

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.

Giáo trình được lưu hành nội bộ Trường Cao Đẳng Nghề Đà Lạt

LỜI GIỚI THIỆU

Giáo trình mạng máy tính dùng để giảng dạy cho sinh viên trình độ trung cấp, cao đẳng nghề Quản trị Mạng máy tính

Việc tổ chức biên soạn giáo trình Cơ sở dữ liệu nhằm phục vụ cho công tác đào tạo của trường Trường Cao đẳng Nghề Đà Lạt - Khoa Công nghệ thông tin – nghề Quản trị mạng máy tính. Giáo trình là sự cố gắng lớn của tập thể Khoa Công nghệ thông tin nhằm từng bước thống nhất nội dung dạy và học môn Cơ sở dữ liệu.

Nội dung của giáo trình đã được xây dựng trên cơ sở thừa kế những nội dung đã được giảng dạy ở các trường kết hợp với những nội dung mới nhằm đáp ứng yêu cầu nâng cao chất lượng phục vụ sự nghiệp công nghiệp hoá, hiện đại hoá. Giáo trình cũng là cẩm nang về Cơ sở dữ liệu riêng cho sinh viên của Trường Cao đẳng Nghề Đà Lạt - Khoa Công nghệ thông tin.

Giáo trình được biên soạn ngắn gọn, dễ hiểu, bổ sung nhiều kiến thức mới phù hợp với ngành nghề đào tạo mà Khoa Công nghệ thông tin đã tự điều chỉnh cho thích hợp và không trái với quy định của chương trình khung đào tạo của trường.

Xin chân trọng cảm ơn Khoa Công nghệ thông tin - Trường Cao đẳng Nghề Đà Lạt cũng như sự giúp đỡ quý báu của đồng nghiệp đã giúp tác giả hoàn thành giáo trình này.

Mặc dù đã rất cố gắng nhưng chắc chắn không tránh khỏi sai sót, tác giả rất mong nhận được ý kiến đóng góp của người đọc để lần xuất bản sau giáo trình được hoàn thiện hơn.

Lâm Đồng, ngày 5 tháng 9 năm 2019

Tham gia biên soạn

Chủ biên: Ngô Thiên Hoàng

Tailieu.vn

MỤC LỤC

Nội dung	Trang
CHƯƠNG 1: MÔ HÌNH QUAN HỆ	10
MÃ CHƯƠNG: MH10_01	10
1. NGUYÊN NHÂN RA ĐỜI CỦA MÔ HÌNH QUAN HỆ (RELATIONAL MODEL)	10
2. HỆ QUẢN TRỊ CƠ SỞ DỮ LIỆU (DATABASE MANAGEMENT SYSTEM)	11
2.1 CSDL LÀ GÌ?	11
2.2 HỆ QUẢN TRỊ CSDL.....	12
2.3 NGƯỜI DÙNG (USER)	13
2.4 CSDLQH VÀ HỆ TẬP TIN THEO LỐI CŨ	13
3. MÔ HÌNH QUAN HỆ (RELATIONAL MODEL)	15
3.1 MÔ HÌNH QUAN HỆ LÀ GÌ ?	15
3.2 CÁC KHÁI NIỆM CƠ BẢN CỦA MÔ HÌNH QUAN HỆ.....	16
3.2.1 Thuộc tính (Attribute, Arity).....	16
3.2.2 Lược Đồ Quan Hệ (Relation Schema).....	16
3.2.3 Bộ (Tuple)	18
3.2.4 Khóa (Key, Candidate Key):	20
3.3 CÁC PHÉP TOÁN TẬP HỢP (SET OPERATION).....	21
3.3.1 Phép hợp (Union operation)	21
3.3.2 Phép Giao (Intersection):.....	21
3.3.3 Phép Trừ (Minus, difference)	22
3.3.4 Tích Descartes (Cartesian Product, product).....	22
3.4 CÁC PHÉP TOÁN QUAN HỆ.....	23
3.4.1 Phép Chiếu (Projection)	23

