

Xin cảm ơn bạn đã dành sự quan tâm cho tài liệu này!

Tài liệu này tôi tổng hợp từ nhiều quyển sách khác nhau. Tôi nghĩ nó có ích cho những người muốn hiểu các hệ quản trị cơ sở dữ liệu làm việc như thế nào. Trên cơ sở đó, người đọc có thể thiết kế các cơ sở dữ liệu và tạo ra các hệ thống sử dụng cơ sở dữ liệu quan hệ tốt hơn. Ngoài ra, tôi còn giới thiệu qua một số hướng đang được phát triển trên thế giới để các bạn tham khảo.

Trong quá trình biên-dịch-soạn dựa trên kiến thức của mình, tôi đã rất cố gắng nhưng chắc chắn vẫn có nhiều hạn chế và sai sót. Nếu bạn phát hiện sai sót, hoặc có ý kiến đóng góp nhằm mục đích giúp tôi nâng cao chất lượng của tài liệu này cũng như hoàn thiện chính mình, mong bạn phản hồi về ntquocvinh@gmail.com, ntquocvinh@yahoo.com và ntquocvinh@due.edu.vn.

Tôi sẽ cố gắng hoàn thiện tài liệu này, và tất nhiên, không thể thiếu những ý kiến phản hồi của các bạn.

Nguyễn Trần Quốc Vinh

MỤC LỤC

Phần I Kiến thức cơ bản	9
Chương I.1 CSDL và hệ thống tệp tin	9
I.1.1 Hệ thống tệp tin	11
I.1.1.1 Cấu trúc tệp tin	11
I.1.1.2 Đặt tên (gán nhãn) cho tệp tin	12
I.1.1.3 Bảo vệ tệp tin	13
I.1.1.4 Chế độ truy cập nhiều người dùng	14
I.1.2 Lĩnh vực ứng dụng của tệp tin	14
I.1.3 Các yêu cầu của các hệ thống thông tin	15
Chương I.2 Các chức năng của HQT CSDL. Tổ chức điển hình của một HQT CSDL hiện đại	18
I.2.1 Các chức năng của HQT CSDL	18
I.2.1.1 Quản trị dữ liệu trực tiếp trong bộ nhớ ngoài	18
I.2.1.2 Quản trị các bộ nhớ đệm (buffers) trong RAM	18
I.2.1.3 Điều khiển các giao tác	19
I.2.1.4 Nhật ký	20
I.2.1.5 Hỗ trợ các ngôn ngữ CSDL	22
I.2.1.6 Các lợi ích khác	23
I.2.2 Tổ chức điển hình của một HQT CSDL hiện đại	24
Phần II Tổ chức bên trong của HQT CSDL	26
Chương II.1 Tổ chức bên trong của HQT CSDL	26
II.1.1 Kiến trúc của một HQT CSDL	26
II.1.2 Các cấu trúc bộ nhớ ngoài, các phương pháp tổ chức chỉ mục (index)	26
II.1.3 Lưu trữ quan hệ	27
II.1.4 Các chỉ mục	30
II.1.4.1 B-tree	31
II.1.4.2 Băm (Hashing)	34
II.1.4.3 Thông tin nhật ký	35
II.1.4.4 Thông tin phục vụ	35
Chương II.2 Giao tác và tính toàn vẹn CSDL (cố định 2 pha?)	36

