

TRƯỜNG ĐẠI HỌC ĐÀ LẠT
KHOA CÔNG NGHỆ THÔNG TIN

CAO THỊ NHẬN

NGUYỄN THỊ THANH BÌNH

GIÁO TRÌNH

CƠ SỞ DỮ LIỆU

Dành cho sinh viên ngành Công nghệ Thông tin

Đà Lạt 2010

LỜI MỞ ĐẦU

Để đáp ứng nhu cầu học tập của sinh viên chuyên ngành Công nghệ Thông tin, bài giảng Cơ sở dữ liệu được biên soạn theo chương trình hệ thống tín chỉ của Trường Đại học Đà Lạt, cung cấp các kiến thức cơ bản về lý thuyết cơ sở dữ liệu.

Giáo trình gồm 8 chương sau:

Chương 1: giới thiệu chung

Chương 2: trình bày mô hình thực thể kết hợp E_R (Entity Relationship) để mô hình hóa các hoạt động trong thế giới thực, nhìn thế giới thực như là một tập các đối tượng căn bản được gọi là các thực thể, và các mối quan hệ ở giữa các đối tượng này. Mô hình được phát triển để làm thuận tiện cho việc thiết kế cơ sở dữ liệu bằng cách đặc tả một tổ chức.

Chương 3: trình bày mô hình dữ liệu quan hệ, các quy tắc chuyển đổi cơ sở dữ liệu biểu diễn dạng lược đồ E_R sang mô hình dữ liệu quan.

Chương 4, 5 và 6: trình bày các ngôn ngữ truy vấn trên cơ sở dữ liệu. Trong chương 4 trình bày ngôn ngữ đại số quan hệ, chương 5 trình bày ngôn ngữ tân từ và chương 6 trình bày ngôn ngữ truy vấn SQL.

Chương 7: trình bày các khái niệm liên quan đến phụ thuộc hàm, khóa, các thuật toán tìm phủ tối thiểu của tập phụ thuộc hàm, tìm khóa và cuối chương trình bày về ràng buộc toàn vẹn dữ liệu.

Chương 8: Chương này giới thiệu các dạng chuẩn, phân rã bảo toàn thông tin, bảo toàn phụ thuộc hàm, qua đó cũng trình bày cách phân rã bảo toàn bảo toàn thông tin và bảo toàn phụ thuộc.

Mặc dù đã rất cố gắng trong quá trình biên soạn bài giảng nhưng chắc chắn bài giảng sẽ còn nhiều thiếu sót và hạn chế. Rất mong nhận được sự đóng góp ý kiến quý báu của sinh viên và các bạn đọc để giáo trình ngày một hoàn thiện hơn.

MỤC LỤC

<i>Lời mở đầu</i>	2
Mục lục	3
Chương 1: Giới thiệu chung	8
1. Giới thiệu chung.....	8
1.1. Giới thiệu	8
1.2. Định nghĩa CSDL	10
1.3. Các đối tượng sử dụng CSDL.	11
1.4. Hệ quản trị CSDL	12
1.5. Các mức biểu diễn một CSDL.....	12
1.6. Sơ đồ tổng quát một hệ quản trị CSDL	14
1.7. Tính độc lập giữa dữ liệu và chương trình.	15
2. Các cách tiếp cận của một CSDL.....	15
2.1. Cách tiếp cận theo mô hình dữ liệu mạng	16
2.2. Mô hình dữ liệu phân cấp	17
2.3. Mô hình dữ liệu quan hệ thực thể.....	18
2.4. Mô hình dữ liệu quan hệ.....	18
2.5. Mô hình dữ liệu hướng đối tượng	19
3. Bài tập.....	19
Chương 2: Mô hình thực thể kết hợp	21
1. Mô hình thực thể kết hợp.	21
1.1. Thực thể - tập thực thể.....	21
1.2. Thuộc tính.....	21
1.3. Mối kết hợp.....	22
1.4. Bản số	23
1.5. Khoá.....	24

1.6. Số chiều của mỗi kết hợp.....	25
1.7. Tổng quát hóa và chuyên biệt hóa.....	25
1.8. Tập thực thể yếu	26
2. Ví dụ	27
3. Bài tập.....	29

Chương 3: Mô hình dữ liệu quan hệ 30

1. Các khái niệm cơ bản	30
1.1. Thuộc tính.....	30
1.2. Quan hệ n ngôi.....	31
1.3. Bộ.....	31
1.4. Lược đồ quan hệ	32
1.5. Khóa của một quan hệ	33
1.6. Ràng buộc toàn vẹn	35
2. Các thao tác cơ bản trên quan hệ.....	35
2.1. Phép thêm	35
2.2. Phép xóa	36
2.3. Phép sửa.....	36
3. Các bước chuyển đổi từ mô hình thực thể kết hợp sang mô hình quan hệ.....	37
3.1. Biến các tập thực thể chuyên biệt hóa về dạng bình thường.....	37
3.2. Chuyển tất cả các tập thực thể thành quan hệ	38
3.3. Mỗi kết hợp.....	39
3.4. Nhập tất cả các quan hệ có cùng khóa.....	39

