

TRƯỜNG ĐẠI HỌC HÀNG HẢI
KHOA CÔNG NGHỆ THÔNG TIN



BÀI GIẢNG CƠ SỞ DỮ LIỆU NÂNG CAO

Biên soạn: Th.S Nguyễn Trung Nghĩa

B GIAO THÔNG V N T I
TR NG I H CHÀNG H I
B MÔN: H TH NG THÔNG TIN
KHOA: CÔNG NGH THÔNG TIN

BÀI GI NG

C S D LI U NÂNG CAO

TÊN H C PH N : C S D LI U NÂNG CAO
MÃ H C PH N : 17406
TRÌNH ÀO T O : I H C CHÍNH QUY
DÙNG CHO SV NGÀNH : CÔNG NGH THÔNG TIN

Tên học phần: Cơ sở dữ liệu nâng cao

Bộ môn phụ trách giảng dạy: Học thuật Thông tin

Mã học phần: 17406

Loại học phần: 2

Khoa phụ trách: CNTT.

Thời gian TC: 2

TS tiết	Lý thuyết	Thực hành/Xemina	Thí nghiệm	Bài tập lớn	Điểm môn học
45	30	15	0	0	0

Điều kiện tiên quyết:

Sinh viên phải học và thi tốt các học phần sau mới có thể 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 chính yếu:

Giới thiệu về các quy trình dữ liệu; Các mô hình mạng, mô hình phân cấp; Điều kiện ràng buộc trong quy trình 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 MÔN HỌC	PHÂN PHỐI SỐ TIẾT				
	TS	LT	TH	BT	KT
Chương 1. Quy trình cơ sở dữ liệu và lịch sử phát triển	3	3			
1.1. Quan niệm về CSDL					
1.2. Ứng dụng của quy trình CSDL					
1.3. Lịch sử của các quy trình 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. Truyền thống hoá trong thiết kế CSDL					
3.3. Các thuộc tính tính năng của các lớp					
3.4. Các mô hình dữ liệu					
3.5. Mô hình thực thể quan hệ					
Chương 4. Điều kiện khai thác thực trạng	3	3			
4.1. Giới thiệu					
4.2. Một số khái niệm					
4.3. Các tính chất khai thác không xung đột					
4.3.1. Một số khái niệm					
4.3.2. Khai thác có thể					
4.3.3. Phân tích thực thể của 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 kiện an toàn dữ liệu					
5.2.1. Điều kiện lưu trữ thông tin					

TÊN CH NG M C	PHÂN PH IS TI T				
	TS	LT	TH	BT	KT
5.2.2. i u khi n suy di n					
5.2.3. i u khi n truy nh p					
Ch ng 6. ánh giá câu h i	3	2			1
6.1. Phân tích câu h i					
6.1.1. Phân tích cú pháp					
6.1.2. Phân tích ng ngh a					
6.2. C u trúc i s câu h i					
6.2.1. Cây i s quan h					
6.2.2. Các lu t bi n i cây i s quan h					
6.2.3. Các b c t i u cây i s quan h					
6.3. Phân rã câu h i					
Ch ng 7. C s d li u h ng it ng	12	5	6		1
7.1. Gi i thi u v h ng it ng					
7.1.1. Các khái ni m h ng it ng					
7.1.2. Mô hình hoá vi c phân tích h ng it ng					
7.1.3. Mô hình hóa d li u					
7.2. Nguyên tác c a các mô hình h ng it ng					
7.2.1. Mô hình hoá các it ng					
7.2.2. Ph ng pháp					
7.2.3. Xác nh d ng d li u					
7.2.4. Các liên k t th a k gi a các l p					
7.2.5. a c u và s áp t					
7.2.6. Xác nh t p các it ng					
7.2.7. Khía c nh ng					
7.2.8. L c CSDL h ng it ng					
7.3. Tính b n v ng các các it ng					
7.3.1. CSDL h ng it ng					
7.3.2. Qu n lý tính b n v ng					
7.3.3. K th a tính b n v ng					
7.3.4. Tính b n v ng do tham chi u					
7.3.5. Tích h p v i ngôn ng l p trình					
7.4. i s v i các it ng ph c t p					
7.4.1. M r ng i s quan h theo ng d n và các ph ng pháp					
7.4.2. Các phép toán i s					
7.4.3. Các phép toán nhóm					
7.4.4. th các phép toán					
Ch ng 8. C s d li u phân tán	15	5	9		1
8.1. C u trúc CSDL phân tán					
8.2. ánh giá CSDL phân tán					
8.3. Thi t k CSDL phân tán					
8.3.1. B n sao d li u					
8.3.2. Chia nh d li u					
8.4. Tính thông su t và tính t tr c a CSDL phân tán					
8.5. X lý câu h i					
8.6. Khôi ph c sai sót trong CSDL phân tán					

Nhi m v c a sinh viên:

Tham d các bu i h c lý thuy t và th c hành, làm các bài t p c giao, làm các bài ki m tra gi a k và bài thi k t thúc h c ph n theo úng quy nh.

Tài liệu h c t p:

1. Nguyễn Xuân Huy, *Giáo trình v c s d li u*, i h c Qu c Gia Hà N i, 2000.
2. Nguyễn Xuân Huy-Lê Hoài B c, *Bài t p c s d li u*, Nhà xu t b n Th ng kê, 2003.
3. Phạm H u Khang, oàn Thi n Ngân, *Qu n tr SQL Server 2000*, Nhà xu t b n Th ng kê, 2004.

Hình th c và tiêu chu n ánh giá sinh viên:

- Hình th c thi: thi vi t.
- Tiêu chu n ánh giá sinh viên: d a trên thái tham d các bu i h c lý thuy t và th c hành, k t qu làm các bài t p c giao, các bài ki m tra gi a k và bài thi k t thúc h c ph n.

Thang i m: Thang i m ch A,B,C,D,F.

i m ánh giá h c ph n: $Z=0,3X+0,7Y$

Bài gi ng này là tài li u **chính th c và th ng nh t** c a B môn H th ng Thông tin, Khoa Công ngh Thông tin và c dùng gi ng d y cho sinh viên.

Ngày phê duy t: / /

Tr ng B môn

M C L C

Ch 1. H qu n tr c s d li u	7
1.1. Quan ni m v CSDL	7
1.2. Các kh n ng c am th qu n tr c s d li u.	7
Ch 2. C s d li u h ng it ng	9
2.1. Nhu c u v h th ng c s d li u h ng it ng	9
2.1.1. Các it ng ph c t p	9
2.1.2. Qu n lý các tri th c	9
2.1.3. Qu n tr các d li u phân tán	10
2.1.4. Nhu c u v h th ng c s d li u h ng it ng	10
2.2. Khái ni m v h ng it ng	11
2.2.1. it ng	12
2.2.2. L p it ng	12
2.2.3. Cá th	13
2.2.4. K th a	13
2.3. C s d li u h ng it ng	13
2.4. Thi t k c s d li u h ng it ng	14
2.4.1. Phân l p	14
2.4.2. T ng quát hóa và c bi t hóa	14
2.4.3. G p	15
2.5. Xây d ng c s d li u h ng it ng	15
Ch 3. C s d li u phân tán	17
3.1. Các ph ng pháp phân tán d li u	17
3.1.1. Khái ni m v phân tán d li u	17
3.1.1.1. Các lý do phân m nh	17
3.1.1.2. Các ki u phân m nh	17
3.1.1.3. M c phân m nh	19
3.1.1.4. Quy t c phân m nh úng n	19
3.1.1.5. Các ki u c p phát	19
3.1.1.6. Các yêu c u thông tin	19
3.1.2. Phân m nh ngang	20
3.1.2.1. Yêu c u thông tin c a phân m nh ngang.	20
3.1.2.2. Phân m nh ngang nguyên th y.	21
3.1.2.3. Phân m nh ngang d n xu t	23
3.1.3. Phân m nh d c	24
3.1.4. C p phát	24
3.2. Kì m soát d li u ng ngh a	26
3.2.1. Qu n lý khung nhìn	26
3.2.1.1. Khung nhìn trong qu n lý t p trung	26
3.2.1.2. C p nh t qua các khung nhìn	26
3.2.1.3. Khung nhìn trong c s d li u phân tán	27
3.2.2. An toàn d li u	27
3.2.2.1. Kì m soát c p quy n t p trung	27
3.2.2.2. Kì m soát c p quy n phân tán	28
3.3. Qu n lý giao d ch và i u khi n ng th i phân tán	28
3.3.1. Các khái ni m c b n v giao d ch	28
3.3.1.1. Tính nguyên t	29
3.3.1.2. M c d li u	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 thi của lịch sử	32
3.3.1.8. Bối cảnh	33
3.3.1.9. Nghi thức	33
3.3.2. Mô hình giao dịch nguyên	33
3.3.2.1. Ý nghĩa của giao dịch – hàm kiểm tra	33
3.3.2.2. Kiểm tra tính khả thi bằng thủ tục	35
3.3.3. Nghi thức khóa 2 pha	35
3.3.4. Mô hình khóa cứng và khóa mềm	36
3.3.4.1. Ý nghĩa của giao dịch với khóa cứng và khóa mềm	36
3.3.4.2. Thủ tục tối ưu hóa trong các giao dịch Rlock và Wlock	36
Chương 4. Hỗ trợ quản lý	38
4.1. Giới thiệu về hỗ trợ quản lý	38
4.2. Thiết kế sơ đồ dữ liệu cho hỗ trợ quản lý	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	40
4.3.1. Kho dữ liệu	41
4.3.2. Kho dữ liệu chuyên	41
4.3.3. Các loại dữ liệu	42
4.4. X lý phân tích truy vấn	43
4.4.1. Giới thiệu	43
4.4.2. Bảng chéo	43
4.4.3. Sơ đồ 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?

Như nghĩa: Một cơ sở dữ liệu (Database) là một tập hợp có cấu trúc các dữ liệu tác nghiệp có lưu trữ lại và các cách ứng dụng thực tiễn.

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

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

Hệ quản trị cơ sở dữ liệu là gì?

Như nghĩa: Hệ quản trị cơ sở dữ liệu (Database Management System - DBMS) là một hệ thống phần mềm cho phép tập hợp CSDL và điều khiển mọi truy cập vào CSDL đó.

Hệ cơ sở dữ liệu là gì?

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

- CSDL hợp nhất: CSDL cơ sở có hai tính chất là tập trung hóa dữ liệu và chia sẻ.
- Như ứng dụng: Ứng dụng cơ sở là một tập hợp nào có nhu cầu truy cập vào CSDL, bao gồm tất cả ứng dụng cụ thể, ứng dụng vì tính năng trình ứng dụng và ứng dụng điều khiển toàn bộ hệ thống hay còn gọi là quản trị CSDL.
- Phần mềm quản trị CSDL.
- Phần cơ sở cơ sở bao gồm các thiết bị như thiết bị phần cứng lưu trữ CSDL.

1.2. Các khái niệm cơ bản về hệ quản trị CSDL.

Có hai khái niệm cho phép phân biệt các hệ quản trị CSDL về các khía cạnh thực tiễn khác:

- Khái niệm quản lý dữ liệu dài hạn;
- Khả năng truy cập các dữ liệu theo nhu cầu.

Cụ thể như thể hiện 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 cập và quản lý. Cụ thể hai phân biệt về hệ quản trị về mặt thực tiễn về tập hợp quản lý dữ liệu dài hạn nói chung không cung cấp các truy cập nhanh chóng đến các bộ phận dữ liệu tùy ý.

Các khái niệm cơ bản về hệ quản trị CSDL hiện tại là cần thiết khi khi lập ứng dụng dữ liệu cần lưu trữ là rất lớn, bởi vì các dữ liệu như thế các kỹ thuật truy cập nên gì, chương trình quét toàn bộ các dữ liệu là thích hợp.

Khi xem xét hai khái niệm trên cơ bản về hệ quản trị CSDL là cần, còn một số các khái niệm khác mà có thể thấy trong hầu hết các hệ quản trị CSDL thông minh, đó là:

- Hệ thống mô hình dữ liệu hay mô hình cấu trúc toán học mà qua đó ứng dụng có thể quan sát dữ liệu.

- M b o tính c l p d li u hay s b t b i n c a các ch ng tr ình ng d ng i v i các thay i v c u trúc trong mô hình d li u.
- H tr các ngôn ng c p cao nh t nh cho phép ng i s d ng nh ngh a c u trúc c a d li u, truy nh p d li u và thao tác d li u.
- Qu n tr giao d ch, có ngh a là kh n ng cung c p các truy c p ng th i, úng n i v i CSDL t nhi u ng i s d ng t i cùng m t th i i m.
- i u khi n truy c p, có ngh a là kh n ng h n ch truy nh p n d li u b i nh ng ng i s d ng không c c p phép và kh n ng ki m tra tính úng n c a d li u.
- Ph c h i d li u, có ngh a là kh n ng ph h i, không làm m t mát d li u i v i các l i c a h th ng.

2.1.2. Quản lý các trị số

Một quan hệ bất đẳng thức nhiều biến, thường hay gọi tắt là bất đẳng thức nhiều biến, là một mệnh đề logic có dạng: $A(x_1, x_2, \dots, x_n)$, trong đó x_1, x_2, \dots, x_n là các biến, A là một mệnh đề logic. Một bất đẳng thức nhiều biến được gọi là đúng nếu mệnh đề A đúng với mọi bộ giá trị của các biến x_1, x_2, \dots, x_n . Một bất đẳng thức nhiều biến được gọi là sai nếu mệnh đề A sai với một bộ giá trị của các biến x_1, x_2, \dots, x_n .

Ng i ta s d ng tri th c này theo hai khía c nh, ng v i hai cách suy lu n:

- a ra s ki n m i, d a trên các s ki n và các tri th c ã bi t.
- S d ng tr l i các câu h i c n n suy lu n.

Do vậy vì cơ quan lý các trị thế trong CSDL là nhu cầu thế t, nh t là i vì các ng d ng c n n các ki n th c chuyên gia.

Hệ quả thứ CSDL sẽ định các điều kiện toàn vẹn. Điều kiện ràng buộc định nghĩa các thuộc tính cần ghi và thuộc tính để liên lạc bị từ là thuộc tính trong các chức năng quan hệ và ngôn ngữ định nghĩa. Khi đưa thuộc tính vào ngôn ngữ hình thức liên lạc hay ngôn ngữ chương trình thì ta cần ghi quy tắc:

- Tri thức cũ mã hóa trong chương trình học để học sinh học lại là một điều vô trùng lặp.
- Việc quy nạp mối liên hệ giữa những kiến thức có dùng tri thức khó có thể tốt nhất vì quy nạp trong trình học phải sử dụng các dữ liệu nhất định.
- Việc suy luận về khả năng liên các thông tin như các số liệu và tri thức trong học tập của học sinh có thể nhận được, dựa trên việc làm một tính hiệu quả toàn bộ học sinh.

2.1.3. Quy trình các dữ liệu phân tán

Mô hình quản trị CSDL thông thường chỉ chú trọng đến phần mềm và cài đặt trên máy tính trung tâm. Hiện nay môi trường tin học phức tạp các công tác quản lý các thông tin, khi cần thì phải dùng nhiều thiết bị không trung tâm và xử lý song song. Như vậy cần có công cụ khác là hệ quản trị CSDL phân tán.

H phân tán t o l p do vì c t p h p các máy n i nhau theo m ng truy n thông dùng cho m t công vi c t ng th và c qu n lý trên a bàn r ng l n. Yêu c u t ra ây là qu n tr và x lý nh ng đ li u phân tán t i các máy c l p. Các tr m c l p y ã có các h qu n tr , nh ng không hoàn toàn gi ng nhau.

khai thác d li u theo h th ng qu n tr CSDL phân tán, ng i ta không th không thay i h qu n tr c . Ít ra c ng ph i m r ng h qu n tr CSDL t p trung. H th ng phân tán c ng òi h i các ch c n ng x lý song song.

Các hình thức vận dụng các chức năng ngữ pháp song song để thể hiện khai thác các khả năng ngữ pháp song song của máy phân tích. Điều kiện vận dụng các thủ pháp cho phép xử lý đa lượng một cách song song, cần có các ngôn ngữ bậc cao cho phép mô tả đa lượng phân tán và mô tả song song các chức năng ngữ pháp.

Như vậy, để tìm các hàm lượng tử CSDL thì hai điều kiện đầu tiên là các hàm lượng tử CSDL phải là hàm nguyên, CSDL suy diễn và CSDL phân tán.

