy mỗi giá trị của K là xác định duy nhất một bộ.

Ví dụ 3.1:

$K_1 = \{\text{Mã_Sách}\}$ là khoá

$K_2 = \{\text{Mã_ĐộcGiả}\}$ là khoá

$K_3 = \{\text{Tên_Sách}\}$ không phải là khoá vì có thể có hai quyển sách có tên giống nhau nhưng nội dung khác nhau (không duy nhất).

3.2. Tính chất

□ Nếu K là khoá của quan hệ $r(A_1, A_2, \dots, A_n)$ thì $K \subseteq K' \subseteq R$, K' cũng là khoá của r , nghĩa là bất kì $t_1, t_2 \in r$ từ $t_1[K] \neq t_2[K]$ luôn có $t_1[K'] \neq t_2[K']$.

□ K là khoá tối thiểu của nếu như K là khoá của r và với mọi $K' \subsetneq K$, K' không là khoá của quan hệ r .

□ Một quan hệ có thể có nhiều khoá tối thiểu, do đó không cần xác định hết khoá tối thiểu mà chỉ cần xác định 1, 2 khoá tối thiểu làm khoá chính, các khoá khác là dự bị.

□ K được gọi là khoá ngoài của quan hệ r nếu như K không phải là khoá chính của quan hệ r nhưng nó lại là khoá chính của quan hệ khác.

Ví dụ 3.2:

Thực thể Sách có khoá chính là Mã_Sách.

Thực thể Độc giả có khoá chính là Mã_Độc_Giả.

Thực thể Mượn có khoá ngoài là Mã_Sách và Mã_Độc_Giả.

Bài 2. Các phép toán đại số quan hệ (3 tiết)

Trong bài này sẽ trình bày chi tiết đại số quan hệ như là cơ sở của một ngôn ngữ bậc cao để thao tác trên các quan hệ. Bao gồm các phép toán tập hợp như: hợp, trừ, giao, tích Đề Các và các phép toán quan hệ như: chọn, chiếu, kết nối, chia.

1. Định nghĩa

a) Định nghĩa 1:

Hai quan hệ r và s gọi là khả hợp nếu chúng được xác định trên cùng tập miền trị.

Ví dụ: quan hệ r xác định trên miền $\{D_1, D_2, \dots, D_n\}$ có bậc n
quan hệ s xác định trên miền $\{D'_1, D'_2, \dots, D'_m\}$ có bậc m
nếu $n = m$ và $\{r\} = \{s\}$ thì r và s được gọi là khả hợp

b) Định nghĩa 2:

Hai quan hệ r và s khả hợp nếu chúng được xác định trên cùng tập thuộc tính, các thuộc tính cùng tên có cùng miền trị và đối sánh với nhau.

2. Các phép toán đại số

2.1. Phép hợp (Union)

Hợp của hai quan hệ r và s khả hợp, kí hiệu là $r \cup s$, là tập các bộ thuộc r hoặc s hoặc thuộc cả hai quan hệ.

Biểu diễn hình thức cho phép hợp có dạng như sau:

Ví dụ 2.1:

r (A B C)	s (A B C)	$r \cup s$ (A B C)
$a_1 \ b_1 \ c_1$	$a_1 \ b_1 \ c_1$	$a_1 \ b_1 \ c_1$
$a_2 \ b_1 \ c_2$	$a_2 \ b_2 \ c_2$	$a_2 \ b_1 \ c_2$
$a_2 \ b_2 \ c_1$		$a_2 \ b_2 \ c_1$
		$a_2 \ b_2 \ c_1$

2.2. Phép giao (Intersect)

Giao của hai quan hệ r và s khả hợp, kí hiệu là $r \cap s$, là tập các bộ thuộc cả hai quan hệ r và s . Biểu diễn hình thức cho phép giao có dạng như sau:

Ví dụ 2.2: Với r và s là hai quan hệ trong ví dụ trên, giao của chúng là:

$$r \cap s = \begin{pmatrix} A & B & C \\ a_1 & b_1 & c_1 \end{pmatrix}$$

2.3. Phép trừ (Minus)

Hiệu của hai quan hệ r và s khả hợp, kí hiệu r - s.

Biểu diễn hình thức phép trừ có dạng như sau:

Ví dụ 2.3: Với r và s là hai quan hệ trong ví dụ trên, phép trừ của chúng là:

$$r - s = \begin{pmatrix} A & B & C \\ a_2 & b_1 & c_2 \\ a_2 & b_2 & c_1 \end{pmatrix}$$

Chú ý: Phép giao của 2 quan hệ $r \cap s$ có thể biểu diễn quan phép trừ:

$$r \cap s = r - (r - s)$$

2.4. Phép chọn (Selection)

Phép chọn trên một quan hệ thực chất là phép toán lọc ra một tập con các bộ của quan hệ thoả mãn điều kiện cho trước. Điều kiện là biểu thức chọn, là một tổ hợp logic của các toán hạng trong đó mỗi một toán hạng là một phép so sánh giữa hai thuộc tính hoặc giữa một thuộc tính và một giá trị hằng.

- Các phép toán logic bao gồm: \wedge (và), \vee (hoặc), \neg (không hay phủ định).
- Các phép toán so sánh trong biểu thức là: $<$, $=$, $>$, $<=$, $>=$ và \neq .

Hình thức hoá phép chọn được định nghĩa như sau:

$$\sigma_F(r) = \{ t \in r \mid F(t) = \text{đúng} \} \quad (2.4)$$

F(t) được hiểu là giá trị của các thuộc tính xuất hiện trong biểu thức F tại bộ t thoả các điều kiện của F.

Ví dụ 2.4: cho quan hệ như sau:

$$r \begin{pmatrix} A & B & C & D \\ a_1 & b_1 & c_1 & d_1 \\ a_1 & b_1 & c_1 & d_2 \\ a_2 & b_2 & c_2 & d_2 \\ a_2 & b_2 & c_3 & d_3 \end{pmatrix}$$

Các phép chọn:

$$\sigma_{A=a_1}(r) = s \begin{pmatrix} A & B & C & D \\ a_1 & b_1 & c_1 & d_1 \\ a_1 & b_1 & c_1 & d_2 \end{pmatrix} \quad \sigma_{A=a_1 \wedge D=d_2}(r) = \begin{pmatrix} A & B & C & D \\ a_1 & b_1 & c_1 & d_2 \end{pmatrix}$$

$$a_1 \quad b_1 \quad c_1 \quad d_1$$

$$a_1 \quad b_1 \quad c_1 \quad d_2$$

$$a_1 \quad b_1 \quad c_1 \quad d_2$$

2.5. Phép chiếu (Projection)

Phép chiếu trên một quan hệ là phép toán loại bỏ đi một số thuộc tính và chỉ giữ lại một số thuộc tính còn lại của quan hệ đó. Giả sử r là một quan hệ n - ngôi (gồm n thuộc tính) ta có:

Gọi X là tập con của thuộc tính $R = \{A_1, A_2, \dots, A_n\}$. Phép chiếu trên tập X của quan hệ r , kí hiệu là $\Pi_X(r)$ được định nghĩa như sau:

với $t[X]$ là giá trị của bộ t tại thuộc tính X .

