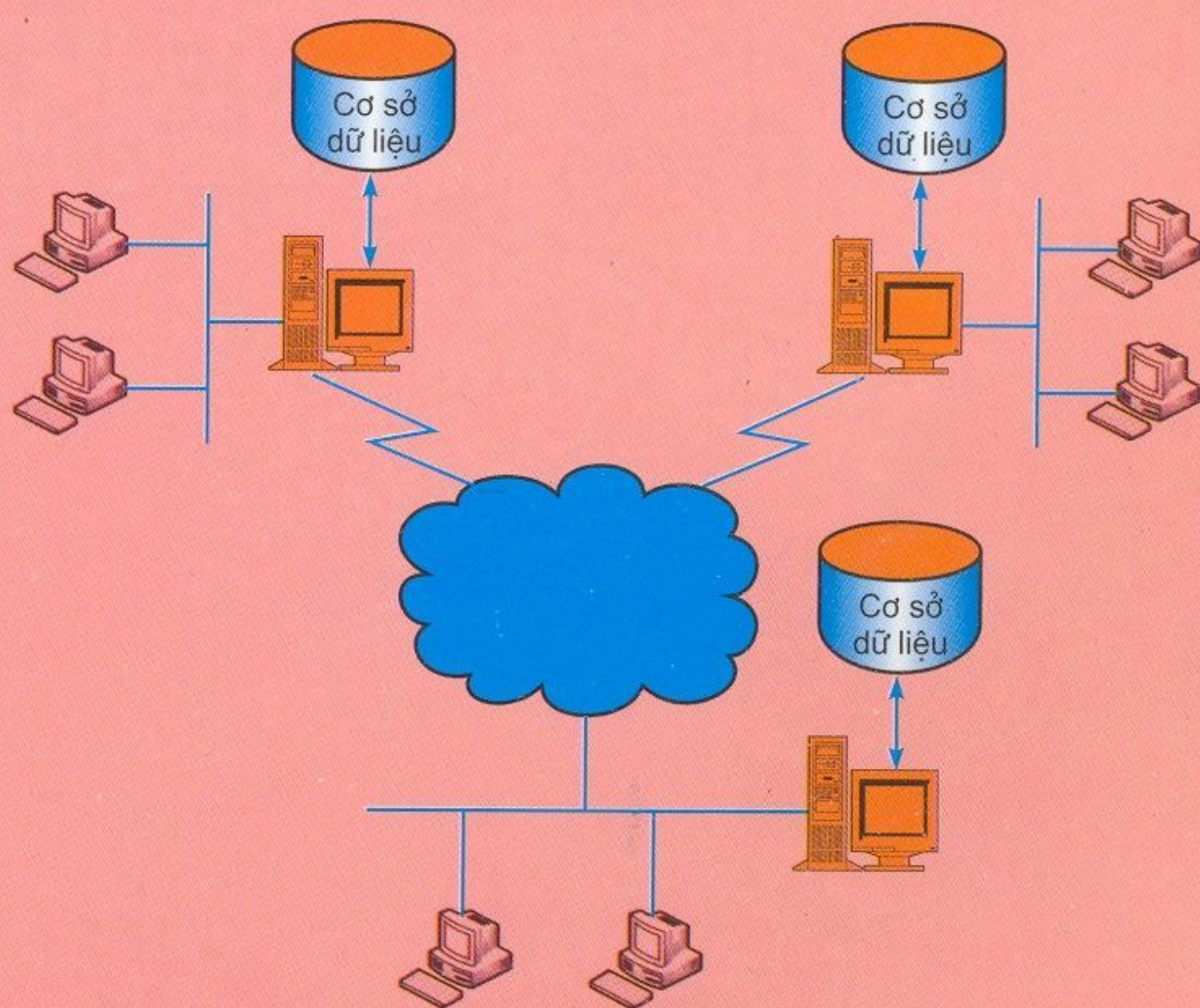


HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG  
TS. Phạm Thế Quế (Chủ biên)

# GIÁO TRÌNH CƠ SỞ DỮ LIỆU PHÂN TÁN



NHÀ XUẤT BẢN THÔNG TIN VÀ TRUYỀN THÔNG



005.74

GI-108TR

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

TS. Phạm Thế Quế (Chủ biên)