Chương 4: Ngôn ngữ đại số quan hệ 40

1. Các phép toán cơ sở	40
1.1. Các phép toán tập hợp	40
1.2. Các phép toán quan hệ	41
2. Các phép toán khác	46
2.1. Phép kết hai quan hệ	46
2.2. Phép kết nối nội	47
2.3. Phép kết nối trái	48
2.4. Phép kết nối phải	49
2.5. Hàm kết hợp và gom nhóm	49
2.6. Các phép toán cập nhật trên quan hệ	50
3. Bài tập	52

Chương 5: Ngôn ngữ tân từ 55

1. Ngôn ngữ tân từ có biến là bộ	55
1.1. Một số khái niệm	55
1.2. Định nghĩa hình thức của phép tính bộ	55
1.3. Lượng từ tồn tại $\exists$ và với mọi $\forall$	57
2. Ngôn ngữ tân từ có biến là miền giá trị	58
3. Bài tập	59

Chương 6: Ngôn ngữ truy vấn SQL 61

1. Các lệnh hỏi	61
1.1. Cú pháp lệnh truy vấn	61
1.2. Phép chiếu	62
1.3. Phép chọn	62
1.4. Phép kết	63

1.5. Một số lưu ý.....	63
2. Truy vấn lồng	65
3. Hàm kết hợp và gom nhóm	68
4. Các lệnh khai báo cấu trúc CSDL	69
5. Các thao tác cập nhật dữ liệu.....	72
6. Bài tập.....	73
Chương 7: Phụ thuộc hàm, khóa, ràng buộc toàn vẹn	74
1. Phụ thuộc hàm	74
1.1. Khái niệm phụ thuộc hàm.....	74
1.2. Hệ luật dẫn Amstrong.....	75
1.3. Thuật toán tìm bao đóng của tập thuộc tính	77
1.4. Bài toán thành viên.....	78
1.5. Phủ tối thiểu của một tập phụ thuộc hàm	78
2. Khóa	82
2.1. Định nghĩa	82
2.2. Thuật toán tìm khóa.....	82
3. Ràng buộc toàn vẹn	85
3.1. Định nghĩa – các yếu tố của ràng buộc toàn vẹn.....	85
3.2. Các loại ràng buộc toàn vẹn	87
4. Bài tập.....	93
Chương 8: Dạng chuẩn và chuẩn hóa CSDL	96
1. Dạng chuẩn của lược đồ quan hệ	96
1.1. Dạng chuẩn 1	96
1.2. Dạng chuẩn 2	97
1.3. Dạng chuẩn 3	100

1.4.	Dạng chuẩn BC	101
1.5.	Kiểm tra dạng chuẩn	101
2.	Phép phân rã	102
2.1.	Phân rã bảo toàn thông tin	102
2.2.	Phân rã bảo toàn phụ thuộc hàm.....	103
3.	Thiết kế CSDL bằng cách phân rã	104
3.1.	Phân rã thành dạng chuẩn BC (hoặc dạng chuẩn 3) bảo toàn thông tin.....	104
3.2.	Phân rã thành dạng chuẩn 3 vừa bảo toàn thông tin vừa bảo toàn phụ thuộc hàm	108
4.	Bài tập.....	108
	<i>Tài liệu tham khảo</i>	<i>109</i>

Chương 1

Giới Thiệu Chung

Chương 1 giới thiệu cho học viên các khái niệm cơ bản về cơ sở dữ liệu (CSDL) và hệ quản trị CSDL.

1. Giới thiệu chung

1.1. Giới thiệu

Xét một hệ thống các tập tin cổ điển: được tổ chức một cách riêng rẽ, phục vụ cho một mục đích của một đơn vị hay một đơn vị con trực thuộc cụ thể. Ứng dụng là một hệ quản lý nhân sự của một công ty.

Tại một công ty đã được trang bị máy vi tính cho tất cả các phòng ban. Trong đó:

- Phòng Tổ chức lưu giữ những thông tin về lý lịch của nhân viên như họ tên, bí danh, giới tính, ngày sinh, ngày vào đoàn, ngày vào đảng, ngày tuyển dụng, hệ số lương, hệ số phụ cấp, hoàn cảnh gia đình ... Phần mềm được sử dụng để quản lý là Microsoft Access.
- Phòng Tài vụ lưu giữ những thông tin về họ tên, hệ số lương, hệ số phụ cấp, đơn vị ... để tính toán lương và phụ cấp dựa trên danh sách do phòng tổ chức gửi sang. Các bảng tính này được làm bởi Microsoft Excel.
- Bộ phận Văn thư sử dụng Microsoft word để thảo các văn bản báo cáo về tình hình của đơn vị trong đó có nêu tổng số công nhân viên và trình độ chuyên môn được đào tạo.
- Trong khi đó tại tổng công ty các phòng ban cũng được trang bị máy vi tính và cũng quản lý những thông tin của các cán bộ chủ chốt của công ty con là giám đốc, phó giám đốc, trưởng phó phòng.

Việc phát triển của hệ thống quản lý như sau:

- Những chương trình ứng dụng mới phải được viết khi một nhu cầu nảy sinh.
- Các tập tin thường trực mới được tạo ra theo yêu cầu.
- Trong một thời gian dài các tập tin có thể ở các dạng thức khác nhau.
- Những chương trình ứng dụng có thể viết bởi các ngôn ngữ khác nhau.

