

**TRƯỜNG ĐẠI HỌC HÀNG HẢI
KHOA CÔNG NGHỆ THÔNG TIN**



BÀI GIẢNG CƠ SỞ DỮ LIỆU NÂNG CAO

Biên soạn: Th.S Nguyễn Trung Đức

**BỘ GIAO THÔNG VẬN TẢI
TRƯỜNG ĐẠI HỌC HÀNG HẢI
BỘ MÔN: HỆ THỐNG THÔNG TIN
KHOA: CÔNG NGHỆ THÔNG TIN**

BÀI GIẢNG CƠ SỞ DỮ LIỆU NÂNG CAO

**TÊN HỌC PHẦN : CƠ SỞ DỮ LIỆU NÂNG CAO
MÃ HỌC PHẦN : 17406
TRÌNH ĐỘ ĐÀO TẠO : ĐẠI HỌC CHÍNH QUY
DÙNG CHO SV NGÀNH : CÔNG NGHỆ THÔNG TIN**

HẢI PHÒNG - 2008

Tên học phần: Cơ sở dữ liệu nâng cao
Bộ môn phụ trách giảng dạy: Hệ thống Thông tin
Mã học phần: 17406

Loại học phần: 2
Khoa phụ trách: CNTT.
Tổng số TC: 2

TS tiết	Lý thuyết	Thực hành/Xemina	Tự học	Bài tập lớn	Đồ án môn học
45	30	15	0	0	0

Điều kiện tiên quyết:

Sinh viên phải học và thi đạt các học phần sau mới được đăng ký học học phần này:
 Cơ sở dữ liệu.

Mục tiêu của học phần:

Cung cấp kiến thức nâng cao về mô hình cơ sở dữ liệu quan hệ, các phương pháp thiết kế cơ sở dữ liệu khác nhau.

Nội dung chủ yếu:

Giới thiệu về các hệ quản trị dữ liệu; Các mô hình mạng, mô hình phân cấp; Điều khiển tương tranh trong hệ quản trị cơ sở dữ liệu; An toàn và xử lý sai sót; Mô hình cơ sở dữ liệu hướng đối tượng; Mô hình cơ sở dữ liệu phân tán.

Nội dung chi tiết:

TÊN CHƯƠNG MỤC	PHÂN PHỐI SỐ TIẾT				
	TS	LT	TH	BT	KT
Chương 1. Hệ quản trị cơ sở dữ liệu và lịch sử phát triển	3	3			
1.1. Quan niệm về CSDL					
1.2. Ứng dụng của hệ quản trị CSDL					
1.3. Lịch sử của các hệ quản trị dữ liệu					
Chương 2. Mô hình mạng, mô hình phân cấp	3	3			
2.1. Mô hình mạng					
2.1.1. Giới thiệu					
2.1.2. Các khái niệm					
2.2. Mô hình phân cấp					
2.2.1. Giới thiệu					
2.2.2. Các khái niệm					
Chương 3. Thiết kế cơ sở dữ liệu khái niệm	3	3			
3.1. Giới thiệu					
3.2. Trừu tượng hoá trong thiết kế CSDL					
3.3. Các thuộc tính tương xứng giữa các lớp					
3.4. Các mô hình dữ liệu					
3.5. Mô hình thực thể quan hệ					
Chương 4. Điều khiển khai thác tương tranh	3	3			
4.1. Giới thiệu					
4.2. Một số khái niệm					
4.3. Đặc tính của khai thác không xung đột					
4.3.1. Một số khái niệm					
4.3.2. Khai thác có thứ tự					
4.3.3. Đồ thị về thứ tự thực hiện các giao tác					
Chương 5. An toàn dữ liệu và xử lý sai sót	3	3			
5.1. An toàn trong CSDL					
5.1.1. Phạm vi an toàn dữ liệu					
5.1.2. Các yêu cầu bảo vệ CSDL					
5.1.3. Các dạng sai sót					
5.2. Các điều khiển an toàn dữ liệu					
5.2.1. Điều khiển luồng thông tin					