2.1.4. Nhu c u v h th ng CSDL h ng i t ng

Các thí dụ về CSDL thể nghiệm lý thuyết lập trình và xử lý dữ liệu truy vấn thông tin. Một ứng dụng xử lý dữ liệu có cấu trúc là dùng các tập dữ liệu, xử lý dữ liệu tập hợp để đáp ứng các yêu cầu. Các công nghệ CSDL quan hệ, mạng hay phân cấp đều thể hiện các tính chất khác nhau như tích hợp các tập dữ liệu để đáp ứng nhu cầu ứng dụng. Các đặc điểm của CSDL quan hệ là khi cần chúng chỉ mô tả trên thực tế CSDL. Tuy nhiên vì tính tổng quát, trừu tượng của các kiểu dữ liệu do máy xử lý lập trình mà ý nghĩa của công nghệ quan hệ. Trong khi tập dữ liệu hình ảnh thì trong

CSDL, bên cạnh các hình thức của CSDL quan hệ, ngày nay còn thấy các hình thức nêu trên và hình thức CSDL nói chung, và phạm vi nghiên cứu cấu trúc dữ liệu, và tính chính xác và tính toàn vẹn dữ liệu.

1. Phạm vi nghiên cứu: Lý thuyết quan hệ chính trị giữa các khái niệm nghiên cứu. Nghiên cứu khái niệm về nghiên cứu không thể hiện qua mô hình quan hệ.

2. Cấu trúc dữ liệu: Các hệ thống quan hệ dữ liệu cần phải có cấu trúc dữ liệu. Trong hệ thống quan hệ các dữ liệu được tổ chức thành bảng gồm các thuộc tính. Nhiệm vụ của thông tin không phù hợp với cách thức hiện nay.

3. Tính thống nhất của đề tài: Trong hệ thống quan hệ, đề tài chủ yếu là thống nhất. Các chương trình thống nhất có vai trò kích hoạt các đề tài này, đề tài từ nên linh hoạt. Trong mô hình hệ thống thống nhất, thống nhất có thể bị tác động hành vi của người học cấu trúc đề tài.

4. Toàn văn nghị quyết: Mối liên hệ giữa toàn văn nghị quyết là có thể gì? Tính bản văn nghị quyết a nghị quyết CSDL để tác động các chính trị. Hình thức quan hệ quản lý hành vi cá nhân dựa trên các chính trị để động. Cùng một thể lệ có thể mang các hành vi khác nhau, nên dựa trên có ý nghĩa khác nhau tùy theo chính trị xã hội, để gây nên tình trạng vi phạm tính toàn văn dựa trên. Như tình huống kém khả năng nhúng vào CSDL quan hệ xã hội CSDL hành vi cá nhân kinh tế, nhất là dựa trên các thể lệ pháp lý và có khả năng liên các thông tin nghị quyết. Các lĩnh vực cá nhân hình thành CSDL hành vi cá nhân là:

- CSDL a ph ng ti n, c n l u tr kh i l ng l n các tín hi u âm thanh, hình nh và v n b n;
c n liên k t nhi u ki u d li u.

- Các hình thức thông tin a lý, c n x lý nhi u lo i d li u th ng kê, b n ; các d li u c thu th p t nhi u vùng khác nhau.

- Các CSDL ph c v thi t k g m các s , nhi u thành ph n đ li u ph c t p liên k t nhau. M t thi t k òi h i x lý qua nhi u giai o n, nhi u th h .

2.2. Khái niệm và hình thức

th y quan i m v h ng i t ng, hăy xét m t v t trong th gi i th c là chi c gh . Gh là m t ph n t , hay là m t th hi n c a l p r ng h n g i là c. M t t p các thu c tính liên k t v i i t ng trong l p c, ch ng h n giá thành, kích th c, tr ng l ng, v trí và m u s c. Nh ng i u này c áp d ng m i khi ng i ta nói v cái bàn hay cái gh , t B i vì gh là thành viên c a l p c, nó th a k t t c các thu c tính ã xác nh cho l p.

M i i t ng trong l p c có th c x lý theo nhi u cách. M i phép x lý này s thay i m t hay nhi u thu c tính c a i t ng và chúng c g i là d ch v hay ph ng pháp trên i t ng.

Các í t ñ g s bao b c:

- D li u thông qua giá tr thu c tính.
- Các phép toán nh các ho t ng có tác d ng thay i giá tr thu c tính.
- Các i t ng khác, nh là các i t ng ph c t p.
- Các h ng s , nh các giá tr m c nh.
- Các thông tin liên quan.

Vì có bao nhiêu thông tin của các biến có nghĩa là tất cả thông tin này được thu gọn lại để tìm tên và có thể dùng như một biến thay thế thành phần chính trong trình.

2.2.1. Biến

Khái niệm về biến là khái niệm sinh ra từ việc nhận thức thế giới thực. Một biến có các tính chất sau:

- Mang tên duy nhất, không thay đổi.
- Thuộc phạm trù.
- Có thể ghi các thông báo về các biến khác.
- Có trạng thái riêng.

nh nghĩa 2.1. Biến là một thực thể có vai trò xác định rõ ràng trong lĩnh vực ứng dụng, có trạng thái, hành vi và các xác định tên.

Ví dụ các biến thuộc lập trình Ng. Chúng liên lạc với nhau thông qua thông báo. Thông báo là dùng các thao tác áp dụng lên biến. Thao tác trong môi trường lập trình biến có nghĩa là phép toán. Chúng nhận phép toán kết hợp tác động lên biến biến lập trình này kết hợp với ai.

nh nghĩa 2.2. Trạng thái bao gồm tính chất của biến, như là thuộc tính và các mối quan hệ, và như giá trị gán cho các tính chất đó.

nh nghĩa 2.3. Phép toán là một hàm số hay một dịch vụ mà tất cả các thể hiện của lập trình chấp nhận.

nh nghĩa 2.4. Phép toán là việc thực hiện một phép toán.

Hành vi và cấu trúc của một biến trong môi trường lập trình lập trình biến hoàn toàn do lập trình biến xác định. Lập trình biến là khái niệm cơ bản trong lập trình biến.

nh nghĩa 2.5. Hành vi là thể hiện cách thức tác động của một biến.

2.2.2. Lập biến

Một lập biến có giao diện và miền riêng. Giao diện của biến là cái mà các biến khác thấy được. Giao diện lập biến gồm hai thành phần sau:

1. Thuộc tính của lập: Trong chương trình nào đó các thuộc tính của lập được coi là biến và các thuộc tính của quan hệ. Như nhiên thông qua thuộc tính lập biến ta có thể thể hiện các liên kết giữa các biến, hay trong lập trình có thuộc tính nào, chúng nhận thuộc tính từ biến có giá trị tùy thuộc vào thuộc tính ngày sinh. Vì vậy lập, không có hạn chế về cách thức cấu trúc thuộc tính hay cách liên kết với nhau. Các thuộc tính có thể là biến, dùng để biến lập trình biến biến. Như vậy ta không xử lý trực tiếp các giá trị thuộc tính của biến mà xử lý thông qua các phép toán liên kết với lập biến.

2. Phép toán gán với lập: Các thông báo chuyển đổi lập biến như phép toán gán với lập biến. Chúng thể hiện có dùng các phép toán, các hàm với các tham số. Các biến trong một lập biến có thể truy cập thông qua các phép toán. Tìm kiếm giao diện, phần hiển thị rõ là tên phép toán và các tham số cần cho phép toán này.

nh nghĩa 2.6. Lập biến là tập các biến có chung cấu trúc và hành vi.

nh nghĩa 2.7. Sự liên kết cho biết cấu trúc của mô hình hệ thống, đó là các liên kết, cấu trúc bên trong của chúng, và mối quan hệ mà chúng tham gia.

Mỗi riêng của liên kết là phần xác định, nhưng không hiển thị ra cho liên kết khác thấy. Mỗi liên kết có thể gồm các thông tin chi tiết về cấu trúc của liên kết. Trong chương trình, người ta có thể cài đặt để ghi lại các giá trị, ngay cả các liên kết để ghi lại, cũng như không thấy các thông báo. Vì vậy, thông tin làm việc bên trong cùng với các giá trị liên kết các sự kiện thông thường này cũng là để che giấu thông tin.

Một khái niệm quan trọng trong OOP là bao bọc, có nghĩa là ẩn đi những gì mà liên kết ẩn đi thông qua những liên kết. Người ta truy cập khái niệm như giao diện liên kết và xác định các hành vi thông qua việc xác định liên kết.

nh nghĩa 2.8. Sự liên kết là thể hiện các liên kết, thể hiện thích hợp liên kết.

nh nghĩa 2.9. Bao gói, hay bao bọc là kỹ thuật để ghi lại, làm nên những chi tiết cài đặt bên trong của liên kết và các truy cập bên ngoài.

2.2.3. Cá thể

Cá thể hóa là quá trình khi nào sự tồn tại của các liên kết trong môi trường hệ thống, bằng việc xác định liên kết của chúng. Mỗi liên kết là một cá thể của liên kết; thể hiện cũng dùng với thuật ngữ thể hiện của liên kết.

2.2.4. Khả năng

Khái niệm khả năng là khái niệm quan trọng trong tính năng của liên kết. Người ta thường dùng thuật ngữ này khi chỉ định liên kết này tính năng, thể hiện các thuộc tính của liên kết khác. Tuy nhiên mỗi liên kết có thể mang một số thuộc tính hay phương pháp riêng.

nh nghĩa 2.10. Khả năng là tính chất cho phép các liên kết có thể như những thuộc tính, phép toán của liên kết.

Việc khả năng của liên kết xảy ra khi một liên kết thể hiện của liên kết.

2.3. CSDL hệ thống

Dù có nhiều ngôn ngữ hệ thống, các CSDL hệ thống dựa trên C++. Lựa chọn này là do tính hiệu quả và thông dụng của C++.

Thuyết CSDL hệ thống có các nội dung:

- Cho phép xét các liên kết liên kết để định các phép toán với các liên kết.
- Các liên kết có phép dùng chung cho nhiều ứng dụng.
- Khả năng phát triển kho trữ thể hiện cách thêm các liên kết mới và các phép xử lý kèm theo.
- Phát triển hệ quản trị CSDL dựa trên việc xử lý các liên kết phức tạp, thiết kế giao diện chương trình, một liên kết và truy vấn.

Hệ quản trị CSDL hệ thống là hệ quản trị cho phép lưu trữ và chia sẻ các liên kết với nhiều ứng dụng. Hệ thống hệ thống là sự kết hợp có ý nghĩa của các trình hệ thống. Trong môi trường OOP các liên kết coi như các biến chương trình, chi tiết nội dung vòng lặp của chương trình để tạo ra và sử dụng liên kết. Còn trong hệ thống CSDL hệ thống, các liên kết là bản vẽ hệ thống. Thuật ngữ bản vẽ hệ thống dùng để chỉ những liên kết nội dung không

thu c vào vòng i c a ch ng trnh t o ra nó. Ch ng trnh khác có th truy c p hay h y b i t ng này. H th ng CSDL h ng i t ng có các nét c tr ng sau:

- Ngôn ng CSDL có kh n ng mô t l p i t ng, tạo sinh, l u tr và xóa i t ng.

- Các i t ng cho phép ch ng trnh ng d ng truy c p. M i i t ng trong kho các i t ng có tên duy nh t OID. Khái ni m OID khác v i khái ni m khóa chính trong c s d li u quan h . Khóa chính là t p các thu c tính xác nh duy nh t b d li u, mang giá tr có th thay i c. Trong h th ng h ng i t ng, OID th ng là chu i 64, 128bit ho c cao h n và mang giá tr không thay i.. M t i t ng c ng có th có các thu c tính mang vai trò nh khóa chính trong CSDL quan h . i u này hoàn toàn ph thu c vào vi c xác nh l p. Vi c gán OID hoàn toàn do h qu n tr CSDL h ng i t ng th c hi n.

Kho i t ng là n i ch a d li u t o nên CSDL h ng i t ng. H qu n tr CSDL h ng i t ng c n có các ch c n ng c b n nh h qu n tr CSDL bình th ng, nh i u khi n t ng tranh, an toàn d li u, toàn v n d li u... Th c ra, h qu n tr CSDL h ng i t ng ho t ng nh h qu n tr th ng, v i s tham giá c a ph n m m h ng i t ng truy c p các i t ng.

2.4. Thi t k CSDL h ng i t ng

xác nh rõ vi c th c hi n CSDL h ng i t ng theo các d ng ti p c n, tr c h t c n xác nh cách th c thi t k m t h th ng h ng i t ng.

M t CSDL h ng i t ng g m các i t ng; t t c các th thu c v l p. xây d ng mô hình khái ni m cho CSDL h ng i t ng, ng i ta c n xác nh t p các l p i t ng.

L p i t ng c xác nh nh các thu c tính và ph ng th c th c. Ng i ta c ng ph i xác nh s t ng tác gi a các l p. Khái ni m v l p c ng t ng t nh ki u d li u ã gi i thi u tr c ây.

2.4.1. Phân l p

Quá trình phân l p liên quan n vi c nh tên i t ng v i các thu c tính, hành vi t ng t nhau và nhóm các i t ng vào cùng m t l p. Theo thí d v s ng i ta xác nh s v i các thu c tính tên, ngày t o, hình v . Các phép toán chung là l u tr tìm ki m, v .

Trong o n ch ng trnh trên, danh sách các tr ng và các ki u d li u n gi n dùng cho các s c li t kê trong m c thu c tính. Ti p theo là các ph ng th c, có tên và các tham s . Có m t s ph ng th c nh t o m i, xóa... áp d ng cho t t c các i t ng trong CSDL.

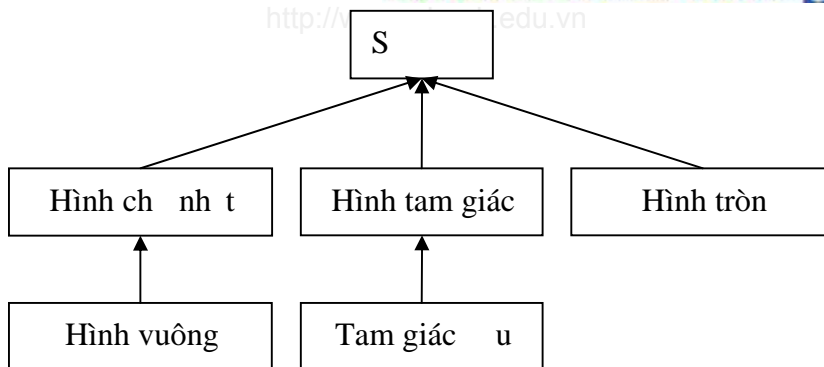
T t c các nh ngh a v giao di n l p i t ng c n có phép toán t o m i và h y b i t ng. Quá trình phân l p s t o l p c a các i t ng có các thu c tính, ph ng th c chung, và m t vài i t ng có thu c tính và ph ng th c riêng. Lúc ó ng i ta c n n khái ni m t ng quát hóa và chuyên bi t hóa.

2.4.2. T ng quát hóa và c bi t hóa

T ng quát hóa là quá trình xác nh l p i t ng mang các thu c tính t ng t và theo s t ng t này ng i ta có th tr u t ng hóa c l p cao h n, hay l p cha. Ch ng h n ban u ng i ta xác nh l p hình tam giác, hình vuông, hình ch nh t, và hình tròn r i tr u t ng hóa thành l p cao h n v s , g m các thu c tính chung c a t t c các s .

nh ngh a 2.11. L p tr u t ng là l p không có th hi n tr c ti p, nh ng các thành phân sau nó có th có th hi n tr c ti p.

nh ngh a 2.12. L p c th là l p có th có các th hi n tr c ti p.



Chuyên biệt hóa là quá trình ngược lại với tổng quát hóa. Bởi vì tập lớp, nên nó có thể xác định lớp con phân biệt các loại khác nhau; mỗi lớp con chia sẻ thuộc tính và phương thức chung trong lớp nhưng có các thuộc tính và phương thức riêng.

Chúng ta dùng cây phân cấp để thể hiện quá trình tổng quát hóa. Phân cấp như vậy có ý nghĩa trong hình thức học máy, để đưa ra các thuật toán. Khi mô tả các lớp, chúng ta cần đưa ra sự tham gia của lớp vào dãy kết quả này.

Hai hình học là lớp để tổng quát này tham gia vào lớp để tổng quát của hình học là hình tam giác. Lớp tam giác là tập các thuộc tính và phương thức của nó. Các thuộc tính bổ sung của nó mô tả ngay. Phương thức tính toán mô tả tính toán các thuộc tính riêng của hình tam giác. Phương thức Di tích của nó là phương thức riêng cho hình tam giác. Còn lớp để tổng quát Hình học thì là tập các thuộc tính của lớp Tam giác. Vì vậy nó có nghĩa là nó là tập các thuộc tính của lớp tam giác và lớp tổng quát. Lớp để tổng quát này có thuộc tính riêng. Chúng ta mô tả các phương thức phù hợp với hình này.