Ưu điểm:

- Việc xây dựng những hệ thống tập tin riêng lẻ tại từng đơn vị quản lý ít tốn thời gian bởi lượng thông tin cần quản lý và khai thác là nhỏ, không đòi hỏi đầu tư nhiều nên triển khai nhanh.
- Thông tin được khai thác chỉ phục vụ cho mục đích hẹp nên khả năng đáp ứng nhanh chóng, kịp thời.

Hạn chế:

Do thông tin được tổ chức ở mỗi phòng ban mỗi khác (thông tin và công cụ triển khai) nên rõ ràng sự phối hợp về mặt tổ chức và khai thác ở các phòng ban rất khó khăn. Thông tin ở phòng ban này không được sử dụng cho phòng ban khác, tại công ty con với tổng công ty. Hơn nữa cùng một thông tin được cập nhật vào máy tại nhiều nơi khác nhau. Do đó có những hạn chế sau:

- **Dữ liệu dư thừa và không nhất quán**
 - ✓ Những thông tin giống nhau có thể bị trùng lặp ở một số nơi.
 - ✓ Tất cả những bản sao dữ liệu có thể không được cập nhật đúng.
- **Khó khăn trong việc truy xuất dữ liệu**
 - ✓ Có thể phải viết một trình ứng dụng mới thỏa yêu cầu mới đặt ra.
 - ✓ Có thể tạo ra dữ liệu này một cách thủ công nhưng mất thời gian.
- **Khó khăn khi viết chương trình ứng dụng đòi hỏi thông tin liên quan đến nhiều đơn vị**
 - ✓ Dữ liệu ở những tập tin khác nhau.
 - ✓ Dữ liệu ở các dạng thức khác nhau.

Ngoài ra còn một số vấn đề chưa được giải quyết như:

- **Nhiều người sử dụng**
 - ✓ Muốn thời gian trả lời nhanh hơn cho các tác vụ đồng thời.
 - ✓ Cần sự bảo vệ đối với những cập nhật đồng thời.
- **Các vấn đề bảo mật**
 - ✓ Mỗi người sử dụng hệ thống chỉ có thể truy xuất những dữ liệu mà họ được phép thấy.

- ✓ Chẳng hạn như những người tính lương và phụ cấp chỉ giải quyết các mẫu tin nhân viên, và không thể thấy được thông tin về lý lịch nhân viên; những thủ quỹ chỉ truy xuất những dữ liệu về các khoản chi...
- ✓ Khó khăn đề ràng buộc điều này trong những chương trình ứng dụng.
- **Các vấn đề toàn vẹn.**
 - ✓ Dữ liệu có thể được yêu cầu thỏa mãn những ràng buộc.
 - ✓ Chẳng hạn như nhân viên đang hưởng chế độ nào đó ở phòng tổ chức thì sẽ không được hưởng chế độ trợ cấp cao nhất của phòng công đoàn.
 - ✓ Với cách tiếp cận xử lý-tập tin, khó khăn để thực hiện hoặc thay đổi những ràng buộc như vậy.
- **Khó khăn khi nâng cấp hệ thống.**
 - ✓ Do hệ thống được tổ chức thành các tập tin riêng lẻ nên thiếu sự chia sẻ thông tin giữa các phòng ban. Bên cạnh đó việc kết nối các hệ thống này với nhau hay nâng cấp ứng dụng sẽ gặp rất nhiều khó khăn.

Những nhược điểm này đã dẫn tới việc phát triển các hệ thống quản trị cơ sở dữ liệu.

1.2. Định nghĩa cơ sở dữ liệu

Cơ sở dữ liệu (database) là một hệ thống các thông tin có cấu trúc được lưu trữ trên các thiết bị lưu trữ thứ cấp (băng từ, đĩa từ...) nhằm thỏa mãn yêu cầu khai thác thông tin đồng thời của nhiều người sử dụng hay nhiều chương trình ứng dụng với nhiều mục đích khác nhau.

Ưu điểm của CSDL:

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

Để đạt được những ưu điểm trên CSDL đặt ra những vấn đề cần giải quyết, đó là:

- **Tính chủ quyền của dữ liệu.** Do sự chia sẻ của CSDL nên tính chủ quyền của dữ liệu có thể bị lu mờ và làm mờ nhạt tinh thần trách nhiệm, được thể hiện trên vấn đề an toàn dữ liệu, khả năng biểu diễn các mối liên hệ ngữ nghĩa của dữ liệu và tính chính xác của dữ liệu. Nghĩa là người khai thác CSDL phải có nghĩa vụ cập nhật các thông tin mới nhất của CSDL.