Ví dụ 2.5:

$$R = \{A, B, C, D\}; \quad X = \{A, B\}; \quad Y = \{A, C\}$$

$r(A \quad B \quad C \quad D)$	$\Pi_X(r) = p(A \quad B)$	$\Pi_Y(r) = p(A \quad C)$
$a_1 \quad b_1 \quad c_1 \quad d_1$	$a_1 \quad b_1$	$a_1 \quad c_1$
$a_1 \quad b_1 \quad c_1 \quad d_2$	$a_2 \quad b_2$	$a_2 \quad c_2$
$a_2 \quad b_2 \quad c_2 \quad d_2$		$a_2 \quad c_3$
$a_2 \quad b_2 \quad c_3 \quad d_3$		

So sánh: $s = \sigma_{A=a_1}(r)$; $p = \Pi_X(r)$ và r :

- $\text{Bậc } p = \text{bậc của } r$, $\text{lực lượng của } p \leq \text{lượng của } r$
- s không khả hợp với r , $\text{bậc của } s \leq \text{bậc của } r$; $\text{lực lượng của } s \leq \text{lực lượng của } r$.

2.6. Tích Đề Các (Times)

Gọi r là quan hệ xác định trên tập thuộc tính (A_1, A_2, \dots, A_n) và s là quan hệ xác định trên tập thuộc tính (B_1, B_2, \dots, B_m) . Tích Đề Các $r \times s$ của r và s là tập $(n+m)$ bộ với n thành phần đầu có dạng một bộ thuộc r và m thành phần sau đó có dạng một bộ thuộc s .

Biểu diễn hình thức của quan hệ $r(U)$ và quan hệ $s(V)$ với $U = (A_1, A_2, \dots, A_n)$ và $V = (B_1, B_2, \dots, B_m)$ có dạng:

$$\text{với } u = (a_1, a_2, \dots, a_n) \text{ và } s = (b_1, b_2, \dots, b_m)$$

Ví dụ 2.6: $r(A \quad B) \quad s(C \quad D) \quad r \times s = p(A \quad B \quad C \quad D)$

$a_1 \quad 1 \quad 1 \quad d_1$	$a_1 \quad 1 \quad 1 \quad d_1$
$a_2 \quad 2 \quad 2 \quad d_2$	$a_1 \quad 1 \quad 2 \quad d_2$

a_2	2	1	d_1
a_2	2	2	d_2

2.7. Phép kết nối (Join)

Biểu thức kết nối được định nghĩa là tổ hợp logic các toán hạng, trong đó mỗi toán hạng là một phép so sánh giữa một thuộc tính của quan hệ r

Phép toán kết nối của hai quan hệ r(U) và s(V) được biểu diễn như sau:

Trong đó F là điều kiện kết nối, $F = (A \theta B)$, với θ là một trong các phép so sánh $\{<, =, >, <=, >=, \neq\}$. Dĩ nhiên, ở đây cần giả thiết rằng mỗi giá trị của thuộc tính A ($A \in U$) đều có thể so sánh được (qua phép θ) với mỗi giá trị của thuộc tính B ($B \in V$)

Ví dụ 2.7: phép kết nối giữa quan hệ r và s với điều kiện kết nối $F = (B \geq C)$ như sau:

$r(A \ B)$	$s(C \ D)$	$r \triangleright_{B \geq C} s = p'(A \ B \ C \ D)$
$a_1 \ 1$	$1 \ d_1$	$a_1 \ 1 \ 1 \ d_1$
$a_2 \ 2$	$2 \ d_2$	$a_2 \ 2 \ 1 \ d_1$
		$a_2 \ 2 \ 2 \ d_2$

Nhận xét:

Tập kết quả của phép kết nối p' là tập con của tập kết quả phép tích Đề Các p:

$$p' \subseteq p$$

$$\Leftrightarrow p' = \sigma_{B \geq C}(p)$$

$$\Leftrightarrow r \triangleright_{B \geq C} s = \sigma_{B \geq C}(r \times s)$$

Do đó có thể định nghĩa phép chiếu thông qua phép chọn và phép tích Đề Các như sau:

$$r \triangleright_{B \geq C} s = \sigma_{B \geq C}(r \times s)$$

Nếu điều kiện kết nối $F = (A \theta B)$ với $A \in U$ và $B \in V$ trong trường hợp phép so sánh θ là "=" thì phép kết nối được gọi là "kết nối bằng".

Ví dụ: với quan hệ r và s trong ví dụ trên ta có:

Bài giảng Cơ sở dữ liệu quan hệ

$$\begin{array}{ccc}
 r(A \ B) & s(C \ D) & r \triangleright \langle s = p''(A \ B \ C \ D) \\
 a_1 \ 1 & 1 \ d_1 & \begin{array}{ccc} B=C & a_1 & 1 \ 1 \ d_1 \\ a_2 & 2 & 2 \ d_2 \end{array}
 \end{array}$$

Trường hợp kết nối bằng tại thuộc tính cùng tên của hai quan hệ và một trong hai thuộc tính đó được loại bỏ qua phép chiếu, thì phép kết nối được gọi là "kết nối tự nhiên" và sử dụng kí hiệu "*". Khi đó phép kết nối tự nhiên của hai quan hệ $r(AB)$ và $s(CD)$ biểu diễn như sau:

$$r(AB)*s(BD) = \{t[ABD] \mid t[AB] \in r \text{ và } t[BD] \in s\}$$

Ví dụ:

$$\begin{array}{ccc}
 r(A \ B) & s'(B \ D) & r \triangleright \langle s = p'''(A \ B \ B \ D) \\
 a_1 \ 1 & 1 \ d_1 & \begin{array}{ccc} B=B & a_1 & 1 \ 1 \ d_1 \\ a_2 & 2 & 2 \ d_2 \end{array}
 \end{array}$$

Hay kí hiệu $r * s' = p'''(A \ B \ D)$

$$\begin{array}{ccc}
 a_1 & 1 & d_1 \\
 a_2 & 2 & d_2
 \end{array}$$

Nhận xét: Phép kết nối tự nhiên có thể được biểu diễn thông qua phép chiếu và kết nối thông thường. Xét ví dụ dưới đây:

Cho hai quan hệ $r(A \ B \ C)$ và $s(B \ C \ D)$

Kết nối tự nhiên $r * s = p(A \ B \ C \ D)$

Hay có thể biểu diễn $r * s = \Pi_{ABCD} (r \triangleright \langle s)$
 $B=B, C=C$

Có nghĩa là tập kết quả của phép kết nối tự nhiên thu được sau khi thực hiện phép chiếu trên các thuộc tính khác nhau đối với tập kết quả của phép kết nối tại các thuộc tính cùng tên.