TÊN CHƯƠNG MỤC	PHÂN PHỐI SỐ TIẾT				
	TS	LT	TH	BT	KT
5.2.2. Điều khiển suy diễn					
5.2.3. Điều khiển truy nhập					
Chương 6. Đánh giá câu hỏi	3	2			1
6.1. Phân tích câu hỏi					
6.1.1. Phân tích cú pháp					
6.1.2. Phân tích ngữ nghĩa					
6.2. Cấu trúc đại số câu hỏi					
6.2.1. Cây đại số quan hệ					
6.2.2. Các luật biến đổi cây đại số quan hệ					
6.2.3. Các bước tối ưu cây đại số quan hệ					
6.3. Phân rã câu hỏi					
Chương 7. Cơ sở dữ liệu hướng đối tượng	12	5	6		1
7.1. Giới thiệu về hướng đối tượng					
7.1.1. Các khái niệm hướng đối tượng					
7.1.2. Mô hình hoá việc phân tích hướng đối tượng					
7.1.3. Mô hình hóa dữ liệu					
7.2. Nguyên tắc của các mô hình hướng đối tượng					
7.2.1. Mô hình hoá các đối tượng					
7.2.2. Phương pháp					
7.2.3. Xác định dạng dữ liệu					
7.2.4. Các liên kết thừa kế giữa các lớp					
7.2.5. Đa cầu và sự áp đặt					
7.2.6. Xác định tập các đối tượng					
7.2.7. Khía cạnh động					
7.2.8. Lược đồ CSDL hướng đối tượng					
7.3. Tính bền vững các đối tượng					
7.3.1. CSDL hướng đối tượng					
7.3.2. Quản lý tính bền vững					
7.3.3. Kế thừa tính bền vững					
7.3.4. Tính bền vững do tham chiếu					
7.3.5. Tích hợp với ngôn ngữ lập trình					
7.4. Đại số với các đối tượng phức tạp					
7.4.1. Mở rộng đại số quan hệ theo đường dẫn và các phương pháp					
7.4.2. Các phép toán đại số					
7.4.3. Các phép toán nhóm					
7.4.4. Đồ thị các phép toán					
Chương 8. Cơ sở dữ liệu phân tán	15	5	9		1
8.1. Cấu trúc CSDL phân tán					
8.2. Đánh giá CSDL phân tán					
8.3. Thiết kế CSDL phân tán					
8.3.1. Bản sao dữ liệu					
8.3.2. Chia nhỏ dữ liệu					
8.4. Tính thông suốt và tính tự trị của CSDL phân tán					
8.5. Xử lý câu hỏi					
8.6. Khôi phục sai sót trong CSDL phân tán					

Nhiệm vụ của sinh viên:

Tham dự các buổi học lý thuyết và thực hành, làm các bài tập được giao, làm các bài kiểm tra giữa kỳ và bài thi kết thúc học phần theo đúng quy định.

Tài liệu học tập:

1. Nguyễn Xuân Huy, *Giáo trình về cơ sở dữ liệu*, Đại học Quốc Gia Hà Nội, 2000.
2. Nguyễn Xuân Huy-Lê Hoài Bắc, *Bài tập cơ sở dữ liệu*, Nhà xuất bản Thống kê, 2003.
3. Phạm Hữu Khang, Đoàn Thiện Ngân, *Quản trị SQL Server 2000*, Nhà xuất bản Thống kê, 2004.

Hình thức và tiêu chuẩn đánh giá sinh viên:

- Hình thức thi: thi viết.
- Tiêu chuẩn đánh giá sinh viên: dựa trên thái độ tham dự các buổi học lý thuyết và thực hành, kết quả làm các bài tập được giao, các bài kiểm tra giữa kỳ và bài thi kết thúc học phần.

Thang điểm: Thang điểm chữ A,B,C,D,F.

Điểm đánh giá học phần: $Z=0,3X+0,7Y$

Bài giảng này là tài liệu **chính thức và thống nhất** của Bộ môn Hệ thống Thông tin, Khoa Công nghệ Thông tin và được dùng để giảng dạy cho sinh viên.

