TẠP CHÍ KHOA HỌC TÍNH TOÁN VÀ ĐIỀU KHIỂN TẬP VI (1990), Số 1 (12 - 19)

ÁNH XẠ CƠ SỞ DỮ LIỆU, CÁC BÀI TOÁN TÁCH VÀ CẬP NHẬT ĐỘC LẬP

HỒ VĂN TÍCH

I - MỞ ĐẦU

Trong bài này chúng ta áp dụng lý thuyết ánh xạ CSDL (data base mapping) vào việc nghiên cứu một vấn đề trung tâm của CSDL là tách CSDL (data base decomposition).

Theo quan điềm của Francois Bancilhon và Nicolas Spyratos ([1] , [2]) việc giới hạn phép tách CSDL ban đầu bằng phép kết nối (Join) là chưa đủ. Chẳng hạn chỉ dùng mỗi phép chiếu để tách một CSDL thành các CSDL địa phương (local data base) của một CSDL phân tán (distributed data base) là không thỏa đáng. Bởi vì trên thực tế đề tạo ra các CSDL địa phương người ta phải dùng cả phép chọn (selection). Do đó hai tác giả này đã xem phép tách một cách tổng quát như là một ánh xạ CSDL. Với quan niệm tổng quát như vậy các tác giả này đã đưa ra khái niệm tách không mất mát thông tin cho trường hợp tách đôi, đã chứng minh điều kiện cần và đủ đề cho phép cập nhật độc lập trên hai CSDL địa phương của một CSDL phân tán.

Ở đây chúng ta nghiên cứu mở rộng hướng trên cho vấn đề n-tách và CSDL phân tán gồm n CSDL địa phương. Trước khi đi vào trình bày các kết quả chính trong phần II ta nhắc lại và mở rộng các kết quả đã được trình bày trong [1] và [2]

II - ÁNH XẠ CSDL

1. Định nghĩa ánh xạ CSDL

Cho $U = \{A_1,..., A_n\}$ là tập thuộc tính.

$R(A_1,..., A_n)$ là lược đồ quan hệ (relation schema).

R là biến quan hệ (relational variable) ứng với lược đồ quan hệ $R(A_1,..., A_n)$, nghĩa là R là một biến lấy giá trị là những quan hệ trên U.

R_t là giá trị của biến quan hệ R tại thời điểm t.

D là biến CSDL (data base variable). Một biến CSDL là một tập các biến quan hệ.

D_t là giá trị của biến CSDL D tại thời điểm t.

Một lược đồ CSDL (data base schema) gồm:

1) Một biến CSDL D.

2) Một tân từ trên các giá trị của D biểu thị các ràng buộc toàn vẹn. Ta ký hiệu một lược đồ CSDL là: $<D|C>$.

Không gian trạng thái CSDL (data base state space) ứng với lược đồ $<D|C>$ ký hiệu S $<D|C>$ (có khi viết tắt là S) được xác định như sau:

$$S\langle D\,|C\rangle = \{D_1 | C(D_1) = \text{true}\}$$

Ta có $S\langle D|C\rangle \subset S\langle D\,|\lambda\rangle$ trong đó λ là tân từ trống tức là $\lambda(D_1) = \text{true}\ \forall D_1$.

Định nghĩa 1. Cho lược đồ CSDL $\langle D|C\rangle$ và một biến CSDL D' ánh xạ $f : S\langle D|C\rangle \to S\langle D'|\lambda\rangle$ gọi là ánh xạ CSDL

Ví dụ 1. Các lược đồ quan hệ:

$R(S, C, G, P)$; $R_1(S, C)$; $R_2(C, P)$; $R_3(S, C, P)$.

Các ràng buộc toàn vẹn:

$C_1: S, C \to G$ trong R; $C_2: C \to P$ trong R; $C = C_1 \wedge C_2$.

Các biến CSDL:

$D = \{R\}$, $D_1 = \{R_1, R_2\}$, $D_2 = \{R_3\}$.

Các ánh xạ CSDL:

$f_1 = R[S, C, P]$; $f_2 = (R[S, C], R[C, P])$

trong đó $f_1: S\langle D|C\rangle \to S\langle D_1|\lambda\rangle$; $f_2: S\langle D|C\rangle \to S\langle D_2|\lambda\rangle$.

2. Cấu trúc của tập các ánh xạ CSDL trên một không gian trạng thái

Chúng ta ký hiệu MAP là tập các ánh xạ CSDL trên không gian trạng thái S. Từ định nghĩa của ánh xạ CSDL ta thấy các vấn đề: hỏi (quering), cập nhật (updating), xác định khung nhìn (view definition)... có thể xem là các ánh xạ CSDL. Do vậy có thể xem ánh xạ CSDL như là một "kênh liên lạc" (communication channel) giữa CSDL và người dùng. Chính vì lẽ đó ta tìm cách "đo" lượng thông tin mỗi ánh xạ CSDL có thể mang. Chúng ta nghiên cứu cấu trúc của tập MAP theo khía cạnh này.