2.4.3. Gp

Gp là quá trình liên kết các lớp để tổng quát với nhau để tạo nên một tập lớp gp. Chúng ta cần CSDL ngân hàng gồm khách hàng, tài khoản, chi nhánh và mối quan hệ giữa chúng. Khi thể hiện CSDL này theo kết cấu quan hệ, chúng ta tạo các quan hệ tách biệt về mặt nội dung và dùng khái niệm khóa ngoài để thể hiện mối quan hệ 1-n. Chúng ta cần dùng quan hệ khác có khóa ngoài thể hiện quan hệ m-n giữa khách hàng và tài khoản. Khi tập lớp khung nhìn về mặt thông tin liên quan đến tài khoản nào đó, chúng ta cần biết 4 bước.

Trong CSDL hình học để tổng quát chúng ta gọi quy tắc này như là lớp để tổng quát gp. Lớp gp ký hiệu là tập, liên kết các lớp để tổng quát khác nhau, tài khoản, và chi nhánh. Mỗi lớp để tổng quát ký hiệu là tập khách hàng và tài khoản, có thuộc tính để tổng quát cho tập thể hiện của hình học này.

2.5. Xây dựng CSDL hình học để tổng quát

Một số liên quan đến CSDL hình học để tổng quát xuất hiện khi phân lớp để tổng quát, tổng quát hóa, chuyên biệt hóa và gp. Các vấn đề này xảy ra song song. Chúng ta không trình bày quá trình chuyển mô hình thực thể E/R sang mô hình hình học để tổng quát không có ý nghĩa gì mà bên ngoài là thực thể và mối quan hệ, mà bên ngoài là lớp để tổng quát. Hơn nữa, mô hình E/R như hình học để tổng quát không cho biết vị trí hành vi của thực thể khi đưa vào hình học của quan hệ mà thực thể tham gia.

Vì c chuyển n hóa mô hình i t ng ng ngh a sang CSDL h ng i t ng c th c hi n nh sau:

- i v i m i i t ng ng ngh a, ng i t a t o m t l p i t ng.
- Khi có quan h cha con gi a hai i t ng ng ngh a, ng i t a th c hi n liên k t th a k gi a l p i t ng th hi n ki u con và l p i t ng th hi n ki u cha.
- T i n hành g p và liên k t các i t ng ng ngh a v ki u i t ng c các l p i t ng g p.
- Gi tính toàn v n v h ng gi a các i t ng ng ngh a, duy trì vi c h p nhóm c a các i t ng ng ngh a và phân rã ki u con b ng các ph ng th c.

Vòng i phát tri n h ng i t ng g m ba pha:

- Pha phân tích: ng i t a phát tri n mô hình ng d ng trên th gi i th c, ch ra các thu c tính quan tr ng. C n xác nh các khái ni m tr u t ng v mô hình xu t phát t l nh v c ng d ng và các mô t v h th ng. Ng i t a xác nh các hành vi ch c n ng c a h th ng, c l p v i môi tr ng s th c hi n h ng i t ng.

Pha thi t k : Cho phép xác nh cách th c s th c hi n mô hình phân tích h ng i t ng.

Pha th c hi n: Ng i t a xác nh và kh o sát t t c các k t qu c a vi c th c hi n thi t k . T t c các quy t nh thi t k chi n l c, nh là cách th c tích h p h qu n tr CSDL, cách liên l c gi a các i t ng, cách x lý sai sót... u c tri n khai. T i p theo ng i t a tích h p các quy t nh này vào mô hình thi t k ban u, r i hình th c hóa mô hình thi t k các i t ng có th t ng tác v i các i t ng khác theo các k ch b n.

Chương 3. Các số liệu phân tán

Thi t k h th ng thông tin có CSDL phân tán bao g m:

- Phân tán và ch n nh ng v trí t đ li u;
- Các ch ng trình ng d ng t i các i m;
- Thi t k t ch c khai thác h th ng ó trên n n m ng.

3.1. Các ph ng pháp phân tán d li u

3.1.1. Khái ni m v phân tán d li u

Khi thi t k các h th ng CSDL phân tán ng i ta th ng t p trung xoay quanh các câu h i?

- T i sao l i c n ph i phân m nh?
- Làm th nào th c hi n phân m nh?
- Phân m nh nên th c hi n n m c nào?
- Có cách gì ki m tra tính úng n c a vi c phân m nh?
- Các m nh s c c p phát trên m ng nh th nào?
- Nh ng thông tin nào s c n thi t cho vi c phân m nh và c p phát?

3.1.1.1. Các lý do phân m nh

Tr c tiên vi c phân tán d li u c th c hi n trên c s c p phát các t p tin cho các nút trên m t m ng máy tính. Các nút m ng th ng n m các v trí a lý khác nhau tr i r ng trên m t di n tích l n. Do v y t i u vi c khai thác thông tin thì d li u không th t p trung mà ph i phân tán trên các nút c a m ng.

H n n a m t quan h không ph i là m t n v truy xu t đ li u t t nh t. Ví d nh , n u ng d ng c th c hi n trên m t b ph n nh các d li u c a quan h mà quan h ó n m t i các v trí khác nhau thì có th gây ra nh ng truy xu t th a và h n th vi c nhân b n các quan h làm t n không gian b nh . Do v y phân rã m t quan h thành nhi u m nh, m i m nh c x lý nh m t n v s cho phép th c hi n nhi u giao d ch ng th i. M t câu truy v n ban u có th c chia ra thành m t t p các truy v n con, các truy v n này có th c th c hi n song song trên các m nh s giúp c i thi n t c ho t ng c a h th ng.

Tuy nhiên chúng ta c ng s g p nh ng r c r i c a vi c phân m nh, ví d n u các ng d ng có nh ng xung t s ng n c n ho c gây khó kh n cho vi c truy xu t đ li u. Phân rã các m nh nói chung làm t ng chi phí trong vi c truy xu t đ li u. M t v n n a liên quan n vi c ki m soát ng ngh a và tính toàn v n đ li u.

3.1.1.2. Các ki u phân m nh

Th hi n c a các quan h chính là các b ng, vì th v n là tìm nh ng cách khác nhau chia m t b ng thành nhi u b ng nh h n. Rõ ràng có hai ph ng pháp khác nhau: Chia b ng theo chi u d c và chia b ng theo chi u ngang. Chia đ c ta c các quan h con mà m i quan h ch a m t t p con các thu c tính c a quan h g c – g i là phân m nh đ c. Chia ngang m t quan h ta c các quan h con mà m i quan h ch a m t s b c a quan h g c - g i là phân m nh ngang.

Ngoài ra còn có m t kh n ng h n h p, ó là phân m nh k t h p cách phân m nh ngang và d c. T t nhiên quá trình phân m nh g n li n v i v n c p phát và bài toán c th .

Ví d 3.1.

Trong ví d này chúng ta s d ng m t CSDL c a m t công ty máy tính th c hi n các d án ph n m m g m các quan h :

DuAn(MaDuAn, Ten, KinhPhi, ViTri);

NhanVien(MaNv, Ten, ChucVu);

TrachNhiem(MaNv, MaDuAn, TrachNhiem, ThoiGianTG);

Luong(ChucVu, Luong).

Hình 3.1. D li u c a các b ng

NhanVien			TrachNhiem			
MaNV	Ten	ChucVu	MaNV	MaDuAn	TrachNhiem	ThoiGianTG
E1	Hoàng Lan	Ks i n	E1	P1	Qu n lý	12
E2	ình V ng	Phân tích h th ng	E2	P1	Phân tích	24
E3	Minh Tài	Ks Máy	E2	P2	Phân tích	6
E4	D ng Hà	L p trình viên	E3	P3	T v n	10
E5	Minh Hoa	Phân tích h th ng	E3	P4	K s	48
E6	V n Hi n	Ks i n	E4	P2	L p trình viên	18
E7	Hoài Nam	Ks Máy	E5	P2	Qu n lý	24
E8	Vân D ng	Phân tích h th ng	E6	P4	Qu n lý	48
			E7	P3	K s	36
			E8	P3	Qu n lý	40

DuAn				Luong	
MaDuAn	Ten	KinhPhi	Vitri	ChucVu	Luong
P1	Trang thi t b	150000	Hà N i	Ks i n	4000
P2	CSDL	135000	H i Phòng	Phân tích h th ng	10000
P3	CAD/CAM	250000	Hà N i	Ks Máy	3500
P4	B o Trì	310000	Qu ng Ninh	L p trình viên	4100

Ng i ta có th chia ngang quan h DuAn Thành các quan h con DuAn1, DuAn2. DuAn1 ch a nh ng thông tin v các d án có ngân sách d i 200000\$, còn DuAn2 l u các thông tin v các d án có ngân sách trên 200000\$.

Hình 3.2. Phân m nh ngang quan h DuAn

DuAn1			
MaDuAn	Ten	KinhPhi	Vitri
P1	Trang thi t b	150000	Hà N i
P2	CSDL	135000	H i Phòng

DuAn2			
MaDuAn	Ten	KinhPhi	Vitri
P3	CAD/CAM	250000	Hà N i
P4	B o Trì	310000	Qu ng Ninh

Ngoài ra c ng có th phân m nh d c quan h DuAn thành hai quan h DuAn3, DuAn4. DuAn3 ch ch a thông tin v ngân sách c a các d án, còn DuAn4 l u thông tin v tên và v trí d án. i u c n l u ý là khóa c a quan h DuAn ph i xu t hi n trong c hai m nh.

Hình 3.3. Phân m nh d c quan h DuAn

DuAn3

MaDuAn	KinhPhi
P1	150000
P2	135000
P3	250000
P4	310000

DuAn4

MaDuAn	Ten	Vitri
P1	Trang thi t b	Hà N i
P2	CSDL	H i Phòng
P3	CAD/CAM	Hà N i
P4	B o Trì	Qu ng Ninh

3.1.1.3. M c phân m nh

Phân m nh CSDL n m c nào là m t quy t nh r t quan trọng có nh h ng n hi u n ng th c hi n v n tin. M c phân m nh có th là t thái c c không phân m nh nào n thái c c phân m nh thành t ng b ho c t ng thu c tính. Tuy nhiên n u phân m nh quá nh s có nh ng tác ng không t t n ho t ng khai thác CSDL. V y c n ph i nh ra c m t m c phân m nh thích h p. M c này s tùy thu c vào t ng CSDL và các ng d ng CSDL c th .

3.1.1.4. Quy t c phân m nh úng n

Chúng ta s ph i tuân th ba qui t c trong khi phân m nh mà chúng s m b o CSDL không có thay i nào v m t ng ngh a sau khi phân m nh.

1. Tính y : N u m t quan h R c phân m nh thành các m nh R_1, R_2, \dots, R_N thì m i m c d li u có trong R ph i có m t trong m t ho c nhi u m nh R_i .

2. Tính tái thi t c: N u m t quan h R c phân m nh thành R_1, R_2, \dots, R_N thì c n ph i nh ngh a m t toán t tái thi t Θ sao cho:

$$R = \Theta R_i, i = 1..N.$$

Toán t Θ thay i tùy theo t ng lo i phân m nh; thông th ng khi phân m nh ngang thì Θ là phép toán h p còn phân m nh d c là phép toán k t n i.

3. Tính tách bi t: N u quan h R c phân m nh ngang thành các m nh R_1, R_2, \dots, R_N và m c d li u t i n m trong m nh r_i , thì nó s không n m trong các m nh R_k v i $k \neq j$. Tiêu chu n này m b o r ng các m nh ngang s tách bi t v i nhau. N u quan h c phân m nh d c, các thu c tính khóa chính ph i c l p l i trong m i m nh. Vì th trong tr ng h p phân m nh d c, tính tách bi t ch c nh ngh a trên các tr ng không ph i là khóa chính c a m t quan h .

3.1.1.5. Các ki u c p phát

Gi s CSDL ã c phân m nh h p lý và c n quy t nh c p phát các m nh cho các v trí trên m ng. Khi d li u c c p phát, nó có th c nhân b n ho c duy trì m t b n duy nh t.

Lý do c n ph i nhân b n là nh m m b o tin c y và hi u qu cho các câu v n tin ch c. N u có nhi u b n sao c a m t m c d li u thì chúng ta v n có c h i truy xu t c d li u ngay c khi h th ng x y ra s c . H n n a các câu v n tin ch c truy xu t n cùng m t m c d li u có th cho th c hi n song song b i vì các b n sao có m t t i nhi u v trí. Ng c l i câu v n tin c p nh t có th gây ra nhi u r c r i b i vì h th ng ph i b o m r ng t t c các b n sao ph i c c p nh t chính xác. Vì v y quy t nh nhân b n c n c cân nh c và ph thu c vào t l gi a các câu v n tin ch c và câu v n tin c p nh t. Quy t nh này có nh h ng n t t c các thu c toán c a h qu n tr CSDL phân tán và các ch c n ng ki m soát khác.

3.1.1.6. Các yêu c u thông tin

M t i u c n l u ý trong vi c thi t k phân tán là quá nhi u y u t có nh h ng n m t thi t k t i u. T ch c logic c a CSDL, v trí các ng d ng, c tính truy xu t c a các ng d ng n

CSDL, và các tính chất hệ thống máy tính tìm vị trí có nhúng các quy tắc phân tán.

3.1.2. Phân minh ngang

Trong phần này chúng ta sẽ bàn về hai chỉ số phân minh:

- Phân minh ngang nguyên thủy: Phân minh ngang nguyên thủy của một quan hệ R thể hiện dựa trên các vị trí của nó trên quan hệ R .

- Phân minh ngang dẫn xuất: Là phân minh một quan hệ dựa vào các vị trí của nó trên một quan hệ khác.

3.1.2.1. Yêu cầu thông tin của phân minh ngang.

1. Thông tin về CSDL

Thông tin về CSDL là tất cả các quan hệ R . Trong ngữ cảnh này chúng ta cần biết các quan hệ R kết nối với nhau bằng phép kết nối hay phép tính khác. Thể hiện ngữ nghĩa của chúng bằng mô hình thể thức biểu diễn các quan hệ và các mối liên kết giữa chúng.

Thông tin về R cần có là tất cả các thuộc tính R , đó là số có trong R ký hiệu là $\text{card}(R)$.

2. Thông tin về ngữ nghĩa

phân tán ngoài thông tin về $\text{Card}(R)$ ta còn cần thông tin về tính chất của các vị trí dùng trong các câu văn tin. Loại thông tin này phụ thuộc vào bài toán cần giải.

Cho một quan hệ $R(U)$, $U = A_1, A_2, \dots, A_N$ trong đó mỗi A_i là một thuộc tính có miền giá trị $\text{dom}(A_i)$. Một vị trí n gọi là P của R có dạng:

$P: A_i \theta \langle \text{giá trị} \rangle$

Trong đó $\theta \in \{=, <, <=, >=, >, <>\}$, A_i là một thuộc tính, $\langle \text{giá trị} \rangle \in \text{dom}(A_i)$.

Như vậy cho trước một R , nếu các miền giá trị $\text{dom}(A_i)$ là hữu hạn chúng ta có thể xác định tập tất cả các vị trí n trên R .

Ví dụ: Với hình 3.1. các vị trí n của quan hệ DuAn:

P1: Ten = "Bồ đề"

P2: KinhPhi \leq 200000

Trong các bài toán thể thức các câu văn tin thể thức của nhiều vị trí phức tạp hơn, là tập hợp của các vị trí n . Ví dụ vị trí n của các vị trí n . Một vị trí n thể thức boolean luôn có thể biến đổi thành dạng chuẩn n cho nên chúng ta sẽ dùng vị trí n để biểu diễn các vị trí phức tạp.

Cho một quan hệ $R(U)$, $U = A_1, A_2, \dots, A_N$ trong đó mỗi A_i là một thuộc tính có miền giá trị $\text{dom}(A_i)$. $P = \{p_1, p_2, \dots, p_t\}$. Tập các vị trí n $M = \{m_1, m_2, \dots, m_k\}$ của R như sau:

$$m_i = \bigwedge_{j=1}^t p_j \text{ với } 1 \leq j \leq t$$

Trong đó $p_j = p_j$ hoặc $p_j = \neg p_j$

Lưu ý: Phép lấy phủ định một vị trí không phải lúc nào cũng thể hiện được.

Theo ngữ nghĩa thông tin về tính vị trí của chúng ta cần biết hai tập dữ liệu:

1.. S l ng các b có quan h s c truy xu t b i câu v n tin c c t theo m t v t h i s c p ã cho.

3. T n s truy xu t và trong m t s tr ng h p là s truy xu t. N u $Q = \{q_1, q_2, \dots, q_N\}$ là các câu v n tin, $\text{acc}(q_i)$ bi u th cho s truy xu t c a q_i trong m t kho ng th i gian ã cho ho c trong quan h c th .

3.1.2.2. Phân m nh ngang nguyên th y.

Phân m nh ngang nguyên th y c nh ngh a b ng m t phép toán ch n trên các quan h ch c a m t l c CSDL. Vì th cho quan h R, các m nh ngang c a R là các R_i v i $R_i = R(E_i)$, $1 \leq i \leq k$

Trong ó E_i là công th c ch n (h i s c p) c s d ng có th có c m nh R_i . Chú ý r ng n u E_i có d ng chu n h i, nó là m t v t h i s c p m_i . R_i là các b c a R th a m ã E_i .

