



ACTIVE COLLABORATIVE e-LEARNING SYSTEM

Pedagogical Technique Department - Faculty of Information Technology - University of Education - Ho Chi Minh City, Vietnam

ACeLS ► CSDL

People

Participants

Administration

Grades

Unenrol me from CSDL

Profile

Courses

CSDL-02-2011/2012

LTCB-02-2011/2012

CSDL - Demo

LTCB - Demo

THDC - Demo

WIKI - Sharing Essay

e-COURSE

All courses ...

Các tiện ích hỗ trợ

- Video Demos - Các Videoclip demo cần thiết
- Template for Report - Mẫu báo cáo nộp [Link] - Hình thức báo cáo.pdf
- Review - Week01~03 - Củng cố và ôn tập [Link]
 - Link download - Review Week 01~03 - [pdf]
- Evaluating effectiveness of ACeLS 01 - Đánh giá mức độ hữu ích của ACeLS
 - Practical Exercise 01 - Bài tập thực hành 01 - Updated 03/2012 NEW !!!
 - Practical Exercise 02 - Bài tập thực hành 02 - Updated 03/2012 NEW !!!

Yêu cầu thực hiện ở Week trước

2 Các hoạt động thực hiện trong: Week 04 ~ Week 09

Bài giảng và bài tập-đồ án

- Activities in Week 04 - Hoạt động tuần 04
 - Lecture 04 - Bài giảng tuần 04 [pdf]
- Activities in Week 05 - Hoạt động tuần 05
 - Lecture 05 - Bài giảng tuần 05 [pdf]
 - Handout 04 - Bài tập đồ án 04 [pdf]

Tự nghiên cứu và thảo luận nhóm

- Forum for Week04~Week09
 - Review02 - Bài học củng cố kiến thức
 - Practice Quiz 02 - Câu hỏi suy luận bổ ích.
 - Database - Thu thập dữ liệu 02
- Useful Documentations - Tài liệu hữu dụng cần xem HOT NEW !!!

THẢO LUẬN CÁC CHỦ ĐỀ LIÊN QUAN

Yêu cầu thực hiện ở Week 04~ Week 09

3 Các hoạt động thực hiện trong: Week 10 ~ Week 13

CỦNG CỐ KIẾN THỨC



Chương 6 (mở rộng) Biểu diễn RBTV bằng phụ thuộc hàm (PTH)

Nội dung trọng tâm

Biểu diễn RBTV bằng phụ thuộc hàm

Khoá - Thuật toán xác định khoá của một lược đồ quan hệ

Nội dung tự nghiên cứu

Không có

Bài tập

Bài tập lý thuyết 4 – Không bắt buộc
(ho09.pdf)

Tham khảo

Slide bài giảng Chương 6 (Chuong 6b.pdf)

Bài giảng chương 6 mở rộng (PTH-Chuong 6.pdf)

*** Chương 3 (trang 107~125). Giáo trình Nhập môn CSDL - Nguyễn An Tế

ĐỌC TÀI LIỆU - TỰ NGHIÊN CỨU - THẢO LUẬN NHÓM



CƠ SỞ DỮ LIỆU

(Introduction to Database System)

Lê Đức Long

Email: longld@math.hcmup.edu.vn

Website: <http://www.2learner.edu.vn>

NỘI DUNG TRÌNH BÀY



Chương 9 Chuẩn hóa lược đồ CSDL

Nội dung trọng tâm

Phụ thuộc hàm

Bao đóng của tập phụ thuộc hàm F – Bao đóng của tập thuộc tính X

Khoá - Thuật toán xác định khoá của một lược đồ quan hệ

Nội dung tự nghiên cứu

KHÔNG CÓ

Bài tập

KHÔNG BẮT BUỘC

Tham khảo

Slide bài giảng Chương 9 - CHỦ ĐỀ 12 (Section12.pdf)

*** Chương 3 (trang 107~125). Giáo trình Nhập môn CSDL - Nguyễn An Tế

*** *Part 2 - Chapter 9. An Introduction to Database Systems – C.J.Date*

*** *Chapter 3. (3.5) A First Course in Database Systems – Jeffrey D. Ullman*

Thực hành

Thực hành cài đặt RBTv (Thuc hanh 8.pdf)

Sử dụng nguồn thông tin từ trong đề cương chi tiết

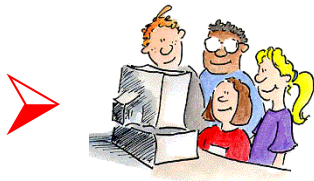
Một số quy ước trên slide



Tắt màn hình máy tính



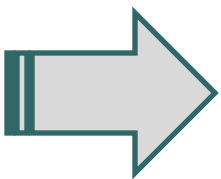
Được dùng máy tính



Làm việc theo nhóm



Ghi chép bằng văn bản



TỰ NGHIÊN CỨU- ĐỌC THÊM



CHUẨN HOÁ LỢC ĐỒ CSDL

Phần lí thuyết

Chương 08

Lê Đức Long

Email: longld@math.hcmup.edu.vn

Website: <http://www.2learner.edu.vn>

ĐẶT VẤN ĐỀ



Xuất phát từ giai đoạn phân tích nhu cầu trước đó, chúng ta có thể có 1 trong 2 kết quả sau:

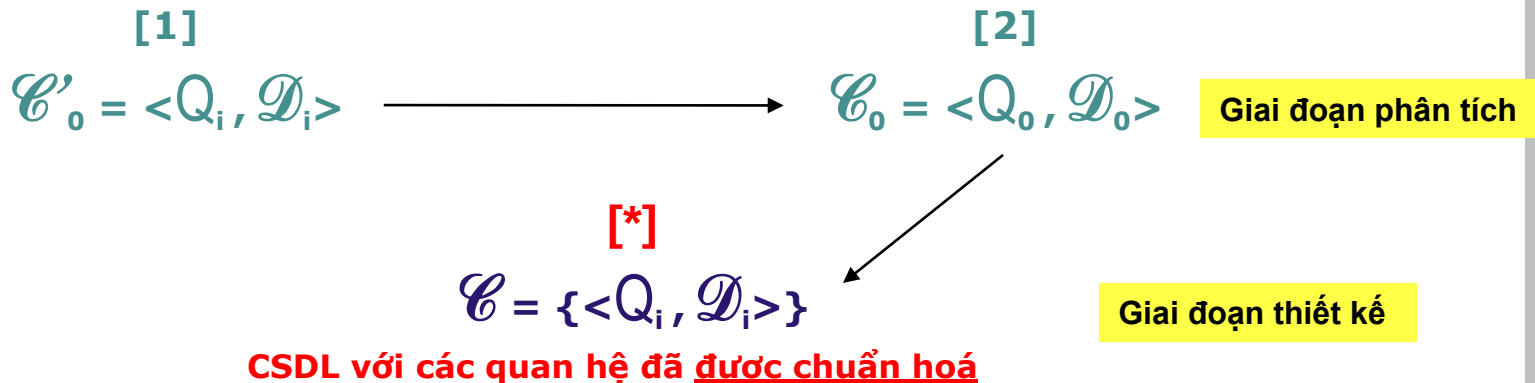
- ❖ [1] bằng cách sử dụng một mô hình dữ liệu nào đấy như mô hình Quan Hệ, hoặc mô hình Thực thể – Kết hợp, để mô hình hóa các thông tin nhận diện được, từ đó chúng ta sẽ có một sơ phác của cấu trúc CSDL gồm các quan hệ con $Q'_1, Q'_2, \dots, Q'_m$ cùng các phụ thuộc dữ liệu định nghĩa trên mỗi quan hệ con (gọi là FD hay PTH) là $\mathcal{D}'_1, \mathcal{D}'_2, \dots, \mathcal{D}'_m$:

$$\mathcal{C}'_0 = \{ \langle Q'_1, \mathcal{D}'_1 \rangle, \langle Q'_2, \mathcal{D}'_2 \rangle, \dots, \langle Q'_m, \mathcal{D}'_m \rangle \}$$

- ❖ [2] Một sơ phác tổng thể hơn gồm một quan hệ duy nhất Q_0 và một tập các phụ thuộc $\mathcal{D}_0$:

$$\mathcal{C}_0 = \langle Q_0, \mathcal{D}_0 \rangle ; \text{thường được gọi là quan hệ phổ quát} \rightarrow \text{CSDL với quan hệ không chuẩn}$$

Giai đoạn thiết kế ở mức quan niệm nhằm biến đổi kết quả đầu tiên này, dựa trên một số tiêu chuẩn thiết kế, để có được một cấu trúc quan niệm CSDL $\mathcal{C}$ được đánh giá “tốt” hơn, “phù hợp” hơn với các yêu cầu của môi trường ứng dụng.



Ví dụ minh họa



Tenlop	Siso	TenHS1	Diem11	Diem12	...	TenHS2	Diem21	Diem22	...
11A2	30	N.V.An	7	8		T.T.Binh	9	5	
12A1	28	L.T.Anh	5	6		H.T.Chi	10	8	
...									

(a) CSDL có quan hệ **chưa được chuẩn hoá**
(Unnormalized relation)

Q1: DS_LOP

Tenlop	Siso	MS_HS	TenHS
11A2	30	06005	N.V.An
11A2	30	06008	T.T.Binh
12A1	28	05001	L.T.Anh
...			

Q2: KQ_LOP

Tenlop	MS_HS	Mon	Diem
11A2	06005	Toan	7
11A2	06005	Van	8
11A2	06005	Ly	10
11A2	06008	Toan	9
11A2	06008	Van	5
12A1	05001	Toan	5
12A1	05001	Van	6
...			

(b) CSDL có các quan hệ con **đã được chuẩn hoá**

(Normalized relation)

Vấn đề cần giải quyết ...



Hầu hết các công trình nghiên cứu về thiết kế CSDL đều thỏa thuận rằng hai tiêu chuẩn quan trọng cần đạt được qua quá trình thiết kế một CSDL ở mức quan niệm là :

- ❖ **Cấu trúc [*] cần đạt một dạng chuẩn (normal form) cao nhất.**
- ❖ **Cấu trúc kết quả ở [*] phải tương đương với cấu trúc ban đầu ở [1] hoặc [2].**



Tiêu chuẩn dạng chuẩn được đề ra nhằm đáp ứng hai yêu cầu cụ thể:

- ❖ (1) giảm tối đa sự trùng lặp thông tin trong CSDL, do đó sẽ tránh được một số bất tiện khi cập nhật CSDL.
- ❖ (2) và tạo điều kiện để kiểm tra các ràng buộc toàn vẹn dưới dạng phụ thuộc dữ liệu được thuận lợi nhất, nghĩa là các phụ thuộc dữ liệu sẽ được kiểm tra dễ dàng nhất, đơn giản nhất và tương đối ít tốn kém nhất.



Tiêu chuẩn *tương đương* nhằm bảo đảm thông tin lấy ra được lưu trữ trong quan hệ phổ quát sẽ được tìm thấy đầy đủ trong CSDL của cấu trúc [*].

Một CSDL thỏa mãn cả hai tiêu chuẩn sẽ bảo đảm cho việc khai thác nó được thuận lợi trên cả ba phương diện :

- ❖ ***Truy vấn : nhờ vào tiêu chuẩn tương đương, các thông tin được truy xuất từ CSDL đúng là những thông tin đã được phân tích.***
- ❖ ***Cập nhật : tiêu chuẩn dạng chuẩn sẽ giảm bớt các tình huống thông tin mâu thuẫn sau những lần cập nhật CSDL.***
- ❖ ***Kiểm tra ràng buộc toàn vẹn: cả hai tiêu chuẩn đều nhằm vào việc kiểm tra ràng buộc toàn vẹn ở dạng phụ thuộc dữ liệu được thuận lợi.***

Dạng chuẩn (Normal Form)



❖ Dạng chuẩn của quan hệ: (E.F.Codd, 1972)

- DC 1 (1NF)
- DC 2 (2NF)
- DC 3 (3NF)
- DC BCK (Boyce-Kott-Kent)



❖ Dạng chuẩn 4: (R.Fagin, 1976)

- Là dạng chuẩn BCK và đồng thời giải quyết sự tồn tại một số bất tiện liên quan đến PTH
 - Dạng chuẩn tốt nhất, được mong đợi để được thiết kế
- ❖ **Tất cả những quan hệ được chuẩn hoá đều ở 1NF; một số quan hệ 1NF thì cũng ở 2NF; một số quan hệ 2NF thì cũng ở 3NF; và một số quan hệ 3NF thì cũng ở DC BCK và 4NF**

Dạng chuẩn 1



❖ Một lược đồ quan hệ Q được gọi là ở **dạng chuẩn 1 (1NF)** nếu **mọi thuộc tính của Q đều là thuộc tính đơn**.