Định nghĩa 2. Cho $f, g \in$ MAP. Ta nói f trội hơn g và ký hiệu là $f \geq g$ nếu và chỉ nếu:

$$\forall D_{11}, D_{12} \in S : f(D_{11}) = f(D_{12}) \Rightarrow g(D_{11}) = g(D_{12}).$$

Từ định nghĩa ta thấy nếu $f \geq g$ thì khả năng phân biệt các giá trị CSDL của f "mạnh" hơn g. Trở lại ví dụ 1 ta thấy $f_1 \geq f_2$.

Định lý 1. Cho $f, g \in$ MAP, $f \geq g$, suy ra tồn tại ánh xạ h sao cho $g = h \circ f$.

Chứng minh (xem [1]).

Với mỗi $f \in$ MAP ta xác định một quan hệ trên S, $\equiv_f$ như sau: $\forall D_{11}, D_{12} \in S$, $D_{11} \equiv_f D_{12} \Leftrightarrow f(D_{11}) = f(D_{12})$. Ta thấy $\equiv_f$ là một quan hệ tương đương trên S, ký hiệu $S/f = S/\equiv_f$ ($S/\equiv_f$ là lớp thương của S theo quan hệ tương đương $\equiv_f$). Từ định nghĩa 1 dễ dàng suy ra:

Định lý 2. $f \geq g \Leftrightarrow\ \equiv_f \subset\ \equiv_g \Leftrightarrow S/f \geq S/g$.

Trong đó $S/f \geq S/g$ hiểu theo nghĩa phân hoạch S/f mịn hơn phân hoạch S/g (tức là mỗi phần tử của S/f được chứa trong một phần tử nào đó của S/g).

Quan hệ "trội hơn" $\geq$ trong MAP không phải là quan hệ thứ tự bộ phận. Nó chỉ phản xạ, bắc cầu nhưng không phản xứng. Do đó ta cần xét một quan hệ khác, quan hệ tương đương trên MAP.

Định nghĩa 3. Cho $f, g \in$ MAP, ta gọi $f \equiv g$ nếu và chỉ nếu $f \geq g$ và $g \geq f$.

Xem lại ví dụ 1 ta thấy $f_1 \geq f_2$ và $f_2 \geq f_1$ do đó $f_1 \equiv f_2$. Dễ dàng thấy $\equiv$ là quan hệ tương đương trên MAP. Do đó ta có phân hoạch của tập MAP

$$MAP/\equiv \; = \{|f| \text{ với } f \in MAP\}$$

trong đó $|f| = \{g \mid g \in MAP, f \equiv g\}$.

Từ định nghĩa của quan hệ tương đương $\equiv$ và định lý 2 ta có:

Định lý 3. $f \equiv g \Leftrightarrow |f| = |g| \Leftrightarrow \equiv_f \; = \; \equiv_g \Leftrightarrow S/f = S/g$.

Bây giờ ta đưa vào $MAP/\equiv$ một quan hệ thứ tự sau:

Định nghĩa 4. Với $|f|, |g| \in MAP/\equiv$, ta nói $|f| \geq |g|$ nếu và chỉ nếu $f \geq g$.

Định lý sau nói lên cấu trúc của $MAP/\equiv$

Định lý 4. $MAP/\equiv$ là một dàn với thứ tự $\geq$.

Chứng minh. a) Rõ ràng quan hệ $\geq$ là quan hệ thứ tự trong $MAP/\equiv$.

b) Với $|f|, |g| \in MAP/\equiv$ luôn luôn tồn tại inf $(|f|, |g|)$ và sup $(|f|, |g|)$ (xem [1]).

Chú ý rằng nếu $|h| = \inf(|f|, |g|)$ thì $\equiv_h \; = (\equiv_f \cup \equiv_g)^*$, trong đó $(\equiv_f \cup \equiv_g)^*$ là bao đóng bắc cầu của $\equiv_f \cup \equiv_g$. Nếu $|h| = \sup(|f|, |g|)$ thì $\equiv_h \; = \; \equiv_f \cap \equiv_g$. Ta ký hiệu

$$|f| \wedge |g| = \inf(|f|, |g|)$$

$$|f| \vee |g| = \sup(|f|, |g|)$$

Từ định lý này suy ra với $f, g \in MAP$ tồn tại inf (f, g) theo sự tương đương, ký hiệu là $f \wedge g$ và tồn tại sup (f, g) theo sự tương đương ký hiệu là $f \vee g$.

