



SỞ GIÁO DỤC VÀ ĐÀO TẠO HÀ NỘI

GIÁO TRÌNH

CƠ SỞ

dữ liệu quan hệ

DÙNG TRONG CÁC TRƯỜNG
TRUNG HỌC CHUYÊN NGHIỆP

công. com



NHÀ XUẤT BẢN HÀ NỘI

SỞ GIÁO DỤC VÀ ĐÀO TẠO HÀ NỘI

PHẠM ĐỨC NHIỆM

**GIÁO TRÌNH
cơ sở dữ liệu quan hệ**

(Dùng trong các trường THCN)

NHÀ XUẤT BẢN HÀ NỘI - 2005

Lời giới thiệu

Nước ta đang bước vào thời kỳ công nghiệp hóa, hiện đại hóa nhằm đưa Việt Nam trở thành nước công nghiệp văn minh, hiện đại.

Trong sự nghiệp cách mạng to lớn đó, công tác đào tạo nhân lực luôn giữ vai trò quan trọng. Báo cáo Chính trị của Ban Chấp hành Trung ương Đảng Cộng sản Việt Nam tại Đại hội Đảng toàn quốc lần thứ IX đã chỉ rõ: “Phát triển giáo dục và đào tạo là một trong những động lực quan trọng thúc đẩy sự nghiệp công nghiệp hóa, hiện đại hóa, là điều kiện để phát triển nguồn lực con người - yếu tố cơ bản để phát triển xã hội, tăng trưởng kinh tế nhanh và bền vững”.

Quán triệt chủ trương, Nghị quyết của Đảng và Nhà nước và nhận thức đúng đắn về tầm quan trọng của chương trình, giáo trình đối với việc nâng cao chất lượng đào tạo, theo đề nghị của Sở Giáo dục và Đào tạo Hà Nội, ngày 23/9/2003, Ủy ban nhân dân thành phố Hà Nội đã ra Quyết định số 5620/QĐ-UB cho phép Sở Giáo dục và Đào tạo thực hiện đề án biên soạn chương trình, giáo trình trong các trường Trung học chuyên nghiệp (THCN) Hà Nội. Quyết định này thể hiện sự quan tâm sâu sắc của Thành ủy, UBND thành phố trong việc nâng cao chất lượng đào tạo và phát triển nguồn nhân lực Thủ đô.

Trên cơ sở chương trình khung của Bộ Giáo dục và Đào tạo ban hành và những kinh nghiệm rút ra từ thực tế đào tạo, Sở Giáo dục và Đào tạo đã chỉ đạo các trường THCN tổ chức biên soạn chương trình, giáo trình một cách khoa học, hệ

thống và cập nhật những kiến thức thực tiễn phù hợp với đối tượng học sinh THCN Hà Nội.

Bộ giáo trình này là tài liệu giảng dạy và học tập trong các trường THCN ở Hà Nội, đồng thời là tài liệu tham khảo hữu ích cho các trường có đào tạo các ngành kỹ thuật - nghiệp vụ và động đảo bạn đọc quan tâm đến vấn đề hướng nghiệp, dạy nghề.

Việc tổ chức biên soạn bộ chương trình, giáo trình này là một trong nhiều hoạt động thiết thực của ngành giáo dục và đào tạo Thủ đô để kỷ niệm “50 năm giải phóng Thủ đô”, “50 năm thành lập ngành” và hướng tới kỷ niệm “1000 năm Thăng Long - Hà Nội”.

Sở Giáo dục và Đào tạo Hà Nội chân thành cảm ơn Thành ủy, UBND, các sở, ban, ngành của Thành phố, Vụ Giáo dục chuyên nghiệp Bộ Giáo dục và Đào tạo, các nhà khoa học, các chuyên gia đầu ngành, các giảng viên, các nhà quản lý, các nhà doanh nghiệp đã tạo điều kiện giúp đỡ, đóng góp ý kiến, tham gia Hội đồng phản biện, Hội đồng thẩm định và Hội đồng nghiệm thu các chương trình, giáo trình.

Đây là lần đầu tiên Sở Giáo dục và Đào tạo Hà Nội tổ chức biên soạn chương trình, giáo trình. Dù đã hết sức cố gắng nhưng chắc chắn không tránh khỏi thiếu sót, bất cập. Chúng tôi mong nhận được những ý kiến đóng góp của bạn đọc để từng bước hoàn thiện bộ giáo trình trong các lần tái bản sau.