# GIÁO TRÌNH CƠ SỞ DỮ LIỆU PHÂN TÁN

ĐẠI HỌC VÙNG VƯƠNG  
THƯ VIỆN  
PHÒNG MƯỢN

HVV- 00005207

NHÀ XUẤT BẢN THÔNG TIN VÀ TRUYỀN THÔNG



# Lời nói đầu

Lý thuyết hệ cơ sở dữ liệu phân tán là sự tích hợp của hai hướng tiếp cận xử lý dữ liệu, đó là lý thuyết các hệ cơ sở dữ liệu (Database System) và công nghệ mạng máy tính (Computer Network). Các phương thức xử lý dữ liệu phát triển từ phương thức truyền thống - mỗi một ứng dụng được định nghĩa, cập nhật và xử lý dữ liệu riêng rẽ trên các tệp riêng chuyển sang xử lý dữ liệu tập trung, nghĩa là dữ liệu được định nghĩa và được quản lý một cách tập trung. Do đó đảm bảo được tính độc lập của dữ liệu, các chương trình ứng dụng không phụ thuộc vào cấu trúc dữ liệu logic hay cấu trúc lưu trữ vật lý và ngược lại.

Nhằm mục đích trang bị kiến thức cơ sở và nâng cao về các hệ cơ sở dữ liệu phân tán, lý thuyết phân mảnh không tổn thất thông tin, vấn đề tương tranh và hiệu năng xử lý phân tán cho học sinh, sinh viên ngành Công nghệ thông tin và các ngành kỹ thuật khác, Học viện Công nghệ Bưu chính Viễn thông đã phối hợp với Nhà xuất bản Thông tin và Truyền thông xuất bản cuốn sách ***“Giáo trình Cơ sở dữ liệu phân tán”*** do TS. Phạm Thế Quế biên soạn.

*“Giáo trình Cơ sở dữ liệu phân tán”* không chỉ đề cập đến những vấn đề cơ sở lý thuyết mà còn trình bày một số kỹ năng cần thiết để thiết kế và cài đặt các hệ cơ sở dữ liệu cụ thể. Nội dung giáo trình gồm 06 chương:

*Chương 1: Khái niệm cơ bản về cơ sở dữ liệu phân tán.* Giới thiệu những khái niệm cơ bản về xử lý truy vấn, mục đích của việc xử lý truy vấn và giới thiệu chức năng các tầng của quá trình xử lý truy vấn. Công cụ để xử lý truy vấn là các phép tính quan hệ và đại số quan hệ. Trong thiết kế cơ sở dữ liệu phân tán, việc phân mảnh dữ liệu và cấp phát dữ liệu có vai trò quan trọng cho việc xử lý truy vấn dữ liệu, làm tăng tính cục bộ tham chiếu, tăng khả năng thực hiện truy vấn đồng thời song song trên nhiều vị trí.

*Chương 2: Thiết kế các hệ cơ sở dữ liệu phân tán.* Trình bày những vấn đề cơ bản về thiết kế cơ sở dữ liệu phân tán. Nhấn mạnh đặc biệt đến các vấn đề phân mảnh và cấp phát dữ liệu trên các node của mạng máy tính.



*Chương 3: Xử lý truy vấn trong cơ sở dữ liệu phân tán.* Giới thiệu những khái niệm cơ bản về xử lý truy vấn, mục đích của việc xử lý truy vấn và giới thiệu chức năng các tầng của quá trình xử lý truy vấn. Công cụ để xử lý truy vấn là các phép tính quan hệ và đại số quan hệ. Trong thiết kế cơ sở dữ liệu phân tán, việc phân mảnh dữ liệu và cấp phát dữ liệu có vai trò quan trọng cho việc xử lý truy vấn dữ liệu, làm tăng tính cục bộ tham chiếu, tăng khả năng thực hiện truy vấn đồng thời song song trên nhiều vị trí.

*Chương 4: Quản lý giao dịch và điều khiển đồng thời phân tán.* Trình bày các khái niệm cơ bản về giao dịch, các tính chất của một giao dịch, các loại giao dịch và các kỹ thuật điều khiển đồng thời phân tán, điều khiển đồng thời bằng khóa chốt và bằng nhãn thời gian. Nội dung của chương cũng giới thiệu các thuật toán điều khiển đồng thời có các tính chất biệt lập và nhất quán của các giao dịch. Cơ chế điều khiển đồng thời phân tán của các hệ quản trị cơ sở dữ liệu phân tán đảm bảo tính nhất quán của các hệ cơ sở dữ liệu được duy trì.