Vì $(MAP/\equiv, \geq)$ là dàn, do đó với $|f_1|, |f_2|, \ldots, |f_n| \in MAP/\equiv$ tồn tại inf $(|f_1|, \ldots, |f_n|)$ ký hiệu là $|f_1| \wedge |f_2| \ldots \wedge |f_n|$ tồn tại $\sup(|f_1|, \ldots, |f_n|)$ ký hiệu là $|f_1| \vee |f_2| \ldots \vee |f_n|$ Do đó với $f_1, f_2, \ldots f_n \in MAP$ tồn tại $\inf(f_1, f_2, \ldots, f_n)$ và $\sup(f_1, f_2, \ldots, f_n)$ theo sự tương đương ký hiệu tương ứng là $f_1 \wedge f_2 \ldots \wedge f_n$ và $f_1 \vee f_2 \ldots \vee f_n$. Từ định lý này ta có hệ quả sau:

Hệ quả 1. Cho $f_1, \ldots, f_n \in MAP$

$$S/(\equiv_{f_1} \cup \equiv_{f_2} \cup \ldots \equiv_{f_n})^* = \inf(S/f_1, S/f_2, \ldots, S/f_n)$$

$$S/(\equiv_{f_1} \cap \equiv_{f_2} \cap \ldots \equiv_{f_n}) = \sup(S/f_1, S/f_2, \ldots, S/f_n)$$

Bây giờ cho $f_1, f_2, \ldots, f_n \in MAP$ ta tìm cách biểu diễn các ánh xạ CSDL $f_1 \vee f_2 \ldots \vee f_n$ và $f_1 \wedge f_2 \ldots \wedge f_n$.

Định nghĩa 5. Cho các ánh xạ CSDL $f_i : S \to S\langle D_i | \lambda \rangle$ $i = 1, \ldots, n$ và $D_i \cap D_j = \varnothing$ với $i \neq j$. Ta gọi phép ghép (Juxtaposition) của các ánh xạ $f_1, f_2, \ldots f_n$, ký hiệu $f_1 \times f_2 \times \ldots f_n$, là ánh xạ

$$f_1 \times f_2 \times \ldots \times f_n : S \to S\langle D_1 \cup \ldots \cup D_n | \lambda \rangle$$

$$D_i \to (f_1(D_i), f_2(D_i), \ldots, f_n(D_i))$$

Định lý 5. Với $f_1, \ldots, f_n \in MAP$ ta có

$$f_1 \times f_2 \times \ldots f_n = f_1 \vee f_2 \vee \ldots f_n$$

Chứng minh. Với $D_{11}, D_{12} \in S\langle D|C \rangle$

$$D_{11} \equiv_{f_1 \times f_2 \times \ldots f_n} D_{12} \Leftrightarrow (f_1(D_{11}), \ldots, f_n(D_{11})) = (f_1(D_{12}), \ldots, f_n(D_{12}))$$

$$\Leftrightarrow f_i(D_{11}) = f_i(D_{12}) \;\; \forall i = 1, \ldots, n$$

$$\Leftrightarrow \equiv_{f_1 \times f_2 \times \ldots \times f_n} \; = \; \equiv_{f_1} \cap \ldots \cap \equiv_{f_n}$$

$$\Leftrightarrow f_1 \times f_2 \times \ldots \times f_n = f_1 \vee f_2$$

$\vee \ldots \vee f_n$

Đối với $f_1 \wedge f_2 \wedge \ldots f_n$ không có biểu diễn tổng quát. Xét ví dụ sau.

Ví dụ 2. Cho các lược đồ CSDL :

$R(A, B, C, D, E)$; $R_1(A, B)$; $R_2(A, C)$; $R_3(A, D)$; $R_4(A, E)$; $R_5(A)$

Các biến CSDL : $D = \{R\}$, $D_i = \{R_i\}$, $i = 1,\ldots,4$.

Các ràng buộc toàn vẹn : $C_1: A \rightarrow B$ trong R

$C_2: A \rightarrow C, D$ trong R

$C_3: E \rightarrow A$ trong R

$C = C_1 \wedge C_2 \wedge C_3$

Các ánh xạ CSDL :

$f_1 = R[A, B]$; $f_2 = R[A, C]$; $f_3 = R[A, D]$; $f_4 = R[A, E]$; $f_5 = R[A]$.

Ta có $f_5 = f_1 \wedge f_2 \wedge f_3 \wedge f_4$. Để kiểm chứng điều này ta có thể kiểm chứng $(\equiv_{f_1} \cup \equiv_{f_2} \cup \equiv_{f_3} \cup \equiv_{f_4})^* = \equiv_{f_5}$

Tập MAP có các phần tử lớn nhất và bé nhất theo sự tương đương.