GIÁM ĐỐC SỞ GIÁO DỤC VÀ ĐÀO TẠO

Lời nói đầu

Trong thực tế việc, ứng dụng công nghệ thông tin (CNTT) hiện nay để quản lý các thông tin của các đối tượng cần quan tâm là một đòi hỏi cấp thiết. Việc xây dựng các hệ thống quản lý các thông tin trên máy tính được sử dụng rộng rãi trên nhiều lĩnh vực như: kinh tế, xã hội, quốc phòng, an ninh, v.v. Một trong những vấn đề quan trọng trong quản lý thông tin là xây dựng cơ sở dữ liệu sao cho các phép toán xử lý trên chúng có hiệu quả cao nhất. Chính vậy, cơ sở dữ liệu (CSDL- Database) là một lĩnh vực phát triển mạnh của công nghệ thông tin. Cùng với sự phát triển của CNTT ở nước ta, việc sử dụng các kiến thức về cơ sở dữ liệu (CSDL) ngày càng trở nên cấp thiết. Ở Việt Nam, các tài liệu về CSDL bằng tiếng Việt còn rất ít ỏi. Hiện nay chúng ta chưa có một tài liệu chung dùng làm giáo trình chuẩn, đầy đủ, cơ bản để giảng dạy cho sinh viên, học sinh. Trong giáo trình này, tôi muốn trình bày một số kiến thức cơ bản nhất về mô hình CSDL quan hệ giúp cho học sinh có thể ứng dụng các kiến thức về CSDL vào thực tiễn, tiếp tục nghiên cứu sâu về lý thuyết CSDL cũng như các môn tin học khác, đồng thời làm tài liệu tham khảo cho các bạn đọc trong công việc nghiên cứu, giảng dạy và học tập.

Giáo trình được chia làm 4 chương:

Chương 1: Sơ lược về cơ sở dữ liệu

Nội dung chương 1 là một số khái niệm chung về cơ sở dữ liệu, hệ quản trị cơ sở dữ liệu, sơ lược về một số mô hình và có trình bày một số kiến thức quan trọng về mô hình thực thể - liên hệ, làm nền tảng cho việc thiết kế các lược đồ cơ sở dữ liệu khái niệm trong các mô hình khác.

Chương 2: Mô hình cơ sở dữ liệu quan hệ

Chương này đi sâu vào trình bày các khái niệm quan trọng cũng như một

số phép toán cơ bản của mô hình CSDL quan hệ. Chương 2 còn đề cập đến cách chuyển sơ đồ thực thể - liên hệ về các lược đồ CSDL quan hệ cùng với các kiến thức nhằm giúp đơn giản hóa và tinh chỉnh các lược đồ này.

Chương 3: Ngôn ngữ con dữ liệu SQL

Chương 3 trình bày SQL như một ngôn ngữ đầy đủ, vừa là ngôn ngữ định nghĩa dữ liệu (DDL), vừa là ngôn ngữ thao tác dữ liệu (DML). Các ví dụ trong chương này là các ví dụ đã chạy thật bằng các câu lệnh SQL chuẩn ISO trên môi trường VISUAL FOXPRO.

Chương 4: Lý thuyết thiết kế cơ sở dữ liệu quan hệ

Trọng tâm của chương này là các kiến thức về phụ thuộc hàm, áp dụng các kiến thức này vào việc tìm khoá và chuẩn hoá các lược đồ CSDL quan hệ.

Nhân đây tôi xin chân thành cảm ơn các thầy và các cán bộ nghiên cứu khoa học PGS.TS. Nguyễn Thiện Luận; TS. Đỗ Năng Toàn; ThS. Lê Văn Năng; ThS. Nguyễn Duy Phương đã đóng góp những ý kiến quý báu cho giáo trình.

Mặc dù đã cố gắng tham khảo các tài liệu trong nước, ngoài nước cũng như nhiều ý kiến tham gia của các thầy, các bạn đồng nghiệp đã dạy và nghiên cứu môn CSDL, song cuốn giáo trình vẫn không tránh khỏi những thiếu sót. Rất mong nhận được sự đóng góp ý kiến của bạn đọc và các em học sinh.

TÁC GIẢ

Chương 1

SƠ LƯỢC VỀ CƠ SỞ DỮ LIỆU

Mục tiêu:

Học sinh hiểu được các khái niệm chung về cơ sở dữ liệu, một số mô hình cơ sở dữ liệu và các kiến thức quan trọng về mô hình thực thể - liên hệ, làm nền tảng cho việc thiết kế các lược đồ cơ sở dữ liệu khái niệm trong các mô hình khác.

Nội dung tóm tắt:

Nội dung chương 1 trình bày một số khái niệm cơ sở dữ liệu, hệ quản trị cơ sở dữ liệu, sơ lược về một số mô hình cơ sở dữ liệu và trình bày các kiến thức quan trọng đối với mô hình thực thể - liên hệ.