Ví d :

Phân rã quan h DuAn thành các m nh ngang DuAn1, DuAn2 trong hình 3.1. c nh ngh a nh sau:

DuAn1 = DuAn(KinhPhi \leq 200000)

DuAn2 = DuAn(KinhPhi $>$ 200000)

M t m nh ngang R_i c a quan h R có ch a t t c các b R th a v t h i s c p m_i . Vì v y cho m t t p các v t h i s c p M, s l ng các m nh ngang c ng b ng s l ng các v t h i s c p (t t nhiên m t m nh ngang có th r ng n u v t không truy xu t n b nào c a quan h , các m nh ngang có th b ng nhau n u các v t t ng ng). T p các m nh ngang này c ng th ng c g i là t p các m nh h i s c p.

Rõ ràng vì c nh ngh a các m nh ngang ph thu c vào cá v t h i s c p. Vì th b c u tiên c a m i thu t toán phân m nh là ph i xác nh t p các v t n gi n s t o ra các v t h i s c p. T p v i t này nói chung ph thu c vào m c tiêu và yêu c u c a bài toán.

M t s khái ni m quan tr ng c a các v t n gi n là tính y và c c ti u c a t p các v t .

T p các v t n gi n Pr c g i là y n u và ch n u xác su t c a m i b c a R (ho c c a các m nh R_i) c truy xu t b i các v t $p \in Pr$ u b ng nhau.

Lý do c n ph i b o m tính y là vì các m nh thu c theo t p v t y s nh t quán v m t logic do t t c chúng u th a v t h i s c p. Chúng c ng ng nh t và y v m t th ng kê theo cách mà ng d ng truy xu t chúng.

c tính th hai c a t p v t là tính c c ti u. ây là m t c tính c m tính. V t n gi n phải có liên i trong vì c xác nh m t phân m nh. M t v t mà không tham gia vào m t t p phân m nh nào thì có th coi v t ó là th a. N u t t c các v t c a Pr u có liên i và không có các v t t ng ng thì Pr là c u ti u.

Chúng ta quy c:

Thu t toán 3.1. COM-MIN Thu t toán tìm t p v t y và c c ti u

Vào: Quan h R, t p các v t n gi n Pr

Ra: Pr' – t p các v t n gi n y và c c ti u

Ph ng pháp:

0. Giả sử tập các mệnh đề đúng

$$Pr' = \emptyset, F = \emptyset$$

1. $\forall p \in Pr$, nếu phân hoạch R theo quy tắc 1

$$- Pr' = Pr' \cup \{p\}$$

$$- Pr = Pr - \{p\}$$

$$- F = F \cup \{f\} \text{ nếu } f \text{ là mệnh đề đúng theo } p$$

2. $\forall p \in Pr$ nếu phân hoạch mệnh đề $f \in F$ theo quy tắc 1

$$- Pr' = Pr' \cup \{p\}$$

$$- Pr = Pr - \{p\}$$

$$- F = F \cup \{f\} \text{ nếu } f \text{ là mệnh đề đúng theo } p$$

Lưu ý bước 2 cho đến khi nào không $\exists p \in Pr'$ phân hoạch mệnh đề $f \in F$

3. $\forall p \in Pr'$ nếu $\exists p'$ mà p t... ng v i p'

$$- Pr' = Pr' - \{p\}$$

$$- F = F - \{f\} \text{ nếu } f \text{ là mệnh đề đúng theo } p$$

Sau bước 3 Pr' là tập v t... y và t i thi u c n tìm.

Bước tiếp theo của thi t k phân mệnh đề ngang nguyên thủy là suy d n ra t p các v t h i s c p có thể c nh ngh a trên các v t trong t p Pr' . Các v t h i s c p này xác nh các mệnh đề cho b c c p phát.

Bước cuối của quá trình thi t k là lo i b m t s mệnh đề vô nghĩa. i u này c th c hi n b ng cách xác nh nh ng v t mâu thu n v i t p các phép kéo theo I. Ch ng h n, nếu $Pr' = (P1, P2)$, trong ó

$$P1: A = \text{giá tr 1}$$

$$P2: A = \text{giá tr 2}$$

Và m i n b i n thiên c a A là {giá tr 1, giá tr 2} rõ ràng I ch a hai phép kéo theo v i kh ng nh:

$$i1: (A = \text{giá tr 1}) \rightarrow \neg (A = \text{giá tr 2})$$

$$i2: \neg (A = \text{giá tr 1}) \rightarrow (A = \text{giá tr 2})$$

B n v t h i s c p sau ây c nh ngh a theo Pr'

$$m1: (A = \text{giá tr 1}) \wedge (A = \text{giá tr 2})$$

$$m2: (A = \text{giá tr 1}) \wedge \neg (A = \text{giá tr 2})$$

$$m3: \neg (A = \text{giá tr 1}) \wedge (A = \text{giá tr 2})$$

$$m4: \neg (A = \text{giá tr 1}) \wedge \neg (A = \text{giá tr 2})$$

Trong tr ng h p này các v t h i s c p m1 và m4 mâu thu n v i phép kéo theo I và vì th b lo i kh i M

Thu t toán 3.2. Thu t toán phân mệnh đề ngang nguyên thủy

Vào: Quan hệ R , tập các vị trí n và Pr

Ra: Tập các vị trí h và c

Phương pháp:

0. $Pr' = \text{COM-MIN}(R, Pr)$

Xác định tập M các vị trí h và c

Xác định tập I các phép kéo theo giữa các $P_i \in Pr'$

1. $\forall m_i \in M$ nếu m_i mâu thuẫn với I

$$M = M - \{m_i\}$$

Sau bước này M là tập các vị trí h và c .

3.1.2.3. Phân mảnh ngang dữ liệu

Phân mảnh ngang dữ liệu được thực hiện dựa trên một quan hệ thành viên của một bảng dữ liệu theo phép toán chọn trên quan hệ chủ nhân của bảng dữ liệu. Nếu cần cho truy vấn liên kết L , trong đó $\text{owner}(L) = S$ và $\text{member}(L) = R$, các mảnh ngang dữ liệu của R được thực hiện là:

$$R_i = R \bowtie S_i, 1 \leq i \leq k$$

Trong đó k là số lượng các mảnh được thực hiện dựa trên R , và $S_i = S(E_i)$, với E_i là công thức thực hiện phân mảnh ngang nguyên thủy S_i .

Ví dụ: Xét liên kết giữa bảng *Luong* và *NhanVien*. Chúng ta có thể nhóm các nhân viên thành hai nhóm tùy theo lương. Giả sử nhóm có lương từ 4000\$ trở xuống và nhóm có lương trên 4000\$. Hai mảnh *NhanVien1* và *NhanVien2* được thực hiện như sau:

$$\text{NhanVien1} = \text{NhanVien} \bowtie \text{Luong1}$$

$$\text{NhanVien2} = \text{NhanVien} \bowtie \text{Luong2}$$

Trong đó

$$\text{Luong1} = \text{Luong}(\text{Luong} \leq 4000)$$

$$\text{Luong2} = \text{Luong}(\text{Luong} > 4000)$$

Kết quả thu được như sau:

NhanVien1			NhanVien2		
MaNV	Ten	ChucVu	MaNV	Ten	ChucVu
E1	Hoàng Lan	Ks i n	E2	ình V ãng	Phân tích h ãnh
E3	Minh Tài	Ks Máy	E4	D ãng Hà	L ãp trình viên
E6	V ãn H ãi n	Ks i n	E5	Minh Hoa	Phân tích h ãnh
E7	Hoài Nam	Ks Máy	E8	V ãn D ãng	Phân tích h ãnh

Muốn thực hiện phân mảnh ngang dữ liệu chúng ta cần ba nguyên lý: Tập các phân hoạch của quan hệ chủ (chẳng hạn *Luong1* và *Luong2* trong ví dụ trên), quan hệ thành viên, và tập các vị trí n và c của quan hệ chủ và quan hệ thành viên (chẳng hạn *NhanVien.ChucVu = Luong.ChucVu*). Công có một vấn đề phức tạp phải chú ý, trong lược đồ CSDL chúng ta hãy giả định rằng n và c của một quan hệ R . Nếu có thể có nhiều cách phân mảnh ngang dữ liệu cho R . Quy trình chọn cách phân mảnh nào cần dựa trên hai tiêu chuẩn sau:

1. Phân mảnh có tính *ít* thay đổi;
2. Phân mảnh có sự đồng nhất trong nhu cầu dữ liệu.

3.1.2.4. Kiểm tra tính đúng đắn của phân minh ngang

1. Tính ý

Tính ý của một phân minh ngang nguyên thủy dựa vào các vectơ chèn c dùng. Với i u ki n các vectơ chèn là y , phân minh thu c cung c b o m là y , b i vì c s c a thu t toán phân minh là m t t p các vectơ c c ti u và y . Tính ý s c b o m v i i u ki n là không có sai sót x y ra khi nh ngh a t p v t y và c c ti u Pr. Tính ý của phân minh ngang d n xu t có h i khác. Khó kh n là do vectơ nh ngh a phân minh có liên quan n hai quan h .

G i R là quan h thành viên c a m t liên k t mà ch là quan h S c phân minh thành $F_s = \{S_1, S_2, \dots, S_N\}$. A là thu c tính n i gi a R và S . V y i v i m i b t c a R , ph i có m t b t' sao cho $t.A = t'.A$

Quy t c này c g i là ràng bu c toàn v n hay toàn v n tham chi u, b o m m i b trong các m nh c a quan h thành viên u n m trong quan h ch .

2. Tính tái thi t c

Tái thi t m t quan h toàn c c t các m nh c th c hi n b ng toán t h p trong c phân minh ngang nguyên thủy l n d n xu t. Vì th m t quan h R v i phân minh $F_R = \{R_1, R_2, \dots, R_N\}$, chúng ta có $R = \cup R_i, \forall R_i \in F_R$.

3. Tính tách bi t

Chúng ta có th d d ng th y r ng tính tách r i c a phân minh nguyên thủy là rõ ràng theo cách phân minh c a ta. V i phân minh d n xu t, tính tách r i s c b o m n u các vectơ h i s c p xác ình phân minh có tính loài tr l n nhau.

3.1.3. Phân minh d c.

Phân minh d c m t quan h r s sinh ra các m nh r_1, r_2, \dots, r_N m i m nh ch a m t t p con thu c tính c a R và c khóa c a r . M c ích c a phân minh d c là phân ho ch m t quan h thành m t t p các quan h nh h n nhi u ng d ng có th ch ch y trên m t quan h . Trong ng c nh này, m t phân minh t i u là m t phân minh sinh ra m t l c phân minh cho phép gi m n t i đa th i gian th c thi các ng d ng ch y trên các m nh ó.

Phân minh d c ã c nghiên c u trong ng c nh c a các h CSDL t p trung. Lý do chính là s d ng nó làm m t công c thi t k cho phép các v n tin làm vi c trên các quan h nh h n vì th gi m b t s truy xu t và t i t ki m b nh . M t trong s các ph ng pháp phân minh d c ã nghiên c u trong mô hình CSDL quan h là vì c chu n hóa các quan h v các d ng chu n c p cao.

Bên c nh ph ng pháp chu n hóa các quan h còn có nh ng ph ng pháp khác và chúng th ng g n v i m c tiêu c a bài toán.

3.1.4. C p phát.

Bài toán c p phát:

Gi s r ng có m t t p các m nh $F = \{F_1, F_2, \dots, F_N\}$ và m t m ng máy tính bao g m các v tr $S = \{S_1, S_2, \dots, S_M\}$ trên ó có m t t p các ng d ng d ng $Q = \{q_1, q_2, \dots, q_k\}$ ang ch y. Bài toán c p phát là tìm m t phân ph i “t i u” c a F cho S .

Tính t i u có th c nh ngh a ng v i hai s o:

- Chi phí nhúng: Hàm chi phí g có chi phí lưu trữ F_i và vị trí S_j , chi phí vận chuyển F_i và vị trí S_j , chi phí cập nhật F_i và vị trí S_j và chi phí truy cập lưu trữ. Vì thế bài toán cập nhật có thể tìm kiếm các cập nhật vị trí hàm chi phí tối ưu.

- Hiệu quả: Chỉ cần các cập nhật tối thiểu duy trì mức hiệu quả lớn nhất đó là hiệu quả thời gian cập nhật và tính toán.

Nói chung bài toán cập nhật tổng quát là bài toán phức tạp, vì thế các nghiên cứu tập trung tìm ra các giải thuật heuristic tối ưu có thể có các lợi ích. Phân biệt bài toán cập nhật tối ưu và bài toán cập nhật trong các thuật toán CSDL phân tán, chúng ta gọi bài toán tối ưu là bài toán cập nhật tối ưu, và bài toán sau là bài toán cập nhật CSDL.

Hệ thống không có mô hình heuristic tổng quát nào nhằm tối ưu các mức và sinh ra mức chỉ cần các cập nhật tối ưu và các loại ràng buộc. Các mô hình đã phát triển nhằm cải thiện các giải thuật tối ưu hóa và đáp ứng cho mức cách tối ưu.

Thông tin cho cập nhật:

Thông tin về CSDL:

Để phân tích mức ngang, chúng ta đã nghiên cứu tùy chỉnh. Bây giờ chúng ta cần mô hình ngang để cho các mức và nghiên cứu tùy chỉnh mức F_j và vị trí q_i . Đây là số lượng các bản F_j cần truy xuất x lần q_i . Giá trị này ký hiệu là $Se_i(F_j)$.

Mô hình thông tin khác trên các mức là kích thước của chúng. Kích thước mức F_j cho bởi: $size(F_j) = card(F_j) * Length(F_j)$

Trong đó $length(F_j)$ là chiều dài tính theo byte của mức F_j

Thông tin về ngữ nghĩa:

Phân tích các thông tin có liên quan đến ngữ nghĩa để biên dịch trong khi phân tích mức ngang cung cấp một số ít nhất cho mô hình cập nhật. Hai số liệu quan trọng là số truy xuất do câu văn bản q_i thể hiện trên mức F_j trong mức liên quan nó – ký hiệu là RR_{ij} . Và ngược lại là các truy cập cập nhật – ký hiệu là UR_{ij} .

Chúng ta nghiên cứu hai ma trận UM và RM về các phần tử ngữ nghĩa u_{ij} và r_{ij} có thể như sau

- $u_{ij} = 1$ nếu văn bản q_i có cập nhật mức F_j , ngược lại bằng 0

- $r_{ij} = 1$ nếu văn bản q_i cần mức F_j , ngược lại bằng 0.

Mô hình vectơ O gồm các giá trị $0(i)$ có nghĩa, và $0(i)$ có vị trí của câu văn bản q_i . Cụ thể, nghiên cứu ràng buộc thời gian cập nhật, thời gian cập nhật tối ưu phép cập nhật ngữ nghĩa cập nhật tối ưu.

Thông tin về vị trí:

Về vị trí chúng ta cần biết vị trí lưu trữ và vị trí của nó. Hệ thống là mức giá trị này có thể tính bằng các hàm thích hợp hoặc bằng các phép toán giá trị. Chi phí lưu trữ lưu trữ dữ liệu tại các vị trí S_j ký hiệu là USC_j . Các cập nhật tối ưu chi phí LPC_j là chi phí xử lý mức ngữ nghĩa vị trí S_j . Ngữ nghĩa cập nhật vị trí S_j và vị trí RR và UR .

Thông tin về mức:

Trong mô hình đang xét, chúng ta giả sử có n tính toán n biến, trong đó chỉ phí truy vấn mỗi bộ ghi hai vị trí S_i và S_j . Có thể tính các số lượng thông báo, chúng ta dùng $fsize$ làm kích thước (tính theo byte) của một bộ dữ liệu.

3.2. Kỹ thuật quản lý ngữ nghĩa

3.2.1. Quản lý khung nhìn

Một trong những nhiệm vụ của mô hình quan hệ là nó cung cấp các tính toán để diễn giải logic. Trong một hệ thống quan hệ, khung nhìn cũng như ngữ nghĩa là một tập hợp các tính toán trên quan hệ các số hoặc quan hệ thực, nhưng không có vị trí thay thế cho một tập quan hệ các số, nghĩa là không có lưu trữ các số trong CSDL. Một khung nhìn là một tập các số theo nghĩa là nó phản ánh tất cả các tính toán trên CSDL. Một lần nữa ngoài có thể cũng như ngữ nghĩa là một tập các khung nhìn. Bên cạnh vì các số đang chúng trong lần nữa ngoài, các khung nhìn còn có tác động bảo đảm tính an toàn dữ liệu bằng một cách rõ ràng. Như chúng ta đã biết, trong một CSDL, các khung nhìn đã che giấu một số dữ liệu. Nhưng nếu sử dụng các phép truy vấn CSDL qua các khung nhìn, hệ thống không thể thay đổi hoặc thao tác trên các dữ liệu này vì thế chúng được bảo vệ.