Định nghĩa 6. Cho $D_{t0} \in S \langle D|C\rangle$ ta xác định hai ánh xạ:

$L_S : S\langle D|C\rangle \rightarrow \langle D|\lambda\rangle: L_S(D_t) = D_t \ \forall t \in S$

$0_S : S\langle D|C\rangle \rightarrow \langle D|\lambda\rangle: 0_S(D_t) = D_{t0} \ \forall t \in S$.

Từ định nghĩa này ta suy trực tiếp ra:

Định lý 2.6. $\forall f \in$ MAP,

(i) $0_S \leq f \leq 1_S$

(ii) $'\equiv_S \subset \equiv_f \subset S \times S$.

3. Hệ các ánh xạ CSDL bù nhau

Bây giờ ta xác định một hệ ánh xạ CSDL mà thông tin của hệ này mang có thể dựng lại được CSDL ban đầu.

Định nghĩa 7. Cho hệ $\{f_i\}$ $i = 1,\ldots,n$, $f_i \in$ MAP. Ta nói hệ $\{f_i\}$ $i = 1,\ldots,n$ là hệ bù nhau (Complement) nếu và chỉ nếu

$$f_1 \times f_2 \times \ldots \times f_n = L_S \ (n \geq 1)$$

Từ định nghĩa này suy trực tiếp ra kết quả sau:

Định lý 7. Cho hệ $\{f_i\}$ $i = 1,\ldots,n$, $f_i \in$ MAP. Các mệnh đề sau tương đương với nhau:

(i) Hệ $\{f_i\}$ $i = 1,\ldots,n$ là một hệ bù nhau.

(ii) $\equiv_{f_1 \times f_2 \times \ldots f_n} = '\equiv_S$.

(iii) $f_1 \times f_2 \times \ldots f_n$ là đơn ánh.

(iv) Lấy các phần tử $f_{i1}, f_{i2},\ldots f_{ik}$ $(k < n)$, thì ghép của $f_{i1}, f_{i2},\ldots f_{ik}$ phân biệt được các cặp giá trị của biến CSDL mà không phân biệt được bởi ghép của các ánh xạ CSDL còn lại trong hệ $\{f_i\}$, $i = 1,\ldots,n$.

4. Tính độc lập của hệ ánh xạ CSDL

Định nghĩa 8. Cho hệ $\{f_i\}$, $i = 1,..,n$, $f_i \in$ MAP

(i) Ta gọi hệ $\{f_i\}$, $i = 1,..,n$ là hệ độc lập, ký hiệu $\sim [f_1,..,f_n]$ nếu và chỉ nếu

$$(f_1 \times f_2 \times \ldots \times f_n)(S) = f_1(S) \times \ldots \times f_n(S)$$

(ii) Với $h \in$ MAP. Ta nói hệ $\{f_i\}$, $i = 1,..,n$ độc lập với điều kiện h, ký hiệu $\sim [f_1,...,f_n]$ mod h nếu và chỉ nếu

$$(f_1 \times f_2 \times \ldots \times f_n)(B) = f_1(B) \times \ldots \times f_n(B) \quad \forall B \in S/h.$$

(iii) Ta nói hệ $\{f_i\}$, $i = 1,...,n$ độc lập yếu (weak independence) nếu $\sim [f_1,..,f_n]$ mod $(f_1 \wedge f_2 \wedge ... f_n)$.

III - TÁCH CSDL

Trước tiên để cho gọn cách viết ta nêu các ký hiệu qui ước như sau : Cho $<D|C>$ là lược đồ CSDL, f_i: $S<D|C> \rightarrow S<D_i|\lambda>$. Qui ước

1) S thay cho $S<D|C>$.

2) S_i thay cho $f_i(S)$.

3) C_i là tân từ trên $S<D_i|\lambda>$ sao cho $C_i(D_{i1}) = true \Leftrightarrow D_{i1} \in S_i$.

4) $D_{12...n}$ thay cho $D_1 \cup D_2 \cup \ldots \cup D_n$.

5) $S_{12...n}$ thay cho $(f_1 \times f_2 \times \ldots \times f_n)(S)$.

6) $C_{12...n}$ là tân từ trên $S<D_{12...n}|\lambda>$ sao cho $C_{12...n}(D_{12...n1}) = true \Leftrightarrow D_{12...n1} \in S_{12...n}$.

Định nghĩa 9. Cho lược đồ CSDL $<D|C>$. Một phép n-tách (n-decomposition) là một bộ n ánh xạ CSDL $(f_1, f_2,...f_n)$ với $f_i \in$ MAP $\forall i = 1, 2,..,n.(n \geq 1)$. Kết quả của phép n-tách là $<D_{12...n}|C_{12...n}>$

Ví dụ 3. Các lược đồ quan hệ

$R(A, B, C, D, E)$; $R_1(A, B)$; $R_2(B, C)$; $R_3(B, D)$; $R_4(B, E)$.