3.4.2	<i>Phép Chọn (Selection)</i>	23
3.4.3	<i>Phép kết, Phép Kết Tự Nhiên (join, natural join):</i>	24
3.4.4	<i>Phép chia (division):</i>	26
3.4.5	<i>Các tính chất của đại số quan hệ</i>	26
4.	MÔ HÌNH THỰC THỂ KẾT HỢP	29
4.1	GIỚI THIỆU MÔ HÌNH THỰC THỂ KẾT HỢP.....	29
4.1.1	<i>Ví dụ – Mỗi quan hệ một-nhiều</i>	30
4.1.2	<i>Ví dụ – mỗi quan hệ một-một</i>	32
4.1.3	<i>Ví dụ – mỗi quan hệ nhiều-nhiều</i>	32
4.2	CHUYỂN TỪ MÔ HÌNH THỰC THỂ KẾT HỢP SANG LƯỢC ĐỒ CSDL.	34
4.2.1	<i>Qui tắc chung</i>	34
4.2.2	<i>Mỗi quan hệ một-một</i>	34
4.2.3	<i>Mỗi quan hệ một-nhiều</i>	35
4.2.4	<i>Mỗi quan hệ nhiều-nhiều đến tập kết hợp</i>	35
4.2.5	<i>Mỗi quan hệ nhiều-nhiều</i>	36
5.	BÀI TẬP	36
5.1	PHÉP TOÁN TẬP HỢP VÀ PHÉP TOÁN QUAN HỆ.....	36
5.2	MÔ HÌNH THỰC THỂ KẾT HỢP.....	38
	CHƯƠNG 2 : NGÔN NGỮ TRUY VẤN SQL	41
	MÃ CHƯƠNG: MH10_02	41
1.	CÁCH TẠO QUAN HỆ BẰNG SQL SERVER	41
1.1	CÁC BƯỚC TẠO QUAN HỆ BẰNG SQL SERVER.....	41
1.1.1	<i>Tạo cơ sở dữ liệu với đường dẫn mặc định</i>	41
1.1.2	<i>Tạo cơ sở dữ liệu với đường dẫn chỉ định</i>	41
1.2	TẠO QUAN HỆ VÀ NHẬP DỮ LIỆU.....	42
1.2.1	<i>Cú pháp tạo bảng</i>	42

1.2.2 Nhập dữ liệu	42
2. CÂU LỆNH TRUY VẤN.....	42
2.1 BIỂU THỨC (EXPRESSION).....	42
2.1.1 Hàm:.....	46
2.1.2 Kiểu dữ liệu :	48
2.2 CÂU LỆNH SQL	48
2.2.1 Truy vấn định nghĩa dữ liệu (ngôn ngữ mô tả dữ liệu) :	48
2.2.2 Ngôn ngữ truy vấn dữ liệu.....	52
3. BÀI TẬP.....	61
CHƯƠNG 3: RÀNG BUỘC TOÀN VỆ QUAN HỆ	67
(ENTEGRITY CONSTRAINT)	67
MÃ CHƯƠNG: MH10_03	67
1. RÀNG BUỘC TOÀN VỆ - CÁC YẾU TỐ CỦA RÀNG BUỘC TOÀN VỆ	67
1.1 RÀNG BUỘC TOÀN VỆ.....	67
1.2 CÁC YẾU TỐ CỦA RÀNG BUỘC TOÀN VỆ	68
1.2.1 Điều kiện.....	68
1.2.2 Bối cảnh.....	68
1.2.3 Tầm ảnh hưởng.....	68
2. PHÂN LOẠI RÀNG BUỘC TOÀN VỆ.....	69
2.1 RÀNG BUỘC TOÀN VỆ LIÊN BỘ.....	71
2.2 RÀNG BUỘC TOÀN VỆ VỀ PHỤ THUỘC TỒN TẠI:.....	71
2.3 RÀNG BUỘC TOÀN VỆ VỀ MIỀN GIÁ TRỊ	71
2.4 RÀNG BUỘC TOÀN VỆ LIÊN THUỘC TÍNH	72
2.5 RÀNG BUỘC TOÀN VỆ LIÊN THUỘC TÍNH LIÊN QUAN HỆ	72

2.6	RÀNG BUỘC TOÀN VỆN VỀ THUỘC TÍNH TỔNG HỢP	73
3.	BÀI TẬP	73
	CHƯƠNG 4: PHỤ THUỘC HÀM	76
	(FUNCTIONAL DEPENDENCY)	76
	MÃ CHƯƠNG: MH10_04	76
1.	KHÁI NIỆM PHỤ THUỘC HÀM.....	76
1.1	ĐỊNH NGHĨA PHỤ THUỘC HÀM	78
1.2	PHỤ THUỘC HÀM HIỆN NHIÊN (TRIVIAL DEPENDENCIES).....	79
1.3	THUẬT TOÁN SATIFIES.....	79
1.4	CÁC PHỤ THUỘC HÀM CÓ THỂ CÓ.....	81
1.4.1	<i>Cách tìm tất cả tập con của Q^+</i>	<i>81</i>
1.4.2	<i>Cách tìm tất cả các phụ thuộc hàm có thể có của Q.....</i>	<i>82</i>
2.	HỆ LUẬT DẪN ARMSTRONG (ARMSTRONG INFERENCE RULE)	
	92	
2.1	PHỤ THUỘC HÀM ĐƯỢC SUY DIỄN LOGIC TỪ F	92
2.2	HỆ LUẬT DẪN ARMSTRONG.....	93
2.2.1	<i>Hệ luật dẫn Armstrong là đúng</i>	<i>95</i>
2.3	BAO ĐÓNG CỦA TẬP THUỘC TÍNH X (CLOSURES OF ATTRIBUTE SETS).....	96
2.3.1	<i>Định nghĩa</i>	<i>96</i>
2.3.2	<i>Các tính chất của bao đóng.....</i>	<i>96</i>
2.3.3	<i>Thuật toán tìm bao đóng</i>	<i>97</i>
2.3.4	<i>Định lý</i>	<i>98</i>
2.3.5	<i>Hệ quả</i>	<i>99</i>
2.3.6	<i>Hệ luật dẫn Armstrong là đầy đủ</i>	<i>99</i>
3.	THUẬT TOÁN TÌM F^+	101