NỘI DUNG

I. CÁC KHÁI NIỆM CƠ BẢN

1. Cơ sở dữ liệu

Cơ sở dữ liệu (Database, viết tắt là CSDL) là một lĩnh vực nghiên cứu các mô hình, nguyên lý, phương pháp tổ chức dữ liệu trên các vật mang tin.

Để dễ dàng cho việc giải thích các khái niệm, trước hết ta thử 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 đảm bảo cho hành khách đi đúng chuyến? Dữ liệu nêu trên được lưu trữ trong máy theo một quy định nào đó và được gọi là cơ sở dữ liệu.

Như vậy, CSDL là tập hợp các thông tin có quan hệ với nhau được lưu trữ

trong máy tính theo một quy định nhất định nhằm phục vụ cho một mục đích nào đó.

2. Hệ quản trị cơ sở dữ liệu (Database Management System - HQTCSQL)

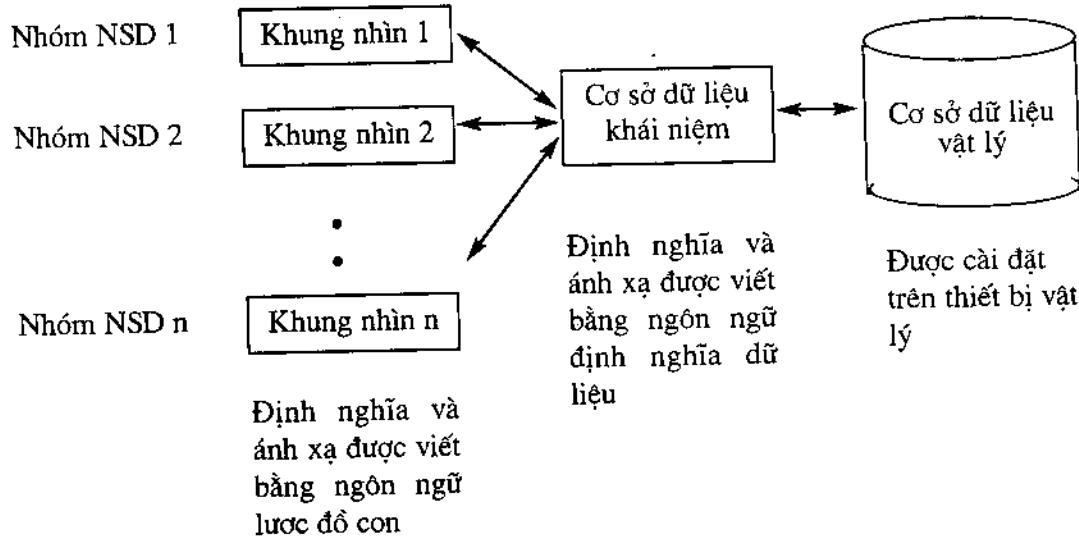
- Hệ quản trị cơ sở dữ liệu là hệ thống các chương trình nhằm tạo lập, xử lý, thay đổi, quản lý và khai thác các CSDL.
- Một số HQTCSQL thông dụng hiện nay: Foxpro, Access, Oracle,... với các phiên bản khác nhau.
- Có hai đặc điểm để phân biệt một HQTCSQL với các hệ thống lập trình khác:
 - (1) **Khả năng quản lý những dữ liệu cố định.**
 - (2) **Khả năng truy xuất có hiệu quả một số lượng lớn dữ liệu.**

Điểm (1) khẳng định rằng có một CSDL tồn tại thường xuyên và nội dung của CSDL này là những dữ liệu được HQTCSQL truy xuất và quản lý. Điểm (2) phân biệt một HQTCSQL với các hệ thống quản lý tập tin cùng quản lý dữ liệu cố định nhưng nói chung không cho phép truy xuất nhanh chóng các thành phần tùy ý của dữ liệu. Ngoài ra, còn có một số khả năng khác thường gặp trong các HQTCSQL trên thị trường:

- + HQTCSQL hỗ trợ ít nhất một mô hình dữ liệu, nhờ đó người sử dụng có thể xem được dữ liệu.
- + HQTCSQL hỗ trợ một số ngôn ngữ bậc cao cho phép người sử dụng định nghĩa các cấu trúc dữ liệu, truy xuất dữ liệu và thao tác dữ liệu.
- + HQTCSQL quản lý các giao dịch, nghĩa là cho phép nhiều người sử dụng truy xuất đồng thời và chính xác đến một CSDL.
- + HQTCSQL điều khiển quá trình truy xuất, là khả năng giới hạn các quá trình truy xuất dữ liệu của những người không được phép và khả năng kiểm tra độ tin cậy của dữ liệu.
- + HQTCSQL có khả năng tự thích ứng là khả năng phục hồi lại dữ liệu do sự cố của hệ thống mà làm mất dữ liệu.