Cần chú ý trong môi trường phân tán, một khung nhìn có thể lý giải các quan hệ phân tán và vị trí truy vấn một khung nhìn đòi hỏi phải thiết lập câu lệnh phân tán theo ngữ nghĩa khung nhìn.

Khung nhìn trong CSDL phân tán

Các khung nhìn trong hệ thống phân tán có thể được xem xét các quan hệ đã phân mảnh các lưu trữ vị trí khác nhau. Bởi vì khung nhìn có thể được sử dụng làm quan hệ các số trong các ứng dụng, như nghĩa của chúng phải có lưu trữ trong các thuộc tính của mô tả quan hệ các số. Tùy vào một tập các vị trí của các bộ nhớ, các khung nhìn có thể có tập trung tính toán vị trí, các nhân bản một phần hoặc toàn bộ. Trong môi trường hệ phân tán, thông tin liên kết khung nhìn vị trí như nghĩa của nó phải có nhân bản. Nếu như nghĩa khung nhìn không có vị trí của câu lệnh thì sẽ phải truy vấn xa và vị trí có như nghĩa khung nhìn đó.

Trong phần tiếp theo chúng ta sẽ xem xét các phân mảnh của quan hệ, xét theo một khía cạnh nào đó các phân mảnh thực sự là như ngữ nghĩa các khung nhìn thực.

3.2.1.1. Khung nhìn trong quản lý tập trung

Phân loại các hệ quản trị SQL quan hệ sử dụng các khung nhìn. Một khung nhìn là một quan hệ được xem xét các quan hệ các số như một tập hợp các tính toán trên quan hệ. Nó cũng như ngữ nghĩa bằng cách gán tên của khung nhìn cho câu lệnh.

Ví dụ: Khung nhìn SYSAN (nhân viên phân tích hệ thống) được xem xét quan hệ NhanVien có thể cũng như ngữ nghĩa như sau:

```
CREATE VIEW SYSAN
```

```
AS SELECT MaNV, TenNV FROM NhanVien WHERE ChucVu = "Phân tích hệ thống"
```

Tác động của câu lệnh này là như ngữ nghĩa khung nhìn SYSAN, và nó có thể được sử dụng như một tập quan hệ các số (bảng).

3.2.1.2. Cập nhật qua các khung nhìn

Khung nhìn có thể cũng như ngữ nghĩa bằng các câu truy vấn phức tạp với các phép chọn, chiếu, kết nối, các hàm tập hợp... Và tất cả các khung nhìn có thể truy vấn như một tập quan hệ các số,

những không phải tất cả chúng có khả năng chấp nhận. Chấp nhận qua khung nhìn chỉ cần lý tưởng nếu chúng có thể lan truyền chính xác đến các quan hệ số. Hiện nay hầu hết các hệ quản trị CSDL quan hệ hiện tại hỗ trợ chấp nhận để lưu qua khung nhìn bằng cách sử dụng các cơ chế xử lý ngầm, ví dụ như trigger.

3.2.1.3. Khung nhìn trong CSDL phân tán

Những khung nhìn tương đương nhau trong các hệ quản trị CSDL tập trung hay phân tán. Tuy nhiên khung nhìn trong các hệ thống phân tán có thể dẫn xuất các quan hệ đã phân mảnh cục bộ như vị trí khác nhau. Khi một khung nhìn được nhúng, tên và câu văn tin truy xuất của nó sẽ lưu trữ cấu trúc của CSDL.

Bởi vì khung nhìn có thể sử dụng làm quan hệ số trong các ứng dụng, những cơ sở dữ liệu trong CSDL giữ nguyên các quan hệ số. Tùy thuộc vào mức độ truy cập vị trí của rập hình thức, các những khung nhìn có thể tập trung tại một vị trí, hoặc nhân bản một phần hoặc toàn bộ. Trong một trường hợp, thông tin liên kết tên khung nhìn và vị trí những cơ sở dữ liệu nó phụ thuộc nhân bản. Nếu những khung nhìn không có vị trí những cơ sở dữ liệu nó phụ thuộc nhân bản. Nếu những khung nhìn không có vị trí của câu văn tin thì sẽ phụ thuộc xa vị trí có những khung nhìn đó.

3.2.2. An toàn dữ liệu

An toàn CSDL là một nhiệm vụ quan trọng của hệ thống CSDL nhằm bảo vệ dữ liệu không bị truy xuất “bất hợp pháp”. An toàn dữ liệu bao gồm hai vấn đề: bảo vệ dữ liệu và kiểm soát truy cập.

Bảo vệ dữ liệu nhằm tránh cho những người “không có phân quyền” hiểu được nội dung vật lý của dữ liệu. Chức năng này do hệ thống tập tin đảm trách trong các hệ điều hành tập trung và phân tán. Phương pháp chính là mã hóa dữ liệu, được dùng cho các thông tin cục bộ trên toàn thông tin trao đổi trên mạng. Dữ liệu đã mã hóa chỉ có thể “giải mã” bởi những người sử dụng được ủy quyền.

Kiểm soát truy cập phụ thuộc vào những chức năng của người sử dụng được phép thực hiện các thao tác trên CSDL. Những người sử dụng khác nhau có thể có quyền truy cập đến một tài nguyên dữ liệu để kiểm soát những hoạt động tập trung hay phân tán. Vì thế các DBMS phân tán hay tập trung phải có khả năng hạn chế truy xuất một phần dữ liệu vì một tập con những người sử dụng.

3.2.2.1. Kiểm soát truy cập tập trung

Ba tác nhân chính có liên quan đến việc kiểm soát truy cập là: người sử dụng, là người kích hoạt các chương trình ứng dụng; các thao tác được gửi vào ứng dụng; và các hệ thống CSDL sẽ thực hiện các thao tác đó.

Kiểm soát truy cập bao gồm việc xem xét ba (người sử dụng, thao tác, hệ thống) có được phép tiến hành hay không? Một quy định xác định rằng người sử dụng có quyền thực hiện một thao tác thu được nào trên một hệ thống.

Khai báo một người sử dụng (hay nhóm người sử dụng) và hình thức thông tin được thực hiện bằng một tập (tên người sử dụng, một thuộc tính). Tên và một thuộc tính phụ thuộc ra khi những người dùng vào hệ thống. Điều này ngăn chặn những người không có thẩm quyền xâm nhập vào hệ thống.

Quy định này liên hệ giữa những người sử dụng và một hệ thống để xem xét các thao tác được thực hiện. Trong các hệ quản trị CSDL dựa trên SQL, một thao tác là một câu lệnh bậc cao

nh SELECT, INSERT, UPDATE, hoặc DELETE... Các quy định về ngôn ngữ SQL và các thủ tục lập trình các loại như:

GRANT <thao tác> ON < đối tượng> TO <Người sử dụng>

REVOKE <thao tác> FROM < đối tượng> TO <Người sử dụng>

3.2.2.2. Kiểm soát cấp quyền phân tán

Các vấn đề về kiểm soát cấp quyền trong môi trường phân tán có nguồn gốc từ sự kiện là các đối tượng và các chức năng phân tán. Nhiệm vụ này bao gồm: cấp quyền cho người sử dụng xa, quản lý các quy tắc cấp quyền phân tán và việc xử lý các khung nhìn và các nhóm người sử dụng. Có hai kỹ thuật cho vấn đề này:

1. Thông tin xác định người sử dụng cá nhân bản địa từ các vị trí trong mạng. Các chức năng trình bày các quy định rõ tên và mật khẩu của người sử dụng.

2. Tất cả các vị trí trong hệ thống phân tán có quyền địa phương và xác định nhau một cách thống nhất về quyền. Giao tiếp giữa các vị trí về cách sử dụng mật khẩu của vị trí. Mật khẩu vị trí đã xác định thì không cần phải xác định người sử dụng của chúng.

Các quy tắc cấp quyền phân tán có liên quan đến cùng phạm vi trong hệ thống trung. Giả sử các nhà nghiên cứu khung nhìn, chúng phải đưa vào trong hệ cấu trúc của CSDL. Chúng có thể cá nhân bản địa hoàn toàn từ vị trí hoặc từ các vị trí của các đối tượng cần truy xuất. Tuy nhiên chính xác địa lý của cá nhân bản địa hoàn toàn là vì cấp quyền có thể xử lý bằng kỹ thuật hữu ích như tin vào lúc biên dịch. Kỹ thuật hai tầng trong trường hợp tính chất của các tham chiếu rất cao. Tuy nhiên vì cấp quyền phân tán không thể kiểm soát được vào lúc biên dịch.

Khung nhìn có thể xem như các đối tượng của các cấp quyền. Khung nhìn là những đối tượng phức tạp, nghĩa là nó có các thuộc tính khác. Vì thế trao quyền truy xuất đến một khung nhìn sẽ dẫn thành trao quyền truy xuất đến các đối tượng. Nếu nhà nghiên cứu khung nhìn và các quy tắc cấp quyền cá nhân bản địa hoàn toàn thì việc biên dịch này khá ngắn gọn và có thể thực hiện được.

Nhóm các người sử dụng dùng cấp quyền chung làm nhiệm vụ công việc quản lý CSDL phân tán. Trong các DBMS trung tâm, khái niệm “mã người sử dụng” có thể xem là nhóm công việc. Trong môi trường phân tán, nhóm công việc bị phân tán cho tất cả mã người sử dụng của hệ thống. Tuy nhiên, người ta thường phân chia các trung gian như mô hình nhóm công việc từ vị trí thực.

Quản lý các nhóm trong môi trường phân tán dựa trên sự phân phối giữa các quy tắc vì các chức năng của một nhóm có thể có tính chất khác nhau và quyền truy xuất đến một đối tượng có thể trao cho nhiều nhóm, mà bản thân chúng lại phân tán. Nếu thông tin của nhóm và các quy tắc cấp quyền cá nhân bản địa hoàn toàn từ vị trí thì việc duy trì quyền truy xuất trong hệ thống trung tâm. Tuy nhiên việc duy trì các bản sao này rất tốn kém.

3.3. Quản lý giao dịch và điều kiện đồng nhất phân tán

3.3.1. Các khái niệm về giao dịch

Giao dịch là một loạt các hành động. Đôi khi biểu thức một giao dịch T ta viết T: begin...end. Giữa begin và end là những bước của giao dịch. Chương trình này có thể là một câu lệnh tin học mà thực hiện chương trình ngôn ngữ lập trình và đưa vào một câu lệnh tin. Có

nhiệm vụ thực hiện các lệnh cùng một chương trình T có thể tiến hành song song với nhau và vị trí khác nhau trên màn hình; mà nhiệm vụ này là một giao dịch khác nhau.

Một giao dịch sẽ được liệt kê và ghi đè lại vào CSDL, qua một vùng làm việc riêng (private) – gọi là vùng bộ nhớ tính toán tạm thời. Các lệnh là các tính toán do giao dịch thực hiện sẽ không có tác động trên CSDL cho đến khi các giá trị mới được ghi vào CSDL.

Ví dụ chúng ta có giao dịch T:

Begin

Read A

$A = A + 100$

Read A

$A = A + 2$

Write A

End

Khi đó trong CSDL giá trị của A chỉ tăng lên 2, vì phép toán $A = A + 100$ chỉ làm việc trong vùng bộ nhớ tính toán tạm thời.

3.3.1.1. Tính nguyên tử

Trên quan niệm về quản lý các, quản lý giao dịch là một có gắng nhằm làm cho các thao tác phức tạp xuất hiện dưới dạng các nguyên tử. Nghĩa là thao tác x y ra trên văn học không xảy ra. Nếu x y ra, không có bị ngăn hay giao dịch nào cùng x y ra trong suốt thời gian tồn tại của nó. Mỗi nguyên tử về sau ta sẽ gọi là một bước của một thao tác cụ thể. Cách thông dụng nhất mà một bộ xử lý tính nguyên tử của các giao dịch là phương pháp tuần tự hóa. Phương pháp này làm cho các giao dịch thực hiện một cách tuần tự. Một giao dịch không có tính nguyên tử như:

1, Trong hệ thống phân chia thời gian, thời gian cho giao dịch T có thể kết thúc trong khi T đang tính toán và các hoạt động của một giao dịch khác sẽ thực hiện trên cơ sở khi T hoàn tất.

2, Một giao dịch có thể không hoàn tất, chẳng hạn có khi nó phải chờ đợi gì đó, có thể vì nó thực hiện một phép tính không hợp lệ (ví dụ phép chia cho 0), hoặc có thể do nó đòi hỏi những dữ liệu không được quy định truy xuất. Bản thân hệ thống CSDL có thể buộc giao dịch này ngừng lại vì những lý do. Chẳng hạn giao dịch đó có thể bị khóa “chết” (deadlock)

Trong trường hợp (1), nhiệm vụ của hệ thống CSDL là phải báo động cho dù bất kỳ ai gây xảy ra ngay giữa một giao dịch, tác động của giao dịch trên CSDL không bị ảnh hưởng của những bị ngăn cản tạm thời này. Trong trường hợp (2), hệ thống phải báo động giao dịch bị hủy bỏ không ảnh hưởng gì trên CSDL hoặc các giao dịch khác

Trong thực tế, mỗi giao dịch đều có một chuỗi các bước cụ thể như: đọc hay ghi một mục dữ liệu vào CSDL hoặc thực hiện các phép toán số học nào đó trong vùng làm việc riêng, hoặc những bước song song khác như các bước khóa, mở khóa, giải thoát hoàn tất giao dịch... Chúng ta luôn giữ sự riêng biệt của những bước này là nguyên tử. Thậm chí thao tác tính toán kết thúc sau khi thời gian dành cho nó đã hết cũng có thể xem là nguyên tử, bởi vì các phép tính toán xảy ra khi đang làm việc trong vùng dữ liệu cục bộ và không thể ảnh hưởng đến vùng làm việc của người khác cho đến khi giao dịch đang thực hiện được phép tính số học được tái hoạt động trở lại.

3.3.1.2. Mục dữ liệu

quản lý các hoạt động nghiệp vụ, CSDL phải được phân chia thành các mô-đun, đó là những mô-đun và dữ liệu cần truy xuất có liên quan. Bên cạnh và kích thước các mô-đun dựa vào nhà thiết kế hệ thống nghiệp vụ. Chẳng hạn trong mô hình dữ liệu quan hệ, chúng ta có thể chia các mô-đun như các quan hệ, hoặc các mô-đun như các bảng hay thành phần của các bảng. Chúng ta cũng có thể chia nghiệp vụ các mô-đun có kích thước trung gian, như một khối của quan hệ. Kích thước của các mô-đun dữ liệu hệ thống sẽ phụ thuộc vào mức độ phân chia hệ thống. Một hệ thống có thể là hệ thống, nhưng nó sẽ được các mô-đun dữ liệu và hệ thống là hệ thống của nó sẽ được các mô-đun dữ liệu.

Phương pháp thông dụng nhất để liên quan vì truy xuất các mô-đun là sử dụng khóa. Bộ quản lý khóa là thành phần của hệ quản trị chịu trách nhiệm theo dõi xem một mô-đun nào đang ghi vào các phần của I hay không. Nếu có thì bộ quản lý khóa sẽ ngăn không cho các giao dịch khác truy xuất I trong trường hợp truy xuất đó có thể gây ra xung đột.

Chọn chế độ hoạt động làm giảm thiểu chi phí cần duy trì các khóa, bởi vì chúng ta cần ít chế độ của các khóa, và chúng ta tiết kiệm được thời gian bởi vì hệ thống chỉ phải thực hiện rất ít hành động khóa. Tuy nhiên hệ thống cho phép nhiều giao dịch hoạt động song song, bởi vì xác suất các giao dịch yêu cầu khóa trên cùng một mô-đun sẽ thấp.

3.3.1.3. Khóa

Như chúng ta đã biết, khóa là một quy tắc truy xuất trên một mô-đun dữ liệu mà bộ quản lý khóa có thể trao cho một giao dịch hay thu hồi lại. Có thể có nhiều kiểu khóa, ví dụ như khóa chính, khóa ghi... Thông thường thì mỗi mô-đun, chỉ có một tập con các mô-đun khóa, vì vậy bộ quản lý khóa có thể liệt kê các khóa hiện hành trong một bảng khóa với các mô-đun tin: (I, L, T) – giao dịch T có một khóa liên quan L trên mô-đun I.

3.3.1.4. Kiểm soát hoạt động nghiệp vụ bằng khóa

Để thấy được nhu cầu phải sử dụng khóa chúng ta xem xét ví dụ sau đây:

Xét hai giao dịch T1 và T2. Một giao dịch truy xuất một mô-đun dữ liệu A có giá trị là mang giá trị nguyên, rồi tăng thêm 1 vào A. hai giao dịch này thực hiện chương trình P.