- **Tính bảo mật và quyền khai thác thông tin của người sử dụng (NSD).** Do có nhiều người cùng khai thác CSDL một cách đồng nên cần phải có một cơ chế bảo mật và phân quyền khai thác CSDL.
- **Tranh chấp dữ liệu.** Nhiều người được phép truy cập vào cùng một tài nguyên dữ liệu (data source) của CSDL với những mục đích khác nhau như xem, thêm, xóa, sửa dữ liệu. Như vậy cần phải có cơ chế ưu tiên truy nhập dữ liệu cũng như cơ chế giải quyết tình trạng khoá chết (DeadLock) trong quá trình khai thác cạnh tranh. Cơ chế ưu tiên có thể được thực hiện dựa trên:
 - ✓ Cấp quyền ưu tiên cho người khai thác CSDL, người nào có quyền ưu tiên cao hơn thì được ưu tiên cấp quyền truy nhập dữ liệu trước.
 - ✓ Thời điểm truy nhập, ai có yêu cầu truy xuất trước thì có quyền truy nhập dữ liệu trước.
 - ✓ Hoặc dựa trên cơ chế lập lịch truy xuất hay cơ chế khoá.
 - ✓ ...
- **Đảm bảo dữ liệu khi có sự cố.** Việc quản lý tập trung có thể làm tăng khả năng mất mát hoặc sai lệch thông tin khi có sự cố như mất điện đột xuất, một phần hay toàn bộ đĩa lưu trữ dữ liệu bị hư... Một số hệ điều hành mạng có cung cấp dịch vụ sao lưu đĩa cứng (cơ chế sử dụng đĩa cứng dự phòng - RAID), tự động kiểm tra và khắc phục lỗi khi có sự cố. Tuy nhiên bên cạnh dịch vụ của hệ điều hành, một CSDL nhất thiết phải có một cơ chế khôi phục dữ liệu khi các sự cố bất ngờ xảy ra để đảm bảo CSDL luôn ổn định.

1.3. Các đối tượng sử dụng CSDL

- Những người sử dụng không chuyên về lĩnh vực tin học và CSDL, do đó CSDL cần có các công cụ để giúp cho người sử dụng không chuyên có thể khai thác hiệu quả CSDL
- Các chuyên viên tin học biết khai thác CSD, đây là những người có thể xây dựng các ứng dụng khác nhau phục vụ cho những yêu cầu khác nhau trên CSDL.
- Những người quản trị CSDL, là những người hiểu biết về tin học, các hệ quản trị CSDL và hệ thống máy tính. Họ là người tổ chức CSDL (khai báo cấu trúc CSDL, ghi nhận các yêu cầu bảo mật cho các dữ liệu cần bảo vệ...), do đó họ cần phải nắm rõ các vấn đề về kỹ thuật để có thể phục hồi dữ liệu khi có sự cố. Họ là những người cấp quyền hạn khai thác CSDL nên họ có thể giải quyết các vấn đề tranh chấp dữ liệu.

1.4. Hệ quản trị CSDL

Để giải quyết tốt những vấn đề đặt ra cho một CSDL như đã trình bày trong 1.2 thì cần có một hệ thống các phần mềm chuyên dụng. Hệ thống này được gọi là hệ quản trị CSDL (Database Management System - DBMS), là công cụ hỗ trợ tích cực cho các nhà phân tích, thiết kế và khai thác CSDL. Hiện nay có nhiều hệ quản trị CSDL trên thị trường như: Visual Foxpro, SQL server, DB2, Microsoft Access, Oracle...

Một hệ quản trị CSDL phải có:

- Ngôn ngữ giao tiếp giữa NSD và CSDL:
 - ✓ Ngôn ngữ mô tả dữ liệu (Data Definition Language - DDL) để cho phép khai báo cấu trúc của CSDL, khai báo các mối liên hệ của dữ liệu (Data Relationship) và các quy tắc (Rules, Constraint) quản lý áp đặt trên dữ liệu.
 - ✓ Ngôn ngữ thao tác dữ liệu (Data Manipulation Language - DML) cho phép NSD có thể thêm, xóa, dữ liệu trong CSDL.
 - ✓ Ngôn ngữ truy vấn dữ liệu (Structured Query Language - SQL) cho phép NSD truy vấn các thông tin cần thiết.
 - ✓ Ngôn ngữ quản lý dữ liệu (Data Control Language - DCL) cho phép những người quản trị hệ thống thay đổi cấu trúc của các bảng dữ liệu, khai báo bảo mật thông tin và cấp quyền khai thác CSDL cho NSD.
- Tự điển dữ liệu (Data dictionary) dùng để mô tả các ánh xạ liên kết, ghi nhận các thành phần cấu trúc của CSDL, các chương trình ứng dụng, mật mã, quyền sử dụng...
- Có biện pháp bảo mật tốt.
- Có cơ chế giải quyết tranh chấp dữ liệu.
- Có cơ chế sao lưu và phục hồi dữ liệu khi có sự cố xảy ra.
- Có giao diện tốt, dễ sử dụng.
- Bảo đảm tính độc lập giữa dữ liệu và chương trình: khi có sự thay đổi dữ liệu (như sửa đổi cấu trúc lưu trữ các bảng dữ liệu, thêm các chỉ mục,...) thì các chương trình ứng dụng đang chạy trên CSDL vẫn không cần phải viết lại hay cũng không ảnh hưởng đến NSD khác.

1.5. Các mức biểu diễn một CSDL

Để hệ thống có thể dùng được thì phải tìm dữ liệu một cách hiệu quả. Điều này dẫn đến việc cần phải thiết kế những cấu trúc dữ liệu phức tạp đối với việc biểu diễn dữ liệu trong

cơ sở dữ liệu. Vì nhiều người sử dụng các hệ thống cơ sở dữ liệu không phải là người hiểu biết nhiều về máy tính nên những người phát triển hệ thống phải che dấu đi sự phức tạp khỏi người sử dụng thông qua một số mức trừu tượng, nhằm làm đơn giản hóa sự tương tác của người sử dụng đối với hệ thống.

