

Chương 10

Phân rã lược đồ (Decomposition)

Trần Thị Kim Chi

1

1

Nội dung

- Mục đích phân rã
- Định nghĩa phân rã
 - Phân rã không mất thông tin
 - Phân rã bảo toàn phụ thuộc
- Phân rã thành BCNF
- Phân rã thành 3NF
 - Phân rã thông thường
 - Tổng hợp

Trần Thị Kim Chi

2

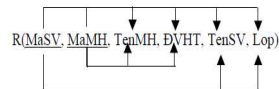
2

Mục đích của phân rã lược đồ quan hệ

Định nghĩa

- Phép phân rã các lược đồ quan hệ $R = \{A_1, A_2, \dots, A_n\}$ là việc thay thế lược đồ quan hệ R thành các lược đồ con $\{R_1, \dots, R_k\}$, trong đó $R_i \subseteq R$ và $R = R_1 \cup R_2 \dots \cup R_k$
- Ví dụ: Cho quan hệ R với các phụ thuộc hàm như sau:

Ví dụ: Cho quan hệ R với các phụ thuộc hàm như sau:



Ta có thể phân rã thành 3 lược đồ $R_1(\text{MaSV}, \text{TenSV}, \text{Lop})$ và $R_2(\text{MaMH}, \text{TenMH}, \text{ĐVHT})$ và $R_3(\text{MaSV}, \text{MaMH})$.

Trần Thị Kim Chi

3

3

Mục đích của phân rã lược đồ quan hệ

- Được xem như 1 công cụ bổ sung vào phương pháp ER để loại trừ dư thừa dữ liệu
 - Phụ thuộc hàm được xem như là sự khái quát hóa các ràng buộc chính (key constraint). Các FD được dùng để xác định các dạng chuẩn (normal form). Việc phân rã lược đồ sẽ dựa theo các dạng chuẩn này
- Lý thuyết phân rã còn được gọi là lý thuyết chuẩn hóa.

Trần Thị Kim Chi

4

4

Tính chất của phân rã lược đồ

- Vì chuẩn BCNF không có dư thừa và chuẩn 3NF tuy có dư thừa nhưng cũng hạn chế, nên việc phân rã lược đồ sẽ chỉ tập trung vào hai dạng này.

Trần Thị Kim Chi

5

Phân rã lược đồ – Decomposition

- Phân rã 1 lược đồ $R = (U, F)$ với U là tập các thuộc tính, F là tập phụ thuộc hàm sẽ cho ra 1 tập hợp các lược đồ

$$R_1 = (U_1, F_1) \quad R_2 = (U_2, F_2) \dots$$

$$R_n = (U_n, F_n)$$

Sao cho thỏa mãn điều kiện sau:

- $U = \bigcup_{i=1}^n U_i$
- F suy dẫn F_i với $i = 1, \dots, n$

Trần Thị Kim Chi

6

Phân rã lược đồ – Decomposition

- Phân rã lược đồ sẽ dẫn đến việc **phân rã quan hệ**.
- Phân rã 1 quan hệ r trên lược đồ R , cho ra 1 tập hợp các quan hệ

$$r_1 = \pi_{U_1}(r) \quad r_2 = \pi_{U_2}(r), \dots$$

$$r_n = \pi_{U_n}(r)$$

Trần Thị Kim Chi

7

Phân rã không mất mát thông tin (Lossless decomposition)

- Khảo sát quan hệ r và các phân rã của nó r_1, \dots, r_n
- Sau phân rã, CSDL không còn lưu trữ quan hệ r nữa mà chỉ lưu lại các quan hệ chiếu của nó r_1, \dots, r_n . CSDL phải có khả năng khôi phục lại quan hệ gốc r từ các quan hệ chiếu này.
- Nếu không khôi phục lại được quan hệ r thì việc phân rã không biểu diễn cùng 1 thông tin với CSDL gốc → Phân rã mất mát thông tin (lossy decomposition)

Trần Thị Kim Chi

8

Phân rã kết nối không mất mát thông tin (Lossless-join decomposition)

- Phân rã lược đồ $R = (U, F)$ thành 1 tập hợp các lược đồ

$$R_1 = (U_1, F_1) \quad R_2 = (U_2, F_2) \dots$$

$$R_n = (U_n, F_n)$$
- Không mất mát (lossless) nếu với mỗi điển hình (instance) hợp lệ r của lược đồ R thì

$$r = r_1 \bowtie r_2 \bowtie \dots \bowtie r_n$$
 Với $r_1 = \pi_{U_1}(r) \quad r_2 = \pi_{U_2}(r), \dots$
 $r_n = \pi_{U_n}(r)$