P: Begin

Read A

$A = A + 1$

Write A

End

Chúng ta thực hiện hai giao dịch T1 và T2 như sau:

A trong csdl	5	5	5	5	6	6
T1	Read A		$A = A + 1$			Write A
T2		Read A		$A = A + 1$	Write A	
A trong vùng làm việc của T1	5	5	6	6	6	6
A trong vùng làm việc của T2	5	5	5	6	6	

Giá trị của A tại thời điểm trong CSDL, với mỗi giao dịch P của A vào vùng làm việc, cũng đưa vào giá trị này rồi ghi kết quả vào trong CSDL. Chúng ta nhận thấy rằng, mặc dù hai giao dịch đều đã tăng thêm 1 vào A, giá trị của A trong CSDL chỉ tăng 1.

Để phương pháp thông dụng nhất cho vấn đề trình bày trong ví dụ trên là cung cấp một khóa trên A. Trước khi của A, một giao dịch T phải khóa A lại, ngăn không cho giao dịch khác truy xuất A cho đến khi T hoàn tất xong thao tác trên A. Hiện nay A khóa của mô-đun A chỉ khi trước đó A

không b khóa b i m t giao d ch khác. N u A ã b khóa, T ph i i n khi giao d ch m khóa cho A.

V y ng n c n nh ng tr ng h p áng t i c x y ra ta ph i dùng khóa. Nh v y trong nh ng ch ng trình giao d ch ph i có khóa và m khóa. Ta gi s r ng m t khóa ph i c t trên m t m c tr c khi c hay ghi nó, và thao tác khóa hành ng nh m t hàm ng b hóa. Ngh a là n u m t giao d ch khóa m t m c ã c khóa, nó không th t i n hành cho n khi khóa này c gi i phóng b ng m t l nh m khóa c th c hi n b i giao d ch ang gi khóa. Ta c ng gi s r ng m i giao d ch u có th m c m i khóa do chính nó khóa. M t l ch bi u ch a cá thao tác c b n c a nhi u giao d ch tuân theo các quy t c c a khóa c gi là h p l .

Ví d :

Ch ng trình P c a ví d trên c vi t l i v i các khóa nh sau:

P: begin

Lock A

Read A

$A = A + 1$

Write A

Unlock A

End

Gi s T1 và T2 là hai giao d ch th c hi n P. N u T1 b t u tr c, nó yêu c u khóa trên A. Gi s r ng không có giao d ch nào ang khóa, b qu n lý khóa s cho nó khóa m c A. Bây gi ch có T1 m i có th truy xu t A, n u T2 b t u tr c khi T1 ch m d t thì khi T2 th c hi n l nh khóa A, h th ng bu c T2 ph i i. Chi khi T1 th c hi n l nh Unlock A, h th ng m i cho phép T2 t i n hành. K t qu là i u b t th ng không x y ra; T1 ho c T2 s hoàn t t tr c khi giao d ch kia b t u, và k t qu chung c a chúng là c ng 2 vào A.

3.3.1.5. Khóa s ng (livelock)

H th ng qu n lý khóa trao và bu c khóa các m c d li u không th ho t ng m t cách tùy t i n. Gi s trong ví d trên, khi T1 gi i phóng khóa trên A, trong khi T2 ang i nh n khóa, m t giao d ch T3 khác cùng xin m t khóa trên A, và T3 c trao khóa này tr c T2. T ng t sau khi T3 m khóa cho A thì l i có giao d ch T4 xin khóa A... Và r t có th T2 ph i i r t lâu ho c không bao gi nh n c khóa trên A.

Tình hu ng này g i là khóa s ng. V y khóa s ng trên m c A c a giao d ch T là T không khóa c A vì A luôn b khóa b i m t giao d ch khác. R t nhi u gi i pháp ã c các nhà thi t k h i u hành xu t gi i quy t v n khóa s ng, ví d nh các giao d ch khi khóa m t m c ph i ng ký th t , và b qu n lý khóa s l n l t trao quy n khóa m c A khi m c này không b khóa theo th t ã xác nh tr c.

3.3.1.6. Khóa “c ng” (deadlock)

M t v n khác có th x y ra trong i u khi n các ho t ng ng th i, ó là v n khóa “c ng” (deadlock). ó là các giao d ch khóa chéo l n nhau các m c c n hoàn thi n giao d ch.

Ví d :

Gi s chúng ta có hai giao d ch ng th i T1 và T2 nh sau:

T1: begin	T2: begin
Lock A	Lock B
Lock B	Lock A
Unlock A	Unlock B
Unlock B	Unlock A
end	end

Giả sử T1 và T2 có thể xảy ra cùng lúc. T1 yêu cầu khóa A và trao quyền khóa trên A, còn T2 yêu cầu và trao quyền khóa trên B. Do đó khi T1 yêu cầu khóa trên B thì nó phải chờ vì T2 đã khóa B. Tương tự khi T2 yêu cầu khóa trên A, nó cũng bị chặn vì T1 đã khóa A. Kết quả là không có giao dịch nào tiếp tục hoạt động, mọi giao dịch bị chặn cho giao dịch kia mượn khóa, và chúng bị chặn như những giao dịch khác nhau.

Như vậy khóa “cứng” là một tình huống mà trong đó mọi thành viên T_i của một tập giao dịch $T = \{T_1, T_2, \dots, T_N\}$ đang giữ một khóa của một tập các khóa khác trong tập T. Bởi vì mọi giao dịch trong tập T đều đang giữ, nên không thể mượn khóa cho một giao dịch khác đang chờ, vì vậy tất cả các giao dịch đều bị chặn mãi mãi.

Một số giải pháp cho vấn đề khóa “cứng”:

1. Yêu cầu các giao dịch phải đạt được yêu cầu khóa cùng một lúc, và bộ quản lý khóa trao tất cả các khóa cho chúng cùng một lúc, hoặc không trao và cho giao dịch này mượn một hay nhiều khóa đang giữ bởi các giao dịch khác.

2. Gán một thứ tự tùy ý tính cho các giao dịch, bắt đầu các giao dịch phải xin khóa theo thứ tự này.

Một cách khác nhằm xử lý các khóa “cứng” là không ngăn chặn chúng. Và cuối cùng khi họ nghĩ rằng gian lận như họ nghĩ sẽ kiểm tra yêu cầu khóa và tìm xem có xảy ra khóa “cứng” hay không. Nếu chúng ta sử dụng thứ tự, vì các nút là các giao dịch và các cung $T_1 \rightarrow T_2$ bị chặn cho T2 đang giữ một khóa trên một tập các khóa của T1 giữ. Khi đó mọi chu trình trong thứ tự bị chặn cho một khóa “cứng”, và nếu không có chu trình thì kết luận là không có khóa “cứng”. Nếu một khóa “cứng” được phát hiện, thì họ nghĩ phải khôi phục lại và các tác động trên CS L của giao dịch khóa “cứng” đó phải bị bỏ. Quá trình hủy bỏ và tái khôi phục có thể gây phiền nhiễu cho chúng ta không bị tắc cách thức các giao dịch ghi vào CSDL trước khi chúng hoàn thành.

3.3.1.7. Tính khả thi của lịch bị chặn.

Giả sử chúng ta có một tập các giao dịch $T = \{T_1, T_2, \dots, T_N\}$. Chúng ta thấy ngay rằng nếu các giao dịch thực hiện tuần tự theo một thứ tự nào đó, giao dịch nào tiếp theo giao dịch kia thì các sự tranh chấp chắc chắn không xảy ra và trong CSDL chúng ta có một kết quả nào đó.

Giả sử chúng ta có tập giao dịch $T = \{T_1, T_2, \dots, T_N\}$, thì tổng cộng với T ta có $N!$ các lịch bị chặn tuần tự khác nhau. Bởi vì chúng ta giả sử rằng hoạt động của các giao dịch độc lập là đúng nên nếu và chỉ nếu tác động của nó giống như tác động có các giao dịch bị chặn tuần tự.

Chúng ta định nghĩa một lịch bị chặn S cho một tập các giao dịch T là một (có thể xen kẽ) các bộ các bộ của các giao dịch (khóa, ghi, ...) có thể xảy ra.

Các bộ của một giao dịch đã cho phải xuất hiện trong lịch bị chặn theo đúng thứ tự xảy ra trong giao dịch đó.

Vậy mệnh đề bị u S có tập giao dịch T đúng là hợp lệ và đúng nên các bước cần làm để giao dịch T_i đã cho phải thực hiện trong lịch bị u S theo đúng thứ tự xảy ra trong giao dịch T_i và các bước cần của các giao dịch tuân theo các quy tắc khóa.

Mệnh đề bị u S của các giao dịch trong T là một hoán vị của các bước cần của $T_i \in T$. Giả sử trong T có k bước cần nên chúng ta có $k!$ hoán vị khác nhau của các bước cần. Vậy với tập giao dịch T ta có $k!$ lịch bị u S khác nhau. Tuy nhiên trong số đó có nhiều lịch bị u vô nghĩa và không đúng nên hoặc không hợp lệ.

Vì cơ quan lý các giao dịch là cơ quan lý các lịch bị u đúng nên và hợp lệ nên vì mệnh đề bị u tu n t nào đó. Lịch bị u đúng là lịch tu n t mà tác động của nó giống với tác động của mệnh đề bị u tu n t nên đúng là lịch tu n t.

Hai lịch bị u đúng là tương đương nên chúng cho kết quả giống nhau.

3.3.1.8. B x p l ch

Chúng ta nhận thấy rằng khi thực hiện hoạt động thì các giao dịch có thể gây ra tình trạng khóa song, khóa “c ng” và vấn đề bất khả thi. Để giải quyết vấn đề này, chúng ta sẽ dùng bước xếp lịch và các nghi thức.

B x p l ch là thành phần của hệ thống CSDL, có vai trò làm trung tài phân xử các yêu cầu có xung đột. Chúng ta đã biết cách giải quyết khác song của một bước xếp lịch “n tr c ph c v tr c”. Một bước x p l ch c ng có thể xử lý các khóa “c ng” và tính bất khả thi bằng cách

1. Bu c m t giao dịch phải, chúng ta cho nên khi khóa đang yêu cầu phải ghi phóng.
2. Bu c m t giao dịch ng ng l i và tái khởi động.

3.3.1.9. Nghi thức

Chúng ta c ng có thể sử dụng một công cụ khác xử lý khóa gài và tính bất khả thi. Công cụ đó là các nghi thức mà tất cả các giao dịch phải tuân theo.

Một nghi thức, theo nghĩa tổng quát nhất, chỉ là một hành vi trên chuỗi các bước cần mà một giao dịch có thể thực hiện.

3.3.2. Mô hình giao dịch nghiêm

3.3.2.1. Ý nghĩa của giao dịch – hàm đặc trưng

Về nguyên tắc, ý nghĩa của giao dịch T_i chính là tác động của chương trình P_i đang thực hiện trên CSDL. Về hình thức, chúng ta gán một hàm đặc trưng cho mỗi P_i Lock A và Unlock A.

Hàm f nh n i là các giá trị của tất cả các mức khóa bị giao dịch T trước khi mở khóa cho A, và giá trị của f là giá trị mức của A sau khi mở khóa A. Chú ý rằng một giao dịch có thể có nhiều hàm nh n i vì một mức A, bởi vì chúng ta có thể khóa và mở khóa một mức A nhiều lần.

Vậy giá trị A_0 là giá trị ban đầu của A trước khi các giao dịch bắt đầu thực hiện, như vậy giá trị của hàm $f(A_0, \dots)$ sau khi Unlock A là giá trị mức của A. Một cách tổng quát giá trị A là giá trị của mức A trước khi lock A thì giá trị mức của A sau khi Unlock A là $f(A, \dots)$ ta ký hiệu: lock A....unlock A $f(A, \dots)$.

Các giá trị mà A có thể nhận trong khi thực hiện giao dịch là những công thức xây dựng bằng cách áp dụng những hàm này cho các giá trị ban đầu của các mức.

Hai công thức khác nhau có thể coi là hai giá trị khác nhau. Hai lịch biểu là hai công thức cho giá trị cuối cùng của mỗi giá trị khác nhau trong hai lịch biểu.

Ta xét ví dụ sau:

T1: begin	T2: begin	T3: begin
Lock A	Lock B	Lock A
Lock A	Lock C	Lock B
Unlock A $f_1(A,B)$	Unlock B $f_3(B,C)$	Unlock C $f_6(A,C)$
Unlock B $f_2(A,B)$	Lock A	Unlock A $f_7(A,C)$
End	Unlock C $f_4(A,B,C)$	End
	Unlock A $f_5(A,B,C)$	
	End	

Chúng ta có ba giao dịch và hai hàm có liên quan với mỗi cặp Lock – Unlock; là hai hàm xuất hiện trên cùng một dòng với Unlock. Chẳng hạn f_1 , hàm A và B làm việc, bởi vì là hai hàm trong Lock A – Unlock A của T_1 . Hàm f_3 chỉ hàm B và C làm việc bởi vì T_2 mở khóa B, và trong Lock B – Unlock B có B và C.

Hình sau trình bày một lịch biểu của hai giao dịch này và tác động của chúng trên các biến A, B, và C.

Biến	Thao tác	A	B	C
1	T1: Lock A	A_0	B_0	C_0
2	T2: Lock B	A_0	B_0	C_0
3	T2: Lock C	A_0	B_0	C_0
4	T2: UnLock B	A_0	$f_3(B_0, C_0)$	C_0
5	T1: Lock B	A_0	$f_3(B_0, C_0)$	C_0
6	T1: UnLock A	$f_1(A_0, f_3(B_0, C_0))$	$f_3(B_0, C_0)$	C_0
7	T2: Lock A	$f_1(A_0, f_3(B_0, C_0))$	$f_3(B_0, C_0)$	C_0
8	T2: UnLock C	$f_1(A_0, f_3(B_0, C_0))$	$f_3(B_0, C_0)$	(i)
9	T2: UnLock A	(ii)	$f_3(B_0, C_0)$	(i)
10	T3: Lock A	(ii)	$f_3(B_0, C_0)$	(i)
11	T3: Lock C	(ii)	$f_3(B_0, C_0)$	(i)
12	T1: UnLock B	(ii)	$f_2(A_0, f_3(B_0, C_0))$	(i)
13	T3: UnLock C	(ii)	$f_2(A_0, f_3(B_0, C_0))$	(iii)
14	T3: UnLock A	(iv)	$f_2(A_0, f_3(B_0, C_0))$	(iii)

Ký hiệu:

$$(i) = f_4(f_1(A_0, f_3(B_0, C_0)), B_0, C_0)$$

$$(ii) = f_5(f_1(A_0, f_3(B_0, C_0)), B_0, C_0)$$

$$(iii) = f_6((ii), (i))$$

$$(iv) = f_7((ii), (i))$$

Chúng ta có thể nhận xét rằng lịch biểu này là không có tính khả thi. Thật vậy, giả sử rằng nó khả thi. Thì thì nếu T_1 thực hiện trước T_2 trong lịch biểu thì, giá trị cuối cùng của B sẽ là: $f_3(f_2(A_0, B_0), C_0)$ chứ không phải $f_2(A_0, f_3(B_0, C_0))$.

Chúng ta thử nghĩ thì các công thức của hàm sinh ra một giá trị duy nhất là một chuỗi lặp lại, thì thật sự là một sự mâu thuẫn:

$$f_3(f_2(A_0, B_0), C_0) = f_3(A_0, f_3(B_0, C_0))$$

Trong trường hợp này chúng ta không thể luận lịch biểu bất khả thi.

Cần lưu ý rằng phép kiểm tra tính khả thi bằng hàm có thể là một phương pháp phức tạp vì hai yêu cầu giao dịch và hai thao tác. Hai phép kiểm tra này là một phương pháp

y u vì hai công th c c a các hàm c tr ng có th khác nhau nh ng giá tr c a chúng có th gi ng nhau.

3.3.2.2. Ki m tra tính kh tu n t b ng th có h ng.

xác nh r ng m t b x p l ch nào ó là úng, chúng ta ph i ch ng minh r ng m i l ch bi u S c nó cho phép u có c tính kh tu n t . Vì v y chúng ta c n có m t phép ki m tra n gi n v tính kh tu n t c a m t l ch bi u.

Thu t toán 3.3. Ki m tra tính kh tu n t m t l ch bi u S

Vào: M t l ch bi u S c a m t t p các giao d ch T_1, T_2, \dots, T_k

Ra: Kh ng nh S có kh tu n t hay không. N u có thì a ra m t l ch bi u tu n t t ng ng v i S.

Ph ng pháp:

Xây d ng m t th có h ng G (c g i là th tu n t hóa) có các nút là các giao d ch T_i . xác nh các cung c a th G , g i (a_1, a_2, \dots, a_N) là t p các b c c b n trong S. Trong ó m i a_i là m t thao tác có d ng

T_j : Lock A ho c T_j : Unlock A

N u a_j là thao tác ki u Unlock A thì tìm thao tác a_p k ti p sau a_j có d ng T_s : Lock A. N u có m t c p thao tác nh th và $s \neq j$, chúng ta v m t cung t T_j n T_s ($T_j \rightarrow T_s$). Cung này có ý ngh a là trong m t l ch bi u tu n t t ng ng v i S, T_j ph i th c hi n tr c tr c T_s .

N u G có m t chu trình thì S là b t kh tu n t , ng c l i n u G không có chu trình thì S là kh tu n t và chúng ta tìm m t th t tuy n tính cho các giao d ch cho T_i b ng m t quá trình g i là s p x p topo c a th G nh sau:

S p x p Topo th có h ng G không có chu trình

Ta bi t r ng trong G ph i có m t nút T_j nào ó không có cung n, n u không G có m t chu trình. Li t kê T_j r i lo i T_j ra kh i G . Sau ó l p l i quá trình này trên th còn l i cho n khi không còn nút nào n a. Th t các nút c li t kê trong danh sách là m t th t tu n t c a các giao d ch. Th t ó t o nên l ch bi u tu n t t ng d ng v i S.

3.3.3. Nghi th c khóa 2 pha

Chúng ta c n ph i hi u rõ nh ng i u ki n m t l ch bi u là kh tu n t nh m tìm c m t b x p l ch và m t nghi th c, m b o r ng m i l ch bi u mà chúng ta cho phép u kh tu n t .

Nghi th c c g i là nghi th c hai pha, n u m i giao d ch th c hi n t t c các thao tác khóa tr c t t c các thao tác m khóa. Các giao d ch tuân th theo nghi th c này c g i là các giao d ch hai pha: pha u là khóa và pha th hai là pha m khóa.

Nghi th c hai pha có c i m là m i t p giao d ch tuân theo nghi th c này u có các l ch bi u kh tu n t . Tr c tiên chúng ta s ch ng minh r ng nghi th c hai pha b o m c tính kh tu n t .

nh lý 3.1. N u S là m t l ch bi u c a các giao d ch hai pha thì S là kh tu n t .

Ch ng minh:

Gi s kh ng nh trên không úng. V y th G c a l ch bi u S s có m t chu trình

$T_{i1} \rightarrow T_{i2} \rightarrow \dots \rightarrow T_{in} \rightarrow T_{i1}$

Do ó có m t thao tác khóa do T_{i2} s i sau m t thao tác m khóa do T_{i1} , m t thao tác kho do T_{i3} s i sau m t thao tác m khóa T_{i2} ... Cu i cùng m t thao tác khóa T_{i1} s i sau m t thao tác m khóa T_{i1} . i u này mâu thu n v i gi thi t T_{i1} là giao d ch hai pha.

V y kh ng nh trên là úng.

3.3.4. Mô hình khóa c và khóa ghi

N u chúng ta phân bi t m t truy xu t ch c và m t truy xu t c – ghi, chúng ta có th phát tri n m t mô hình chi ti t h n cho các giao d ch lo i c ghi này. Trong m t s tr ng h p mô hình này cho phép s d ng m t s ho t ng ng th i b c m ví d nh nhi u giao d ch có th cùng gi khóa c m t m c A.

Chúng ta phân bi t hai lo i khóa.

1. Khóa c Rlock. M t giao d ch T ch c m t m c A s th c hi n l nh Rlock A, ng n không cho b t k giao d ch khác ghi giá tr m i lên A, tuy nhiên các giao d ch khác v n có th gi m t khóa c trên A cùng lúc v i T.

2. Khóa ghi Wlock. Khi m t giao d ch mu n thay i giá tr c a m c A u tiên s l y khóa ghi b ng cách th c hi n l nh Wlock A, khi m t giao d ch ang gi khóa ghi trên m t m c, nh ng giao d ch khác không th l y c khóa c ho c khóa ghi trên m c ó.

Nh v y khóa c m c A ch cấm các giao d ch khác ghi d li u vào a, còn khóa ghi trên m c c m các giao d ch khác c ghi ho c c m c A.

3.3.4.1. Ý ngh a c a giao d ch v i khóa c và khóa ghi

Chúng ta gi s m i l n áp d ng khóa ghi cho m t m c A s có m t hàm duy nh t i kèm v i khóa ó và nó t o ra m t giá tr m i cho A; hàm ó ph thu c vào t t c các m c b khóa tr c khi m khóa A. Tuy nhiên chúng ta gi s r ng m t khóa c trên A không làm thay i A.

Gi s r ng m c A có m t giá tr ban u A_0 và tác d ng c a m t l ch bi u trên CSDL có th c di n t b ng nh ng công th c c a các hàm c t ng f, là nh ng giá tr c a m i m c c ghi ít nh t là m t l n b i các giao d ch. Tuy nhiên vì có th có m t giao ch c các m c nh ng không ghi gì ho c c m t s m c ch sau khi ghi vào l n cu i cùng, vì th nh ng giá tr mà m i m c ang có khi m t giao d ch ch c nó c ng c x lý nh giá tr c . Vì v y chúng ta có th nói hai l ch bi u là t ng ng n u:

1. Chúng sinh ra cùng giá tr cho m i m c.

2. M i khóa c c áp d ng b i m i giao d ch trong c hai l ch bi u vào nh ng lúc m c b khóa có cùng giá tr .

3.3.4.2. th tu n t hóa trong các giao d ch Rlock và Wlock

Chúng ta xét nh ng i u ki n mà trong ó, t ý ngh a c a các giao d ch và các l ch bi u, ta có th suy ra c khi nào m t giao d ch ph i i tr c m t giao d ch khác trong m t l ch bi u tu n t t ng ng. Gi s r ng có m t l ch bi u S trong ó m t khóa ghi m c A b i giao d ch T_1 , và g i f là hàm i kèm v i khóa ghi ó. Sau khi T_1 m khóa A, g i T_2 là m t trong nh ng giao d ch k ti p nh n khóa c A tr c khi m t giao d ch khác nh n khóa ghi A. Ch c ch n r ng T_1 ph i th c hi n tr c T_2 trong m t l ch bi u tu n t t ng ng v i S. N u không thì T_2 c m t tía tr c a A không ch a hàm f. T ng t , n u T_3 là giao d ch k ti p sau T_1 nh n khóa ghi A thì T_1 ph i th c hi n tr c T_3 .

Bây giờ ta giả sử rằng T_4 là một giao dịch nhúng khóa c của A trước khi T_1 khóa ghi A . Nếu T_1 xuất hiện trước T_4 trong lịch biên tập thì T_4 có một giá trị của A có chứa hàm f , còn trong lịch biên tập S , giá trị của c bởi T_4 không chứa f . Vì vậy, T_4 phải thực hiện trước T_1 trong một lịch biên tập tổng quát của S . Suy luận duy nhất không thể thực hiện được là nếu trong S hai giao dịch cùng nhúng khóa c của A theo một thứ tự nào đó thì những giao dịch này phải xuất hiện theo thứ tự đó trong một lịch biên tập. Kết quả thực tế tổng quát này là các khóa c không tạo ra sự khác biệt nào trên các giá trị của c ở các giao dịch thực hiện theo thứ tự.

Thuật toán 3.4. Kiểm tra tính khả thi của các lịch biên tập với các khóa c /ghi.

Vào: Một lịch biên tập S cho một tập các giao dịch T_1, T_2, \dots, T_N

Ra: Khả năng nhúng S có khả thi hay không, nếu không thì các lịch biên tập tổng quát.

Phương pháp:

Chúng ta xây dựng một đồ thị G như sau. Các nút tổng quát là các giao dịch. Các cung của xác định bằng quy tắc sau:

1. Giả sử trong S , giao dịch T_i nhúng khóa c hoặc khóa A . T_j là giao dịch kết tiếp khóa ghi A , và $i \neq j$. Khi đó chúng ta thêm một cung từ T_i đến T_j .

2. Giả sử trong S , giao dịch T_i khóa ghi A . Gọi T_m với $m \neq i$ là một giao dịch khác khóa c của A sau khi T_i mở khóa A trong các giao dịch khác khóa ghi A . Chúng ta thêm một cung từ T_i đến T_m .

Nếu G có chu trình thì S là bất khả thi, ngược lại thì một sắp xếp topo của G là thứ tự thực hiện cho các giao dịch này.

Chương 4. Hệ thống giúp ra quyết định

4.1. Giới thiệu về hệ thống giúp ra quyết định

Các hệ thống trợ giúp ra quyết định là các hệ thống giúp việc quản lý xác định đúng vấn đề, ra quyết định một cách thông minh. Hệ thống này dựa trên lý thuyết tính toán, lý thuyết hành vi quản lý, dựa trên quá trình thực tế...

Cuối thế kỷ 60, người ta dùng máy tính trong quá trình ra quyết định. Ban đầu là thông tin hóa công việc lập các báo cáo và đôi lúc là tính toán, phân tích. Các hệ thống đó có thể gọi là hệ thống quyết định quản lý, và sau đó xây dựng thành những hệ thống thông tin quản lý.

Vào những năm 70, người ta phát triển những ngôn ngữ lập trình, và xây dựng trên những ngôn ngữ này một số hệ thống trợ giúp ra quyết định chuyên dùng. Các hệ thống này là những bộ công cụ tiên cho phép người dùng có kiến thức truy vấn trực tiếp dữ liệu, tổ chức thành các câu hỏi liên quan đến kinh doanh mà không cần sự trợ giúp của các chuyên gia tin học.

Ngày nay các sản phẩm của ngôn ngữ lập trình SQL có dùng một cách riêng rẽ, ý tưởng phân tích dữ liệu là vì sao chọn các dữ liệu từ môi trường đang hoạt động sang một môi trường khác, và rút quan trọng. Ý tưởng này cho phép người ta thao tác với các dữ liệu có trích ra theo các yêu cầu mà không cần quá trình suy luận trong môi trường đang hoạt động. Đồng thời trong quá trình ra quyết định, người ta có thể sử dụng một phần chỉ số để đưa ra quyết định không cần tính toán.

Khác với hệ thống CSDL, hệ thống trợ giúp ra quyết định chưa có lý thuyết riêng và hoàn chỉnh. Người ta thường yêu cầu mô hình logic để với các hệ thống này trợ giúp ra quyết định. Nhưng không phân biệt rõ mô hình logic với mô hình vật lý.

Các khía cạnh về trợ giúp ra quyết định

Các CSDL trợ giúp ra quyết định thể hiện các đặc tính của hệ thống CSDL. Tuy nhiên đây không áp dụng cho tất cả các hệ thống trợ giúp ra quyết định. Do vậy, các phép toán xử lý dữ liệu như cập nhật dữ liệu, rút trích khi cần sử dụng; người ta thường tìm kiếm dữ liệu hay làm việc với dữ liệu thay cho việc cập nhật trực tiếp. Đôi khi trên những bảng trung gian, người ta có thể thể hiện các phép toán cập nhật dữ liệu, nhưng quá trình ra quyết định bình thường không bao gồm cập nhật CSDL trợ giúp quyết định.

Có sáu đặc tính của hệ thống trợ giúp quyết định, ba đặc tính liên quan đến khía cạnh logic và ba đặc tính còn lại liên quan đến khía cạnh vật lý. Các khái niệm này đồng, các đặc tính dùng theo nghĩa của CSDL quan hệ.

1. Các đặc tính dùng theo nghĩa tập hợp của nhiều thuộc tính.
2. Nói chung hệ thống trợ giúp ra quyết định không có phần tính toán với dữ liệu, do các dữ liệu đã có sẵn để sử dụng.
3. Các khóa thông tin mang ý nghĩa thực tiễn.
4. Kích thước CSDL có xu hướng ngày càng lớn.
5. Các bảng chỉ số dùng trong CSDL thường có kích thước lớn và nhiều.

6. CSDL thường sử dụng nhiều loại dữ liệu, nhưng các dữ liệu này có thể thống kê kiểm soát.

Các câu hỏi trong hệ thống trợ giúp quy tắc thường có các tính riêng, thường là rất phức tạp.

1. Tính phức tạp của biểu thức logic. Các câu hỏi trợ giúp ra quy tắc thường yêu cầu các biểu thức phức tạp trong câu lệnh nên khó viết, khó hiểu.

2. Tính phức tạp của phép toán kết nối quan hệ. Các câu hỏi trợ giúp quy tắc thường truy cập nhiều loại dữ liệu. Trong một CSDL có chuẩn hóa thì các câu hỏi loại này thường yêu cầu truy cập nhiều phép kết nối. Do vậy, người thiết kế hệ thống phi chuẩn CSDL bằng cách tiến hành truy cập các phép kết nối trên các bảng.

3. Tính phức tạp của các hàm số. Các câu hỏi trong hệ thống trợ giúp ra quy tắc thường sử dụng các hàm thống kê, hàm toán học. Chỉ một số ít các hệ thống áp dụng các loại câu hỏi dạng này. Thường thì người thiết kế cho phép câu hỏi ghi nhận những trình diễn vì trong một ngôn ngữ lập trình hoặc ngôn ngữ khai thác dữ liệu khác.

4. Tính phức tạp về phân tích. Việc liệt kê các câu hỏi trong hệ thống trợ giúp ra quy tắc, người ta cần phải sử dụng nhiều câu hỏi con. Điều này gây khó khăn cho người sử dụng và hệ thống.

4.2. Thiết kế CSDL cho hệ thống trợ giúp ra quy tắc

Vì thiết kế CSDL cho hệ thống trợ giúp ra quy tắc có thể nhìn nhận qua hai bước là bước thiết kế logic và thiết kế vật lý:

1. Thiết kế logic: tập trung vào việc chuyển lý các quan niệm theo lý thuyết quan hệ. Các bảng dữ liệu dựa vào các quan hệ. Người ta xác định các miền giá trị của các kiểu dữ liệu, xác định các mối liên kết giữa các thực thể. Tập trung vào việc chuẩn hóa các bảng và xác định các ràng buộc dữ liệu.

2. Bước thiết kế vật lý: tập trung vào tính hiệu quả của lưu trữ và hiệu suất của hệ thống. Về nguyên tắc, người ta có thể lựa chọn bất kỳ một cách thức nào, nhưng việc chuyển hóa mô hình logic sang mô hình vật lý theo các quan hệ là hợp lý hơn.

4.2.1. Thiết kế logic.

Các quy luật thiết kế logic không phụ thuộc vào việc sử dụng CSDL, tức là không phụ thuộc vào loại ngôn ngữ trên CSDL. Do vậy thiết kế cho hệ thống tác nghiệp hay cho hệ thống trợ giúp quy tắc như sau:

1. Tập hợp các thực thể và ít phụ thuộc: Các câu hỏi trợ giúp quy tắc thường xử lý tập hợp các thực thể không dùng các thực thể. Có thể gộp tập hợp các thực thể thành một tập hợp. Theo quan niệm thiết kế logic, các tập hợp có thể cách nhau thành bình đẳng. Điều này có tác động gì đến các phụ thuộc hàm giữa các thực thể. Việc này dẫn đến thiết kế logic nghiêm ngặt, cần nhận xét kỹ lưỡng về việc thiết kế kiểu dữ liệu do người dùng xác định.

2. Các ràng buộc về tính toàn vẹn: hệ thống trợ giúp quy tắc là hệ thống cho phép chuyển đổi dữ liệu và tính toán trên dữ liệu chính xác khi kiểm tra tính đúng đắn của dữ liệu cho nên trong logic người ta không đưa ra các kiểm tra tính toán vẹn.

3. Các khóa vật lý gián: Các CSDL tác nghiệp thường chỉ cần các dữ liệu hiện tại. Các CSDL trợ giúp quy tắc thì ngược lại cần các dữ liệu lịch sử, nên người ta cần tính toán về ảnh hưởng vật lý gián trên dữ liệu các dữ liệu hay trên một vài dữ liệu. Quản lý các mức độ gián, các thuộc tính

khóa c ng c n c ánh d u th i gian. Vì c làm này c n c quá trình thi t k logic l u ý khi xác nh các ph thu c hàm gi a các c t ánh d u m i và các c t trong b ng.

4.2.2. Thi t k v t lý

CSDL trong h th ng tr giúp quy t nh có xu h ng n ng n , kích th c l n, và òi h i nhi u lo i d th a d li u có ki m soát. ng i ta th ng c p khái ni m phân o n và ch s hoá. Vì c phân o n nh m vào CSDL l n, chia các b ng d li u thành t p các ph n, hay các o n r i nhau, phù h p v i công vi c l u tr v t lý. Công vi c này giúp ng i ta qu n lý và tr l i các câu h i trên các b ng d h n. i n hình là m i o n d li u g n v i m t s tài nguyên tin h c nh ph n c ng, th i gian, b x lý trung tâm. i u này có th gi m thi u các c nh tranh v tài nguyên trên m i o n d li u.

Vì c dùng các b ng ch s gi m c th i gian vào/ra khi c n truy c p d li u. Nhi u s n ph m h qu n tr CSDL cung c p cho ng i dùng ch m t lo i ch s là cây cân i, nh ng còn có nhi u lo i ch s khác c dùng trong h th ng tr giúp quy t nh là b ng bit, hàm b m.