- Q có cấu trúc phẳng, không có thuộc tính kép, thuộc tính lặp

❖ Ví dụ: Thuộc tính đơn $\neq$ thuộc tính kép

- Thuộc tính Ngaysinh là thuộc tính kép $\rightarrow$ tích hợp của ngày, tháng, năm
- Thuộc tính Diachi là thuộc tính kép $\rightarrow$ số nhà, đường, quận/huyện, tỉnh/thành
- Thuộc tính Diem11, Diem12, ..., Diem21, Diem22, .. $\rightarrow$ là thuộc tính lặp vì Diem11, Diem21 đều là điểm Toán, ...

Ví dụ 1



CSDL có các quan hệ con **đã được chuẩn hoá**

DS_LOP(Tenlop, Siso, MS_HS, TenHS)

KQ_LOP(Tenlop, MS_HS, Mon, Diem)

ĐƯỢC GỌI LÀ CÁC QUAN HỆ CÓ DC 1

T_{DS_LOP}

Tenlop	Siso	MS_HS	TenHS
11A2	30	06005	N.V.An
11A2	30	06008	T.T.Binh
12A1	28	05001	L.T.Anh
...			

T_{KQ_LOP}

Tenlop	MS_HS	Mon	Diem
11A2	06005	Toan	7
11A2	06005	Van	8
11A2	06005	Ly	10
11A2	06008	Toan	9
11A2	06008	Van	5
12A1	05001	Toan	5
12A1	05001	Van	6
...			

HẠN CHẾ CỦA DC 1

Mỗi quan tâm về giảm thiểu sự trùng lặp thông tin chưa được giải quyết. Sự trùng lặp thông tin dẫn đến một số vấn đề (bất tiện) trong quá trình khai thác (Thêm, Xoá, Sửa)

Ví dụ 2



Cho lược đồ **LICH_CT**(GV_CT, N, G, P, M, GV_PT)

và $\mathcal{F} = \{f_1: \text{GV_CT} \rightarrow \text{N, G, P}: \text{một GV coi thi vào 1 ngày, 1 giờ trong 1 phòng duy nhất};$
 $f_2: \text{M} \rightarrow \text{GV_PT}: \text{mỗi môn thi chỉ có một GV phụ trách ra đề thi}\}$

XÁC ĐỊNH KHOÁ CỦA QUAN HỆ TRÊN ???

Ta có: GV_CT, M $\rightarrow$ Q⁺. Vậy GV_CT, M là khoá

QUAN HỆ LICH_CT Ở DẠNG CHUẨN 1 (1NF)

T _{LICH_CT}	GV_CT	N	G	P	M	GV_PT
	N.V.An	2	8.00-10.00	101	KTLT	Viet
	T.T.Binh	2	10.00-12.00	101	CSDL	Long
	L.V.Chien	2	10.00-12.00	101	CSDL	Long
	V.T.Hoa	2	8.00-10.00	103	KTLT	Viet

CSDL này có sự trùng lặp thông tin trên các tập thuộc tính nào?

N, G, P; M, GV_PT

Dạng chuẩn 2



❖ Một lược đồ quan hệ Q được gọi là ở **dạng chuẩn 2 (2NF)** nếu và chỉ nếu

- Q ở dạng chuẩn 1, VÀ
- *Mọi thuộc tính không khoá của Q đều phụ thuộc đầy đủ vào khoá của Q*

☞ Mọi thuộc tính không khoá phụ thuộc đầy đủ vào khoá, cũng có thể xảy ra khi có phụ thuộc bắc cầu → phụ thuộc được nhận thấy ở các thuộc tính không khoá trong quan hệ có phụ thuộc qua lại với nhau
Ví dụ: $f_1: ABC \rightarrow D$; $f_2: D \rightarrow E \Rightarrow$ có phụ thuộc bắc cầu $f_3: ABC \rightarrow E$
(luật bắc cầu)

Bổ sung kiến thức ...



- ❖ **Thuộc tính khoá (không khoá)**: A là một khoá (không khoá) của Q nếu: $A \in Q^+ \wedge A \in (\notin) K$
- ❖ **Phụ thuộc hàm đầy đủ**: cho PTH $f: X \rightarrow Y \in \mathcal{F}$, f được gọi là phụ thuộc hàm đầy đủ trong $\mathcal{F}$ nếu: $\nexists X' \subseteq X : \mathcal{F} \models X' \rightarrow Y$
Ví dụ: Cho $\mathcal{F} = \{A \rightarrow BCD; BCD \rightarrow E; CD \rightarrow E\}$
 $BCD \rightarrow E$ là một phụ thuộc hàm không đầy đủ vì $\mathcal{F} \models (CD \rightarrow E)$
- ❖ **Thuộc tính phụ thuộc đầy đủ vào X**: A là một thuộc tính phụ thuộc đầy đủ vào X nếu $X \rightarrow A$ là một phụ thuộc hàm đầy đủ.
- ❖ **Thuộc tính phụ thuộc bắc cầu vào X**: A phụ thuộc bắc cầu vào X nếu có 4 điều kiện sau:
 - (i) $X \rightarrow Y \in \mathcal{F}^+$
 - (ii) $Y \rightarrow A \in \mathcal{F}^+$
 - (iii) $Y \rightarrow X \notin \mathcal{F}^+$
 - (iv) $A \notin (X \cup Y)$

Ví dụ: Cho $\mathcal{F} = \{MN \rightarrow OPRX; NO \rightarrow M; P \rightarrow RY\}$ Ta có thuộc tính R phụ thuộc bắc cầu vào NO.

Ví dụ 1



Cho lược đồ **LICH_CT**(GV_CT, N, G, P, M, GV_PT)

và $\mathcal{F} = \{f_1: \text{GV_CT} \rightarrow \text{N, G, P}: \text{một GV coi thi vào 1 ngày, 1 giờ trong 1 phòng duy nhất};$
 $f_2: \text{M} \rightarrow \text{GV_PT}: \text{mỗi môn thi chỉ có một GV phụ trách ra đề thi};$
 $f_3: \text{N, G, P} \rightarrow \text{M}: \text{mỗi ngày, vào một giờ, trong một phòng chỉ có bố trí 1 môn duy nhất thi}\}$

XÁC ĐỊNH KHOÁ CỦA QUAN HỆ TRÊN ???