Các ràng buộc toàn vẹn:

$C_1: A \rightarrow B$, $C_2: B \rightarrow C$, $C_3: B \rightarrow D$, $C_4: C \rightarrow E$ trong R;

$C_5: R_1[B] = R_2[B] = R_3[B] = R_4[B]$; $C = C_1 \wedge C_2 \wedge C_3 \wedge C_4$; $C_{1234} = C \wedge C_5$.

Các biến CSDL:

$D = \{R\}$, $D_1 = \{R_1\}$, $D_2 = \{R_2\}$, $D_3 = \{R_3\}$, $D_4 = \{R_4\}$.

Các ánh xạ CSDL:

$f_1 = R[A, B]$, $f_2 = R[B, C]$, $f_3 = R[B, D]$, $f_4 = R[B, E]$.

Bộ (f_1, f_2, f_3, f_4) là một phép 4-tách của $<D|C>$ với kết quả là $<D_{1234}|C_{1234}>$

Bây giờ chúng ta xét vấn đề tách không tổn thất.

Định nghĩa 10. Phép n-tách $(f_1, f_2,...f_n)$ gọi là không tổn thất nếu và chỉ nếu hệ $\{f_i\}$, $i = 1, 2,..,n$ là hệ bù nhau.

Từ định nghĩa suy ra nếu $(f_1, f_2,...f_n)$ là n-tách không tổn thất thì có thể khôi phục lại được CSDL ban đầu qua CSDL tách. Bởi vì lúc đó hệ $\{f_i\}$, $i = 1, 2,...,n$ bù nhau, nên theo định lý 2.7 ánh xạ $f_1 \times f_2 \times ... \times f_n$ là một song ánh từ $<D|C>$ lên $<D_{12...n}|C_{12...n}>$. Ta thấy các kết quả về n-tách cho trường hợp chiếu và kết nối của J. Biskup, U. Dayal, P. A. Bernstein [3] cũng hoàn toàn phù hợp với khái niệm chúng ta đã nêu. Chẳng hạn trong [3] có kết quả sau:

Định lý [3]. Cho $D = \{<X_1, >, <X_2, >,...,<X_n, >\}$ là một lược đồ CSDL và $<U, F>$ là một lược đồ quan hệ sao cho $\cup X_i = U$, $(i = 1,..,n)$.

Giả sử F có phủ $H (F^+ = H)$ sao cho với mỗi $R \to S \in H$ đều tồn tại $<X_i, > \in D$ với $RUS \subset X_i$. Nếu có một thành phần $<X_{i0}, > \in D$ sao cho $X_{i0} \to U \in F^+$ thì D là tách không tổn thất đối với $<U, F>$.

Bây giờ chúng ta nghiên cứu vấn đề: Khi nào có thể cập nhật một cách độc lập lên các thành phần của CSDL tách, mà không phải kiểm tra ảnh hưởng của nó lên các thành phần khác. Điều này rất có ý nghĩa đối với một CSDL phân tán gồm n thành phần, phân bố ở n vị trí khác nhau về mặt địa lý.

Cho $(D_{1t}, D_{2t},..,D_{nt}) \in S_{12...n}$. Một cập nhật U lên D_{it}, $i = 1,2,...,n$ gọi là chấp nhận được đối với $(D_{1t}, D_{2t},..., D_{i-1,t}, D_{i+1t},...,D_{nt})$ nếu và chỉ nếu $(D_{1t}, D_{2t},...,UD_{it},..,D_{nt}) \in S_{12...n}$.

Ta đặt

$ADM(D_{it} | D_{1t},...,D_{i-1t}, D_{i+1t},..,D_{nt}) = \{U \mid U$ chấp nhận được đối với $(D_{1t},...,D_{i-1t}, D_{i+1t},.., D_{nt})\}$.

Do đó ta đã thiết lập được một ánh xạ ứng mỗi bộ $(D_{1t},...,D_{nt}) \in S_{12...n}$ với tập ADM $(D_{it}|D_{1t},.., D_{i-1t}, D_{i+1t},..,D_{nt})$. Chúng ta hãy quan sát ý nghĩa của ánh xạ này trong trường hợp CSDL phân tán, trong đó $D_1, D_2,..,D_n$ ở các vị trí khác nhau, và người dùng muốn cập nhật lên D_i tại thời điểm t chỉ biết D_{it} và không được biết $D_{1t},..D_{i-1t}, D_{i+1t},..D_{nt}$. Có hai điều xảy ra:

1/ Ánh xạ ADM $(D_i | D_1,..,D_{i-1},D_{i+1},..,D_n)$ phụ thuộc vào $D_1,..D_{i-1}, D_{i+1}, D_n$. Khi đó người sử dụng, do không biết $D_1,..D_{i-1}, D_{i+1},..D_n$, nên không biết những cập nhật gì có thể tác dụng lên D_{it} mà không ảnh hưởng đến $D_{1t},...,D_{i-1t},...,D_{i+1t},...,D_{nt}$.