2.8. Phép chia (Divide)

Gọi r là quan hệ n ngôi và s là quan hệ m ngôi (với $n > m$, và $s \neq \emptyset$). Phép chia $r \div s$ là tập của tất cả $(n - m)$ bộ t sao cho với mọi bộ $u \in s$ thì bộ $t \cup u \in r$.

Biểu diễn hình thức như sau:

$$r \div s = \{t \mid \forall u \in s \Rightarrow t \cup u \in r.\} \quad (2.8)$$

Bài giảng Cơ sở dữ liệu quan hệ

Ví dụ 2.8:

r(A B C D)	S (C D)	r ÷ s = (A B)
a b c d	c d	a b
a b e f	e f	e d
a c e f		
e d c d		
e d e f		

3. Các ví dụ về tìm kiếm bằng đại số quan hệ

Ví dụ có ba quan hệ:

S (S#, SNAME, STATUS, CITY) : các hãng cung ứng,

P (P#, PNAME, COLOR, WEIGHT, CITY) : các mặt hàng

SP(S#, P#, QTY) : các mặt hàng đã cung cấp

- Tìm số hiệu của những hãng đã cung cấp mặt hàng P2

$$\Pi_{S\#}(\sigma_{P\# = 'P2'}(SP))$$

- Tìm số hiệu của những hãng đã cung ứng ít nhất là một mặt hàng màu đỏ

$$\Pi_{S\#}(\sigma_{COLOR = 'RED'}(P*SP))$$

hoặc $\Pi_{S\#}((\sigma_{COLOR = 'RED'}(P))*SP)$

4. Kết luận chung và bài tập:

4.1. Kết luận

Nói chung các phép tính của đại số quan hệ là khá đơn giản, mạnh và là một đại số có tính đầy đủ, không cần thủ tục. Tuy nhiên đây là ngôn ngữ khá gần ngôn ngữ lập trình thủ tục, dễ dàng mở rộng và chủ yếu là làm cơ sở cho việc thiết lập các ngôn ngữ con dữ liệu bậc cao hơn. Trong bài sau sẽ trình bày ngôn ngữ con dữ liệu SQL, đây ngôn ngữ con dữ liệu quan hệ được xác nhận là rất mạnh, phổ dụng và lại dễ sử dụng.

4.2. Tài liệu tham khảo

- Lê Tiến Vương, *Nhập môn Cơ sở dữ liệu quan hệ*, NXB Giáo dục, Chương 1
- P. O'Neil, *Database - Principles, Programming, Performance*, Chương 3

4.3. Bài tập

Bài 1 (trang 175), Nhập môn Cơ sở dữ liệu quan hệ, Lê Tiến Vương, NXB Thống Kê.

Chương 4. NGÔN NGỮ VẤN TIN CÓ CẤU TRÚC SQL (STRUCTURE QUERY LANGUAGE)

Hệ QTCSDL cung cấp cho người sử dụng công cụ để thực hiện các yêu cầu thông qua ngôn ngữ giao diện. Trong chương này sẽ trình bày ngôn ngữ dữ liệu SQL (Structure Query Language). Đây là ngôn ngữ con dữ liệu quan hệ được xác nhận là rất mạnh, thông dụng, và dễ sử dụng.

Bộ 1. Ngôn ngữ định nghĩa dữ liệu (2 tiết)

1. Giới thiệu về ngôn ngữ SQL

Ngôn ngữ SQL được phát triển từ ngôn ngữ SEQUEL-2, thử nghiệm và cài đặt tại Trung tâm nghiên cứu của hãng IBM ở San Jose, California cho hệ thống QTCSDL lớn điển hình là System -R. Trong System -R, SQL vừa đóng vai trò là một ngôn ngữ có thể thao tác độc lập của người sử dụng đầu cuối, đồng thời lại có khả năng là một ngôn ngữ con được nhúng trong ngôn ngữ chủ PL/1.

Khác với ngôn ngữ đại số quan hệ, là ngôn ngữ dựa trên các phép toán của đại số quan hệ. SQL là một ngôn ngữ phi thủ tục, chuẩn mực và điển hình. DO vậy hiện nay rất nhiều sản phẩm phần mềm thương mại đều được cài đặt SQL như Access, SQL Server, DB2, Oracle,...

Phép toán cơ bản trong SQL là phép ánh xạ được miêu tả như một khối SELECT - FROM - WHERE. Trong phần sau, các mệnh đề của ngôn ngữ SQL sẽ được trình bày chi tiết bằng các ví dụ.

Các thuật ngữ trong CSDL quan hệ như quan hệ, thuộc tính, bộ,... sẽ được thay bằng các thuật ngữ như bảng (table), cột (column), bản ghi (record) hoặc hàng (row).

2. Các mệnh đề của ngôn ngữ SQL

2.1. Tạo bảng

Mệnh đề tạo bảng tạo một bảng quan hệ mới thông qua tên bảng, xác định các thuộc tính và các ràng buộc logic của bảng.

Mệnh đề tạo bảng có dạng tổng quát như sau:

```
CREATE TABLE <tên bảng> (<tên cột> <kiểu dữ liệu> [NOT  
NULL] ... )
```

Bài giảng Cơ sở dữ liệu quan hệ

Trong đó:

- tên bảng: là xâu kí tự bất kì không có ký hiệu trống và không trùng với các từ khoá.
- tên cột: là xâu kí tự bất kì không chứa ký hiệu trống, trong một bảng tên cột là duy nhất. Thứ tự của cột trong bảng là không quan trọng.
- kiểu dữ liệu: trong mệnh đề tạo bảng dùng một số kiểu dữ liệu sau:
 - integer: kiểu số nguyên, từ - 2.147.483.648 đến 2.147.483.647
 - smallinteger: số nguyên nhỏ từ -32.768 đến 32.767
 - decimal (n,p): số thập phân với độ dài tối đa là n kể cả p chữ số phần thập phân (không tính dấu chấm thập phân).
 - char(n): xâu kí tự có độ dài cố định n. Một số HQTCSDL cho phép kích thước tối đa của char là 2000 bytes. Nếu mỗi kí tự tương ứng một byte (mã ASCII như đối với hệ QTCSDL Access) thì số kí tự tối đa là 2000. Trong trường hợp sử dụng bộ mã UCS2 (như đối với hệ QTCSDL SQL Server) thì số kí tự biểu diễn tối đa là 1000 kí tự (sử dụng 2 byte để biểu diễn 1 kí tự).
 - varchar(n): xâu kí tự có độ dài biến đổi, độ dài xâu có thể từ 0 đến n và được xác định tại thời điểm đưa dữ liệu vào lưu trữ. Một số HQTCSDL cho phép kích thước tối đa của char là 4000 bytes.
 - date: dữ liệu dạng ngày tháng, định dạng ngầm định: DD - MON - YY.
 - Ngoài ra mỗi hệ QTCSDL cũng đưa ra các kiểu dữ liệu riêng.
- NULL là giá trị ngầm định khi không biết chính xác giá trị. Do đó cột đóng vai trò khoá chính không được nhận giá trị NULL, các cột khác có thể tùy chọn.

Ví dụ 2.1: Cho CSDL gồm 3 bảng S (Supplier), P (Part) và SP như sau:

S (S#, SNAME, STATUS, CITY) : các hãng cung ứng, với S# là khoá chính

P (P#, PNAME, COLOR, WEIGHT, CITY) : các mặt hàng, P# là khoá chính

SP(S#, P#, QTY,SDATE) : các mặt hàng đã cung cấp.

Tạo bảng S:

Hình 9 - Cấu trúc lệnh tạo bảng nhà cung cấp

Tạo bảng P:

Bài giảng Cơ sở dữ liệu quan hệ

Tạo bảng SP:

Nhận xét:

Trong mệnh đề tạo bảng có thể sử dụng thêm các yếu tố ràng buộc để hạn chế các giá trị cho một hay nhiều cột trong bảng, như: ràng buộc khoá chính, khoá ngoài, ràng buộc toàn vẹn tham chiếu, ràng buộc miền giá trị tương ứng như sau:

- NULL: cột có thể không chứa giá trị.
- NOT NULL: cột phải chứa một giá trị nào đó
- PRIMARY KEY: ràng buộc khoá chính
- REFERENCE: ràng buộc khoá ngoài
- ON DELETE CASCADE: ràng buộc toàn vẹn tham chiếu

2.2. Xoá bảng

Mệnh đề xoá bảng xoá một bảng quan hệ (cả cấu trúc và nội dung của bảng) khỏi CSDL. Bảng này chỉ có thể được tạo lại bằng cách sử dụng mệnh đề CREATE TABLE.

Mệnh đề xoá bảng có dạng tổng quát như sau:

Ví dụ 2.2: Xoá bảng SP trong ví dụ trên: DROP TABLE SP

2.3. Thêm dữ liệu

Đây là mệnh đề thêm một bản ghi vào một bảng và thường được sử dụng để tổ chức vào dữ liệu. Dạng tổng quát như sau:

Có 3 cách biểu diễn mệnh đề thêm dữ liệu:

Cách 1: INSERT INTO S(S#, SNAME, STATUS, CITY)
VALUES (1,'VINH',30,'HA NOI')

Cách 2: Nếu vị trí của các cột trong bảng là cố định mệnh đề có thể viết:
INSERT INTO S
VALUES (1,'VINH',30,'HA NOI')

Cách 3: Nếu người vào dữ liệu quên vị trí của các cột, khi đó có thể biểu diễn như sau:

```
INSERT INTO S(S#, STATUS, CITY, SNAME)
VALUES (1, 30,'HA NOI', 'VINH')
```

Bụi 2. Ngôn ngữ thao tác dữ liệu (3 tiết)

1. Khối SELECT

Cấu trúc đơn giản nhất trong SQL là khối SELECT được miêu tả về cú pháp như một khối **select - from - where**.

Một cách tổng quát khối select bao gồm 3 mệnh đề chính:

- Select: Xác định nội dung của các cột cần đưa ra kết quả
- From: Xác định các bảng cần lấy thông tin ra.
- Where: Xác định các bản ghi thoả mãn yêu cầu chọn lọc để đưa ra kết quả.

Ngoài ra, để mở rộng khả năng của ngôn ngữ, khối SELECT còn được bổ sung thêm các mệnh đề *group by*, *having*, *order by*, các hàm mẫu,... Trong các phần sau sẽ trình bày chi tiết từng mệnh đề.

Dạng tổng quát của khối select được biểu diễn như sau:

Trong đó mệnh đề WHERE biểu diễn dưới một số dạng sau:

- WHERE [NOT] <biểu thức> phép_so_sánh <biểu thức>
- WHERE [NOT] <tên cột> [NOT] LIKE <xâu kí tự>
- WHERE [NOT] <biểu thức> [NOT] BETWEEN <biểu thức> AND <biểu thức>
- WHERE [NOT] <biểu thức > [NOT] IN ({danh sách / mệnh đề con})
- WHERE [NOT] <tên cột> phép_kết_nối <tên cột>
- WHERE [NOT] <biểu thức logic>
- WHERE [NOT] <biểu thức> {AND | OR} [NOT] <biểu thức>

2. Các mệnh đề tìm kiếm

2.1. Tìm kiếm theo câu hỏi đơn giản

a) Tìm kiếm không điều kiện

Trước hết, để đơn giản ta làm quen với các câu hỏi chỉ liên quan tới một bảng. Trong mệnh đề select có danh sách cột, danh sách này xác định tên các cột cần có trong bảng kết quả. Nếu sử dụng giá trị '*' có nghĩa là chọn toàn bộ các cột của bảng.

Bài giảng Cơ sở dữ liệu quan hệ

Ví dụ 1: Cho biết mã số các nhà cung cấp ứng với một mặt hàng nào đó

```
SELECT S#                SP (S#, P#, QTY, SDATE)
FROM SP
```

Kết quả

(trong đại số quan hệ: $\Pi_{S\#}(SP)$)

Chú ý: Sau khi thực hiện một lệnh SQL, để bảng kết quả đúng là một quan hệ (có nghĩa là không có bộ trùng nhau), trong mệnh đề select cần thêm từ khoá DISTINCT.