Có 3 mức biểu diễn dữ liệu, cụ thể như sau:

Mức trong (mức vật lý - Physical level)

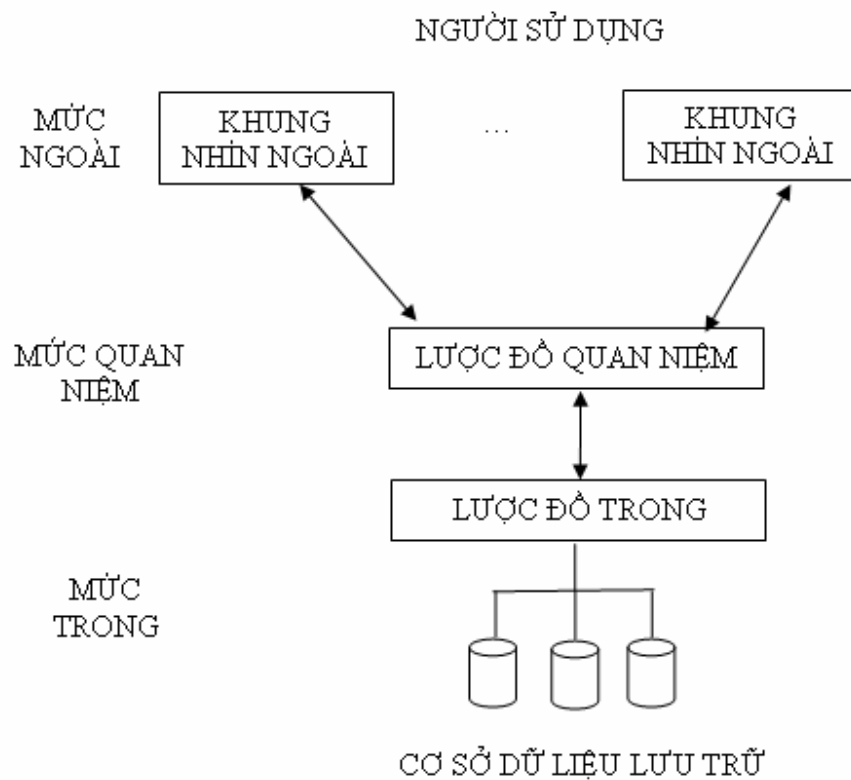
- Nói đến cách thức dữ liệu được lưu trữ thực sự. Trả lời câu hỏi dữ liệu gì và lưu trữ chúng như thế nào? Cần các chỉ mục gì?
- Ví dụ như chỉ mục, B-cây, băm.
- Những người thao tác tại mức này là người quản trị CSDL và những NSD chuyên môn.

Mức quan niệm hay mức logic (conception level, logical level)

- Trả lời câu hỏi cần phải lưu trữ bao nhiêu loại dữ liệu? Đó là những dữ liệu gì? Mối quan hệ giữa chúng như thế nào?
- Như vậy CSDL mức vật lý là sự cài đặt cụ thể của CSDL mức quan niệm.

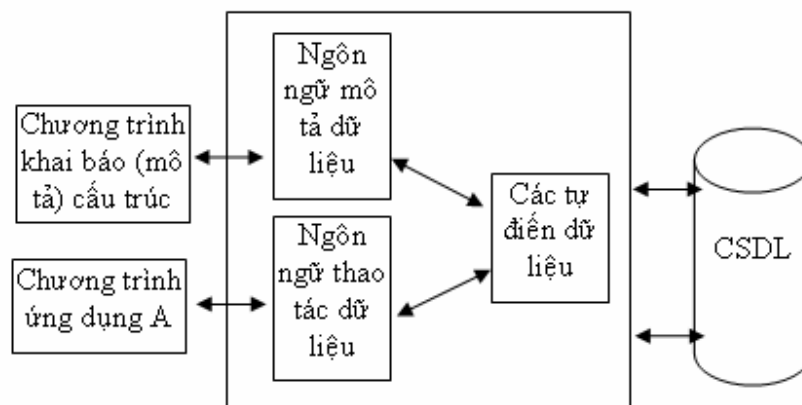
Mức ngoài hay mức nhìn (view level)

- Là mức của NSD và các chương trình ứng dụng.
- Làm việc tại mức này là các nhà chuyên môn, kỹ sư tin học, NSD không chuyên.
- Mỗi NSD hay mỗi chương trình ứng dụng có thể được nhìn CSDL theo một góc độ khác nhau. Tùy thuộc vào vai trò của NSD mà có thể thấy tất cả hoặc một phần của CSDL. NSD hay chương trình ứng dụng có thể không được biết về cấu trúc tổ chức lưu trữ thông tin trong CSDL. Họ chỉ có thể làm việc trên một phần CSDL theo cách nhìn do người quản trị hay chương trình ứng dụng quy định, gọi là khung nhìn.



Hình 1.1. Ba mức trừu tượng dữ liệu

1.6. Sơ đồ tổng quát một hệ quản trị CSDL



Hình 1.2. Sơ đồ tổng quát của một hệ quản trị CSDL

Hình 1.2 minh họa sơ đồ tổng quát của một hệ quản trị CSDL. Có 3 mức: mức chương trình khai báo cấu trúc và chương trình ứng dụng; mức mô tả CSDL, thao tác CSDL và các tự điển dữ liệu; mức CSDL.