3.1	THUẬT TOÁN CƠ BẢN	101
3.2	THUẬT TOÁN CẢI TIẾN	102
4.	BÀI TẬP	103
CHƯƠNG 5: PHỦ CỦA TẬP PHỤ THUỘC HÀM.....		105
MÃ CHƯƠNG: MH10_05		105
1.	ĐỊNH NGHĨA	105
2.	PHỦ TỐI THIỂU CỦA MỘT TẬP PHỤ THUỘC HÀM (MINIMAL COVER).....	106
2.1	PHỤ THUỘC HÀM CÓ VỀ TRÁI DƯ THỪA	106
2.2	TẬP PHỤ THUỘC HÀM CÓ VỀ PHẢI MỘT THUỘC TÍNH (THE RIGHT SIDES OF DEPENDENCIES HAS A SINGLE ATTRIBUTE)	107
2.3	TẬP PHỤ THUỘC HÀM KHÔNG DƯ THỪA	107
2.4	TẬP PHỤ THUỘC HÀM TỐI THIỂU (MINIMAL COVER)	107
3.	KHÓA CỦA LƯỢC ĐỒ QUAN HỆ (KEY).....	109
3.1	ĐỊNH NGHĨA.....	109
3.2	THUẬT TOÁN TÌM TẤT CẢ KHÓA	110
3.2.1	<i>Thuật toán cơ bản</i>	<i>110</i>
3.2.2	<i>Thuật toán cải tiến</i>	<i>111</i>
4.	BÀI TẬP	113
CHƯƠNG 6: CHUẨN HÓA CƠ SỞ DỮ LIỆU.....		117
MÃ CHƯƠNG: MH10_06		117
1.	DẠNG CHUẨN CỦA LƯỢC ĐỒ QUAN HỆ (NORMAL FORMS FOR RELATION SCHEMES).....	117
1.1	DẠNG CHUẨN MỘT (FIRST NORMAL FORM).....	117
1.2	DẠNG CHUẨN 2 (SECOND NORMAL FORM)	119

1.3	DẠNG CHUẨN 3 (THIRD NORMAL FORM)	121
1.4	DẠNG CHUẨN BC (BOYCE-CODD NORMAL FORM)	125
2.	PHÉP TÁCH KẾT NỐI BẢO TOÀN.....	128
2.1	PHÉP TÁCH KẾT NỐI BẢO TOÀN THÔNG TIN (LOSSLESS-JOIN DECOMPOSITION)	128
2.1.1	<i>Định nghĩa phép tách Q thành 2 lược đồ con.....</i>	<i>129</i>
2.1.2	<i>Tính chất.....</i>	<i>129</i>
2.1.3	<i>Phép tách Q thành n lược đồ con.....</i>	<i>130</i>
2.1.4	<i>Thuật toán kiểm tra phép tách kết nối bảo toàn thông tin.....</i>	<i>131</i>
2.2	PHÉP TÁCH BẢO TOÀN PHỤ THUỘC HÀM (DECOMPOSITIONS THAT PRESERVE DEPENDENCIES)	136
2.2.1	<i>Tập phụ thuộc hàm F_i của Q_i</i>	<i>136</i>
2.2.2	<i>Thuật toán kiểm tra bảo toàn phụ thuộc hàm</i>	<i>137</i>
3.	BÀI TẬP.....	138
	TÀI LIỆU THAM KHẢO.....	142

CHƯƠNG TRÌNH MÔN HỌC CƠ SỞ DỮ LIỆU

Tên môn học: CƠ SỞ DỮ LIỆU

Mã môn học: MH10

I. Vị trí, tính chất của môn học:

1. Vị trí: Cơ sở dữ liệu là môn học cơ sở đào tạo trình độ Cao đẳng Quản trị mạng máy tính. Môn học này được bố trí sau môn học: Tin học căn bản.

2. Tính chất: Cơ sở dữ liệu là môn học cơ sở bắt buộc.

II. Mục tiêu môn học:

1. Về kiến thức:

- Trình bày được nguyên lý thiết kế cơ sở dữ liệu quan hệ;
- Trình bày được các mô hình dữ liệu và các công cụ mô tả dữ liệu;
- Trình bày được các khái niệm, tính năng và các phương thức xử lý dữ liệu của hệ quản trị cơ sở dữ liệu SQL;

2. Về kỹ năng:

- Sử dụng thành thạo cách xây dựng các ràng buộc, các phụ thuộc hàm, cách chuẩn hóa các cơ sở dữ liệu quan hệ;
- Thiết kế được một số cơ sở dữ liệu quan hệ thông dụng: quản lý nhân sự, quản lý bán hàng, ...;

3. Về năng lực tự chủ và trách nhiệm:

- Có khả năng tự nghiên cứu, tự học, tham khảo tài liệu liên quan đến môn học để vận dụng vào hoạt động học tập.
- Vận dụng được các kiến thức tự nghiên cứu, học tập và kiến thức, kỹ năng đã được học để hoàn thiện các kỹ năng liên quan đến môn học một cách khoa học, đúng quy định.

