

**Nội dung**

- **Dẫn nhập- Những mục tiêu chính**
- **Dạng chuẩn**
- **Cấu trúc CSDL tương đương**
- **Hai cách tiếp cận để thiết kế CSDL**

19

**Quan hệ được chuẩn hóa – Định nghĩa dạng chuẩn 1 (DC1)**

- **Trước khi Codd đề nghị khái niệm dạng chuẩn (1970), công nghệ lưu trữ dữ liệu khuyến khích cấu trúc dữ liệu phân cấp, trong đó có nhiều trường lặp.**

- **Ví dụ 3.3**

DANH\_SÁCH\_LỚP(TênLớp, SiSố, (TênHS,(Điểm)\*15)\*30)

Điểm được lặp lại 15 lần (mỗi học sinh có tối đa 15 cột điểm)

Thông tin về mỗi học sinh (tên học sinh, các điểm) được lặp lại 30 lần (lớp học có tối đa 30 học sinh)

20

**Quan hệ được chuẩn hóa – Định nghĩa dạng chuẩn 1**

- **Ví dụ 3.3 (tt):**

– Hình sau minh họa dữ liệu của quan hệ DANH\_SÁCH\_LỚP:

TênLớp	SiSố	TênHS_1	Điểm_11	Điểm_12	...	Điểm_15	TênHS_2	Điểm_21	...
11A1	30	NguyễnVânA	7	8	...	9	TrầnThịB	6	...
11A2	28	VõVânC	8	6	...	10	...	...	...

21

***Quan hệ được chuẩn hóa – Định nghĩa dạng chuẩn 1***

- E.F. Codd đề nghị cấu trúc lưu trữ chuẩn hóa, theo đó, cấu trúc CSDL trong ví dụ 3.1 trở thành:

DANH\_SÁCH\_HS(TênLớp, SiSố, STT<sub>i</sub>, TênHS)

ĐIỂM\_HS\_LỚP(TênLớp, STT<sub>i</sub>, STT<sub>j</sub>, Điểm)

- Các thuộc tính STT<sub>1</sub>, STT<sub>2</sub> sau đó được thay bằng tên gọi có tính gợi nhớ, phù hợp với ngữ nghĩa của chúng:

STT<sub>1</sub>: mã số học sinh trong lớp, thay bằng “MãHS”.

STT<sub>2</sub>: môn học, thay bằng “Môn”

- Hai quan hệ trên trở thành:

DANH\_SÁCH\_HS(TênLớp, SiSố, MãHS, TênHS)

ĐIỂM\_HS\_LỚP(TênLớp, MãHS, Môn, Điểm)

22

***Quan hệ được chuẩn hóa – Định nghĩa dạng chuẩn 1***

- Minh họa cho dữ liệu của cấu trúc đã chuẩn hóa:

DANH\_SÁCH\_HS

ĐIỂM\_HS\_LỚP

Tên_Lớp	SiSố	MãHS	TênHS
11A1	30	0001	NguyễnVănA
11A1	30	0002	TrầnThịB
11A2	28	0031	VõVănC

TênLớp	MãHS	Môn	Điểm
11A1	0001	Toán	7
11A1	0001	Lý	9
11A1	0002	Lý	10
11A1	0002	Toán	6
11A1	0001	Văn	7
11A2	0031	Lý	6
...	...	...	...

23

***Quan hệ được chuẩn hóa – Định nghĩa dạng chuẩn 1***

- Các quan hệ trong cấu trúc CSDL được chuẩn hóa như trên gọi là quan hệ ở **dạng chuẩn 1**
- **Định nghĩa dạng chuẩn 1:** Một quan hệ ở dạng chuẩn 1 không có các trường lặp và các trường kép, còn được gọi là **cấu trúc phẳng**.

24

**Quan hệ được chuẩn hóa – Định nghĩa dạng chuẩn 1**

- **Cấu trúc chuẩn hóa cho phép truy vấn chỉ dựa trên ngữ nghĩa của CSDL, không dựa vào cách thức lưu trữ thông tin**

Vi dụ: Cho biết điểm môn Văn của học sinh có mã 0001 lớp 11A1:

ĐIỂM\_HS\_LỚP[TênLớp = '11A1', Mã\_HS = '0001', Môn = 'Văn', Điểm]

Với cấu trúc chưa chuẩn hóa, câu truy vấn này sẽ là:

DANH\_SÁCH\_LỚP(11A2)(1,2)