*Chương 5: Các hệ cơ sở dữ liệu song song.* Giới thiệu về những khái niệm cơ bản của các hệ cơ sở dữ liệu song song. Mạng máy tính có khả năng thực hiện các ứng dụng và quản lý cơ sở dữ liệu, nền tảng cơ bản cho các nguyên lý quản lý dữ liệu phân tán.

*Chương 6: Hệ quản trị cơ sở dữ liệu đối tượng phân tán.* Giới thiệu các khái niệm cơ bản về đối tượng và mô hình dữ liệu phân tán đối tượng, các phương pháp thiết kế phân tán đối tượng. Các kỹ thuật phân mảnh và cấp phát dữ liệu đối tượng phân tán. Giới thiệu các phương pháp quản lý đối tượng, quản lý giao dịch đối tượng và xử lý vấn đề tin đối tượng.

Sau mỗi chương đều có phần câu hỏi và bài tập để bạn đọc củng cố lại kiến thức của mình.

Tác giả đã dành nhiều công sức cho việc biên soạn, song giáo trình sẽ khó tránh khỏi những thiếu sót. Rất mong nhận được các ý kiến góp ý của các bạn đồng nghiệp và bạn đọc để giáo trình được hoàn thiện hơn trong lần tái bản sau.

Xin trân trọng cảm ơn./.