Trần Thị Kim Chi

9

9

Phân rã mất mát thông tin (Lossless decomposition)

- Cho lược đồ quan hệ $Q(\text{TENNCC}, \text{DIACHI}, \text{SANPHAM}, \text{DONGIA})$ có quan hệ tương ứng là r
- Đặt r_1 là quan hệ có được bằng cách chiếu r lên $Q_1(\text{TENNCC}, \text{SANPHAM}, \text{DONGIA})$,
- Đặt r_2 là quan hệ có được bằng cách chiếu r lên $Q_2(\text{TENNCC}, \text{DIACHI})$
- Đặt r' là quan hệ có được bằng cách kết tự nhiên giữa r_1 và r_2 qua TENNCC .

Trần Thị Kim Chi

10

10

Phân rã mất mát thông tin (Lossless decomposition)

r	TENNCC	DIACHI	SANPHAM	DONGIA
	Hung	12 Nguyễn Kiệm	Gạch ống	200
	Hung	12 Nguyễn Kiệm	Gạch thẻ	250
	Hung	40 Nguyễn Oanh	Gạch ống	200

$r_2 = r \cdot Q_2^+$	TENNCC	DIACHI
	Hung	12 Nguyễn Kiệm
	Hung	40 Nguyễn Oanh

$r_1 = r \cdot Q_1^+$	TENNCC	SANPHAM	DONGIA
	Hung	Gạch ống	200
	Hung	Gạch thẻ	250

$r' = r_1 \bowtie r_2$	TENNCC	DIACHI	SANPHAM	DONGIA
	Hung	12 Nguyễn Kiệm	Gạch ống	200
	Hung	12 Nguyễn Kiệm	Gạch thẻ	250
	Hung	40 Nguyễn Oanh	Gạch ống	200
	Hung	40 Nguyễn Oanh	Gạch thẻ	250

Kết quả là $r \neq r'$ hay $r \neq r_1 \bowtie r_2$.
 → phép tách $p(Q_1, Q_2)$ tách Q thành Q_1, Q_2 là tách-kết nối (phân rã) mất mát thông tin.

Trần Thị Kim Chi

11

11

Ví dụ phân rã kết nối mất mát thông tin

Cho quan hệ r

S	P	D
s1	p1	d1
s2	p2	d2
s3	p1	d3

Instance r

Kết nối tự nhiên 2 quan hệ phân rã này:

S	P	D
s1	p1	d1
s2	p2	d2
s3	p1	d3
s1	p1	d3
s3	p1	d1

 $\pi_{SP}(r) \bowtie \pi_{PD}(r)$

Phân rã thành 2 quan hệ

S	P
s1	p1
s2	p2
s3	p1

 $\pi_{SP}(r)$

P	D
p1	d1
p2	d2
p1	d3

 $\pi_{PD}(r)$

→ Có những bộ không thuộc quan hệ gốc ban đầu

Trần Thị Kim Chi

13

13

Thuật toán kiểm tra không mất mát thông tin (Lossless-join decomposition)

- **Input:**
 - Lược đồ quan hệ $R = \{A_1, A_2, \dots, A_n\}$
 - Tập các phụ thuộc hàm F
 - Phép tách $\rho(R_1, R_2, \dots, R_k)$
- **Output:** Kết luận phép tách ρ không mất mát thông tin.

Trần Thị Kim Chi

15

15

Phân rã không mất mát thông tin (Lossless-join decomposition)

- ☀ Ví dụ 10: cho $Q(SAIP)$, $Q_1 = (SA)$, $Q_2 = (SIP)$
 $F = \{S \rightarrow A, SI \rightarrow P\}$. Hỏi việc tách Q thành Q_1 và Q_2 có gây ra mất mát thông tin không?
- Áp dụng tính chất trên, ta có

$$Q_1^+ \cap Q_2^+ = S$$

$$Q_1^+ \cup Q_2^+ = SAIP = Q^+$$

$$S \rightarrow SA = Q_1^+$$
- Theo tính chất trên, với mọi quan hệ r của Q ta luôn có

$$r = r.Q_1 \bowtie r.Q_2.$$
- Suy ra phép tách trên là phép tách kết nối bảo toàn thông tin.