Mỗi hệ quản trị CSDL có một ngôn ngữ khai báo hay mô tả (Data Definition Language - DDL) cấu trúc CSDL riêng. Những người thiết kế và quản trị CSDL thực hiện các công việc khai báo cấu trúc CSDL.

Các chương trình khai báo cấu trúc CSDL được viết bằng ngôn ngữ mà hệ quản trị CSDL cho phép. Hai công việc khai báo cấu trúc logic (là việc khai báo các loại dữ liệu và các mối liên hệ giữa các loại dữ liệu đó, cùng các ràng buộc toàn vẹn dữ liệu) và khai báo vật lý (dữ liệu được lưu trữ theo dạng nào, có bao nhiêu chỉ mục)

Các chương trình ứng dụng được viết bằng ngôn ngữ thao tác CSDL (Data Manipulation Language - DML) với mục đích:

- Truy xuất dữ liệu
- Cập nhật dữ liệu (thêm, xóa, sửa)
- Khai thác dữ liệu
- Ngôn ngữ thao tác CSDL còn được sử dụng cho những người sử dụng thao tác trực tiếp với CSDL.

Tự điển dữ liệu (Data Dictionary - DD) là một CSDL của hệ quản trị CSDL sử dụng để lưu trữ cấu trúc CSDL, các thông tin bảo mật, bảo đảm an toàn dữ liệu và các cấu trúc ngoài. Tự điển dữ liệu còn được gọi là siêu CSDL (Meta-Database)

1.7. Tính độc lập giữa dữ liệu và chương trình

- **Độc lập dữ liệu vật lý**
 - ✓ Khả năng cập nhật lược đồ vật lý không làm cho các chương trình ứng dụng phải bị viết lại.
 - ✓ Việc cập nhật ở mức này thường cần thiết để cải tiến hiệu suất.
- **Độc lập dữ liệu logic**
 - ✓ Khả năng cập nhật lược đồ logic mà không làm cho các chương trình ứng dụng phải bị viết lại.
 - ✓ Những cập nhật thường được thực hiện khi cấu trúc logic của cơ sở dữ liệu thay đổi.

2. Các cách tiếp cận của một CSDL

Nền tảng của cấu trúc cơ sở dữ liệu là mô hình dữ liệu. Mô hình dữ liệu được định nghĩa là một sưu tập các công cụ khái niệm dùng cho việc mô tả dữ liệu, các mối quan hệ dữ liệu, các ngữ nghĩa dữ liệu và các ràng buộc dữ liệu.

2.1. Cách tiếp cận theo mô hình dữ liệu mạng

Mô hình dữ liệu mạng (Network Data Model) hay gọi tắt là mô hình mạng (Network Model) là mô hình được biểu diễn bởi một đồ thị có hướng. Mô hình này được xây dựng bởi Honeywell vào năm 1964-1965. Trong mô hình này, có 3 khái niệm được sử dụng: mẫu tin hay bản ghi (record), loại mẫu tin (record type) và loại liên hệ (set type).

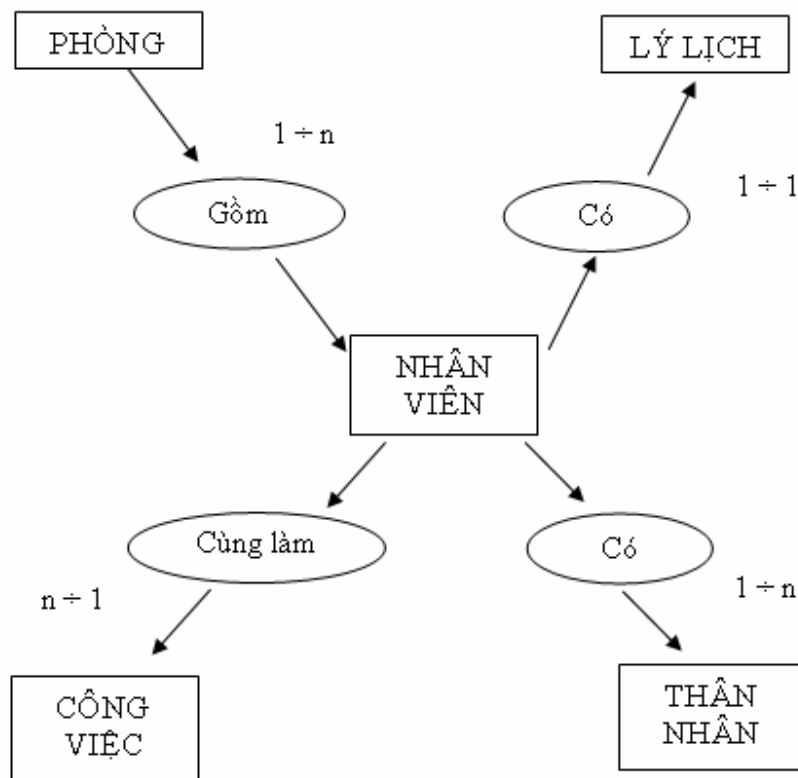
Loại mẫu tin: đặc trưng cho một loại đối tượng riêng biệt. Chẳng hạn như trong quản lý nhân sự tại một đơn vị thì đối tượng cần phản ánh của thế giới thực có thể là Phòng, nhân viên, công việc, lý lịch... do đó có các loại mẫu tin đặc trưng cho từng đối tượng này. Trong mô hình mạng, mỗi loại mẫu tin được thể hiện bằng một hình chữ nhật, một thể hiện (instance) của một loại mẫu tin được gọi là bản ghi (record). Ở ví dụ trên loại mẫu tin Phòng có các mẫu tin là các phòng ban trong công ty, loại mẫu tin Nhân viên có các mẫu tin là các nhân viên đang làm việc tại các phòng ban của công ty...

Loại liên hệ: là sự liên kết giữa một loại mẫu tin chủ với một mẫu tin thành viên. Trong mô hình mạng, một loại liên hệ được thể hiện bằng hình bầu dục và sự liên kết giữa hai loại mẫu tin được thể hiện bằng các cung có hướng đi từ loại mẫu tin chủ đến loại liên hệ và từ loại liên hệ đến loại mẫu tin thành viên.

Trong loại liên kết người ta còn chỉ ra số lượng các mẫu tin tham gia vào trong mỗi kết hợp. Có một số loại sau:

- $1 \div 1$ (One to One): Mỗi mẫu tin của loại mẫu tin chủ kết hợp với đúng 1 mẫu tin của loại mẫu tin thành viên. Ví dụ, mỗi nhân viên có duy nhất một lý lịch cá nhân.
- $1 \div n$ (One to Many): Mỗi mẫu tin của loại mẫu tin chủ kết hợp với 1 hay nhiều mẫu tin của loại mẫu tin thành viên. Ví dụ, mỗi phòng ban có thể có một hoặc nhiều nhân viên.
- $n \div 1$ (Many to One): Nhiều mẫu tin của loại mẫu tin chủ kết hợp với đúng 1 mẫu tin của loại mẫu tin thành viên. Ví dụ, nhiều nhân viên cùng làm một công việc.
- **Đệ quy (Recursive):** Một loại mẫu tin chủ cũng có thể đồng thời là mẫu tin thành viên với chính nó. Ví dụ, trưởng phòng cũng là một nhân viên thuộc phòng trong công ty.

Hình 1.4 minh họa một mô hình dữ liệu mạng.



Hình 1.4. Mô hình dữ liệu mạng

Ưu điểm: đơn giản, dễ sử dụng.

Hạn chế: không thích hợp trong việc biểu diễn CSDL lớn do hạn chế về khả năng diễn đạt ngữ nghĩa của dữ liệu, đặc biệt là các dữ liệu và mối liên hệ phức tạp của dữ liệu trong thực tế là rất hạn chế.

2.2. Mô hình dữ liệu phân cấp

Mô hình dữ liệu phân cấp (Hierarchical Data Model), còn gọi là mô hình phân cấp (Hierarchical Model), được thực hiện thông qua sự kết hợp giữa IBM và North American Rockwell vào khoảng năm 1965. Mô hình là một cây, trong đó mỗi nút của cây biểu diễn một thực thể, giữa nút con với nút cha được liên hệ với nhau theo một mối quan hệ xác định.

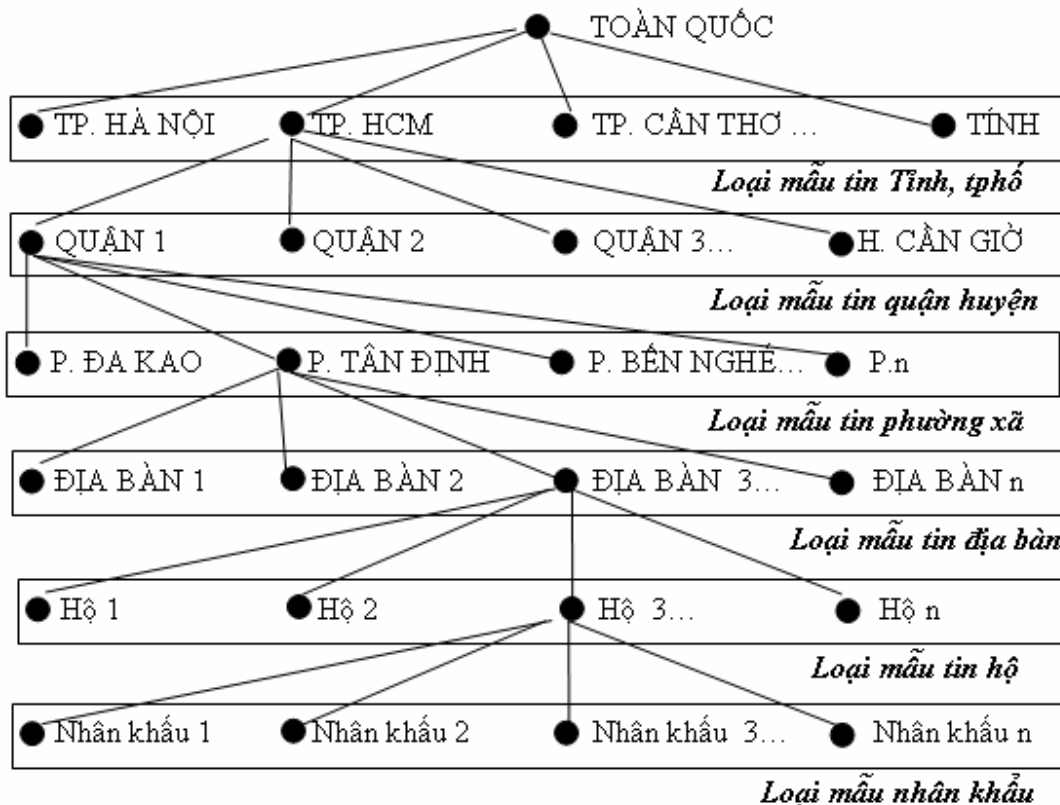
Loại mẫu tin: giống trong mô hình dữ liệu mạng