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG



# MỤC LỤC

<i>Lời nói đầu.....</i>	3
-------------------------	---

<b>Chương 1: KHÁI NIỆM CƠ BẢN VỀ CƠ SỞ DỮ LIỆU PHÂN TÁN.....</b>	13
--	----

<i>1.1 Mở đầu.....</i>	13
------------------------	----

<i>1.2 Xử lý phân tán và các hệ thống xử lý phân tán .....</i>	14
--	----

1.2.1 Khái niệm xử lý phân tán.....	14
-------------------------------------	----

1.2.2 Hệ thống xử lý phân tán.....	15
------------------------------------	----

<i>1.3 Hệ cơ sở dữ liệu phân tán là gì? .....</i>	16
---	----

1.3.1 Giới thiệu .....	16
------------------------	----

1.3.2 Hệ cơ sở dữ liệu phân tán và hệ quản trị cơ sở dữ liệu phân tán .....	17
--	----

<i>1.4 Triển vọng của các hệ cơ sở dữ liệu phân tán .....</i>	18
---	----

<i>1.5 Các đặc điểm của hệ cơ sở dữ liệu phân tán .....</i>	20
---	----

1.5.1 Điều khiển tập trung.....	20
---------------------------------	----

1.5.2 Độc lập dữ liệu.....	21
----------------------------	----

1.5.3 Giảm dư thừa dữ liệu .....	21
----------------------------------	----

1.5.4 Độ tin cậy qua các giao dịch phân tán .....	22
---	----

1.5.5 Cải tiến hiệu năng .....	23
--------------------------------	----

1.5.6 Dễ dàng mở rộng hệ thống.....	24
-------------------------------------	----

1.5.7 Tính toàn vẹn, phục hồi và điều khiển tương tranh .....	24
---	----

<i>1.6 Các mô hình cơ sở dữ liệu máy khách/chủ.....</i>	25
---	----

1.6.1 Mô hình cơ sở dữ liệu tập trung.....	26
--	----

1.6.2 Mô hình cơ sở dữ liệu theo kiểu file - server.....	26
--	----

1.6.3 Mô hình xử lý từng phần cơ sở dữ liệu.....	26
--	----

1.6.4 Mô hình cơ sở dữ liệu Client/Server.....	27
--	----

1.6.5 Mô hình cơ sở dữ liệu phân tán .....	29
--	----



1.7 Mô hình tham chiếu cơ sở dữ liệu phân tán .....	29
1.7.1 Lược đồ toàn cục .....	30
1.7.2 Lược đồ phân mảnh .....	31
1.7.3 Lược đồ cấp phát .....	33
1.7.4 Lược đồ ánh xạ cục bộ .....	32
1.7.5 Hệ quản trị cơ sở dữ liệu DBMS ở các site cục bộ độc lập .....	34
1.8 Kiến trúc logic hệ quản trị cơ sở dữ liệu phân tán .....	34
1.8.1 Kiến trúc tổng quát .....	34
1.8.2 Chức năng của hệ quản trị cơ sở dữ liệu quan hệ .....	35
1.8.3 Các thành phần hệ quản trị cơ sở dữ liệu phân tán .....	37
1.8.4 Các phương thức truy cập dữ liệu từ xa .....	38
1.8.5 Hệ quản trị cơ sở dữ liệu phân tán không thuần nhất .....	39
1.8.6 Hệ quản trị cơ sở dữ liệu phân tán thuần nhất .....	41
1.8.7 Hệ quản trị cơ sở dữ liệu phân tán kiểu Client/Server .....	42
1.8.8 Các hệ quản trị cơ sở dữ liệu phân tán ngang hàng .....	44
1.8.9 Các thành phần hệ quản trị cơ sở dữ liệu phân tán .....	46
1.9 Một số vấn đề cơ bản về cơ sở dữ liệu phân tán .....	49
1.9.1 Thiết kế cơ sở dữ liệu phân tán .....	49
1.9.2 Xử lý truy vấn phân tán .....	49
1.9.3 Quản lý thư mục phân tán .....	49
1.9.4 Điều khiển đồng thời phân tán .....	50
1.9.5 Quản lý bế tắc phân tán .....	50
1.9.6 Độ tin cậy của hệ quản trị cơ sở dữ liệu phân tán .....	50
1.9.7 Cơ sở dữ liệu hỗn hợp .....	51
Câu hỏi .....	51

## **Chương 2: THIẾT KẾ CÁC HỆ CƠ SỞ DỮ LIỆU PHÂN TÁN.....**

2.1 Các chiến lược thiết kế .....	54
2.1.1 Thiết kế phân tán dữ liệu theo mô hình từ trên xuống .....	55
2.1.2 Thiết kế phân tán dữ liệu theo mô hình từ dưới lên .....	56



2.2 Các vấn đề về phân mảnh dữ liệu .....	57
2.2.1 Lý do phân mảnh .....	58
2.2.2 Các kiểu phân mảnh.....	59
2.2.3 Mức độ phân mảnh .....	60
2.2.4 Các quy tắc phân mảnh .....	60
2.2.5 Các kiểu cấp phát .....	61
2.2.6 Các yêu cầu thông tin.....	62
2.3 Phương pháp phân mảnh ngang .....	63
2.3.1 Khái niệm phân mảnh ngang .....	63
2.3.2 Thông tin cần thiết của phân mảnh ngang .....	64
2.3.3 Phân mảnh ngang nguyên thủy .....	68
2.3.4 Tính đầy đủ và tính cực tiểu của vị từ đơn giản.....	70
2.3.5 Thuật toán xác định tập vị từ đầy đủ và cực tiểu từ .....	71
2.3.6 Thuật toán phân mảnh ngang nguyên thủy .....	74
2.3.7 Phân mảnh ngang dẫn xuất .....	77
2.4 Kiểm tra tính đúng đắn của phân mảnh ngang .....	83
2.4.1 Tính đầy đủ .....	83
2.4.2 Tính phục hồi .....	84
2.4.3 Tính tách biệt .....	85
2.5 Phương pháp phân mảnh dọc.....	86
2.5.1 Khái niệm phân mảnh dọc .....	86
2.5.2 Thông tin cần thiết của phân mảnh dọc .....	87
2.5.3 Thuật toán tụ nhóm .....	91
2.5.4 Thuật toán phân mảnh.....	97
2.5.5 Kiểm tra tính đúng đắn của thuật toán phân mảnh dọc ..	100
2.6 Phương pháp phân mảnh hỗn hợp (Hybrid Fragmentation) ..	101
2.7 Cấp phát dữ liệu .....	102
2.7.1 Bài toán cấp phát.....	103
2.7.2 Thông tin cần thiết cho bài toán cấp phát .....	106
2.7.3 Mô hình cấp phát .....	107



2.8 Kiểm soát ngữ nghĩa dữ liệu .....	110
---------------------------------------	-----

2.8.1 Quản lý khung nhìn.....	110
-------------------------------	-----

2.8.2 An toàn dữ liệu.....	112
----------------------------	-----

2.8.3 Kiểm soát tính toàn vẹn ngữ nghĩa .....	115
---	-----

Câu hỏi .....	118
---------------	-----

Bài tập .....	119
---------------	-----

### **Chương 3: XỬ LÝ TRUY VẤN TRONG CƠ SỞ DỮ LIỆU PHÂN TÁN.....**

123

3.1 Khái niệm truy vấn .....	123
------------------------------	-----

3.2 Vấn đề xử lý truy vấn.....	124
--------------------------------	-----

3.2.1 Bài toán xử lý truy vấn.....	124
------------------------------------	-----

3.2.2 Mục đích của việc xử lý truy vấn.....	130
---	-----

3.3 Quy trình xử lý truy vấn .....	131
------------------------------------	-----

3.3.1 Tầng phân rã truy vấn .....	132
-----------------------------------	-----

3.3.2 Tầng cục bộ hóa dữ liệu.....	133
------------------------------------	-----

3.3.3 Tầng tối ưu hóa truy vấn toàn cục .....	134
---	-----

3.3.4 Tầng tối ưu hóa truy vấn cục bộ .....	135
---	-----

3.4 Phân rã truy vấn.....	135
---------------------------	-----

3.4.1 Chuẩn hoá câu truy vấn.....	136
-----------------------------------	-----

3.4.2 Bước phân tích .....	137
----------------------------	-----

3.4.3 Bước loại bỏ dư thừa.....	141
---------------------------------	-----

3.4.4 Bước xây dựng lại câu truy vấn .....	143
--	-----

3.4.5 Các quy tắc biến đổi cây đại số quan hệ .....	144
---	-----

3.5 Cục bộ hoá dữ liệu phân tán .....	148
---------------------------------------	-----

3.5.1 Rút gọn phân mảnh ngang nguyên thuỷ .....	149
---	-----

3.5.2 Rút gọn phân mảnh dẫn xuất .....	153
--	-----

3.5.3 Rút gọn phân mảnh dọc .....	156
-----------------------------------	-----

3.5.4 Rút gọn phân mảnh hỗn hợp .....	157
---------------------------------------	-----



3.6 Tối ưu hóa truy vấn phân tán .....	159
3.7 Các thuật toán tối ưu hóa truy vấn phân tán .....	161
3.7.1 Thuật toán INGRES .....	161
3.7.2 Thuật toán System R.....	166
3.7.3 Đặc trưng các thuật toán tối ưu hoá truy vấn phân tán...	170
3.7.4 Thuật toán INGRES phân tán .....	171
3.7.5 Thuật toán System R* .....	175
3.7.6 Thuật toán SDD-1 .....	177
Câu hỏi .....	181
Bài tập .....	183

## **Chương 4: QUẢN LÝ GIAO DỊCH VÀ ĐIỀU KHIỂN ĐỒNG THỜI PHÂN TÁN .....**

4.1 Giới thiệu.....	185
4.2 Tổng quan về giao dịch .....	187
4.2.1 Các khái niệm cơ bản về giao dịch .....	187
4.2.2 Điều kiện kết thúc giao dịch .....	191
4.2.3 Đặc tính của giao dịch.....	193
4.2.4 Đặc trưng hóa khái niệm giao dịch .....	194
4.3 Các tính chất giao dịch .....	196
4.3.1 Tính nguyên tố .....	196
4.3.2 Tính nhất quán .....	198
4.3.3 Tính cô lập .....	199
4.3.4 Tính bền vững .....	203
4.4 Các loại giao dịch .....	203
4.4.1 Các loại giao dịch theo thời gian hoạt động.....	203
4.4.2 Các loại giao dịch dựa trên việc tổ chức các hành động đọc và ghi.....	204
4.4.3 Luồng công việc - work flows .....	204