Trần Thị Kim Chi

16

16

Thuật toán kiểm tra không mất mát thông tin (Lossless-join decomposition)

Các bước của thuật toán kiểm tra phép tách kết nối bảo toàn thông tin:

- Dữ liệu vào: lược đồ quan hệ $Q(A_1, A_2, \dots, A_n)$, tập phụ thuộc hàm F , phép tách $\rho = (Q_1, Q_2, \dots, Q_k)$.
- Dữ liệu ra: kết luận phép tách ρ có phải là phép tách bảo toàn thông tin ?
- **Bước 1:**
 - Thiết lập một bảng với n cột (tương ứng với n thuộc tính) và k dòng (tương ứng với k quan hệ), trong đó cột thứ j ứng với thuộc tính A_j , dòng thứ i ứng với lược đồ R_i .
 - Tại dòng i và cột j , ta điền ký hiệu a_j nếu thuộc tính $A_j \in R_i$. Ngược lại ta điền ký hiệu b_{ij} .

Trần Thị Kim Chi

17

17

Thuật toán kiểm tra không mất mát thông tin (Lossless-join decomposition)

Các bước của thuật toán:

- **Bước 2:**
 - Xét các phụ thuộc hàm trong F và áp dụng cho bảng trên.
 - Giả sử ta có phụ thuộc hàm $X \rightarrow Y \in F$, xét các dòng có giá trị bằng nhau trên thuộc tính X **thì làm bằng** các giá trị của chúng trên Y .
 - Ngược lại làm bằng chúng bằng ký hiệu b_{ij} . Tiếp tục áp dụng các pth cho bảng (kể cả việc lặp lại các phụ thuộc hàm đã áp dụng) cho tới khi không còn áp dụng được nữa.

Trần Thị Kim Chi

18

18

Thuật toán kiểm tra không mất mát thông tin (Lossless-join decomposition)

Các bước của thuật toán:

- Bước 3:
 - Xem xét bảng kết quả. Nếu xuất hiện một dòng chứa toàn giá trị **a1, a2, ..., an** thì kết luận phép tách ρ không mất mát thông tin.

Trần Thị Kim Chi

19

19

Thuật toán kiểm tra không mất mát thông tin (Lossless-join decomposition)

Ví dụ:

Ví dụ: Cho quan hệ:
EMP_DEPT

ENAME	SSN	BDATE	ADDRESS	DNUMBER	DNAME	DMGRSSN
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5	Research	333445555
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5	Research	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4	Administration	987654321
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire, TX	4	Administration	987654321
Narayan, Ramesh K.	666884444	1962-09-15	975 Fire Oak, Humble, TX	5	Research	333445555
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5	Research	333445555
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4	Administration	987654321
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1	Headquarters	888665555

Hình 6.15. Minh họa dữ liệu của quan hệ EMP_DEPT

Trần Thị Kim Chi

20

20

Thuật toán kiểm tra không mất mát thông tin (Lossless-join decomposition)

Ví dụ:

- Tách thành 2 quan hệ:
EMPLOYEE

ENAME	SSN	BDATE	ADDRESS	DNUMBER
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire, TX	4
Narayan, Ramesh K.	666884444	1962-09-15	975 Fire Oak, Humble, TX	5
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1

DEPARTMENT

DNAME	DNUMBER	DMGRSSN
Research	5	333445555
Administration	4	987654321
Headquarters	1	888665555

Hình 6.16. Quan hệ EMPLOYEE được phân rã (tách) thành 2 quan hệ

21

21

Thuật toán kiểm tra không mất mát thông tin (Lossless-join decomposition)

Ví dụ:

Tập phụ thuộc hàm F:

ENAME	SSN	BDATE	ADDRESS	DNUMBER	DNAME	DMGRSSN
↑	↑	↑	↑	↑	↑	↑

Kiểm tra phép tách trên là không mất mát thông tin:

Bước 1:

	EName	SSN	BDate	Address	DNumber	DName	DMgrSsn
EMPLOYEE	a1	a2	a3	a4	a5	b16	b17
DEPARTMENT	b21	b22	b23	b24	a5	a6	a7

Trần Thị Kim Chi

22

22

Thuật toán kiểm tra không mất mát thông tin (Lossless-join decomposition)

Ví dụ:

- Bước 2: Xét phụ thuộc hàm Dnumber \rightarrow DName, DMgrSsn. Ta nhận thấy có giá trị a5 ở dòng thứ 2, nên ta sẽ làm bằng giá trị a6, a7 cho dòng thứ 1.
- Bước 3: Tồn tại một dòng chứa giá trị a1, a2,...a7. Kết luận, phép phân rã trên không mất mát thông tin.

	EName	SSN	BDate	Address	DNumber	DName	DMgrSsn
EMPLOYEE	a1	a2	a3	a4	a5	a6	a7
DEPARTMENT	b21	b22	b23	b24	a5	a6	a7

Ghi chú: Sinh viên thực hiện phép nối tự nhiên 2 quan hệ EMPLOYEE và DEPARTMENT trên để kiểm tra có bằng quan hệ ban đầu EMP DEPT