2/ Ánh xạ ADM $(D_i|D_1,...,D_{i-1}, D_{i+1},...,D_n)$ không phụ thuộc vào $D_1,..,D_{i-1}, D_{i+1},..,D_n$, nghĩa là ánh xạ này chỉ có một biến D_i. Khi đó người sử dụng nhìn thấy D_{it} và biết được tập hợp tất cả các cập nhật có thể tác dụng lên nó, không phụ thuộc vào các thành phần còn lại.

Định nghĩa 11. Phép n-tách $(f_1,...,f_n)$ gọi là độc lập yếu nếu và chỉ nếu các ánh xạ ADM $(D_i|D_1,..,D_n)$ độc lập đối với $D_1,..,D_{i-1}, D_{i+1},...,D_n$ với $\forall i=1,2,...,n$.

Vấn đề đặt ra:

1/ Xác định khi nào phép n-tách$(f_1,..,f_n)$ là độc lập yếu.

2/ Trong trường hợp đó, hãy xác định các tập ADM $(D_{it}|D_{1t},..,D_{nt})$ với $i=1, 2,...,n$. (Tức là xác định các cập nhật có thể tác dụng lên D_{it}).

Định lý sau đây trả lời câu hỏi thứ nhất.

Định lý 8. Cho phép n-tách $(f_1,..,f_n)$ của một CSDL. Các mệnh đề sau tương đương.

1/ Phép n-tách $(f_1,..f_n)$ độc lập yếu.

2/ Hệ các ánh xạ CSDL $\{f_i\}$, $i = 1,2,..,n$ là hệ độc lập yếu.

Để chứng minh định lý này trước tiên ta chứng minh các bổ đề sau:

Cho hệ các ánh xạ CSDL $\{f_i\}$, $i = 1, 2,..,n$. Theo định lý 1 từ

$f_1 \wedge f_2 \wedge ... \wedge f_n \leq f_i$ với mọi $i = 1, 2,..., n$, suy ra tồn tại các ánh xạ f_i, $i = 1, 2,...,n$ sao cho:

$$f_1 \wedge f_2 \wedge \dots \wedge f_n = f_1' f_1 = f_2' f_2 = \dots = f_n' f_n$$

Bổ đề 1. Cho hệ $\{f_i\}$, $i = 1, 2, \dots, n$ là hệ độc lập yếu. Ta có $\forall (D_{1t}, D_{2t}, \dots, D_{nt})$ mà

$f_1'(D_{1t}) = f_2'(D_{2t}) = \dots = f_n'(D_{nt})$ nếu và chỉ nếu $(D_{1t}, D_{2t}, \dots, D_{nt}) \in S_{12\dots n}$.

Chứng minh. " $\Rightarrow$ " Do $D_{it} \in f_i(S) \Rightarrow \exists D_t^{(i)} \in S$ sao cho $f_i(D_t^{(i)}) = D_{it}$

$$\Rightarrow (f_1 \wedge f_2 \wedge \dots \wedge f_n)(D_t^{(i)}) = f_i' f_i(D_t^{(i)}) = f_i'(D_{it}) = f_1'(D_{1t}) \quad \forall i = 1, 2, \dots, n$$

$\Rightarrow \exists T \in S/_{f_1 \wedge f_2 \wedge \dots \wedge f_n}$ sao cho $D_{it}^{(i)} \in T \ \forall i = 1, \dots, n \Rightarrow (D_{1t}, \dots, D_{nt}) \in f_1(T) \times \dots \times f_n(T)$

Do hệ $\{f_i\}$, $i = 1, \dots, n$ độc lập yếu nên $(D_{1t}, \dots, D_{nt}) \in f_1(T) \times \dots \times f_n(T) = (f_1 \times \dots \times f_n)(T) \subset S_{12\dots n}$.