(Học sinh có mã 0001 lớp 11A1 được lưu thứ 1 trong danh sách, điểm môn Văn được lưu thứ 2 trong bảng điểm của HS)

25

**Quan hệ được chuẩn hóa – Định nghĩa dạng chuẩn 1**

- **Nhận xét:**

– Với dạng chuẩn 1, yêu cầu về sự giảm thiểu trùng lặp thông tin chưa được quan tâm

– Các vấn đề gây ra do trùng lặp thông tin:

Bất tiện khi thêm/cập nhật : để đảm bảo dữ liệu nhất quán, phải kiểm tra và cập nhật hàng loạt trên các thông tin trùng lặp

Vi dụ: Thêm một học sinh (11A1, 31, 0003, Võ Thị H) vào danh sách: phải đảm bảo thông tin về sĩ số lớp của các học sinh khác cùng lớp 11A1 cũng phải là 31. Tương tự khi cập nhật.

Mất thông tin khi xóa

Vi dụ: giả sử ta chỉ mới thêm một học sinh mã 0031 trong lớp 11A2. Khi đó, nếu xóa học sinh này đi, ta sẽ mất thông tin về tên lớp và sĩ số của lớp 11A2 trong CSDL.

26

**Quan hệ được chuẩn hóa – Định nghĩa dạng chuẩn 1**

- **Ví dụ 3.4: Cho quan hệ sau:**

LỊCH\_COI\_THI (GV\_CT, N,G,P,M,GV)

$F = \{f_1: GV\_CT \rightarrow N,G,P: \text{một giảng viên coi thi chỉ coi vào một ngày(N), một giờ (G) và trong một phòng(P) duy nhất}\}$

$f_2: M \rightarrow GV$  : mỗi môn thi (M) có một giảng viên (GV) duy nhất phụ trách

$f_3: N,G,P \rightarrow M$  : mỗi ngày, vào một giờ, trong một phòng, chỉ có một môn thi duy nhất  
}

27

**Quan hệ được chuẩn hóa – Định nghĩa dạng chuẩn 1**• **Ví dụ 3.4 (tt):**

- Một thể hiện của quan hệ LỊCH\_COI\_THI:

GV_CT	N	G	P	M	GV
Nguyễn Văn A	2	8:00–10:00	101	Giải thuật	X
Nguyễn Thị B	2	10:00–12:00	101	CSDL	Y
Trần Văn C	2	10:00–12:00	101	CSDL	Y
Võ Thị D	2	8:00–10:00	103	Giải thuật	X

- Nhận xét về sự trùng lặp thông tin và những bất tiện khi thêm/ xóa/ cập nhật?

28

**Quan hệ được chuẩn hóa – Định nghĩa dạng chuẩn 1**

**Cần tách nhỏ quan hệ theo các tiêu chuẩn nghiêm ngặt hơn để giải quyết các vấn đề về trùng lặp thông tin, để CSDL hoạt động hiệu quả hơn**

29

**Dạng chuẩn 2 (DC2)**• **Một số khái niệm bổ sung**

- Thuộc tính khóa (không khóa): A là một thuộc tính khóa (không khóa) của Q nếu:

$$A \in Q^+$$

A có (không có) tham gia vào một khóa của Q

Ví dụ 3.5:  $Q_1(MNO PX); Q_2(PY)$

M, N, O là các thuộc tính khóa của  $Q_1$ , P là thuộc tính khóa của  $Q_2$  nhưng là thuộc tính không khóa của  $Q_1$

- Thuộc tính phụ thuộc đầy đủ (không phụ thuộc đầy đủ) vào một tập thuộc tính:

A là một thuộc tính phụ thuộc đầy đủ (không phụ thuộc đầy đủ) vào X nếu  $PTH X \rightarrow A$  là một phụ thuộc đầy đủ (không đầy đủ)

30

**DC2 – Định nghĩa**

- Một quan hệ  $Q$  ở dạng chuẩn 2 nếu và chỉ nếu tất cả thuộc tính không khóa đều phụ thuộc đầy đủ vào khóa
- Ví dụ 3.6:  
 $COI\_THI(\underline{GV\_CT}, N, G, P); F_1 = \{GV\_CT \rightarrow N, G, P\}$   
 $MON\_KT(\underline{N, G, P}, M, GV); F_2 = \{N, G, P \rightarrow M; M \rightarrow GV\}$   
 Các quan hệ này có đạt DC2?
- Ví dụ 3.7:  
 $Q(\underline{BCDZ}) F = \{B \rightarrow C; BD \rightarrow Z\}$   
 $Q$  có đạt DC2?