Loại mối liên hệ: kiểu liên hệ là phân cấp:

Mẫu tin thành viên chỉ đóng vai trò thành viên của một mối liên hệ duy nhất, tức là nó thuộc một chủ duy nhất. Như vậy, mối liên hệ từ mẫu tin chủ tới các mẫu tin thành viên là $1 \div n$ và từ mẫu tin thành viên với mẫu tin chủ là $1 \div 1$

Giữa hai loại mẫu tin chỉ tồn tại một mối quan hệ duy nhất.

Hình 1.5 minh họa mô hình dữ liệu phân cấp một CSDL về điều tra dân số.



Hình 1.5. Mô hình dữ liệu phân cấp

2.3. Mô hình dữ liệu quan hệ

Mô hình dữ liệu quan hệ (Relational Data Model) còn được gọi là mô hình quan hệ (Relational Model) do E.F.Codd đề xuất năm 1970. Nền tảng cơ bản là khái niệm lý thuyết tập hợp trên các quan hệ, tức là tập của các bộ giá trị (value tuples). Trong mô hình dữ liệu này những khái niệm được sử dụng là thuộc tính (attribute), quan hệ (relation), lược đồ quan hệ (relation schema), bộ (tuple), khóa (key).

Mô hình quan hệ là mô hình được nghiên cứu nhiều nhất, và có cơ sở lý thuyết vững chắc nhất. Mô hình quan hệ cùng với mô hình dữ liệu thực thể kết hợp đang được sử dụng rộng rãi nhất hiện nay trong việc phân tích và thiết kế CSDL. Chúng ta sẽ nghiên cứu chi tiết về mô hình này trong chương sau.

2.4. Mô hình dữ liệu thực thể - kết hợp

Mô hình dữ liệu thực thể - kết hợp (Entity – Relationship Model) do Peter Pin-Shan Chen đề xuất năm 1976. Mô hình quan hệ-thực thể được dựa trên sự nhận thức thế giới gồm có

một sưu tập các đối tượng căn bản, được gọi là các tập thực thể, và các mối quan hệ ở giữa các đối tượng này. Trong mô hình này các khái niệm được sử dụng là tập thực thể (entity set), thực thể (entity), thuộc tính của loại thực thể (entity attribute), khóa của loại thực thể (entity key), loại mối kết hợp (entity relationship), số ngôi của mối kết hợp (relationship degree), thuộc tính của mối kết hợp (relationship attribute), bản số của mối kết hợp (relationship cardinal).

Chi tiết của mô hình này sẽ được nghiên cứu trong chương sau.

2.5. Mô hình dữ liệu hướng đối tượng

Mô hình dữ liệu hướng đối tượng (Object Oriented Data Model) ra đời vào cuối những năm 80 và đầu những năm 90. Đây là loại mô hình tiên tiến nhất hiện nay dựa trên cách tiếp cận hướng đối tượng. Mô hình này sử dụng các khái niệm như lớp (class), sự kế thừa (inheritance), kế thừa bội (multi - inheritance). Đặc trưng cơ bản của cách tiếp cận này là tính đóng gói (encapsulation), tính đa hình (polymorphism) và tính tái sử dụng (reusability).

3. Bài tập

Bài 1:

Dựa vào những khái niệm đã học, hãy biểu diễn CSDL có các loại mẫu tin Phòng, Nhân viên, Công việc, lý lịch đã trình bày trong mô hình mạng theo cách tiếp cận phân cấp. Biết:

- Loại liên hệ là phân cấp.
- Phòng có nhiều nhân viên, mỗi nhân viên chỉ thuộc một phòng duy nhất
- Công việc có nhiều nhân viên cùng làm; mỗi nhân viên chỉ làm một công việc duy nhất
- Mỗi nhân viên có một lý lịch, mỗi lý lịch chỉ thuộc duy nhất một nhân viên.

Bài 2:

Dựa vào những khái niệm đã học, hãy biểu diễn CSDL về tổng điều tra dân số toàn quốc có các loại mẫu tin tỉnh – thành phố, quận huyện, phường xã, địa bàn, hộ điều tra và nhân khẩu đã trình bày trong mô hình phân cấp theo cách tiếp cận mạng. Biết:

- Nhân khẩu thuộc một hộ điều tra
- Hộ điều tra thuộc một địa bàn
- Địa bàn điều tra thuộc một phường xã

- Phường xã thuộc một quận huyện
 - Quận huyện thuộc một tỉnh, thành phố
-

TaiLieu.vn