Ta có: $\text{GV_CT} \rightarrow Q^+$. Vậy GV_CT là khoá

QUAN HỆ LICH_CT Ở DẠNG CHUẨN 2 (2NF)

$T_{\text{LICH_CT}}$	GV_CT	N	G	P	M	GV_PT
	N.V.An	2	8.00-10.00	101	KTLT	Viet
	T.T.Binh	2	10.00-12.00	101	CSDL	Long
	L.V.Chien	2	10.00-12.00	101	CSDL	Long
	V.T.Hoa	2	8.00-10.00	103	KTLT	Viet

CSDL này có sự trùng lặp thông tin trên các tập thuộc tính nào?

N, G, P, M; M, GV_PT

Ví dụ 2

- ❖ Giả sử lđqh LICH_CT(GV_CT, N, G, P, M, GV_PT) được tách thành 2 lđqh sau:

COI_THI (GV_CT, N, G, P)

MÔN_KT (N, G, P, M, GV_PT)

và $\mathcal{F} = \{f_1: GV_CT \rightarrow N, G, P; f_2: M \rightarrow GV_PT; f_3: N, G, P \rightarrow M\}$

XÁC ĐỊNH KHOÁ CỦA CÁC QUAN HỆ TRÊN ???

DẠNG CHUẨN ???

*COI_THI ở **DC2** vì cả N,G,P đều phụ thuộc đầy đủ vào GV_CT.
MON_KT ở **DC2** vì có phụ thuộc bắc cầu.*

T_{COI_THI}					T_{MON_KT}				
	GV_CT	N	G	P	N	G	P	M	GV_PT
	N.V.An	2	8.00-10.00	101					
	T.T.Binh	2	10.00-12.00	101	2	8.00-10.00	101	KTLT	Viet
	L.V.Chien	2	10.00-12.00	101	2	10.00-12.00	101	CSDL	Long
	V.T.Hoa	2	8.00-10.00	103	2	8.00-10.00	103	KTLT	Viet

Vấn đề tồn tại ở DC2

❖ Trùng lặp thông tin trên COI_THI và MON_KT
→ còn bất tiện khi THÊM, XOÁ, SỬA

- Hơn DC1 ở việc thêm thông tin, nhưng vẫn còn phiền phức (xem lại slide 13)
- Xóa có phiền phức do mất thông tin
- Sửa phải sửa đồng loạt từ trên xuống dưới

T _{COI_THI}					T _{MON_KT}				
	GV_CT	N	G	P	N	G	P	M	GV_PT
N.V.An		2	8.00-10.00	101	2	8.00-10.00	101	KTLT	Viet
T.T.Binh		2	10.00-12.00	101	2	10.00-12.00	101	CSDL	Long
L.V.Chien		2	10.00-12.00	101	2	8.00-10.00	103	KTLT	Viet
V.T.Hoa		2	8.00-10.00	103					

COI_THI (GV_CT, N, G, P)

MÔN_KT (N, G, P, M, GV_PT)

và $\mathcal{F} = \{f_1: GV_CT \rightarrow N, G, P; f_2: M \rightarrow GV_PT; f_3: N, G, P \rightarrow M\}$

Ví dụ 3

❖ Cho Idqh $LICH_DAY(GV, N, G, M, P)$

và $\mathcal{F} = \{f_1: N, G \rightarrow M, P: \text{ngày, giờ xác định cho môn học ở một phòng};$
 $f_2: GV \rightarrow M: \text{mỗi giáo viên dạy một môn học duy nhất}\}$

XÁC ĐỊNH KHOÁ CỦA QUAN HỆ TRÊN ???

Ta có: $GV, N, G \rightarrow Q^+$. Vậy GV, N, G là khoá

DẠNG CHUẨN ???

M phụ thuộc hàm bởi GV, mà GV lại là một thành phần của khoá $\Rightarrow$ **Tồn tại phụ thuộc hàm không phụ thuộc đầy đủ vào khoá.**

$LICH_DAY$ không thoả dạng chuẩn 2 - **VẬY CÓ DC 1**

KẾT LUẬN

Q là dạng chuẩn 1 (không ở DC 2) nếu tìm được một pth có các thuộc tính không khoá không phụ thuộc đầy đủ vào khoá.

Hay viết là: $\exists$ pth $f, f: X \rightarrow A \wedge X \subset K$ với $A \in \{Q^+ \setminus K\}$

Dạng chuẩn 3



❖ Định nghĩa 1:

Q ở DC3 nếu và chỉ nếu tất cả thuộc tính không khóa phụ thuộc đầy đủ và **không có** phụ thuộc bắc cầu vào khóa

❖ Định nghĩa 2:

Q ở DC3 nếu và chỉ nếu $\forall X \rightarrow A$ là phụ thuộc hàm không hiển nhiên định nghĩa trên Q (A là thuộc tính đơn, X là tập thuộc tính), và **1 trong 2 điều kiện** sau được thỏa:

- (i) hoặc X chứa (là) 1 khóa của Q ($K_i \subseteq X$)
- (ii) hoặc A là thuộc tính của một khóa của Q ($A \in K_i$)

LƯU Ý: HIỂN NHIÊN Q THỎA DẠNG CHUẨN 2

Ví dụ 1



❖ Cho Idqh **Q(ABCD)**

và $\mathcal{F} = \{f_1: \mathbf{ABC} \rightarrow \mathbf{D}; f_2: \mathbf{D} \rightarrow \mathbf{C}\}$

XÁC ĐỊNH KHOÁ VÀ DẠNG CHUẨN CỦA Q ???