31

**DC2\_Các vấn đề còn tồn tại**

- Vẫn còn thông tin trùng lặp.
- Các bất tiện trong thao tác thêm/xóa/cập nhật do trùng lặp thông tin gây ra.  
 Các bảng sau minh họa cho dữ liệu của các quan hệ trong ví dụ 3.6 (đều đạt DC2):

**T\_COI\_THI**

GV_CT	N	G	P
Nguyễn Văn A	2	8:00–10:00	101
Nguyễn Thị B	2	10:00–12:00	101
Trần Văn C	2	10:00–12:00	101
Võ Thị D	2	8:00–10:00	103

**T\_MON\_KT**

N	G	P	M	GV
2	8:00–10:00	101	Giải thuật	X
2	10:00–12:00	101	CSDL	Y
2	8:00–10:00	103	Giải thuật	X

Nhận xét về sự trùng lặp thông tin và các bất tiện khi thêm/ xóa/ sửa?

32

**Dạng chuẩn 3 (DC3)**

- **Khái niệm bổ sung: Thuộc tính phụ thuộc bắc cầu vào một tập thuộc tính:**
  - A phụ thuộc bắc cầu vào X nếu có 4 điều kiện sau:
    - $\exists X \rightarrow Y \in F^+$
    - $Y \rightarrow A \in F^+$
    - $Y \rightarrow X \notin F^+$
    - $A \notin (X \cup Y)$

33

### DC3 – Khái niệm bổ sung (tt)

• Ví dụ 3.8 :

Cho  $F = \{MN \rightarrow OPRX; NO \rightarrow M; P \rightarrow RY\}$

– P có phụ thuộc bắc cầu vào NO?

$NO \rightarrow M \Rightarrow NO \rightarrow MN$  : thỏa (i)

$MN \rightarrow P$ :thỏa (ii)

$MN \rightarrow O \Rightarrow MN \rightarrow ON$  : không thỏa (iii)

P không phụ thuộc bắc cầu vào NO.

– R có phụ thuộc bắc cầu vào NO?

$NO \rightarrow MN$  và  $MN \rightarrow P$  nên  $NO \rightarrow P$  (i)

$P \rightarrow R$  (ii)

$P \rightarrow NO \notin F$  (iii)

$R \not\Rightarrow NOP$

$\rightarrow R$  phụ thuộc bắc cầu vào NO

34

### Dạng chuẩn 3 (DC 3)

• Định nghĩa 1:

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

• Định nghĩa 2:

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

i. Hoặc X chứa một khóa của Q

ii. Hoặc A là thuộc tính khóa của Q

35

### DC3 – Ví dụ

– Ví dụ 3.9: Xét cấu trúc CSDL ở ví dụ 3.6 được cải tiến để các quan hệ đều đạt DC3:

COI\_THI (GV, CT, N, G, P)

LỊCH\_KT (NG, PM)

GD (GV)

T\_COI\_THI

T\_LỊCH\_KT

GV_CT	N	G	P
Nguyễn Văn A	2	8:00–10:00	101
Nguyễn Thị B	2	10:00–12:00	101
Trần Văn C	2	10:00–12:00	101
Võ Thị D	2	8:00–10:00	103

N	G	P	M
2	8:00–10:00	101	Giải thuật
2	10:00–12:00	101	CSDL
2	8:00–10:00	103	Giải thuật

T\_GD

M	GV
Giải thuật	X
CSDL	Y

Không còn trùng lặp thông tin

36

**DC3- Ví dụ - Vấn đề còn tồn tại**

– Ví dụ 3.10: môi trường ứng dụng của ví dụ 3.9 có thêm một ràng buộc mới:  $f_4: M \rightarrow P$

Quan hệ LỊCH\_KT có hai khóa (N,G,P) và (N,G,M), vẫn đạt chuẩn 3. Xét thể hiện sau:

**T\_LỊCH\_KT**

N	G	P	M
2	8:00-10:00	101	Giải thuật
2	10:00-12:00	101	CSDL
Thêm: 3	8:00-10:00	101	CSDL
Thêm: 2	8:00-10:00	105	CSDL

Tạo thông tin trùng lặp vì phạm  $f_4$

37

**Dạng chuẩn Boy – Codd – Kent (DC BCK)**

- Định nghĩa:**

$Q$  ở DC BCK nếu và chỉ nếu: với mỗi PTH không hiển nhiên  $X \rightarrow A$  định nghĩa trên  $Q$  thì  $\nexists B \in Q$ , ta luôn có  $(X \rightarrow B)$  là một PTH thuộc  $F$ . Hay nói cách khác,  $X$  chứa một khóa của  $Q$

- Ví dụ 3.11:**

Quan hệ LỊCH\_CT với hai khóa ở ví dụ 3.10 không đạt dạng chuẩn BCK vì có PTH  $M \rightarrow P$ .