Ngày phê duyệt: / /

Trưởng Bộ môn

MỤC LỤC

Chương 1. Hệ quản trị cơ sở dữ liệu	7
1.1. Quan niệm về CSDL	7
1.2. Các khả năng của một hệ quản trị cơ sở dữ liệu.	7
Chương 2. Cơ sở dữ liệu hướng đối tượng	9
2.1. Nhu cầu về hệ thống cơ sở dữ liệu hướng đối tượng	9
2.1.1. Các đối tượng phức tạp	9
2.1.2. Quản lý các tri thức	9
2.1.3. Quản trị các dữ liệu phân tán	10
2.1.4. Nhu cầu về hệ thống cơ sở dữ liệu hướng đối tượng	10
2.2. Khái niệm về hướng đối tượng	11
2.2.1. Đối tượng	12
2.2.2. Lớp đối tượng	12
2.2.3. Cá thể	13
2.2.4. Kế thừa	13
2.3. Cơ sở dữ liệu hướng đối tượng	13
2.4. Thiết kế cơ sở dữ liệu hướng đối tượng	14
2.4.1. Phân lớp	14
2.4.2. Tổng quát hóa và đặc biệt hóa	14
2.4.3. Gộp	15
2.5. Xây dựng cơ sở dữ liệu hướng đối tượng	15
Chương 3. Cơ sở dữ liệu phân tán	17
3.1. Các phương pháp phân tán dữ liệu	17
3.1.1. Khái niệm về phân tán dữ liệu	17
3.1.1.1. Các lý do phân mảnh	17
3.1.1.2. Các kiểu phân mảnh	17
3.1.1.3. Mức độ phân mảnh	19
3.1.1.4. Quy tắc phân mảnh đúng đắn	19
3.1.1.5. Các kiểu cấp phát	19
3.1.1.6. Các yêu cầu thông tin	19
3.1.2. Phân mảnh ngang	20
3.1.2.1. Yêu cầu thông tin của phân mảnh ngang.	20
3.1.2.2. Phân mảnh ngang nguyên thủy.	21
3.1.2.3. Phân mảnh ngang dẫn xuất	23
3.1.3. Phân mảnh dọc	24
3.1.4. Cấp phát	24
3.2. Kiểm soát dữ liệu ngữ nghĩa	26
3.2.1. Quản lý khung nhìn	26
3.2.1.1. Khung nhìn trong quản lý tập trung	26
3.2.1.2. Cập nhật qua các khung nhìn	26
3.2.1.3. Khung nhìn trong cơ sở dữ liệu phân tán	27
3.2.2. An toàn dữ liệu	27
3.2.2.1. Kiểm soát cấp quyền tập trung	27
3.2.2.2. Kiểm soát cấp quyền phân tán	28
3.3. Quản lý giao dịch và điều khiển đồng thời phân tán	28
3.3.1. Các khái niệm cơ bản về giao dịch	28
3.3.1.1. Tính nguyên tử	29
3.3.1.2. Mục dữ liệu	29

3.3.1.3. Khóa	30
3.3.1.4. Kiểm soát hoạt động đồng thời bằng khóa	30
3.3.1.5. Khóa sống (livelock)	31
3.3.1.6. Khóa “cứng” (deadlock)	31
3.3.1.7. Tính khả tuần tự của lịch biểu.	32
3.3.1.8. Bộ xếp lịch	33
3.3.1.9. Nghi thức	33
3.3.2. Mô hình giao dịch đơn giản.....	33
3.3.2.1. Ý nghĩa của giao dịch – hàm đặc trưng.....	33
3.3.2.2. Kiểm tra tính khả tuần tự bằng đồ thị có hướng.	35
3.3.3. Nghi thức khóa 2 pha	35
3.3.4. Mô hình khóa đọc và khóa ghi	36
3.3.4.1. Ý nghĩa của giao dịch với khóa đọc và khóa ghi.....	36
3.3.4.2. Đồ thị tuần tự hóa trong các giao dịch Rlock và Wlock	36
Chương 4. Hệ trợ giúp ra quyết định	38
4.1. Giới thiệu về hệ trợ giúp ra quyết định.....	38
4.2. Thiết kế cơ sở dữ liệu cho hệ trợ giúp ra quyết định	39
4.2.1. Thiết kế logic.	39
4.2.2. Thiết kế vật lý	40
4.3. Kho dữ liệu và kho dữ liệu chuyên đề	40
4.3.1. Kho dữ liệu	41
4.3.2. Kho dữ liệu chuyên đề.	41
4.3.3. Các lược đồ về chiều.....	42
4.4. Xử lý phân tích trực tuyến.....	43
4.4.1. Giới thiệu.....	43
4.4.2. Bảng chéo	43
4.4.3. Cơ sở dữ liệu nhiều chiều.....	44
4.5. Khai phá dữ liệu.....	44

Chương 1. Hệ quản trị cơ sở dữ liệu

1.1. Quan niệm về Cơ sở dữ liệu

Cơ sở dữ liệu (CSDL) là gì?