Ta có: $K_1 = \mathbf{ABC}$; $K_2 = \mathbf{ABD}$

$f_1: \mathbf{ABC} \rightarrow \mathbf{D}$: thoả đn [2.(i)]

$f_2: \mathbf{D} \rightarrow \mathbf{C}$: thoả đn [2.(ii)]. Vậy Q ở DC3

❖ Cho Idqh **Q'(CSZ)**

và $\mathcal{F} = \{f_1: \mathbf{CS} \rightarrow \mathbf{Z}; f_2: \mathbf{Z} \rightarrow \mathbf{C}\}$

Ta có: $K_1 = \mathbf{CS}$; $K_2 = \mathbf{ZS}$

Vậy Q' thoả DC3

Ví dụ 2



Giả sử có các lđqh sau:

COI_THI (GV_CT, N, G, P)

LICH_KT (N, G, P, M)

PT_DE (M, GV_PT)

với tập $\mathcal{F} = \{f_1: GV_CT \rightarrow N, G, P; f_2: N, G, P \rightarrow M; f_3: M \rightarrow GV_PT\}$

T_{COI_THI}

GV_CT	N	G	P
N.V.An	2	8.00-10.00	101
T.T.Binh	2	10.00-12.00	101
L.V.Chien	2	10.00-12.00	101
V.T.Hoa	2	8.00-10.00	103

T_{LICH_KT}

N	G	P	M
2	8.00-10.00	101	KTLT
2	10.00-12.00	101	CSDL
2	8.00-10.00	103	KTLT

T_{PT_DE}

M	GV_PT
KTLT	Viet
CSDL	Long

XÁC ĐỊNH KHOÁ VÀ DẠNG CHUẨN CỦA CÁC QUAN HỆ TRÊN

Nhận xét ...



- ❖ Các thể hiện của các quan hệ trong VD 2 không còn chứa thông tin trùng lặp, do đó sẽ không còn vấn đề trong các thao tác thêm, xóa, sửa
- ❖ Tuy nhiên, trong một số quan hệ có DC3 vẫn tồn tại vấn đề như trùng lặp thông tin, bất tiện trong khai thác. Chẳng hạn trong VD 3 như sau:
 - Với các quan hệ và tập PTH ở VD 2, ta có thêm một ràng buộc mới $f_4: M \rightarrow P$

Ví dụ 3

Giả sử có các lđqh sau:

COI_THI (GV_CT, N, G, P)

LICH_KT (N, G, P, M)

PT_DE (M, GV_PT)

với tập $\mathcal{F} = \{f_1: GV_CT \rightarrow N, G, P; f_2: N, G, P \rightarrow M; f_3: M \rightarrow GV_PT; f_4: M \rightarrow P\}$

XÁC ĐỊNH KHOÁ VÀ DẠNG CHUẨN CỦA CÁC QUAN HỆ TRÊN

KHOÁ CỦA LICH_KT LÀ N, G, P VÀ N, G, M - DẠNG CHUẨN CỦA CÁC QUAN HỆ VẪN LÀ DC3 (3NF)

T_{LICH_KT}

N	G	P	M
2	8.00-10.00	101	KTLT
2	10.00-12.00	101	CSDL
3	8.00-10.00	101	CSDL
2	8.00-10.00	105	CSDL

Các PTH đều được thoả

Thêm vào 1 bộ

Hoặc thêm vào 1 bộ

Không vi phạm PTH nhưng tạo thông tin trùng lặp

Vi phạm f_4

Dạng chuẩn BCK (Boyce-Codd-Kent)

❖ Là **DC3 cải tiến**

❖ Không có tính chất [ĐN2.(ii)]

❖ **Nghĩa là:**

- $\forall A \in Q^+, X \rightarrow A \in \mathcal{F}^+$ hay nói cách khác X chứa 1 khóa của Q ($K_i \subseteq X$)

❖ **Ví dụ:**

- Cho $Q(ABCD)$ và $\mathcal{F} = \{AB \rightarrow CD\}$
Vậy Q ở DC BCK vì $K = AB$

Hạn chế của DC BCK

- ❖ DC BCK quan tâm giải quyết vấn đề trùng lặp thông tin, đây là mẫu chốt cho mọi bất tiện thường gặp trong khi khai thác CSDL. Nhưng BCK lại coi nhẹ một tiêu chuẩn khác, không kém phần quan trọng đó là làm thế nào kiểm tra PT dữ liệu thuận lợi nhất

Ví dụ cấu trúc CSDL trong các ví dụ trước được sửa đổi thành:

COI_THI (GV_CT, N, G, P)

LICH_KT1(M, P) LICH_KT2(M, G, N)

PT_DE (M, GV_PT)

với tập $\mathcal{F} = \{f_1: GV_CT \rightarrow N, G, P; f_2: N, G, P \rightarrow M; f_3: M \rightarrow GV_PT; f_4: M \rightarrow P\}$

T_{COI_THI}

GV_CT	N	G	P
N.V.An	2	8.00-10.00	101
T.T.Binh	2	10.00-12.00	101
L.V.Chien	2	10.00-12.00	101
V.T.Hoa	2	8.00-10.00	103

T_{LICH_KT1}

M	P
KTLT	101
CSDL	101

T_{LICH_KT2}

M	G	N
KTLT	8.00-10.00	2
CSDL	10.00-12.00	2

T_{PT_DE}

M	GV_PT
KTLT	Viet
CSDL	Long

f_2 không còn định nghĩa trên quan hệ nào cả. Muốn kiểm tra thì phải kết 2 bảng LICH_KT1, LICH_KT2

Dạng chuẩn 4



❖ Là DC BCK và

❖ $\forall X \rightarrow Y$: phụ thuộc đa trị không hiển nhiên được định nghĩa trên Q và $\forall B \in Q^+, X \rightarrow B \in \mathcal{D}^+$ hay nói cách khác X chứa 1 khóa của Q ($K_i \subseteq X$)