III. Nội dung môn học:

CHƯƠNG 1: MÔ HÌNH QUAN HỆ

Mã chương: MH10_01

Giới thiệu:

Bài học giúp sinh viên Thực hiện đúng các bước chuyển đổi từ lược đồ cơ sở dữ liệu sang mô hình quan hệ dữ liệu, áp dụng các phép toán đại số quan hệ để biểu diễn trên lược đồ quan hệ.

Mục tiêu:

- Trình bày được các khái niệm về cơ sở dữ liệu, hệ quản trị cơ sở dữ liệu và mô hình quan hệ;
- Thực hiện đúng các bước chuyển đổi từ lược đồ cơ sở dữ liệu sang mô hình quan hệ dữ liệu;
- Áp dụng các phép toán đại số quan hệ để biểu diễn trên lược đồ quan hệ;
- Nghiêm túc, tỉ mỉ trong việc học và làm bài tập.

1. NGUYÊN NHÂN RA ĐỜI CỦA MÔ HÌNH QUAN HỆ (RELATIONAL MODEL)

Trong nhiều năm, công nghệ tính toán và thông tin phát triển từ những hệ thống lớn, đắt tiền, độc quyền đến các hệ thống mở mạnh và không đắt tiền. Sự phát triển này mang lại lợi ích to lớn cho người dùng cuối bởi sự phát triển của các gói ứng dụng số như xử lý văn bản, bảng tính điện tử, văn phòng xuất bản, hệ quản lý cơ sở dữ liệu, máy tính trợ giúp công nghệ phần mềm....

Trước khi máy tính hóa cơ sở dữ liệu được giới thiệu, dữ liệu được lưu trữ theo kiểu điện tử thành nhiều tập tin riêng biệt sử dụng hệ tập tin (từ đây về sau ta gọi hệ tập tin theo lối cũ). Những tập tin này được xử lý bằng các ngôn ngữ thế hệ thứ ba như COBOL, FORTRAN, PASCAL và ngay cả BASIC để tạo ra các giải pháp cho các vấn đề của doanh nghiệp. Mỗi ứng dụng, chẳng hạn như hệ tính lương, hệ kho hay hệ thống kế toán sẽ có một tập các tập tin riêng chứa dữ liệu riêng. Các ứng dụng như vậy tạo ra ba vấn đề sau:

- ✓ Có sự liên kết chặt chẽ giữa cấu trúc luận lý và cấu trúc vật lý của các tập tin và chương trình ứng dụng khai thác chúng. Điều này khiến việc

tạo nên các ứng dụng này rất khó khăn, tốn nhiều thời gian và do vậy mà tồn kém trong bảo trì hệ thống.

- ✓ Có sự dư thừa dữ liệu rất lớn qua việc trùng lặp các tập tin trong các ứng dụng khác nhau. Điều này tạo ra những vấn đề như: dữ liệu thiếu nhất quán, không gian đĩa bị lãng phí, thời gian bảo trì và lưu phòng hồ các tập tin gia tăng, vấn đề về quản trị như không chú trọng bảo mật và tổ chức dữ liệu thiếu thống nhất.

Một ví dụ điển hình về sự trùng lặp dữ liệu là:

Hệ quản lý nguồn nhân lực bao gồm ba hệ chính:

- ✓ Hệ lương, hệ này duy trì ngày công và lương cho tất cả nhân viên.
- ✓ Hệ nhân sự, hệ này duy trì lý lịch cá nhân, dữ liệu về tổ chức, công việc đào tạo và vị trí thăng tiến.
- ✓ Hệ hưu, hệ này quản trị các qui tắc liên quan đến nghỉ hưu, loại nghỉ hưu. Chi tiết về hưu của từng nhân viên.

Vấn đề phức tạp là Hệ lương thông thường được quản lý bởi phòng tài chính, trong khi Hệ nhân sự và Hệ hưu được quản lý bởi phòng tổ chức. Rõ ràng, có nhiều dữ liệu về nhân viên là chung cho cả ba hệ. Thường những hệ này thực hiện và giữ gìn riêng biệt và chúng tạo sự trùng dữ liệu nhân viên mà chúng dùng.

Người sử dụng có ít khả năng khai thác trực tiếp dữ liệu.

2. HỆ QUẢN TRỊ CƠ SỞ DỮ LIỆU (DATABASE MANAGEMENT SYSTEM)

Khởi đầu, sự giới thiệu CSDL và HQTCSĐL nhằm giải quyết các vấn đề của hệ thông tin dựa trên các tập tin theo lối cũ (C1.I). Điều này tạo ra việc phát triển trên hai mươi lăm năm qua một hệ CSDL quan hệ thương mại xuất hiện cuối những năm thập niên 70 và các năm đầu của thập niên 80. Trước khi xem xét CSDL và hệ QTCSĐLQH giải quyết một vài vấn đề của hệ thông tin theo lối cũ như thế nào chúng ta cần làm rõ vài khái niệm.