38

**Giới hạn của DC BCK**

- DC BCK quan tâm giải quyết vấn đề trùng lặp thông tin, nhưng xem nhẹ một tiêu chuẩn khác: sự thuận lợi khi kiểm tra phụ thuộc dữ liệu.**

- Ví dụ 3.12:**

Thay đổi cấu trúc CSDL của ví dụ 3.10 để tất cả quan hệ đạt DC BCK:

COI\_THI(GV, CT, N,G,P)

LICH\_KT\_1(M,P) ; LICH\_KT\_2(M,N,G)

GD(M,GV)

39

**Giới hạn của DC BCK – Ví dụ (tt)**

– Minh họa thể hiện của các quan hệ:

**T\_COI\_THI**

GV_CT	N	G	P
Nguyễn Văn A	2	8:00–10:00	101
Nguyễn Thị B	2	10:00–12:00	101
Trần Văn C	2	10:00–12:00	101

**T\_GD**

M	GV
Giải thuật	X
CSDL	Y

**T\_LICH\_KT\_1**

P	M
101	Giải thuật
101	CSDL

**T\_LICH\_KT\_2**

N	G	M
2	8:00–10:00	Giải thuật
2	10:00–12:00	CSDL

→ Không còn trùng lặp thông tin, nhưng bất tiện khi kiểm tra PTH  
 $N, G, P \rightarrow M$ : phải kết LICH\_KT\_1 và LICH\_KT\_2

40

**Dạng chuẩn 4 (DC 4)**

- Định nghĩa:**

Q ở dạng chuẩn 4 nếu và chỉ nếu Q ở DC BCK và: với mỗi phụ thuộc đa trị ( $X \twoheadrightarrow Y$ ) không hiển nhiên được định nghĩa trên Q thì  $\forall T \in Q$ : ta luôn có  $X \twoheadrightarrow B$  là một phụ thuộc hàm thuộc  $F$ . Hay nói cách khác, X chứa một khóa của Q.

- Ví dụ:**

Môi trường ứng dụng của ví dụ 3.9 có thêm phụ thuộc đa trị  $M \twoheadrightarrow >$  P. Khi đó quan hệ LICH\_KT(N, G, P, M) không đạt DC4.

Tách quan hệ này thành LICH\_KT\_3(M, P) VÀ LICH\_KT\_2(M, N, G), ta có các quan hệ đều đạt DC 4

- Vì kiểm tra phụ thuộc đa trị rất tốn kém (có thể kép theo một chuỗi hiệu ứng thêm/xóa), mục tiêu của DC4 là tách nhỏ các quan hệ, nhằm biến các phụ thuộc đa trị thành phụ thuộc đa trị hiển nhiên.**

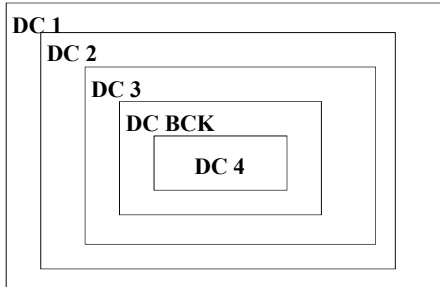
41

**DC4 – Giới hạn**

- Tương tự như DC BCK, việc tách nhỏ các quan hệ để đạt DC 4 có thể làm cho việc kiểm tra một số PTH trở nên tốn kém hơn (phải kết trên nhiều quan hệ).**

42



**Tính chất lồng nhau của các dạng chuẩn**

43

---

---

---

---

---

---

---

**Dạng chuẩn của cấu trúc CSDL**• **Định nghĩa:**

Cho một cấu trúc CSDL  $C = \{ \langle Q_i, F_i \rangle, i = 1..n \}$

Dạng chuẩn của  $C$  là dạng chuẩn thấp nhất trong các dạng chuẩn của các  $Q_i$

• **Ví dụ 3.14:**

- Cấu trúc CSDL ở ví dụ 3.9 đạt BCK
- Cấu trúc CSDL ở ví dụ 3.10 đạt DC 3

44

---

---

---

---

---

---

---