Định nghĩa: Một cơ sở dữ liệu (Database) là một tập hợp có cấu trúc các dữ liệu tác nghiệp được lưu trữ lại và được các hệ ứng dụng cụ thể sử dụng.

Ngày nay CSDL tồn tại trong hầu hết các ứng dụng, ví dụ:

- Ứng dụng quản lý kho hàng;
- Hệ thống đặt chỗ máy bay;
- Quản lý nguồn nhân lực...

Hệ quản trị cơ sở dữ liệu là gì?

Định nghĩa: Hệ quản trị cơ sở dữ liệu (Database Management System - DBMS) là một hệ thống phần mềm cho phép tạo lập CSDL và điều khiển mọi truy nhập đối với CSDL đó.

Hệ cơ sở dữ liệu là gì?

Hệ CSDL là một hệ thống bao gồm 4 thành phần:

- CSDL hợp nhất: CSDL của hệ có hai tính chất tối thiểu hóa dư thừa và được chia sẻ.
- Những người sử dụng: Người sử dụng của hệ là bất kỳ một người nào có nhu cầu truy nhập vào CSDL, bao gồm tất cả những người sử dụng cuối, những người viết chương trình ứng dụng và những người điều khiển toàn bộ hệ thống hay còn gọi là người quản trị CSDL.
- Phần mềm quản trị CSDL.
- Phần cứng của hệ bao gồm các thiết bị nhớ thứ cấp được sử dụng để lưu trữ CSDL.

1.2. Các khả năng của một hệ quản trị CSDL.

Có hai khả năng cho phép phân biệt các hệ quản trị CSDL với các kiểu hệ thống lập trình khác:

- Khả năng quản lý dữ liệu tồn tại lâu dài;
- Khả năng truy nhập các khối lượng dữ liệu lớn một cách hiệu quả.

Đặc điểm thứ nhất chỉ ra rằng có một CSDL tồn tại trong một thời gian dài, nội dung của CSDL này là các dữ liệu mà một hệ quản trị CSDL truy nhập và quản lý. Đặc điểm thứ hai phân biệt một hệ quản trị với một hệ thống xử lý tệp cũng quản lý dữ liệu tồn tại lâu dài nhưng nói chung không cung cấp các truy nhập nhanh chóng đến các bộ phận dữ liệu tùy ý.

Các khả năng của một hệ quản trị CSDL hầu hết là cần thiết khi khối lượng dữ liệu cần lưu trữ là rất lớn, bởi vì các khối lượng dữ liệu nhỏ thì các kỹ thuật truy nhập đơn giản, chẳng hạn quét tuần tự các dữ liệu là thích hợp.

Khi xem xét hai đặc điểm trên của một hệ quản trị CSDL là cơ bản, còn một số các khả năng khác mà có thể thấy trong hầu hết các hệ quản trị CSDL thương mại, đó là:

- Hỗ trợ ít nhất một mô hình dữ liệu hay một sự trừu tượng toán học mà qua đó người sử dụng có thể quan sát dữ liệu.

- Đảm bảo tính độc lập dữ liệu hay sự bất biến của các chương trình ứng dụng đối với các thay đổi về cấu trúc trong mô hình dữ liệu.
- Hỗ trợ các ngôn ngữ cấp cao nhất định cho phép người sử dụng định nghĩa cấu trúc của dữ liệu, truy nhập dữ liệu và thao tác dữ liệu.
- Quản trị giao dịch, có nghĩa là khả năng cung cấp các truy cập đồng thời, đúng đắn đối với CSDL từ nhiều người sử dụng tại cùng một thời điểm.
- Điều khiển truy cập, có nghĩa là khả năng hạn chế truy nhập đến dữ liệu bởi những người sử dụng không được cấp phép và khả năng kiểm tra tính đúng đắn của dữ liệu.
- Phục hồi dữ liệu, có nghĩa là khả năng phục hồi, không làm mất mát dữ liệu đối với các lỗi của hệ thống.