Ví dụ 2: như trong ví dụ trên

```
SELECT DISTINCT S#
FROM SP
```

Tập kết quả chỉ gồm S1 và S2

b) Tìm kiếm với điều kiện đơn giản

Tìm mã số những nhà cung cấp đã cung cấp mặt hàng P2:

```
SELECT DISTINCT S#
FROM SP
WHERE P# = 'P2'
```

Trong SQL các phép so sánh được sử dụng bao gồm >, <, >=, <=, = và <>. Các phép trên dùng cho mọi loại dữ liệu.

c) Tìm kiếm có xử lý xâu kí tự

Xử lý xâu kí tự gần đúng được dùng trong trường hợp người sử dụng không nhớ rõ tên người hoặc địa danh,... ví dụ là Hoa hay Hoan khi đó có thể viết:

```
SELECT *
FROM S
WHERE SNAME LIKE 'HOA%'
```

Trong SQL sử dụng kí hiệu '%' là thay thế cho một xâu con, dấu phân cách '_' để thay thế cho một kí tự.

Ví dụ 3:

- A%B: xâu ký tự bắt đầu bằng chữ A và kết thúc bằng chữ B
- %A: xâu ký tự bất kì có kết thúc là A

Bài giảng Cơ sở dữ liệu quan hệ

- A_B: xâu gồm ba ký tự có ký tự thứ hai là bất kì.

d) Xử lý ngày tháng

Ngoài các phép tính thông thường, SQL còn có thể xử lý dữ liệu dạng ngày tháng.

Ví dụ 4: Tìm những mặt hàng bán trước ngày 11-10-02 là 10 ngày

```
SELECT P#  
FROM SP  
WHERE '11-10-02' - sdate = 10
```

e) Tìm kiếm nhờ sử dụng IN và BETWEEN

- Tìm những mặt hàng cung cấp có số lượng từ 100 đến 200

```
SELECT P#  
FROM SP  
WHERE QTY BETWEEN 100 AND 200
```

- Tìm mã số những nhà cung cấp đã cung cấp ít nhất một trong các mặt hàng P1, P2, P3

```
SELECT S#  
FROM SP  
WHERE P# IN ('P1', 'P2', 'P3')
```

2.2. Các hàm thư viện

Cũng giống như các ngôn ngữ CSDL khác, trong SQL có các hàm mẫu gồm *count*, *max*, *min*, *sum*, *avg*. Riêng hàm *count* khi có đối số là '*' có nghĩa là đếm số bản ghi thoả mãn yêu cầu tìm kiếm mà không cần quan tâm tới bất kì một cột nào.

Ví dụ 2.1: Cho biết số lần mặt hàng P2 đã được cung cấp

```
SELECT COUNT(*)  
FROM SP  
WHERE P# = 'P2'
```

Tìm số mặt hàng P1 bán một lần nhiều nhất

```
SELECT MAX (QTY)  
FROM SP  
WHERE P# = 'P1'
```

2.3. Tìm kiếm nhờ mệnh đề GROUP BY

Tìm mã số những mặt hàng mà mỗi nhà cung cấp đã cung cấp cho khách hàng

```
SELECT S#,P#
```

```
FROM SP
```

```
SP (S#, P#, QTY, SDATE)
```

```
GROUP BY S#
```

Group S1 →

Group S2 →

SS	PP	-	-
1	1		
SS	PP	-	-
1	2		
SS	PP	-	-
1	3		
SS	PP	-	-
2	2		

Trong mệnh đề này bảng dữ liệu SP được lấy ra, sau đó phân thành nhóm

theo mã số của người cung cấp (S#). Có nghĩa là các bộ có cùng giá trị S# được sắp xếp liên tiếp nhau, hết nhóm này đến nhóm khác.

a) Tìm kiếm có sử dụng mệnh đề HAVING

Mệnh đề Having thường được sử dụng cùng mệnh đề Group by. Sau Having là biểu thức điều kiện. Biểu thức điều kiện này không tác động vào toàn bảng được chỉ ra ở mệnh đề From mà chỉ tác động lần lượt từng nhóm các bản ghi đã chỉ ra tại mệnh đề Group by.

Ví dụ 2.3: Tìm mã số những nhà cung cấp đã cung cấp ít nhất hơn hai mặt hàng.

```
SELECT S#  
FROM SP  
GROUP BY S#  
HAVING COUNT (DISTINCT P#) > 2
```

b) Tìm kiếm có sắp xếp

Tìm tên các mặt hàng màu đỏ và sắp xếp theo thứ tự giảm dần của mã số mặt hàng.

```
SELECT PNAME, P#  
FROM P  
WHERE COLOUR = 'ĐỎ'  
ORDER BY P# ASC
```

Bài giảng Cơ sở dữ liệu quan hệ

Sau mệnh đề Order by là tên cột cần sắp xếp rồi đến chiều sắp xếp tăng hoặc giảm (ASC hoặc DESC). Có thể sắp xếp nhiều cột và nếu không chỉ ra chiều sắp xếp thì hệ thống ngầm định là chiều tăng (ASC).

2.4. Tìm kiếm với câu hỏi phức tạp

Trong phần này trình bày việc tìm kiếm với nhiều bảng qua việc sử dụng ánh xạ lồng nhau hoặc qua phép kết nối.

a) Phép kết nối

Trong phép kết nối, các cột tham gia kết nối phải có miền trị là sánh (so sánh) được với nhau (ví dụ: cùng là kiểu integer, char, hay giữa kiểu char và varchar,...). Tên cột của các bảng khác nhau có thể được viết tường minh qua tên bảng.

Trong ví dụ quan hệ người cung cấp - mặt hàng, bảng S chứa cột số hiệu mặt hàng S#, bảng SP cũng chứa số hiệu mặt hàng S#. Để phân biệt, có thể viết tường minh tên cột S# qua tên bảng là S.S# và SP.S#.

Ví dụ 2.4: Với mỗi mặt hàng đã được cung cấp, cho biết mã số của mặt hàng và địa chỉ của hãng đã cung cấp mặt hàng đó.

Ta nhận thấy, trong bảng SP (bảng mặt hàng đã được cung cấp) không chứa địa chỉ của hãng đã cung cấp. Do đó cần kết nối với bảng S để thu được địa chỉ của hãng cung cấp thông qua phép kết nối.

```
SELECT P#, CITY
FROM SP,S
WHERE SP S# = S S#
```

PK				FK			
SS #	SNAME ME	STATU SUS	CTY TY	SS #	PP#	QTYTY	SDATE TE
SS 1	OMOM O	Goodod	HNH N	SS 1	PP1	10	-
SS 2	DASOS O	Goodod	HCM CM	SS 1	PP2	20	-
SS 3	-	kết nối (quan hệ 1 n)	-	SS 2	PP2	50	-
SS 4	-	-	-	SS 2	PP3	-	-

Hình 10 – Ví dụ bảng quan hệ S- SP

Chú ý: Trong các phép tìm kiếm có nhiều hơn một bảng, nếu tên các cột là không duy nhất thì bắt buộc phải viết tên cột dạng tường minh.