2.1 CSDL là gì?

Một cơ sở dữ liệu có thể định nghĩa tạm như sau: một chỗ chứa có tổ chức tập hợp các tập tin dữ liệu có tương quan, các mẫu tin và các cột. Hay nói cách

khác, cơ sở dữ liệu là một sưu tập dữ liệu có cấu trúc, có liên quan với nhau tạo thành cơ sở dữ liệu nhằm phục vụ cho một mục đích quản lý nào đó.

Ngày nay CSDL tồn tại trong mỗi ứng dụng thông dụng.

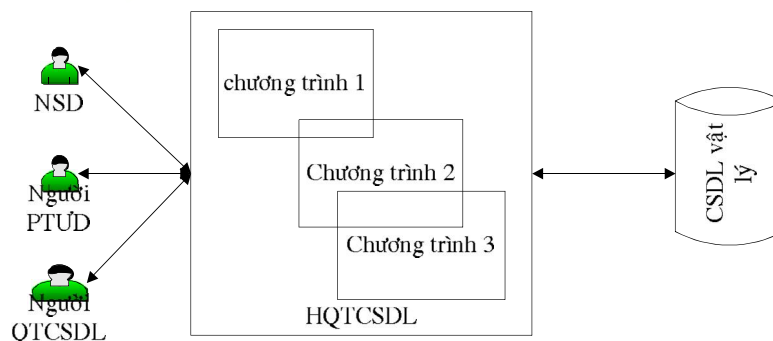
Ví dụ: - Hệ kho và kiểm kê. - Hệ đặt chỗ máy bay - Hệ nguồn nhân lực.

- ✓ Hệ dịch vụ công cộng như cấp nước, điện, khí đốt
- ✓ Điều khiển quá trình chế tạo và sản xuất

2.2 Hệ quản trị CSDL

Một hệ quản trị CSDL (HQTCSDL) là:

- ✓ Một tập các phần mềm quản lý CSDL và cung cấp các dịch vụ xử lý CSDL cho những người phát triển ứng dụng và người dùng cuối.
- ✓ HQTCSDL cung cấp một giao diện giữa người sử dụng và dữ liệu. - HQTCSDL biến đổi CSDL vật lý thành CSDL logic.



Hình 1.2.1 - HQTCSDL cung cấp phần giao diện giữa NSD và CSDL

Dựa vào cách tổ chức dữ liệu, HQTCSDL được chia thành năm loại:

- ✓ Loại phân cấp như hệ IMS của IBM
- ✓ Loại mạng như IDMS của Cullinet Software
- ✓ Loại tập tin đảo như ADABAS của Software AG
- ✓ Loại quan hệ như ORACLE của Oracle, DB2 của IBM, ACCESS của Microsoft Access
- ✓ Loại đối tượng là một tiếp cận khá mới trong thiết kế HQTCSDL và việc sử dụng loại này sớm trở nên phổ biến

Hiện tại, loại HQTCSDL chính được sử dụng trong công nghệ là loại HQTCSDL quan hệ (RDBMS). Loại này đã chiếm lĩnh trong công nghệ trên 10-15 năm cuối cùng khi đánh bật loại HQTCSDL phân cấp và gần đây là HQTCSDL mạng.

2.3 Người dùng (User)

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

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

- ✓ Sự chính xác và toàn vẹn của dữ liệu và ứng dụng trong CSDL, sự bảo mật của CSDL
- ✓ Lưu phòng hồ và phục hồi CSDL
- ✓ Giữ liên lạc với người phát triển ứng dụng, người lập trình và người dùng cuối.
- ✓ Bảo đảm sự hoạt động trôi chảy và hiệu quả của CSDL và HQTCSDL

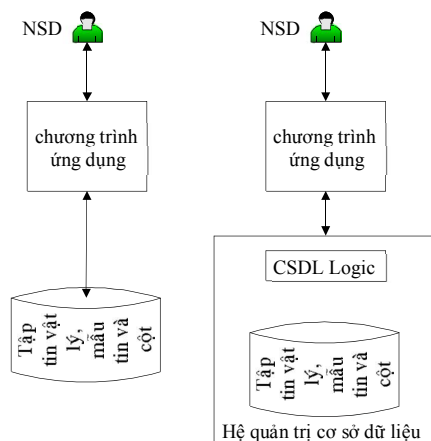
Người phát triển và lập trình ứng dụng là những người chuyên nghiệp về máy tính có trách nhiệm thiết kế, tạo dựng và bảo trì hệ thống tin cho người dùng cuối.

Người dùng cuối là những người không chuyên về máy tính nhưng họ là các chuyên gia trong các lĩnh vực khác có trách nhiệm cụ thể trong tổ chức. Họ khai thác CSDL thông qua hệ được phát triển bởi người phát triển ứng dụng hay các công cụ truy vấn của HQTCSDL.