D th a d li u có i u khi n là công c quan tr ng gi m các thao tác vào/ra và t i thi u hoá các n i dung. Tính d th a d li u c h qu n tr CSDL qu n lý mà ng i dùng không c n quan tâm:

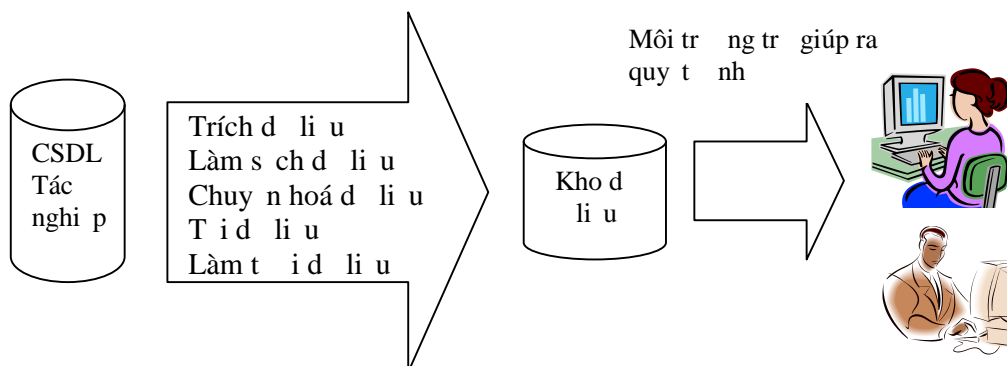
1. Qu n lý các b n sao chính các c a d li u c b n và qu n lý các b n sao.
2. Qu n lý các d li u suy di n bên c nh d li u c b n. Thông th ng ng i ta s d ng các b ng t ng h p và /ho c các c t thu c b ng cách tính toán hay suy di n.

4.3. Kho d li u và kho d li u chuyên

Các h th ng i u hành th ng xuyên yêu c u ch t ch v hi u su t, v kích th c giao tác, v tính ho t ng theo k ho ch, tính ng d ng cao. Ng c l i các h th ng tr giúp quy t nh có nh ng yêu c u a d ng v hi u su t, không bi t tr c v công vi c s làm, kích th c giao tác l n, c s d ng tu theo c m tính c a ng i qu n lý. Nh ng khác bi t này gây nên không ít khó kh n cho vi c tích h p các x lý i u hành v i vi c tr giúp quy t nh trong m t h th ng n mà v n ph i áp ng các yêu c u v k ho ch, qu n lý tài nguyên, và i u ch nh c hi u su t c a h th ng. Do v y, ng i qu n tr h th ng i u hành th ng mi n c ng ch p nh n các ho t ng tr giúp quy t nh trong h th ng c a h .

Ph n này c p m t quan tâm c a các h tr giúp ra quy t nh. ó là vì c truy c p d li u t nhi u ngu n khác nhau, t nhi u h th ng i u hành khác nhau. Nh ng d li u dùng ra quy t nh c l u tr theo cách th c riêng trong kho d li u

Hình 4.1. Ki n trúc t ng quan c a kho d li u



4.3.1. Kho dữ liệu

Thuật ngữ kho dữ liệu được dùng từ cuối những năm 80 của thế kỷ XX. Kho dữ liệu được dùng cho công tác ra quyết định trong quản lý, là tập hợp các dữ liệu thay đổi theo thời gian, không cho phép cập nhật, tích hợp hàng chục. Khái niệm không thể phủ nhận gì thích hợp hơn là một khi được bổ sung, dữ liệu sẽ không thay đổi, dù rằng nó có thể bị xóa. Lí do xuất phát từ khái niệm mới về kho dữ liệu là:

1. Người ta cần sử dụng nguồn dữ liệu, sẵn có, sẵn có và trợ giúp việc ra quyết định
2. Hệ thống trợ giúp ra quyết định không chịu tác động của các hệ thống dữ liệu hành.

/N. Kho dữ liệu (dataware house). Kho dữ liệu là nơi chứa các dữ liệu có ý nghĩa thời gian, không thay đổi, hàng chục, dùng để trợ giúp việc ra quyết định.

/N. Kho dữ liệu xí nghiệp (Enterprise datawarehouse). Kho dữ liệu xí nghiệp là kho dữ liệu tích hợp, tập trung, là nguồn dữ liệu chính khi cần dùng cho việc ra quyết định của các cấp quản lý.

Mục đích của kho dữ liệu là ra quyết định, nên cần tập trung các chức năng xử lý dữ liệu và kích thích các kho dữ liệu thành có xu hướng liên.

/N. Tính hạn chế phạm vi. Tính hạn chế phạm vi khi xét nhu cầu cho phép xét một yếu tố mà không quan tâm đến các yếu tố khác, tức là gì thì các yếu tố khác là không thay đổi.

Người ta thường có một số vấn đề cần xem xét và dùng khi hạn chế phạm vi trong hệ thống trợ giúp ra quyết định

- Các sai sót về mặt kỹ thuật CSDL;
- Sẽ không hiểu được các phép toán quan hệ;
- Hạn chế không nắm bắt được các hệ quả của CSDL theo mô hình quan hệ;
- Các lỗi về mặt kỹ thuật khi cấu trúc làm hạn chế khả năng của hệ thống.

4.3.2. Kho dữ liệu chuyên

Các kho dữ liệu nhằm cung cấp nguồn chính xác các dữ liệu dùng cho các hoạt động trợ giúp ra quyết định. Tuy nhiên, khi các kho dữ liệu trở nên thông dụng, người ta thường nghĩ đến những hạn chế của các thao tác phân tích dữ liệu và ra báo cáo trên một phần nhỏ của kho dữ liệu. Hạn chế này có thể xảy ra hay lặp lại cùng một thao tác trên các phần nhỏ của kho dữ liệu. Vì vậy cần hạn chế những hạn chế thao tác trên toàn bộ kho dữ liệu là không hiểu được, cho nên người ta cần có một kho dữ liệu chuyên dụng, có người xây dựng theo yêu cầu cụ thể riêng. Có thể thấy vì vậy truy cập các dữ liệu sẽ nhanh hơn kho dữ liệu thông thường.

/N. Kho dữ liệu chuyên (data mart). Kho dữ liệu chuyên có vai trò như kho dữ liệu, nhưng các dữ liệu trong đó cho phép cập nhật và dùng để trợ giúp quyết định về mặt chức năng bị hạn chế.

Kho dữ liệu chuyên là kho dữ liệu hạn chế, gồm các dữ liệu cần tập trung và tăng cường phần kho dữ liệu của xí nghiệp.

Để tạo ra một kho dữ liệu chuyên, người ta thường theo ba cách tiếp cận sau:

1. Trích dữ liệu từ kho dữ liệu. Các dữ liệu cần trích từ kho dữ liệu để thể hiện sự tập trung vào và có tính hạn chế phạm vi. Thông thường các dữ liệu trích ra này sẽ đi vào CSDL có

l c v t lý g n gi ng v i ph n ng d ng c a kho d li u. Do tính c bi t h n c a kho d li u chuyên so v i kho d li u, l c v t lý c a d li u có th n gi n h n.

2. T o ra kho d li u chuyên riêng bi t. Tì p c n này xu t phát t tính n th c a kho d li u, không trích d li u t kho d li u và không truy c p kho d li u do m t vài nguyên nhân.

3. Coi kho d li u chuyên là n n t ng c a kho d li u. M t vài phát tri n h th ng tr giúp ra quy t nh ã xây d ng các kho d li u chuyên tr c tiên, m i khi c n thi t. Kho d li u s c tạo ra b ng cách t p h p các kho d li u chuyên .

/N. Tính chia h t. Tính chia h t trong c d li u c p kh n ng l u tr c ph n t nh nh t c u d li u g p trong CSDL.

Liên quan n vi c thi t k kho d li u chuyên , ng i ta nh n th y m t y u t quan tr ng i v i b t kì CSDL tr giúp ra quy t nh nào là tính chia nh thành h t c a CSDL. S m hay muôn thì các kho d li u dùng ra quy t nh u yêu c u truy c p d li u chi ti t nh t, nên yêu c u chi thành h t i v i kho d li u không gây ra v n l n nh i v i kho d li u chuyên . N u kho d li u chuyên c xây d ng b ng cách trích các d li u t kho d li u mà không bi t các ng d ng có nhu c u th ng xuyên v các d li u m c h t hay không, thì vi c trích d li u và c p nh t các d li u m c h t s t n kém nhi u.

4.3.3. Các l c v chi u.

Các h th ng tr giúp ra quy t nh th ng c n n k t qu phân tích v l ch s c a các giao d ch tác nghi p. Thông tin này c l u trong các t p và c truy c p tu n t . Do nhu c u, n m t lúc nào ó ng i ta c n tr c ti p truy c p các thông tin này ch theo m t s góc c nh c n quan tâm. Ch ng h n i v i thông tin v s n l ng r u vang, ng i ta c n bi t v s n l ng, v ng i s n xu t, v tu i c a r u... tr giúp nhu c u truy c p này, ng i ta dùng CSDL có nhi u b ng tra c u. C s d li u nh v y có t p d li u trung tâm ch a cá d li u v các ho t ng tác và nhi u b ng tra c u v s n l ng, ng i s n xu t, tu i c a r u. Các b ng này t a nh b ng ch s vì chúng có con tr tr n các b n ghi trong t p d li u, nh ng khác v i b ng ch s ch ng i dùng có th tác ng n các b ng tra c u theo cách t ng minh và b ng tra c u có th mang các thông tin ph , ch ng h n nh a ch c a nhà s n xu t.

Các t ch c nhi u b ng tra c u có u i m h n so v i vi c dùng m t t p tra c u, c v không gian nh l n th i gian vào/ra. Khi dùng tì p c n này trong c s d li u quan h , t p d li u và các t p tra c u tr thành các b ng, t c là nh c a các t p; các con tr trong t p tra c u tr thành khoá chính c a b ng tra c u; nh ng tên trong t p d li u tr thành cá khoá ngoài trong b nh nh c a t p d li u. Tr ng h p i n hình là các khoá chính và khoá ngoài u c ch s hoá. Theo ph ng th c này, nh c a t p d li u c g i là b ng s ki n và các nh c a t p tra c u c g i là các b ng v chi u. Thi t k t ng th phù h p c g i là l c hình sao, hay l c v chi u, vì trong thi t k th c th quan h ng i ta n i r ng các b ng s ki n, n i v i cá b ng v chi u.

Ví d CSDL RUOU(TenRuou, NhaSX, NamSX, SoLuong) trong ó thu c tính NamSX c mô t b ng kho ng t n m_{t1} n n m_{t2}. Theo thu t ng l c hình sao thì b ng RUOU c g i là b ng s ki n còn b ng SanXuat(Nam, NamBD, NamKT) c g i là b ng v chi u.

Ruou	TenRuou	NhaSX	NamSX	SoLuong	SanXuat	Nam	NamBD	NamKT
	Lúa m i	Hà N i	1980	200		1980	1980	1985
	Th ng Long	Hà N i	1990	300		1990	1986	1992

/N. L c hình sao. L c hình sao là thi t k CSDL n gi n, trong ó các d li u v chi u c tách kh i các d li u v s ki n. Mô hình v chi u là tên g i khác c a l c hình sao.

Câu hỏi trên CSDL theo lược đồ hình sao cần các bảng view chỉ để phát hiện lỗi sai sót trong kết quả truy vấn ngoài cần thì, rồi dùng kết quả này truy cập bảng sự kiện. Vì vậy truy cập các bảng view chỉ để và truy cập bảng sự kiện để hiển thị trong màn hình câu hỏi, thì cách tốt nhất để hiển thị câu hỏi này thì sẽ theo kết quả hình sao.

Kết quả hình sao là chỉ nhận thu thập dữ liệu thì hiển thị phép kết quả thì hiển thị theo hai bước.

1. Tiến hành phép tích các dữ liệu với các bảng view chỉ để. Lưu ý rằng khi tiến hành hoá câu hỏi, người ta thường tránh sử dụng phép tích các. Trong trường hợp này, các bảng kích thước nhỏ tham gia trực tiếp vào phép tích các.

2. Dùng kết quả tích các để quản lý bảng sự kiện theo kết quả thu thập hoá. Kết quả thu thập hoá cho phép chỉ nhận thu thập dữ liệu này.

Một biểu diễn dạng lược đồ hình sao là lược đồ hoa tuyết, thì hiển thị vì cần chuyển hoá các bảng chỉ để.

/N. Lược đồ hoa tuyết. Lược đồ hoa tuyết là biểu diễn dạng lược đồ hình sao, trong đó các bảng cần chuyển hoá.

4.4. X lý phân tích truy vấn

4.4.1. Giới thiệu.

Thuật ngữ xử lý phân tích truy vấn OLAP nhằm vào quá trình tác động và hình thành, bao gồm các thao tác tổ hợp, quản lý, phân tích và lập các báo cáo dữ liệu, thông qua các câu lệnh yêu cầu tìm kiếm, và xử lý dữ liệu trong bảng dữ liệu nhiều chiều.

/N. Xử lý phân tích truy vấn. Xử lý phân tích truy vấn là một dạng của các CSDL quan hệ giúp các mô hình kinh doanh. Quá trình OLAP dùng các quy tắc truy cập nhanh và tìm kiếm các dữ liệu cho các hình thức quản lý thông tin hay trợ giúp ra quyết định. OLAP là vì cần dùng tập các công cụ hoặc ngôn ngữ để nhiều chiều dữ liệu, cho phép phân tích các dữ liệu.

Quá trình phân tích yêu cầu dữ liệu, theo nhiều cách, từ theo cách thức nhóm dữ liệu khác nhau. Một trong những khía cạnh của xử lý phân tích truy vấn là sự phân nhóm dữ liệu nhiều chiều lên rất nhanh, và người dùng cần xem xét tất cả các khía cạnh.

4.4.2. Bảng chéo

Các kết quả của quá trình phân tích truy vấn sẽ ra thành không theo dạng bảng quan hệ, mà theo dạng hai chiều, cũng là bảng chéo. Xét câu hỏi “Tìm tổng sản lượng do nhà công ty sản xuất cung cấp trên bảng dữ liệu này:

Ruou

Tên	CongTy	Nam	SoLuong
Lúa mì	Hà Nội	1980	200
Vang	Hà Nội	1995	350
Vang	Huda Hu	1990	450
Lúa mì	Hà Bắc	1990	320

Người ta xây dựng một bảng mới gọi là bảng chéo để hiển thị câu trả lời. Bảng chéo khác với bảng quan hệ, các cột phụ thuộc vào dữ liệu thực, tức là cấu trúc của bảng liên hệ giữa các dòng phụ thuộc vào giá trị thực của dữ liệu. Do vậy bảng chéo không phải là quan hệ, mà chỉ là một báo cáo có hình thức như ma trận hai chiều. Ví dụ:

	Lúa mì	Vang	Tổng
Hà Nội	200	350	550
Huda Hu	0	450	450

Hà B c	320	0	320
T ng	520	800	1320

4.4.3. CSDL nhi u chi u

Quá trình x lý phân tích tr c tuy n có th c t trong môi tr ng quan h , và quá trình này còn c g i là OLAP quan h vì t t t là ROLAP. Trong th c t nhi u ng i nh n th y có cách t p c n t t h n, là OLAP nhi u chi u vì t t t là MOLAP.

MOLAP c n n CSDL nhi u chi u. Vì c l u tr này c g i là l u tr theo quan ni m. hình dung th y nhi u chi u, nh ng trong các h th ng th c t thì t ch c v t lý c a MOLAP r t g n v i t ch c logic. H qu n tr CSDL trong tr ng h p này c g i là h qu n tr CSDL nhi u chi u. Các CSDL có th c th hi n trong b ng hai, ba chi u.

4.5. Khai phá d li u

Khai phá d li u, m c ích c a nó là nhìn vào phân áng quan tâm c a d li u, nh ng ph n c dùng thì t l p chi n l c kinh doanh hay xác nh hành vi khác th ng.

/N. Khai phá d li u. Khai phá d li u là quá trình trích ra nh ng thông tin dùng c, úng và ch a bi t tr c t CSDL l n, r i dùng thông tin này ra các quy t nh.

Các công c khai phá d li u dùng các k thu t th ng kê i v i kh i l ng d li u l n tìm ra nh ng ph n d li u c n thi t. Các CSDL trong khai phá d li u th ng r t l n, do v y có xu h ng n gi n hoá các thu t toán. M t s thu t ng dùng trong khai phá d li u nh “dân s ” ch các thao tác có th th c hi n trong m t b ng d li u, “lu t liên k t” ch s ph thu c gi a các d li u khi xét các giao tác... Lu t liên k t c phát hi n do áp d ng các phép toán g p phù h p. M t s qui lu t khác c xác nh nh lu t “t ng quan tu n t”, lu t “phân lo i”.