Bài giảng Cơ sở dữ liệu quan hệ

b) ánh xạ lồng

Tìm tên những hãng đã cung cấp mặt hàng P2.

Phép lồng nhau có thể được lồng nhiều mức hoặc sử dụng ánh xạ lồng với sự dẫn trở giữa các khối mỗi khi hướng tới một bảng khác nhau.

Ví dụ: Tìm số hiệu và tên các hãng không cung ứng mặt hàng P1.

c) Tìm kiếm có sử dụng lượng từ ANY và ALL

Tìm tên những mặt hàng có mã số mặt hàng là mặt hàng nào đó mà hãng S1 đã bán.

Tìm mã số những hãng cung cấp số lượng một lần một mặt hàng nào đó lớn hơn hoặc bằng số lượng mỗi lần cung cấp của các hãng

Mệnh đề trên hoàn toàn tương đương với:

d) Tìm kiếm có chứa phép tính tập hợp

Tìm mã số những hãng hiện thời chưa cung cấp một mặt hàng nào cả

Tìm tên các hãng cung cấp tất cả các mặt hàng.

3. Các mệnh đề cập nhật dữ liệu

Trong các phần trước đã trình bày các mệnh đề để tạo bảng, vào dữ liệu. Phần này sẽ trình bày chi tiết các mệnh đề cập nhật dữ liệu trong SQL

3.1. Thêm một bộ dữ liệu

Mệnh đề thêm một bộ dữ liệu có dạng tổng quát như sau:

```
INSERT INTO <tên bảng[(danh sách tên cột)]>  
VALUES (bộ giá trị)  
[câu hỏi con]
```

Ví dụ 3.1: Có thể bổ sung một tập các bản ghi là kết quả xử lý của một câu hỏi nào đó, chẳng hạn:

Nếu như bảng W và bảng S có cùng lược đồ.

W#	WNAME	STATUS	CTY
w1	Hai Ha	Good	HN
W2	Halida	Normal	HN

SS	WNAME	STATU	CTY
#	ME	SUS	TY
SS	OMOMO	Goodod	HNH
1			N
SS	DASOSO	Goodod	HCM
2			CM
W	Haiai	Goodod	HNH
W1	HaHa		N
W	Kinhnh	Goodod	HCM
W3	Đô§«		CM

Bài giảng Cơ sở dữ liệu quan hệ

W3	Kinh Đô	Good	HCM
W4	-	-	-

Hình 11 – Ví dụ thêm bộ dữ liệu

3.2. Xoá bản ghi

Mệnh đề xoá bản ghi có thể được thực hiện cho một hoặc nhiều bản ghi thoả mãn một điều kiện nào đó. Dạng tổng quát là:

```
DELETE  
[FROM] <tên bảng>  
[WHERE <biểu thức điều kiện>]
```

Ví dụ 3.2: Loại bỏ hãng S1 khỏi bảng S

```
DELETE  
FROM S  
WHERE S# = 'S1'
```

Loại bỏ các mặt hàng đã được cung cấp sau ngày 20 - 5 -2000

```
DELETE FROM SP WHERE SDATE > '20 - 5 -2000'
```

Loại bỏ những hãng chưa cung cấp mặt hàng nào cả

```
DELETE FROM S WHERE S# NOT IN  
( SELECT S#  
FROM SP)
```

3.3. Sửa đổi dữ liệu

Mệnh đề sửa đổi các giá trị của các bản ghi trong bảng của CSDL theo một điều kiện nào đó có dạng tổng quát như sau:

Ví dụ 3.3: Đổi màu các mặt hàng P2 thành màu vàng.

```
UPDATE P SET COLOUR = 'Vàng' WHERE P# = 'P2'
```

```
ALTER TABLE <tên bảng>  
ADD <tên cột mới> <kiểu dữ liệu>
```

3.4. Thêm cột mới

Ví dụ 3.4: Thêm cột PRICE (giá) cho bảng SP với kiểu số liệu dạng số thập phân.

```
ALTER TABLE SP  
ADD PRICE DECIMAL(8,2)
```

4. Kết luận và bài tập

4.1. Kết luận

Chương này trình bày về ngôn ngữ dữ liệu SQL. Đây là ngôn ngữ rất mạnh, phổ dụng và dễ sử dụng. Hiện nay, các hệ QTCSDL thông dụng như: Access, SQL Server, Oracle, DB2,... đều hỗ trợ tốt ngôn ngữ này. Do đó để có thể thao tác tốt trên các hệ QTCSDL này, cần hiểu và nắm vững ngôn ngữ SQL.

4.2. Bài tập

Bài 2, 3, 4 (trang 175, 176),
Sách Nhập môn Cơ sở dữ liệu quan hệ, Lê Tiến Vương, NXB Thống Kê.

Tổng quan về Cơ sở dữ liệu quan hệ

Lời mở đầu

Thuật ngữ Cơ sở Dữ liệu (Database) không mấy xa lạ với những người làm tin học. Đây là một trong những lĩnh vực được tập trung nghiên cứu và phát triển của Công nghệ thông tin, nhằm giải quyết các bài toán quản lí, tìm kiếm thông tin trên các hệ thống lớn, phức tạp, nhiều người sử dụng.

Từ những năm 70, mô hình dữ liệu quan hệ do Codd đưa ra với cấu trúc hoàn chỉnh đã tạo cơ sở toán học cho các vấn đề nghiên cứu dữ liệu. Với cấu trúc đơn giản và khả năng hình thức hoá phong phú, CSDL quan hệ dễ dàng mô phỏng các hệ thống thông tin đa dạng trong thực tế. Lưu trữ thông tin tiết kiệm, có tính độc lập dữ liệu cao, dễ sửa đổi, bổ sung cũng như khai thác dữ liệu là những ưu điểm nổi bật của CSDL quan hệ.

Sau đây chúng ta sẽ đề cập tới những khái niệm cơ bản của CSDL quan hệ.

I - Khái quát về CSDL & CSDL quan hệ

I.1. Khái niệm CSDL

1. Định nghĩa

Một cơ sở dữ liệu là một tập hợp dữ liệu về một xí nghiệp được lưu giữ trên máy tính, được người sử dụng, có cách quản lí bằng một mô hình.

VD: Quản lí thi tuyển sinh thì CSDL bao gồm:

- + thí sinh (tên, ngày sinh, địa chỉ, số báo danh ...).
- + phách (số báo danh, số phách).
- + điểm (số phách, điểm).

2. Các tiêu chuẩn của một CSDL

Một CSDL cần :

- + phản ánh tốt xí nghiệp cần quản lí.
- + không dư thừa thông tin: mỗi thông tin chỉ nên có mặt một lần trong hệ thống thông tin để tiết kiệm lưu trữ, đảm bảo truy cập duy nhất.
- + độc lập giữa CSDL và chương trình: sự sửa đổi chương trình không làm kéo theo việc sửa đổi CSDL.
- + tính an toàn: không bị hỏng khi có nhiều người sử dụng hoặc có các sự cố.
- + hiệu suất sử dụng tốt: dù nhiều người sử dụng một lúc, CSDL vẫn đảm bảo hiệu suất như chỉ có một người sử dụng.

3. Hệ QTCSDL

Là tập hợp có thứ tự các phần mềm cho phép mô tả lưu giữ thao tác các dữ liệu trên một CSDL, đảm bảo tính an toàn, bí mật trong môi trường có nhiều người sử dụng.

4. Hệ thống thông tin

Là tập hợp các thông tin được lưu giữ, và một tập hợp các xử lý cho phép xây dựng lại một hình ảnh trung thành về một xí nghiệp.

5. Kiến trúc một CSDL

Gồm 3 thành phần cơ bản

+ **Thực thể**: là đối tượng có trong thực tế mà chúng ta cần mô tả các đặc trưng của nó, đối tượng có thể là cụ thể hoặc trừu tượng .

+ **Thuộc tính**: là các dữ liệu thể hiện các đặc trưng của thực thể.

+ **Ràng buộc**: là các mối quan hệ Logic của thực thể.

I.2.Các mô hình CSDL

6. Mô hình phân cấp

Mô hình dữ liệu là một cây, trong đó các nút biểu diễn các tập thực thể, giữa các nút con và nút cha được liên hệ theo mối quan hệ xác định chủ-thành viên (1-n).

7. Mô hình mạng

Mô hình được biểu diễn là một đồ thị có hướng .

8. Mô hình quan hệ

Mô hình dựa trên cơ sở khái niệm lý thuyết tập hợp của các quan hệ .

I.3.Các khái niệm cơ bản về CSDL quan hệ

9. Thuộc tính

+ là một lô thông tin nhỏ nhất được sử dụng một cách tự do và có ý nghĩa, độc lập với các lô khác .

+ trong một mô hình , thuộc tính là một định vị cơ sở thông tin, một thuộc tính được định nghĩa bằng tên và miền giá trị của nó .

10. Quan hệ

+Quan hệ được định nghĩa là tập con của tích ĐỀ các $D_1 * D_2 * \dots * D_n$ trong đó D_i là miền giá trị của thuộc tính i .

Ta có thể xem quan hệ như một bảng gồm nhiều cột chứa tên các thuộc tính, mỗi cột chứa miền giá trị của thuộc tính, mỗi cột chứa miền giá trị của thuộc tính đó.

+ Mỗi hàng trong bảng đó là một bộ giá trị của quan hệ, là một hình ảnh của thực thể.

11. Phụ thuộc hàm

Xét tập quan hệ R_i và hai tập thuộc tính G_1, G_2 luôn có mặt cùng nhau trong các R_i . Ta nói có sự phụ thuộc của hàm giữa G_1, G_2 nếu với mọi giá trị của G_1 chỉ có thể kết hợp với một và chỉ một giá trị của G_2 tại một thời điểm cho trước. Tính chất đúng trong mọi R_i mà ở đó có mặt G_1, G_2 .

Kí hiệu $G_1 \rightarrow G_2$.

12. Khoá

Là một hoặc một tập các thuộc tính là nguồn của một phụ thuộc hàm có đích lần lượt là các thuộc tính khác của quan hệ.

VD: + Sinh viên(Số thẻ Sv, Họ tên, Ngày sinh, Quê quán)

Số thẻ Sv --> Họ tên

Số thẻ Sv --> Ngày sinh

Số thẻ Sv --> Quê quán

Như vậy Số thẻ SV là khoá của quan hệ Sinh viên

+ Giảng dạy (Số phòng học, Thời gian, Tên giảng viên, Tên môn học)

Số phòng học, Thời gian --> Tên giảng viên, Tên môn học

Như vậy (Số phòng học, Thời gian) là khoá của quan hệ Giảng dạy

I.4. Các phép toán trên các quan hệ

13. Phép chiếu

Xét các tập các thuộc tính C và R là tập các quan hệ định nghĩa trên C. Phép chiếu của R trên tập G thuộc C là sự thu hẹp của R đến các phần tử của G. Kí hiệu $\Pi_G(R)$.

14. Phép nối

Nối hai quan hệ có chứa các thuộc tính hoặc tập các thuộc tính như nhau, quan hệ thu được gồm có các hàng ở hai quan hệ ban đầu đặt nối nhau bằng các thuộc tính giống nhau. Phép nối hai quan hệ S và T kí hiệu là $S \bowtie T$

15. Phép tách

Một quan hệ R được gọi là phân rã thành hai quan hệ S, T nếu thoả mãn :

+ S, T là các phép chiếu của R

+ $R = S \bowtie T$

Chú ý rằng không phải bất kì quan hệ nào cũng tách được

16. Phép chọn : lấy ra các hàng có các thuộc tính thoả mãn một số tiêu chuẩn cho trước.

Ngoài ra còn có phép hợp, phép giao, phép lấy hiệu các quan hệ

II-Thiết kế hệ thống CSDL quan hệ

II.1. Phân tích tư liệu của XN

1. Xác định danh sách các thuộc tính

Từ việc khảo sát thực tế, từ các tin tức trao đổi với người dùng, phân tích các bài toán nghiệp vụ, ta được một danh sách các thuộc tính phục vụ cho xây dựng CSDL. Mỗi một thuộc tính bao gồm :

+ tên thuộc tính

+ kiểu thuộc tính (là tính toán hay không tính toán)

+ miền giá trị

Bài giảng Cơ sở dữ liệu quan hệ

- + quy tắc tính toán (nếu thuộc tính là tính toán)
- + ds các tham số

VD : Xét quan hệ :

Bảng lương (Họ tên, Đơn vị, Hệ số, Lương, Phụ cấp, Tổng lĩnh)
Ta thấy ở đây Lương, Tổng lĩnh là các thuộc tính tính toán với các qui tắc
 $Lương = Hệ số * 14400$
 $Tổng lĩnh = Lương + Phụ cấp$

2.

3. Tìm phụ thuộc hàm giữa các thuộc tính

Sau khi có được một danh sách các thuộc tính từ việc thu thập thông tin và khảo sát thực tế, việc cần làm tiếp theo là xác định các phụ thuộc hàm giữa các thuộc tính. Chúng ta xét một ví dụ sau :

Quản lí thư viện (Số thẻ, Số sách, Tên sách, Loại sách, Ngày mượn, Tên độc giả, Địa chỉ độc giả)

Ta thấy có các phụ thuộc hàm sau :

Số thẻ mượn, Số sách --> Ngày mượn

Số sách --> Tên sách, Loại sách

Số thẻ mượn --> Tên độc giả, Địa chỉ độc giả

II.2. Chuẩn hoá các quan hệ

Coi danh sách các thuộc tính thu được sau các bước trên cùng các phụ thuộc hàm giữa chúng là một quan hệ, chúng ta thực hiện việc chuẩn hoá quan hệ này.

Mục đích của quá trình này là giảm bớt sự dư thừa thông tin, bảo đảm tính duy nhất của mỗi thông tin, do đó tiện lợi cho việc truy nhập và cập nhật cho CSDL. Quá trình chuẩn hoá có thể được tiến hành qua nhiều bước.

Ta xét ví dụ sau :

Xe máy (Số xe, Số máy, Loại xe, Ngày đăng kí, Tên chủ xe, Số điện thoại, Địa chỉ)

+ Nếu một người có nhiều xe thì lặp lại (Số điện thoại, Địa chỉ) của chủ xe, như vậy dư thừa thông tin.

+ Nếu một người chủ xe thay đổi địa chỉ thì phải sửa nhiều lần, như vậy cập nhật rời rạc

Như vậy biên pháp để khắc phục là tách ra quan hệ

Chủ xe (Tên, Số điện thoại, Địa chỉ)

4.

5. Đưa quan hệ về dạng chuẩn 1 NF

Quan hệ chưa ở dạng chuẩn 1NF là quan hệ còn chứa các nhóm lặp lại. Ta đưa về dạng 1NF bằng cách như sau:

+ bỏ nhóm lặp lại ra khỏi quan hệ, chuyển nhóm đó thành một quan hệ mới

+ cộng thêm vào khoá của nó khoá của quan hệ ban đầu để tạo ra khoá phức hợp

VD: xét quan hệ :

R (Số hoá đơn, Ngày bán, Số khách hàng, Tên khách hàng, Số sản phẩm, Tên sản phẩm, Lượng yêu cầu)

Nhóm (Số sản phẩm, Tên sản phẩm, Lượng yêu cầu) là nhóm lặp lại, ta có thể tách R thành R1 và R2 như sau :

+ R1 (Số hoá đơn, Ngày bán, Số khách hàng, Tên khách hàng, Số sản phẩm)

+ R2 (Số hoá đơn, Số sản phẩm, Tên sản phẩm, Lượng yêu cầu)

6.

7. Đưa quan hệ về dạng chuẩn 2NF

Quan hệ đã ở dạng 1NF nhưng chưa ở dạng 2NF là có tồn tại phụ thuộc hàm có nguồn là tập con của khoá. Ta đưa về dạng 2NF bằng cách như sau:

+ nhóm vào một quan hệ các thuộc tính phụ thuộc hoàn toàn vào khoá và giữ lại khoá của quan hệ đó

+ nhóm vào một quan hệ khác các thuộc tính phụ thuộc vào một phần của khoá, lấy phần đó làm khoá chính cho quan hệ.

VD : trong quan hệ R2 (Số hoá đơn, Số sản phẩm, Tên sản phẩm, Lượng yêu cầu) có phụ thuộc hàm : Số sản phẩm --> Tên sản phẩm

Trong đó Số sản phẩm là một phần của khoá, ta tách R2 thành R3 và R4 như sau :

+ R3 (Số hoá đơn, Số sản phẩm, Lượng yêu cầu)

+ R4 (Số sản phẩm, Tên sản phẩm)

8.

9. Dạng chuẩn 3NF

Quan hệ đã ở dạng 2NF nhưng chưa ở dạng 3NF là có tồn tại các phụ thuộc hàm gián tiếp. Đưa về dạng 3NF ta làm như sau :

+ giữ lại trong quan hệ ban đầu các thuộc tính phụ thuộc trực tiếp vào khoá

+ nhóm vào một quan hệ khác các thuộc tính bắc cầu, lấy thuộc tính bắc cầu làm khoá

VD : trong quan hệ

R1 (Số hoá đơn, Ngày bán, Số khách hàng, Tên khách hàng, Số sản phẩm)

Có các phụ thuộc hàm bắc cầu :

Số hoá đơn --> Số khách hàng --> Tên khách hàng

Ta có thể tách R1 thành R5 và R6 như sau :

+ R5 (Số hoá đơn, Ngày bán, Số khách hàng, Số sản phẩm)

+ R6 (Số khách hàng, Tên khách hàng)

10. Dạng chuẩn BCNF

Quan hệ đã ở dạng 3NF nhưng chưa ở dạng BCNF là có tồn tại phụ thuộc hàm có nguồn là thuộc tính không thuộc khoá nhưng có đích là thuộc tính thuộc khoá.

Ta xét ví dụ : R (Học sinh, Môn học, Giáo viên, Điểm)

Phụ thuộc hàm Giáo viên --> Môn học có nguồn không thuộc khoá nhưng đích thuộc khoá, ta tách R thành R1, R2 như sau :

+ R1 (Học sinh, Giáo viên, Điểm)

+ R2 (Giáo viên, Môn học)

Các phụ thuộc hàm đơn trị dừng lại ở dạng chuẩn BCNF (Boyce - Codd). Đến đây chúng ta có thể kết thúc công việc chuẩn hoá .

Lời kết :

Mô hình CSDL quan hệ là một công cụ rất tiện lợi để mô tả cấu trúc logic của các CSDL . Như vậy ở mức logic mô hình này bao gồm các quan hệ được biểu diễn bởi các bảng. Do đó đơn vị của CSDL quan hệ là bảng, trong đó các dòng của bảng là các bản ghi dữ liệu cụ thể, còn các cột là các thuộc tính .

Đối với người sử dụng có thể nói CSDL quan hệ là một tập hợp các bảng biến đổi theo thời gian .

Đối với công việc thiết kế một CSDL thì các công việc phân tích tư liệu của xí nghiệp và chuẩn hoá là hết sức quan trọng, phục vụ cho việc cài đặt thực tế .

Tài liệu tham khảo :

- Cơ sở dữ liệu kiến thức và thực hành
PGS Vũ Đức Thi - NXB thống kê 1997
- Nhập môn cơ sở dữ liệu quan hệ
Lê Tiến Vương - NXB Khoa học và kỹ thuật 1997
- Bài giảng môn Cơ sở dữ liệu I - Khoa CNTT - Đại học KHTN