❖ Định nghĩa về phụ thuộc đa trị

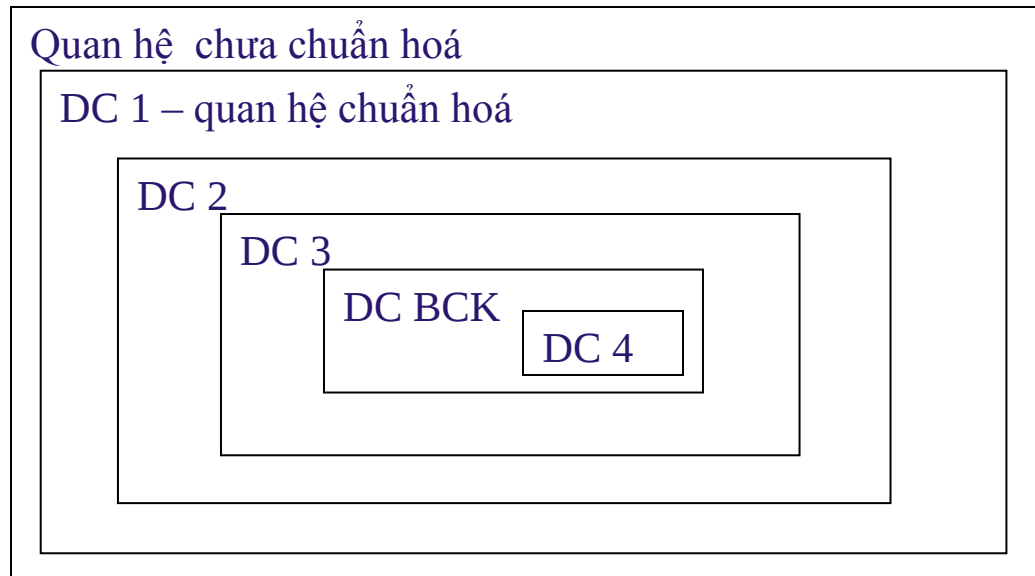
Cho $Q(XYZ)$, $X \subseteq Q^+$, $Y \subseteq Q^+$, $X \cap Y \neq \emptyset$, và $Z = Q^+ - X \cup Y$, $X \rightarrow Y$ là 1 phụ thuộc đa trị được định nghĩa trên Q nếu mỗi giá trị x của X xác định duy nhất một tập giá trị $\{y_1, y_2, \dots\}$ của Y , và tập giá trị này không phụ thuộc vào các giá trị của Z trong các bộ có liên quan đến $x, y_1, y_2, \dots$

- DC BCK và DC4 là những dạng chuẩn nhằm giảm tối thiểu trùng lặp thông tin và giải quyết tương đối hiệu quả việc kiểm tra các PTH (và PT đa trị). Tuy nhiên, đôi khi vẫn còn tồn tại một số PTH mà việc kiểm tra không được thuận lợi vì phải thực hiện trên nhiều quan hệ.

TÓM TẮT



Ta nhận thấy DC BCK là DC3 với chỉ điều kiện (ii) theo định nghĩa 2, do đó mọi quan hệ thỏa DC BCK đều thỏa DC3. Tương tự như vậy mọi quan hệ ở DC3 thỏa DC2, và mọi quan hệ ở DC2 thỏa DC1



DC 1, DC 2 : có sự trùng lặp thông tin

DC 3: khử được trùng lặp thông tin, tuy nhiên vẫn còn tồn tại vấn đề này ở một số trường hợp cụ thể

DC BCK: là dạng chuẩn 3 được cải tiến (*loại bỏ hạn chế của DC 3 ở một số trường hợp*)

DC BCK, DC 4: giảm được tối thiểu sự trùng lặp thông tin

Dạng chuẩn của lđ CSDL



- ❖ $\underline{\mathcal{C}} = \{\underline{Q}\}$

- ❖ Ý nghĩa: tập các lược đồ quan hệ

- ❖ Thông thường được đn là 2 phần:

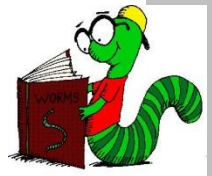
$$\underline{\mathcal{C}} = \langle Q, \mathcal{D} \rangle$$

Q: Tập quan hệ
 $\mathcal{D}$: Tập RBTV



- ❖ Vậy có thể viết:

$$\underline{\mathcal{C}} = \{ \langle Q_i, \mathcal{D}_i \rangle \}_{i=1}^n$$



- ❖ Nếu ta chỉ quan tâm đến pth thì: $\mathcal{D}_i = \mathcal{F}_i$

- ❖ **Dạng chuẩn của lđ CSDL**: xét dạng chuẩn của từng quan hệ trong lược đồ, sau đó suy ra **dạng chuẩn của lđ CSDL = dạng chuẩn của quan hệ có DC thấp nhất**

Cho lược **LICH_CT**(GV_CT, N, G, P, M, GV_PT)

và $\mathcal{F} = \{f_1: GV_CT \rightarrow N, G, P: \text{một GV coi thi vào 1 ngày, 1 giờ trong 1 phòng duy nhất};$
 $f_2: M \rightarrow GV_PT: \text{mỗi môn thi chỉ có một GV phụ trách ra đề thi}\}$ **1NF**

Cải tiến lần 1:

COI_THI (GV_CT, N, G, P) BCK

MÔN_KT (N, G, P, M, GV_PT)

2NF



2NF

và $\mathcal{F} = \{f_1: GV_CT \rightarrow N, G, P; f_2: M \rightarrow GV_PT; f_3: N, G, P \rightarrow M\}$



Cải tiến lần 2:

COI_THI (GV_CT, N, G, P) BCK

LICH_KT (N, G, P, M)

3NF



3NF

PT_DE (M, GV_PT) BCK

với tập $\mathcal{F} = \{f_1: GV_CT \rightarrow N, G, P; f_2: N, G, P \rightarrow M; f_3: M \rightarrow GV_PT; f_4: M \rightarrow P\}$



Cải tiến lần 3:

COI_THI (GV_CT, N, G, P)

BCK

LICH_KT 1(M, P)

LICH_KT 2(M, G, N)

BCK



BCK

PT_DE (M, GV_PT)

BCK

với tập $\mathcal{F} = \{f_1: GV_CT \rightarrow N, G, P; f_2: N, G, P \rightarrow M; f_3: M \rightarrow GV_PT; f_4: M \rightarrow P\}$

Ví dụ



Cho lược đồ quan hệ phổ quát sau:

$$\underline{C}_0 = \langle Q_0(ABCDEFGXYZTV), \mathcal{F}_0 \rangle \text{ với}$$

$$\mathcal{F}_0 = \{A \rightarrow BCEGD\mathbf{X}; GE \rightarrow C\mathbf{Y}; CE \rightarrow \mathbf{Z}; BCE \rightarrow T\mathbf{D}; D \rightarrow VCE\}$$

Và lđ CSDL sau:

$$\underline{C}_1 = \{ \langle Q_1(ABCEGD\mathbf{X}), \mathcal{F}_1 \rangle ;$$

$$\langle Q_2(GEC\mathbf{Y}), \mathcal{F}_2 \rangle ;$$

$$\langle Q_3(CE\mathbf{Z}), \mathcal{F}_3 \rangle ;$$

$$\langle Q_4(BCED\mathbf{T}), \mathcal{F}_4 \rangle ;$$

$$\langle Q_5(DVCE), \mathcal{F}_5 \rangle \}$$

XÁC ĐỊNH KHOÁ VÀ DẠNG CHUẨN CỦA CÁC QUAN HỆ CON ?

ĐÁNH GIÁ CHẤT LƯỢNG CỦA $\underline{C}_1$ (DẠNG CHUẨN CỦA $\underline{C}_1$) ?

Bài tập 1



Cho lược đồ quan hệ phổ quát sau:

$$\underline{C}_0 = \langle Q_0(ABCDEFG), \underline{F}_0 \rangle \text{ với}$$

$$\underline{F}_0 = \{A \rightarrow B; B \rightarrow CG; C \rightarrow D; AC \rightarrow E; ECG \rightarrow B\}$$

Và Id CSDL sau:

$$\underline{C}_1 = \{ \langle Q_1(ABCD), \underline{F}_1 \rangle ;$$

$$\langle Q_2(ACE), \underline{F}_2 \rangle ;$$

$$\langle Q_3(EBCG), \underline{F}_3 \rangle \}$$

XÁC ĐỊNH KHOÁ VÀ DẠNG CHUẨN CỦA CÁC QUAN HỆ CON ?

ĐÁNH GIÁ CHẤT LƯỢNG CỦA $\underline{C}_1$ (DẠNG CHUẨN CỦA $\underline{C}_1$) ?

Bài tập 2



Cho lược đồ quan hệ phổ quát sau:

$$\underline{C}_0 = \langle Q_0(\text{ABCGEDRXYZTV}), \mathcal{F}_0 \rangle \text{ với}$$

$$\mathcal{F}_0 = \{ED \rightarrow Y; B \rightarrow \text{CGZD}; \text{CG} \rightarrow B; A \rightarrow \text{BCDEGRX}; G \rightarrow \text{RV}; D \rightarrow T\}$$

Và lữ CSDL sau:

$$\begin{aligned} \underline{C}_1 = \{ & \langle Q_1(\text{BCGZD}), \mathcal{F}_1 \rangle; \\ & \langle Q_2(\text{EDY}), \mathcal{F}_2 \rangle; \\ & \langle Q_3(\text{DT}), \mathcal{F}_3 \rangle; \\ & \langle Q_4(\text{ABCGEDRX}), \mathcal{F}_4 \rangle; \\ & \langle Q_5(\text{GRV}), \mathcal{F}_5 \rangle \} \end{aligned}$$

XÁC ĐỊNH KHOÁ VÀ DẠNG CHUẨN CỦA CÁC QUAN HỆ CON ?

ĐÁNH GIÁ CHẤT LƯỢNG CỦA $\underline{C}_1$ (DẠNG CHUẨN CỦA $\underline{C}_1$) ?

Bài tập 3



Cho lược đồ quan hệ phổ quát sau:

$$\underline{C}_0 = \langle Q_0(\text{AIMNLTVXZY}), \mathcal{F}_0 \rangle \text{ với}$$

$$\mathcal{F}_0 = \{A \rightarrow \text{VIXLN}; N \rightarrow \text{LY}; L \rightarrow \text{TIX}; \text{MN} \rightarrow \text{ZA}\}$$

Và lữ CSDL sau:

$$\begin{aligned} \underline{C}_1 = \{ & \langle Q_1(\text{MNZA}), \mathcal{F}_1 \rangle; \\ & \langle Q_2(\text{AVIXLN}), \mathcal{F}_2 \rangle; \\ & \langle Q_3(\text{NYL}), \mathcal{F}_3 \rangle; \\ & \langle Q_4(\text{LTIX}), \mathcal{F}_4 \rangle \} \end{aligned}$$

XÁC ĐỊNH KHOÁ VÀ DẠNG CHUẨN CỦA CÁC QUAN HỆ CON ?

ĐÁNH GIÁ CHẤT LƯỢNG CỦA $\underline{C}_1$ (DẠNG CHUẨN CỦA $\underline{C}_1$) ?

Chuẩn hoá dữ cơ sở dữ liệu



- ❖ Đã biết: đánh giá chất lượng thiết kế của một dữ CSDL dựa trên các tiêu chuẩn về dạng chuẩn của các dữ quan hệ con
- ❖ **Chuẩn hoá dữ CSDL nhằm mục đích nâng cao chất lượng thiết kế hay cụ thể hơn là đưa các dữ quan hệ con từ dạng chuẩn thấp (DC1, DC2) lên dạng chuẩn cao hơn mà tối thiểu phải là DC3 (BCK, DC4)**
- ❖ Có 2 phương pháp chính: **phân rã** và tổng hợp

Vài khái niệm cần biết ...



❖ Hai CSDL tương đương nếu:

- Tương đương về mặt thông tin $Q_0^+ = \bigcup_{i=1}^n Q_i^+$
- Tương đương về mặt PTH

$$F^+ = \bigcup_{i=1}^n F_i^+ \Leftrightarrow F = \bigcup_{i=1}^n F_i$$

❖ Đánh giá lđqh (CSDL) sau khi đã cải tiến dựa trên các tiêu chuẩn:

- Dạng chuẩn: 3NF, BCK, 4NF
- Bảo toàn thông tin
- Bảo toàn phụ thuộc hàm



Định lý Delobel



❖ Định lý về sự bảo toàn thông tin

❖ Giả sử có một lđqh $Q(XYZ)$ và tập pth $\mathcal{F}$

❖ Nếu $X \rightarrow Y \in \mathcal{F}^+$ thì phép phân rã Q thành 2 lđqh con $Q_1(XY)$ và $Q_2(XZ)$ là bảo toàn thông tin, nghĩa là:

$$\forall T_Q : T_Q = T_Q[XY] \boxtimes T_Q[XZ]$$

❖ Vận dụng định lý ta có:

$$\forall f: X \rightarrow A \in \mathcal{F},$$

Q được phân rã thành $Q_1(XA)$ và $Q_2(X, (Q^+ - A))$

Một thuật toán phân rã

Procedure **Phan_ra**($Q, \mathcal{F}$)

Begin

$$\mathcal{F}^* = \mathcal{F} \setminus \{ f \in \mathcal{F} / VT(f) \cup VP(f) = Q^+ \}$$

If ($\mathcal{F}^* \neq \emptyset$) then

Begin

B1. chọn một $f_0: X \rightarrow Y \in \mathcal{F}$

B2. Tạo các lđqh con Q_1 và Q_2

$$Q_1^+ = X \cup Y;$$

$$\mathcal{F}_1 = \{ f \in \mathcal{F} / VT(f) \cup VP(f) \subseteq Q_1^+ \};$$

$$Q_2^+ = Q^+ \setminus Y;$$

$$\mathcal{F}_2 = \{ f \in \mathcal{F} / VT(f) \cup VP(f) \subseteq Q_2^+ \};$$

B3. Phân rã đệ quy Q_1 và Q_2

Phan_ra ($Q_1, \mathcal{F}_1$);

Phan_ra ($Q_2, \mathcal{F}_2$);

End;

End;



Ví dụ minh họa



Cho $Q (ABCD); \mathcal{F} = \{A \rightarrow B; B \rightarrow C; C \rightarrow D\}$

(1) Hãy xác định dạng chuẩn cho lđqh này

(2) Hãy xác định một lđqh khác trong đó tập pth $\mathcal{F}$ được bảo toàn, các quan hệ con đạt dạng chuẩn cao nhất, lược đồ được bảo toàn thông tin

Các bước giải

- Xác định khoá của Q
- Xác định dạng chuẩn của Q
- Chuẩn hoá Q thành Q' bằng phương pháp phân rã Q

Một cách phân rã...



$Q_0(\underline{A}BCD), \mathcal{F} = \{A \rightarrow B; B \rightarrow C; C \rightarrow D\}$ **DC2**

$Q_1(\underline{C}D)$

$\mathcal{F}_1 = \{C \rightarrow D\}$

DC BCK

$C \rightarrow D$

$Q_2(C\underline{A}B)$

DC2

$\mathcal{F}_2 = \{A \rightarrow B; B \rightarrow C\}$

$B \rightarrow C$

$Q_{21}(\underline{B}C)$

$\mathcal{F}_{21} = \{B \rightarrow C\}$

DC BCK

$Q_{22}(B\underline{A})$

$\mathcal{F}_{22} = \{A \rightarrow B\}$

DC BCK

$Q' = \{Q_1, Q_{21}, Q_{22}\}$



- Dạng chuẩn BCK
- **Bảo toàn phụ thuộc hàm**
- **Bảo toàn thông tin**

Một cách phân rã khác ...



$Q_0(\underline{A}BCD), \mathcal{F} = \{A \rightarrow B; B \rightarrow C; C \rightarrow D\}$ **DC2**

$Q_1(\underline{A}B)$

$A \rightarrow B$

$Q_2(\underline{A}CD)$ **DC2**

$\mathcal{F}_1 = \{A \rightarrow B\}$

$\mathcal{F}_2 = \{A \rightarrow CD; C \rightarrow D\}$

DC BCK

$C \rightarrow D$

$Q_{21}(\underline{C}D)$

$Q_{22}(C\underline{A})$

$\mathcal{F}_{21} = \{C \rightarrow D\}$

$\mathcal{F}_{22} = \{A \rightarrow C\}$

DC BCK

DC BCK

$Q' = \{Q_1, Q_{21}, Q_{22}\}$



- Dạng chuẩn BCK
- **Không bảo toàn phụ thuộc hàm**
- Bảo toàn thông tin



- ❖ Cách tiếp cận phân rã cho DC cao (BCK, 4NF), bảo toàn thông tin. Tuy nhiên, hạn chế là có thể phát sinh những quan hệ vô nghĩa, có thể không bảo toàn PTH
- ❖ Để phân rã cho ra 1 lđqh có chất lượng tốt hơn. Ta nên sắp xếp thứ tự theo chiều PTH:

$$\begin{array}{ccccc} A & \xrightarrow{f_1} & B & \xrightarrow{f_2} & C & \xrightarrow{f_3} & D \end{array}$$

Sau đó, chọn PTH nằm cuối (phía bên phải)

- ❖ Với một lđ CSDL sau khi xác định dạng chuẩn, nếu trường hợp có dạng chuẩn 1NF, 2NF thì sẽ chọn lđqh nào gây ảnh hưởng đối với dạng chuẩn của lđ CSDL để phân rã → cải tiến lđ CSDL

Bài tập nâng cao

❖ Cho lược đồ quan hệ phổ quát sau:

$$\underline{C}_0 = \langle Q_0(ABCDEFGXYZTV), \mathcal{F}_0 \rangle \text{ với}$$

$$\mathcal{F}_0 = \{A \rightarrow BCEGDX; GE \rightarrow CY; CE \rightarrow Z; BCE \rightarrow TD; D \rightarrow VCE\}$$

Và lđ CSDL sau:

$$\underline{C}_1 = \{\langle Q_1(ABCEGDX), \mathcal{F}_1 \rangle;$$

$$\langle Q_2(GECY), \mathcal{F}_2 \rangle;$$

$$\langle Q_3(CEZ), \mathcal{F}_3 \rangle;$$

$$\langle Q_4(BCEDT), \mathcal{F}_4 \rangle;$$

$$\langle Q_5(DVCE), \mathcal{F}_5 \rangle\}$$

CẢI TIẾN CHẤT LƯỢNG CỦA $\underline{C}_1$ THÀNH $\underline{C}_2$ ĐỂ CÓ DC CAO NHẤT, BẢO TOÀN THÔNG TIN, BẢO TOÀN PTH ?

Bài tập nâng cao



❖ Với các bài tập 1, 2, 3 đã thực hiện ở trên

❖ **CẢI TIẾN CHẤT LƯỢNG CỦA $\mathcal{C}_1$ THÀNH $\mathcal{C}_2$ ĐỂ CÓ DC CAO NHẤT, BẢO TOÀN THÔNG TIN, BẢO TOÀN PTH ?**



Cám ơn đã theo dõi ...

Lê Đức Long

Email: longld@math.hcmup.edu.vn

Website: <http://www.2learner.edu.vn>