II.2.1	Ví dụ sự phá huỷ tính toàn vẹn CSDL	37
II.2.2	Khái niệm giao tác	38
II.2.3	Phân loại giao tác	40
II.2.3.1	Giao tác tự chấp nhận (Autocommit mode)	40
II.2.3.2	Giao tác được khai báo tường minh (Explicit)	40
II.2.3.3	Giao tác ngầm (Implicit)	41
II.2.3.4	Giao tác tự phục hồi (Auto rollback)	42
II.2.3.5	Giao tác lồng (Nested Transaction)	44
II.2.3.5.1	Khái niệm	44
II.2.3.5.2	Các quy tắc	44
II.2.3.5.3	Biến @@trancount trong HQT CSDL SQL Server	46
II.2.4	Ràng buộc tính toàn vẹn CSDL	47
II.2.5	Phân loại ràng buộc tính toàn vẹn CSDL	49
II.2.5.1	Bằng các cách triển khai (realization)	49
II.2.5.2	Phân loại ràng buộc tính toàn vẹn theo thời điểm kiểm tra	52
II.2.5.3	Phân loại ràng buộc tính toàn vẹn theo không gian ứng dụng	53
II.2.5.3.1	Ràng buộc Miền giá trị	53
II.2.5.3.2	Ràng buộc thuộc tính	53
II.2.5.3.3	Ràng buộc bản ghi	54
II.2.5.3.4	Ràng buộc quan hệ	58
II.2.5.3.5	Ràng buộc CSDL	59
II.2.6	Triển khai ràng buộc tính toàn vẹn dạng định nghĩa bằng công cụ SQL	60
II.2.6.1	Các nguyên tắc chung trong triển khai ràng buộc bằng các công cụ SQL	60
II.2.6.2	Cú pháp ràng buộc theo chuẩn SQL	61
Chương II.3	Giao tác và tính song song (tính đối ngẫu)	71
II.3.1	Đôi điều về công việc của giao tác	72
II.3.2	Các vấn đề thực hiện song song các giao tác	74
II.3.2.1	Vấn đề mất kết quả cập nhật dữ liệu	74
II.3.2.2	Vấn đề phụ thuộc chưa cố định (đọc dữ liệu bản, đọc không chuẩn...)	75
II.3.2.3	Vấn đề phân tích xung khắc	75
II.3.2.3.1	Đọc không lặp lại	75
II.3.2.3.2	Các phần tử ảo (phantom)	76
II.3.2.3.3	Phân tích xung khắc thực chất	77
II.3.2.3.4	Xung đột giữa các giao tác	78
II.3.3	Đảm bảo thực thi song song (tính tuần tự) các giao tác	82
II.3.3.1	Định lý Esvaran về tính tuần tự	82
II.3.3.2	Chiếm lĩnh đồng bộ	83
II.3.3.2.1	Các phong toả	83
II.3.3.2.2	Giải quyết các vấn đề của tính song song bằng sự giúp đỡ của các phong toả	85
II.3.3.2.2.1	Vấn đề mất kết quả cập nhật	85
II.3.3.2.2.2	Vấn đề phụ thuộc không được cố định (đọc dữ liệu bản, đọc không chuẩn)	86
II.3.3.2.2.3	Vấn đề phân tích tính xung khắc	87
II.3.3.2.2.4	Phân tích xung khắc thực chất	88
II.3.3.2.2.5	Giải quyết (gỡ bỏ) các tình huống bế tắc	89
II.3.3.2.2.6	Phong toả có chủ định (Chiếm lĩnh đồng bộ bằng cách tán nhỏ)	91
II.3.3.2.3	Phong toả theo mệnh đề	96
II.3.3.2.4	Cơ chế tách phiên bản dữ liệu	97
II.3.3.3	Phương pháp mốc thời gian	100
II.3.3.4	Triển khai tính cô lập các giao tác bằng các công cụ SQL	101

II.3.3.4.1	Các mức độ cô lập	101
II.3.3.4.2	Cú pháp các lệnh SQL để định nghĩa mức độ cô lập	102
Chương II.4	Các giao tác và phục hồi dữ liệu	103
II.4.1	Nhật ký và ghi nhớ trung gian (buffering)	104
II.4.2	Các kiểu phục hồi dữ liệu	106
II.4.2.1	Rollback giao tác riêng biệt	107
II.4.2.2	Phục hồi sau sự cố “mềm”	107
II.4.2.3	Tính nhất quán vật lý của CSDL	110
II.4.2.4	Phục hồi sau sự cố “cứng”	111
II.4.2.5	Phục hồi dữ liệu và chuẩn SQL	112
Chương II.5	Các trình biên dịch SQL. Các vấn đề tối ưu hoá	112
II.5.1	Sơ đồ chung trong xử lý các truy vấn	113
II.5.2	Tối ưu hoá cú pháp các truy vấn	114
II.5.2.1	Biến đổi logic đơn giản các truy vấn	114
II.5.2.2	Biến đổi các truy vấn bằng cách thay đổi thứ tự các giao dịch quan hệ	114
II.5.2.3	Chuyển các truy vấn chứa các truy vấn lồng thành các truy vấn chứa các phép nối	114
II.5.3	Tối ưu ngữ nghĩa các truy vấn	116
II.5.3.1	Biến đổi các truy vấn trên cơ sở thông tin mạng ngữ nghĩa	116
II.5.3.2	Sử dụng các thông tin mạng ngữ nghĩa trong tối ưu hoá các truy vấn	116
II.5.4	Chọn và đánh giá các kế hoạch thực thi các truy vấn khác nhau	116
II.5.4.1	Tạo ra các kế hoạch	116
II.5.4.2	Đánh giá trị giá các kế hoạch thực thi truy vấn	116
II.5.4.3	Các đánh giá chính xác hơn	116
Phần III	HQT CSDL trong kiến trúc “Client–Server”	117
Chương III.1	Kiến trúc “Client–Server”	117
III.1.1	Các hệ thống mở	117
III.1.2	Các máy khách và máy chủ trong mạng cục bộ	118
III.1.3	Kiến trúc hệ thống “khách-chủ”	119
III.1.3.1	Khách thông minh (Smart Client)	121
III.1.4	Các máy chủ CSDL	123
III.1.4.1	Các nguyên tắc tương tác giữa phần khách và phần chủ	123
III.1.4.2	Các ưu điểm của giao thức gọi thủ tục từ xa	123
III.1.4.3	Sự phân bố chức năng điển hình giữa khách và chủ	124
III.1.4.4	Các yêu cầu khả năng phần cứng và phần mềm nền tảng đối với khách và chủ	124
Phần IV	Các khung nhìn thực	125
Phần VKho dữ liệu – DataWarehouse		126
Phần VI	CSDL phân tán	127
Chương VI.1	Các loại hình hệ thống phân tán	128
Chương VI.2	Các kiến trúc của HQT CSDL PT	129
VI.2.1	Các hệ thống khách-chủ	129
VI.2.2	Các hệ thống máy chủ hợp tác	130
VI.2.3	Các hệ thống tầng trung gian	130
Chương VI.3	Lưu trữ dữ liệu trên HQT CSDL PT	131
VI.3.1	Phân mảnh (fragmentation)	131
VI.3.2	Nhân bản replication	132
VI.3.2.1	Nhân bản đồng bộ	133
VI.3.2.2	Nhân bản không đồng bộ	134
VI.3.2.2.1	Nhân bản theo nút chính hay là peer-to-peer	134