4.5 Điều khiển các giao dịch đồng thời phân tán .....	207
4.5.1 Đặt vấn đề .....	207
4.5.2 Tính khả tuần tự lịch biểu .....	209
4.5.3 Phân loại các cơ chế điều khiển đồng thời.....	211
4.6 Các thuật toán điều khiển đồng thời bằng khoá chốt.....	218
4.6.1 Thuật toán quản lý khóa cơ bản .....	219
4.6.2 Thuật toán khóa chốt 2 pha (2PL).....	223
4.6.3 Thuật toán quản lý giao dịch 2PL tập trung (C2PL TM).....	228
4.6.4 Thuật toán 2PL bản chính .....	234
4.6.5 Thuật toán 2PL phân tán .....	234
4.7 Các thuật toán điều khiển đồng thời bằng nhãn thời gian....	235
4.7.1 Đặt vấn đề .....	235
4.7.2 Thuật toán bộ quản lý giao dịch TO cơ bản.....	237
4.7.3 Thuật toán TO bảo toàn .....	242
4.7.4 Thuật toán TO đa phiên .....	244
4.8 Các thuật toán điều khiển đồng thời lạc quan .....	245
4.9 Quản lý bế tắc.....	248
4.9.1 Ngăn chặn bế tắc .....	250
4.9.2 Tránh bế tắc.....	251
4.9.3 Phát hiện và giải tỏa bế tắc .....	252
Câu hỏi .....	255
Bài tập .....	256
<b>Chương 5: CÁC HỆ CƠ SỞ DỮ LIỆU SONG SONG .....</b>	<b>258</b>
5.1 Mục tiêu của xử lý song song .....	258
5.1.1 Giới thiệu chung.....	258
5.1.2 Mục tiêu của xử lý song song .....	259
5.2 Ưu điểm của cơ sở dữ liệu song song .....	260
5.2.1 Hiệu năng cao .....	261
5.2.2 Tính sẵn sàng cao.....	261



5.2.3 Khả năng mở rộng.....	261
5.3 Kiến trúc hệ cơ sở dữ liệu song song .....	262
5.3.1 Bộ quản lý phiên .....	262
5.3.2 Bộ quản lý yêu cầu.....	262
5.3.3 Bộ quản lý dữ liệu.....	262
5.4 Các kiến trúc hệ thống song song .....	263
5.4.1 Tổng quan về kiến trúc song song và hệ thống song song.....	263
5.4.2 Kiến trúc chia sẻ bộ nhớ .....	264
5.4.3 Kiến trúc chia sẻ đĩa.....	265
5.4.4 Kiến trúc không chia sẻ.....	267
5.4.5 Các kiến trúc phân cấp .....	268
5.5 Kỹ thuật hệ quản trị cơ sở dữ liệu song song .....	269
5.5.1 Sắp đặt dữ liệu.....	270
5.5.2 Truy vấn song song.....	278
5.5.3 Xử lý dữ liệu song song .....	279
5.6 Tối ưu hóa truy vấn song song .....	288
5.6.1 Mở đầu .....	288
5.6.2 Không gian tìm kiếm .....	288
5.6.3 Mô hình chi phí.....	292
5.6.4 Chiến lược tìm kiếm.....	293
Câu hỏi .....	295

<b>Chương 6: HỆ QUẢN TRỊ CƠ SỞ DỮ LIỆU ĐỐI TƯỢNG PHÂN TÁN.....</b>	<b>296</b>
6.1 Giới thiệu.....	296
6.2 Khái niệm cơ bản về đối tượng và mô hình dữ liệu đối tượng..	297
6.3 Thiết kế phân tán đối tượng.....	298
6.3.1 Phân hoạch ngang lớp .....	299



6.3.2	Phân hoạch dọc lớp .....	302
6.3.3	Phân hoạch đường dẫn .....	303
6.3.4	Các thuật toán phân hoạch .....	303
6.3.5	Cấp phát .....	304
6.3.6	Nhân bản .....	305
6.4	<i>Các mô hình kiến trúc đối tượng phân tán</i> .....	306
6.4.1	Các kiểu kiến trúc máy khách/chủ .....	306
6.4.2	Lưu trữ đối tượng phân tán .....	309
6.5	<i>Quản lý đối tượng</i> .....	310
6.5.1	Quản lý định danh đối tượng .....	310
6.5.2	Quản lý con trỏ .....	311
6.5.3	Di trú đối tượng .....	312
6.6	<i>Xử lý truy vấn đối tượng</i> .....	314
6.6.1	Kiến trúc xử lý truy vấn đối tượng .....	316
6.6.2	Các vấn đề xử lý truy vấn đối tượng .....	318
6.6.3	Thực thi truy vấn đối tượng .....	321
6.7	<i>Quản lý giao dịch đối tượng phân tán</i> .....	322
6.7.1	Các tiêu chuẩn quản lý .....	322
6.7.2	Mô hình giao dịch và cấu trúc đối tượng .....	324
6.7.3	Quản lý giao dịch trong các hệ quản trị đối tượng phân tán .....	326
	<i>Câu hỏi</i> .....	335
	<i>Bài tập</i> .....	336
	<b>Tài liệu tham khảo</b> .....	337