TaiLieu.vn

Chương 2. Cơ sở dữ liệu hướng đối tượng

2.1. Nhu cầu về hệ thống CSDL hướng đối tượng

Nhìn chung hệ quản trị CSDL quan hệ được sử dụng nhiều nhưng chưa đáp ứng được hết các yêu cầu của thực tế. Bên cạnh mô hình quan hệ, mô hình mạng và phân cấp vẫn tồn tại. Một số hạn chế của hệ quản trị CSDL quan hệ.

2.1.1. Các đối tượng phức tạp

Một cách hình thức, một đối tượng nhằm xác định một cấu trúc phức tạp. Ví dụ: Các đối tượng có cấu trúc phức tạp thường thấy là một siêu văn bản, một lược đồ, bức ảnh hay chương trình. Sự phức tạp của các đối tượng này thể hiện qua: - Cấu trúc của đối tượng; - Mô hình hóa các đối tượng; - Ngôn ngữ hỏi trên các đối tượng.

Bình thường, CSDL quan hệ xử lý các loại dữ liệu quen thuộc như số, chữ, ngày tháng, logic. Với các loại dữ liệu này, chưa thể thể hiện các loại dữ liệu định tính hay một danh sách.

Chính vì vậy người ta đòi hỏi mô hình hóa các đối tượng phức tạp và xử lý chúng trong hệ quản trị nhờ ngôn ngữ chương trình. Các đối tượng phức tạp được coi như các kí tự, các dữ liệu phức. Trong chương trình, chúng được mô tả theo các kiểu đặc biệt. Giải pháp này đụng chạm đến các khái niệm cũng quan trọng khác là hiện tượng dư thừa mã khi mô tả các đối tượng phức tạp trong chương trình ứng dụng, đụng chạm đến sự phụ thuộc chương trình/dữ liệu. Như vậy việc xử lý các dữ liệu phức, có kích thước lớn theo giải pháp đó là không hiệu quả.

Hơn nữa, do cấu trúc dữ liệu hiện tại là quá đơn giản, không thể dùng cho mô hình hóa các đối tượng phức tạp, thí dụ đối tượng trong hệ thống phân cấp hay đồ thị.

Ví dụ 2.1. Một đối tượng phức tạp, chẳng hạn trong CAD, thường được phân rã và đặt trong các quan hệ nhỏ; chính vì vậy mà thông tin ngữ nghĩa sẵn có trong một vấn đề bị chia nhỏ ra, phân tán dưới dạng các giá trị trong các quan hệ. Điều này khiến người sử dụng phải nhìn thể giới của bài toán theo cách nhìn của CSDL quan hệ. Để có thể khôi phục các ngữ nghĩa ban đầu, không tránh được việc yêu cầu các phép kết nối quan hệ, là phép toán tốn kém tài nguyên để thực hiện.

Ngoài việc mô hình hóa, việc truy vấn CSDL thông qua ngôn ngữ cũng cần phải xem xét. Khó có thể có ngôn ngữ lý tưởng trên các đối tượng phức tạp. Cũng có thể sử dụng các ngôn ngữ lập trình để giải quyết nhưng sẽ gặp khó khăn về sự khác biệt giữa kiểu dữ liệu và cách khai thác dữ liệu của ngôn ngữ lập trình và ngôn ngữ CSDL. Ngôn ngữ truy vấn CSDL thì dựa trên cơ sở tập hợp, trong khi ngôn ngữ lập trình dựa trên các thủ tục.

Các ứng dụng về CSDL được nhìn nhận theo hai khía cạnh:

- Khía cạnh tĩnh: thể hiện qua các dữ liệu.
- Khía cạnh động: thể hiện qua các phép xử lý tác động lên dữ liệu.

Người ta thấy những đối tượng được hệ quản trị CSDL quan hệ xử lý đặc biệt là dữ liệu tĩnh. Phần tác động của chúng, tức các hoạt động, được mô tả riêng biệt thông qua các chương trình ứng dụng tác động lên các dữ liệu. Người lập trình phải biết cấu trúc quan hệ của các đối tượng, bởi lẽ hệ quản trị CSDL quan hệ không đáp ứng những nhu cầu về các đối tượng động.

2.1.2. Quản lý các tri thức