2.4 CSDLQH và Hệ tập tin theo lối cũ

Tiếp cận CSDL đã giải quyết 3 vấn đề của hệ tập tin theo lối cũ:

2.4.1.1 Vấn đề 1: cấu trúc logic và cấu trúc vật lý



Tiếp cận theo lối cũTiếp cận theo csdl

Hình 1.2.2 - Tiếp cận theo csdl tạo ra một lớp phần mềm giữa dữ liệu vật lý và chương trình ứng dụng

Kiến trúc bên trong HQTCSDL quan hệ tách biệt rõ ràng giữa:

- ✓ Cấu trúc luận lý của tất cả tập tin và chương trình ứng dụng khai thác tập tin này và
- ✓ Cấu trúc vật lý của csdl và phần lưu trữ các tập tin.

Tiếp cận này tạo cho người quản trị CSDL có thể thay đổi cấu trúc vật lý hay nơi lưu trữ của tập tin mà không ảnh hưởng đến chương trình ứng dụng.

2.4.1.2 Vấn đề 2: Dư thừa dữ liệu

Khi HQTCSDLQH được giới thiệu, nhiều tổ chức mong tích hợp các tập tin đã phân tán khắp trong tổ chức vào một CSDL tập trung. Dữ liệu có thể chia sẻ cho nhiều ứng dụng khác nhau và người sử dụng có thể khai thác đồng thời các tập con dữ liệu liên quan đến họ. Điều này làm hạn chế sự dư thừa dữ liệu.

2.4.1.3 Vấn đề 3: Sự khai thác dữ liệu của người sử dụng

Trong hệ QTCSDLQH người dùng có thể trực tiếp khai thác dữ liệu thông qua việc sử dụng các câu truy vấn hay các công cụ báo cáo được cung cấp bởi hệ QTCSDL.

3. MÔ HÌNH QUAN HỆ (RELATIONAL MODEL)

3.1 Mô hình quan hệ là gì ?

Mô hình Cơ sở dữ liệu Quan hệ (gọi tắt là Mô hình Quan hệ) do E.F Codd đề xuất năm 1971. Mô hình này 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 phép toán tập hợp, phép toán quan hệ.
- ✓ Ràng buộc toàn vẹn quan hệ.

Các hệ HQTCSĐLQH ngày nay được xây dựng dựa vào lý thuyết của mô hình quan hệ. Mục đích của môn học này giúp cho sinh viên nắm được kiến trúc tổng quát về mô hình quan hệ và áp dụng nó để lập mô hình dữ liệu quan hệ có hiệu quả trong lưu trữ và khai thác (nghĩa là: xây dựng cấu trúc CSDL và truy vấn dữ liệu trong CSDL).

MASV	HOTEN	MONHOC	TENKHOA	DIEMTHI
99001	TRAN DAN THU	CO SO DU LIEU	CÔNG NGHỆ THỐNG TIN	3.0
99002	NGUYEN HA DA THAO	CO SO DU LIEU	CÔNG NGHỆ THỐNG TIN	8.0
99001	TRAN DAN THU	TIN HOC VAN PHONG	CÔNG NGHỆ THỐNG TIN	6.0
99005	LE THANH TRUNG	TIN HOC VAN PHONG	ANH VAN	5.0

Chẳng hạn với bài toán quản lý điểm thi của sinh viên, nếu lưu trữ dữ liệu theo dạng bảng với các cột MASV, HOTEN, MONHOC, TENKHOA, DIEMTHI 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.

Việc lưu trữ dữ liệu như trên không đúng với mô hình quan hệ. Để lưu trữ đúng với mô hình quan hệ ta phải thay MONHOC bằng MAMH, thay TENKHOA bằng MAKHOA, tách một bảng dữ liệu lớn đó ra thành nhiều bảng con, như mô hình dưới.

MASV	MAMH	DIEMTHI
99001	CSDL	3.0
99002	CSDL	8.0
99001	THVP	6.0
99005	THVP	5.0

MASV	HOTEN	MAKHOA
99001	TRAN DAN THU	CNTT
99002	NGUYEN HA DA THAO	CNTT
99005	LE THANH TRUNG	AV

MAMH	TENMH	SOTIET
CSDL	CO SO DU LIEU	90
THVP	TIN HOC VAN PHONG	90

MAKHOA	TENKHOA
CNTT	CONG NGHE THONG TIN
AV	ANH VAN

3.2 Các khái niệm cơ bản của mô hình quan hệ

3.2.1 Thuộc tính (Attribute, Arity)

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, nữ, tỉnh thường trú, học bổng, lớp mà sinh viên theo học, . . . các đặc trưng này gọi là thuộc tính.

Các thuộc tính được phân biệt qua tên gọi và phải thuộc vào một kiểu dữ liệu nhất định (số, chuỗi, ngày tháng, logic, hình ảnh, ...). Kiểu dữ liệu ở đây là kiểu đơn. 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. Thuộc tính ngày trong tháng thì có kiểu dữ liệu là số nguyên, miền giá trị của nó là 1 đến (tối đa là) 31. Hoặc điểm thi của sinh viên chỉ là các số nguyên từ 0 đến 10 (đối với thang điểm 10).

Thường người ta dùng các chữ cái hoa A, B, C, ... để biểu diễn các cột thuộc tính, hoặc A1, A2, ..., An để biểu diễn một số lượng lớn các thuộc tính.

3.2.2 Lược Đồ Quan Hệ (Relation Schema)

Tập tất cả các thuộc tính cần quản lý của một đối tượng cùng với mối liên hệ giữa chúng được gọi là lược đồ quan hệ. Lược đồ quan hệ Q với tập thuộc tính {A1, A2, ..., An} được viết là Q(A1, A2, ..., An). Tập các thuộc tính của Q được ký hiệu là Q+. Chẳng hạn lược đồ quan hệ sinh viên (Đặt tên là Sv) với các thuộc tính như trên là:

SV (MASV, HOSV, TENSU, NU, NGAYSINH, MALOP, HOCBONG, TINH)

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 sinh viên (MASV) duy nhất, mỗi mã sinh viên xác định tất cả các thuộc tính của sinh viên đó như họ tên (HOTEN), nữ (NU), ngày sinh (NGAYSINH), lớp theo học (MALOP), học bổng (HOCBONG), tỉnh cư trú (TINH).

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 lược đồ cơ sở dữ liệu để quản lý điểm sinh viên có thể gồm các lược đồ quan hệ sau:

Sv(MASV, HOSV, TENSv, NU, NGAYSINH, MALOP, TINH, *HOCBONG*)

Lop(MALOP, TENLOP, *SISO*, MAKHOA)

Kh(MAKHOA, TENKHOA, *SOCBGD*)

Mh(MAMH, TENMH, SOTIET)

Kq(MASV, MAMH, DIEMTHI)

Phần giải thích các thuộc tính:

MASV	Mã sinh viên	HOCBONG	học bổng
HOTEN	Tên sinh viên	TINH	tỉnh
NU	Nữ	TENKHOA	tên khoa
NGAYSINH	ngày sinh	SOCBGD	số cán bộ giảng dạy
LOP	lớp	MAMH	mã môn học
TENLOP	tên lớp	TENMH	tên môn học
SISO	sĩ số lớp	SOTIET	số tiết
MAKHOA	mã khoa	DIEMTHI	điểm thi

Sự thể hiện của lược đồ quan hệ Q tại một thời điểm nào đó được gọi là quan hệ, rõ ràng là trên một lược đồ quan hệ có thể định nghĩa 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

3.2.3 Bộ (Tuple)

Bộ là tập mỗi giá trị liên quan của tất cả các thuộc tính của một lược đồ quan hệ. Chẳng hạn quan hệ sau có 2 bộ.

MASV	HOTEN	NU	NGAYSINH	MALOP	TINH	HOCBONG
99001	TRAN DAN THU	TRUE	15-03-1977	CĐTH2B	TIEN GIANG	120000
99002	NGUYEN HA DA THAO	TRUE	25-04-1986	TCTH29C	TPHCM	120000

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 thông tin về một bộ. Chẳng hạn sau đây là các thể hiện của các quan hệ định nghĩa trên lược đồ cơ sở dữ liệu quản lý sinh viên ở trên:

rSv

MASV	HOTEN	NU	NGAYSINH	MALOP	TINH	HOCBONG
99001	TRAN DAN THU	TRUE	15-03-1977	CĐTH2B	TIEN GIANG	120000
99002	NGUYEN HA DA THAO	TRUE	25-04-1986	CĐTH2B	TPHCM	120000
99003	PHAM ANH HUY	FALSE	16-08-1977	CĐTH2B	BAC LIEU	
99004	NGUYEN NGOC THUAN	FALSE	24-12-1980	CĐTH2B	CA MAU	
99005	LE THANH TRUNG	FALSE	20-11-1978	CĐAV1	CA MAU	120000
99006	NGUYEN HONG VAN	FALSE	19-09-1979	CĐAV1	SOC TRANG	
99007	VU THI LOAN	TRUE	15-03-1975	CĐAV1	CA MAU	
99008	TRUONG KIM QUANG	FALSE	15-05-1975	CĐTH2B	HA NOI	
99009	TON THAT QUYEN	FALSE	26-06-1976	CĐTH2B	VUNG TAU	60000
99010	HA VAN LONG	FALSE	14-04-1973	CĐAV1	BAC LIEU	
99011	BUI VAN ANH	FALSE	22-12-1972	CĐAV1	AN GIANG	
99012	LE HUU CHI	FALSE	28-08-1977	CĐĐT2	CAN THO	60000
99013	VU THANH CONG	FALSE	29-03-1979	CĐTH2B	KIEN GIANG	60000
99014	TRAN QUANG CUONG	FALSE	30-05-1981	CĐĐT2	DONG THAP	120000
99015	PHAM VAN HAI	FALSE	30-06-1976	CĐĐT2	CA MAU	
99016	HUYNH THANH HOANG	FALSE	29-07-1982	CĐĐT2	TPHCM	80000
99017	TRAN MINH LAM	FALSE	21-08-1977	CĐTH2B	TRA VINH	
99018	PHAN VAN SANG	FALSE	19-05-1979	CĐDL1	DONG THAP	120000
99019	PHAM THI HUYEN	FALSE	16-06-1982	CĐDL1	CAN THO	120000
99020	NGUYEN THI NGAN	TRUE	11-11-1981	CĐTH2B	CA MAU	120000
99021	PHAM TAN QUANG	FALSE	01-01-1980	CĐDL1	CA MAU	
99022	TRAN PHUOC QUYEN	FALSE	12-12-1979	CĐTH2B	BAC LIEU	60000
99023	LE THI THANH VAN	TRUE	11-11-1980	CĐDL1	TPHCM	120000

rKh

MAKHOA	TENKHOA	SOCBGD
CNTT	CÔNG NGHỆ THÔNG TIN	60
AV	ANH VAN	60
HOA	HÓA CHẤT	20
MTRUONG	MÔI TRƯỜNG	10
XD	XÂY DỰNG	10
DL	DU LỊCH	5
TTIN	TOÁN - TIN HỌC	30
SINH	CÔNG NGHỆ SINH HỌC	30
VL	VẬT LÝ	20
ĐT	ĐIỆN TỬ	20

rMh

MAMH	TENMH	SOTIET
CSDL	CƠ SỞ DỮ LIỆU	90
FOX	FOXPPO	120
THVP	TIN HỌC VĂN PHÒNG	90
AVTH	ANH VAN TIN HỌC	60
KTS	KỸ THUẬT SỐ	60
CTDL	CAU TRUC DU LIEU	60
TTNT	TRI TUE NHAN TAO	45
MANG	MANG MÁY TINH CB	45
VB	VI SUAL BASIC	90
AC	ACCESS	180
LR	LAP RAP MAY TINH	60
INTER	CAC DICH VU INTERNET	45
HDH	HE DIEU HANH	60
KTLT	KY THUAT LAP TRINH	45
VIFOX	VISUAL FOXPRO	60

rKq

MASV	MAMH	DIEMTHI
99001	CSDL	3.0
99002	CSDL	8.0
99003	CSDL	8.0
99004	CSDL	3.0
99005	CSDL	2.0
99001	THVP	6.0
99002	THVP	3.0
99003	THVP	8.0
99004	THVP	9.0
99005	THVP	5.0

99006	MANG	6.0
99007	MANG	2.0
99008	MANG	7.0
99009	MANG	3.0
99010	TTNT	5.0
99011	FOX	4.0
99012	FOX	5.0
99013	FOX	7.0
99014	VB	7.0
99015	VB	3.0

99016	KTS	7.0
99017	KTS	7.0
99017	FOX	4.0
99017	MANG	8.0
99017	CSDL	8.0
99017	TTNT	6.0
99002	MANG	8.0
99004	MANG	4.0
99018	TTNT	7.0
99019	CSDL	8.0

99020	THVP	7.0
99021	MANG	7.0
99022	MANG	6.0
99023	CSDL	8.0
99023	MANG	6.0

99023	TTNT	3.0
99023	THVP	6.0
99023	FOX	8.0
99023	VB	9.0
99023	KTS	6.0

99021	CSDL	8.0
99021	THVP	9.0
99022	FOX	5.0
99022	TTNT	6.0

rLop

MALOP	TENLOP	SISO	MAKHOA
CĐTH2B	CAO ĐẲNG TIN HỌC KHOÁ 2000B	60	cntt
TCTH29C	TRUNG CẤP TIN HỌC KHOÁ 29 C	121	cntt
CĐAV1	CAO ĐẲNG ANH VĂN 1	120	av
CĐĐT2	CAO ĐẲNG ĐIỆN TỬ 2	80	dt
CĐDL1	CAO ĐẲNG DU LỊCH 1	42	dl

3.2.4 Khóa (Key, Candidate Key):

Cho lược đồ quan hệ R, $S \sqsubseteq R^+$. S được gọi là một siêu khóa (superkey) của lược đồ quan hệ R nếu với hai bộ tùy ý trong quan hệ R thì giá trị của các thuộc tính trong S là khác nhau.

Một lược đồ quan hệ có thể có nhiều siêu khóa. Siêu khóa chứa ít thuộc tính nhất được gọi là khóa chỉ định (hay khóa đề nghị), trong trường hợp lược đồ quan hệ có nhiều khóa chỉ định, thì khóa được chọn để cài đặt gọi là khóa chính (Primary key) (phần sau khóa chính được gọi tắt là khóa)

Các thuộc tính tham gia vào một khóa được gọi là thuộc tính khóa (prime key), ngược lại được gọi là thuộc tính không khóa (non prime key).

Một 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.