" $\Leftarrow$ " Giả sử $(D_{1t}, \dots, D_{nt}) \in S_{12\dots n} \Rightarrow \exists D_t \in S$ sao cho $f_i(D_t) = D_{it}$ $(\forall i = 1, 2, \dots, n)$.

$$\Rightarrow f_i'(D_{it}) = f_i' f_i(D_t) = (f_1 \wedge \dots \wedge f_n)(D_t) \Rightarrow f_i'(D_{it}) = f_j'(D_{jt}) \ \forall i, j = 1, 2, \dots, n.$$

Giả sử phép n-tách $(f_1, f_2, \dots, f_n)$ độc lập yếu, tức là ánh xạ ADM $(D_i \mid D_1, \dots, D_n)$ với $i = 1, 2, \dots, n$ chỉ là hàm một biến D_i trên S_i. Do đó ta có thể phân hoạch S_i theo hàm ADM $(D_i \mid D_1, \dots, D_n)$.

$$S_i/_{ADM(D_i|D_1,\dots,D_n)} = \{|ADM(D_{it}|D_{1t}, \dots, D_{nt}), \ (D_{1t}, \dots, D_{nt}) \in S_{12\dots n}\}$$

trong đó

$$|ADM(D_{it}|D_{1t}, \dots, D_{nt})|$$

$$= \{D_{it}' \in S_i | ADM(D_{it}|D_{1t}, \dots, D_{nt}) = ADM(D_{it}'|D_{1t}, \dots, D_{nt})\}$$

Bổ đề 2. Cho phép n-tách $(f_1, \dots, f_n)$ độc lập yếu, khi đó ta có các phép tương ứng 1-1 sau:

$$|ADM(D_{1t})|D_{2t}, \dots, D_{nt})| \leftrightarrow |ADM(D_{2t}|D_{1t}, \dots, D_{nt})| \leftrightarrow \dots |ADM(D_{nt}|D_{1t}, \dots, D_{n-1t})|$$

nếu và chỉ nếu $(D_{1t}, D_{2t}, \dots, D_{nt}) \in S_{12\dots n}$.

Chứng minh. Ta chỉ cần chứng minh $|ADM(D_{1t}|D_{2t}, \dots, D_{nt})| \leftrightarrow |ADM(D_{2t}|D_{1t}, \dots, D_{nt})$ là 1-1. Còn lại các tương ứng khác làm tương tự.

Giả sử $|ADM(D_{2t}|D_{1t}, \dots, D_{nt})| = |ADM(D_{2t}'|D_{1t}, \dots D_{nt})|$, $(D_{1t}, D_{2t}, \dots D_{nt})$

và $(D_{1t}, D_{2t}', \dots, D_{nt}) \in S_{12\dots n}$ mà $|ADM(D_{1t}|D_{2t}, \dots, D_{nt})| \neq |ADM(D_{1t}, D_{2t}', \dots, D_{nt}|$

$$\Rightarrow ADM(D_{1t}|D_{2t}, \dots, D_{nt}) \neq ADM(D_{1t}|D_{2t}', \dots, D_{nt})$$

Vậy ánh xạ ADM $(D_1 \mid D_2, \dots, D_n)$ phụ thuộc D_2, điều này mâu thuẫn với giả thiết.

Bổ đề 3. Cho $f_1, \dots, f_n, k \in$ MAP, nếu $\sim [f_1, \dots, f_n] \bmod k$ thì $k \geq h = f_1 \wedge f_2 \wedge \dots \wedge f_n$.

Chứng minh. Do $\sim [f_1, \dots, f_n] \bmod k$ suy ra

$$(f_1 \times f_2 \times \dots f_n)(T) = f_1(T) \times f_2(T) \times \dots f_n(T) \quad \forall t \in S/k.$$

Do đó $\forall D_{it} \in f_i(T)$ với $i = 1, 2, \dots, n$ ta có $\displaystyle\bigcap_{i=1}^{n} f_i^{-1}(D_{it}) \neq \emptyset$

Có thể viết cách khác là $\forall A_i \in T/f_i|_T$ $i = 1, 2, \dots, n$, trong đó $f_i|_T$ là hạn chế của f_i trên T, ta có

$$\bigcap_{i=1}^{n} A_i \neq \emptyset \qquad (*)$$

Mặt khác, từ $h = f_1 \wedge \dots \wedge f_n$ suy ra $f_i|_T \geq h|_T$ (trong đó $h|_T$ là hạn chế của h trên T). Theo định lý

2, $\forall C_i \in T/h|_T$, $\exists A_i \in T/f_i|_T$ sao cho $A_i \subset C_i$ với $i = 1, 2,..,n$. Do đó từ (*) suy ra $\cap C_i \neq \varnothing$ $(i = 1,..., n)$. Vậy $C_1 = C_2 = ... = C_n = T$. Điều này suy ra tồn tại $C \in S/h$ sao cho $T \subset C$, vậy chứng tỏ rằng $k \geq h$.

Bây giờ trở lại chứng minh định lý 8.

Chứng minh: 2) $\Rightarrow$ 1) Do hệ $\{f_i\}$, $i = 1, 2,..., n$ độc lập yếu, nên theo bổ đề 1, với mỗi $(D_{1i}, D_{2i},..,D_{ni}) \in S_{12...n}$ suy ra $f_1'(D_{1i}) = f_2'(D_{2i}) = ... = f_n'(D_{ni})$.

Ta có $ADM(D_{ii}|D_{1i},..., D_{ni}) = \{U|(D_{1i},..., U(D_{ii}),.., D_{ni}) \in S_{12...n}\}$

$= \{U|f_1'(D_{1i}) = ... = f_i'(U(D_{ii})) = ... = f_n'(D_{ni})\} = \{U|f_i'(D_{ii}) = f_i'(U(D_{ii}))\}$.

Điều này chứng tỏ rằng ánh xạ $ADM(D_i|D_1,..., D_n)$ độc lập đối với $(D_1,.., D_n)$. Vậy phép n-tách $(f_1, f_2,... f_n)$ độc lập yếu.

1) $\Rightarrow$ 2) Do phép n-tách$(f_1, f_2,..., f_n)$ độc lập yếu, do đó theo bổ đề 2 ta có các phép tương ứng 1-1 sau:

$|ADM(D_{1i}|D_{2i},..., D_{ni})| \leftrightarrow |ADM(D_{2i}|D_{1i},..., D_{ni})| \leftrightarrow ... |ADM(D_{ni}|D_{1i},..., D_{n-1t})|$ nếu và chỉ nếu $(D_{1i},..., D_{ni}) \in S_{12...n}$. Do đó ta xác định được các ánh xạ $f_1', f_2',.., f_n'$ sao cho

$f_1'(D_{1i}) = f_2'(D_{2i}) = ... = f_n'(D_{ni})$ nếu và chỉ nếu $|ADM(D_{1i}|D_{2i},..., D_{ni})|$ $|ADM(D_{2i}|D_{1i},..., D_{ni})|$ ... , $|ADM(D_{ni}|D_{1i},..., D_{n-1t})|$ tương ứng với nhau. Vậy ta có:

$$(D_{1i}, D_{2i},..., D_{ni}) \in S_{12...n} \Leftrightarrow f_1'(D_{1i}) = f_2'(D_{2i}) = ... = f_n'(D_{ni}).$$

Đặt $h = f_1' f_1 = ... = f_n' f_n$ ta có

$$f_1(B) \times f_2(B) \times ... \times f_n(B) = (f_1 \times f_2 \times ... \times f_n)(B), \forall B \in S/h$$

Điều này chứng tỏ $\quad \sim [f_1,..., f_n] \bmod h \quad$ (a)

Theo các xác định h ta có $h \leq f_i$ với $i = 1, 2,..., n$, suy ra

$$h \leq f_1 \wedge f_2 \wedge ... \wedge f_n \qquad (b)$$

Mặt khác từ (a) và bổ đề 3 suy ra

$$h \geq f_1 \wedge f_2 \wedge ... \wedge f_n \qquad (c)$$

Vậy từ (b) và (c) suy ra

$$h = f_1 \wedge f_2 \wedge ... \wedge f_n \qquad (e)$$

Từ (a) và (e) suy ra hệ $\{f_i\}$, $i = 1, 2,..., n$ độc lập yếu.

Bây giờ chúng ta giải quyết vấn đề thứ hai, tức là "tính toán" tập các cập nhật mà chúng ta có thể tác dụng lên một thành phần, không cần để ý đến các thành phần khác của CSDL tách. Định lý sau giải quyết vấn đề này.

Định lý 9. Cho hệ $\{f_i\}$, $i = 1, 2,..., n$ độc lập yếu, các ánh xạ $f_1', f_2',.., f_n'$ sao cho $f_1' f_1 = f_2' f_2 = ... = f_n' f_n = f_1 \wedge ... \wedge f_n$. Khi đó cập nhật U trên D_i là chấp nhận được nếu và chỉ nếu $f_i' U = f_i'$.

Chứng minh. Do $\sim [f_1,..., f_n] \bmod (f_1 \wedge ... \wedge f_n)$ nên U chấp nhận được $\Leftrightarrow (D_{1i},..., U(D_{ii}),..., D_{ni}) \in S_{12...n} \Leftrightarrow f_1'(D_{1i}) = ... = f_i'(U(D_{ii})) = ... = f_n'(D_{ni}) \Leftrightarrow f_i'(U(D_{ii}) = f_i'(D_{ii}) \Leftrightarrow f_i' U = f_i'$.

Từ định lý này suy ra các cập nhật chấp nhận được là các cập nhật bất biến trên $f_1 \wedge f_2 \wedge ... \wedge f_n$.

Trở lại ví dụ 2 ta thấy $\sim [f_1, f_2, f_3, f_4] \bmod f_5$ và $f_5 = f_1 \wedge f_2 \wedge f_3 \wedge f_4$, do đó cho phép cập nhật độc lập trên các thành phần D_1, D_2, D_3, D_4 và các cập nhật chấp nhận được là các cập nhật bất biến trên thuộc tính A.

(Xem tiếp trang 26)