VI.3.3	Quản trị danh mục phân tán	137
VI.3.3.1	Đặt tên cho các đối tượng	137
VI.3.3.2	Cấu trúc danh mục phân tán	138
VI.3.4	Tính độc lập của dữ liệu phân tán	139
Chương VI.4	Xử lý các truy vấn phân tán	140
VI.4.1	Các thuật toán nối join	140
VI.4.1.1	Các truy vấn nonjoin	140
VI.4.1.2	Thực thi phép nối trên HQT CSDL PT	141
VI.4.1.2.1	Rút ra những gì cần thiết	141
VI.4.1.2.2	Chuyển đến một nút	141
VI.4.1.2.3	Semijoins và Bloomjoins	142
VI.4.2	Biên dịch phân tán các truy vấn	143
Chương VI.5	Quản trị các giao tác và quá trình đồng bộ hoá (Thành viên hay là cấp dưới???)	145
VI.5.1	Giao thức cố định hai pha (2PC – Two-Phases Commit Protocol)	145
VI.5.2	Các phương pháp quản trị	146
VI.5.3	Các bế tắc phân tán	146
Chương VI.6	Khởi động lại sau sự cố	148
VI.6.1	Khởi động lại hệ thống sau sự cố	148
VI.6.2	Giao thức cố định 3 pha (3PC – Three Phases Commit)	149
Chương VI.7	Các hệ thống tích hợp (hay liên bang) và đa CSDL	151
Phần VII	Các hướng nghiên cứu và phát triển hiện đại	153
Chương VII.1	Các HQT CSDL thế hệ mới	153
VII.1.1	Theo hướng mở rộng mô hình quan hệ	154
VII.1.2	Các kiểu dữ liệu trừu tượng	156
VII.1.3	Tạo ra các hệ CSDL theo hướng trình ứng dụng	157
VII.1.4	Tối ưu các truy vấn được điều khiển bởi các quy luật	158
VII.1.5	Hỗ trợ các thông tin lịch sử và các truy vấn biểu thị thời gian	159
Chương VII.2	Các HQT CSDL hướng đối tượng	161
VII.2.1	Mối liên hệ giữa HQT CSDL hướng đối tượng với khái niệm chung về cách tiếp cận hướng đối tượng	162
VII.2.2	Các mô hình dữ liệu hướng đối tượng	164
VII.2.3	Các ngôn ngữ lập trình các CSDL HĐT	167
VII.2.3.1	Sự đánh mất mối tương quan giữa các ngôn ngữ lập trình và các ngôn ngữ truy vấn trong HQT CSDL quan hệ	167
VII.2.3.2	Các ngôn ngữ lập trình CSDL HĐT như là các ngôn ngữ HĐT hỗ trợ các đối tượng bền vững (persistent)	168
VII.2.3.3	Các ví dụ ngôn ngữ lập trình CSDL HĐT	169
VII.2.4	Các ngôn ngữ truy vấn CSDL HĐT	170
VII.2.4.1	Navigation tường minh như là một công cụ vượt qua sự mất mát tính tương ứng	170
VII.2.4.2	Các ngôn ngữ truy vấn không-navigation	170
VII.2.4.3	Các vấn đề tối ưu hoá truy vấn	172
VII.2.5	Các ví dụ HQT CSDL HĐT	174
VII.2.5.1	Dự án ORION	175
VII.2.5.2	Project O2	176
Chương VII.3	Các hệ CSDL trên nền tảng các quy luật	177
VII.3.1	Phần mở rộng và phân tầng cường của CSDL	177
VII.3.2	Các CSDL năng động (chủ động – active DB)	178

VII.3.3 CSDL suy diễn	179
Phần VIII Phụ lục	182
A. MS SQL Server	182
A.1 Kiến trúc của SQL Server	184
A.2 Cài đặt và khôi phục trong SQL Server	188
B. Oracle	188
B.1. Kiến trúc của Oracle	188
B.2. Cài đặt và khôi phục trong Oracle	188
C. Các hình vẽ	189
D. Các bảng	189
Tài liệu tham khảo	190