3. Kiến trúc một hệ cơ sở dữ liệu

Một CSDL được phân thành 3 mức trừu tượng khác nhau: Mức CSDL vật lý, mức CSDL khái niệm và mức khung nhìn.



Hình 1.1

* Mức CSDL vật lý

Một tập hợp các tệp dữ liệu, các chỉ mục hoặc những cấu trúc lưu trữ khác dùng để truy xuất dữ liệu một cách có hiệu quả gọi là CSDL vật lý. CSDL vật lý tồn tại thường xuyên trong thiết bị lưu trữ như đĩa từ, nhiều CSDL có thể được quản lý bởi cùng một HQTCSDL.

* Mức CSDL khái niệm

CSDL khái niệm là một sự trừu tượng hoá của thế giới thực khi nó được gắn với người sử dụng. HQTCSDL cung cấp ngôn ngữ định nghĩa dữ liệu (Data Definition Language viết tắt là DDL) để mô tả lược đồ khái niệm (Conceptual Scheme) và nó sẽ được cài đặt bằng lược đồ vật lý (Physical Scheme). DDL cho phép ta mô tả CSDL khái niệm nhờ các thuật ngữ của “mô hình dữ liệu”. Ví dụ: Chẳng hạn trong mô hình quan hệ, dữ liệu được trình bày qua các bảng với các cột là các thuộc tính, các hàng là các bộ.

CSDL vật lý là sự cài đặt cụ thể của CSDL mức khái niệm.

* Mức CSDL khung nhìn

Khung nhìn (view) hay lược đồ con (SubScheme) là một phần của CSDL khái niệm hoặc là sự trùu tượng hoá một phần CSDL khái niệm.

Hay nói một cách khác, View là cách nhìn của người sử dụng đối với CSDL, là một đối tượng tương ứng được xây dựng từ CSDL khái niệm nhưng thực sự không tồn tại trong CSDL đó.

Hầu hết các HQTCSDL đều cung cấp những phương tiện để khai báo khung nhìn được gọi là ngôn ngữ định nghĩa dữ liệu lược đồ con (SubScheme Data Definition Language) và các phương tiện để diễn đạt các câu vấn tin và thao tác trên khung nhìn gọi là ngôn ngữ thao tác dữ liệu lược đồ con (SubScheme Data Manipulation Language). Theo một nghĩa nào đó, việc xây dựng khung nhìn ngược lại với quá trình tích hợp CSDL. Đối với một tập hợp dữ liệu tham gia vào CSDL khái niệm, ta có thể xây dựng một khung nhìn chứa dữ liệu này. Các khung nhìn có vai trò quan trọng trong việc bảo mật hệ thống CSDL, chỉ những người có nhu cầu và quyền hạn mới có thể đọc được các tệp dữ liệu dành cho họ.

Ví dụ 1.1: Để phân tích sự khác nhau giữa các mức vật lý, mức khái niệm và mức khung nhìn nhờ tính tương tự của chúng đối với các ngôn ngữ lập trình cụ thể. Hãy xét một mảng hai chiều có kích thước $m \times n$.

Ở mức khái niệm có thể khai báo mảng như sau:

A: array[1...m,1...n] of Integer;

Ở mức vật lý mảng A được lưu trữ trong một vùng nhớ liên tục nhờ quy tắc: $A[i, j]$ sẽ ở vị trí $a_0 + 2(n(i-1) + j-1)$.

Một khung nhìn của mảng A có thể tạo ra bằng cách khai báo một hàm $f(i)$ là tổng của $A[i, j]$ với ($i = 1..m$, $j = 1..n$). Trong khung nhìn này không chỉ nhìn thấy A vừa có liên hệ vừa tách biệt, là hàm chứ không phải là mảng, mà còn che dấu được một số thông tin, bởi vì chỉ có thể thấy được tổng của các hàng mà không phải bản thân các hàng này.

Ví dụ 1.2: Có thể trùu tượng hoá các sinh viên qua các thuộc tính: MaSV, TenSV, NgaySinh, Gioi, Quequan. Điểm của các sinh viên và mối liên hệ giữa các sinh viên và điểm của họ qua các thuộc tính: MaSV, Mon1, Mon2, Mon3. Chẳng hạn dùng HQTCSDL Visual Foxpro thì: