### Chương 5 Các loại phụ thuộc dữ liệu

Phạm Thị Ngọc Diễm Bộ môn HTTT - ĐHCT

### Nội dung

- Giới thiệu vấn đề
- Phụ thuộc hàm

### Nội dung

- Giới thiệu vấn đề
- Phụ thuộc hàm

#### Dư thừa dữ liệu

- Cho quan hệ:
  - DULIEU (MSSV, hoten, diachi, MM, tenmon, diem)
- Cho các quan hệ:
  - SINHVIEN (MSSV, hoten, diachi)
  - MONHOC (MM, tenmon)
  - DANGKY (MSSV, MM, diem)

### Dư thừa dữ liệu

• Xét quan hệ CÁ\_NHÂN (id, hoten, diachi, sothich) với các thể hiện:

id	hoten	diachi	sothich
10110100	John Doe	123 Lý Tự Trọng	Bơi lội
10110100	John Doe	123 Ly Tự Trọng	Bida
21345622	Huỳnh Huy	2 Võ Thị Sáu	Cầu Lông
21345622	Huỳnh Huy	2 Võ Thị Sáu	Bóng chuyền
5555555	Lê Văn Tám	411 30/4	Leo núi

- → (21345622, Huỳnh Huy, 1 Ly tu trong, Bơi lội)
- Xét 4 bộ đầu tiên:
  - Nhiều thông tin lặp lại (id, hoten, diachi) => lưu trữ dư thừa cho cùng thông tin => Đây không là vấn đề chính
  - Vấn đề chính là giữ cho các bản sao dư thừa luôn nhất quán trong CSDL và điều này phải được thực hiện một cách hiệu quả.
  - => Dư thừa có thể dẫn đến bất thường (anomaly) dữ liệu

# **Dư thừa dữ liệu** CÁ\_NHÂN (<u>id</u>, hoten, diachi) THICH (<u>id,sothich</u>)

• Xét quan hệ CÁ\_NHÂN (id, hoten, diachi, sothich) với các thể hiện:

id	hoten	diachi	sothich
10110100	John Doe	123 Lý Tự Trọng	Bơi lội
10110100	John Doe	123 Ly Tự Trọng	Bida
21345622	Huỳnh Huy	2 Võ Thị Sáu	Cầu Lông
21345622	Huỳnh Huy	2 Võ Thị Sáu	Bóng chuyền
	Lê Văn Tám	411 30/4	Leo núi
→ (21345622, Huỳnh Huy, 1 Ly tu trong, Bơi lội)			

- Xét 4 bộ đầu tiên:
  - Nhiều thông tin lặp lại (id, hoten, diachi) => lưu trữ dư thừa cho cùng thông tin => Đây không là vấn đề chính
  - Vấn đề chính là giữ cho các bản sao dư thừa luôn nhất quán trong CSDL và điều này phải được thực hiện một cách hiệu quả.
  - => Dư thừa có thể dẫn đến bất thường (anomaly) dữ liệu

### Dị thường dữ liệu

- Dị thường dữ liệu là
  - Những mâu thuẫn trong các dữ liệu được lưu trữ trong cơ sở dữ liệu,
  - Kết quả của một thao tác như cập nhật, thêm, và / hoặc xóa.
- Sự mâu thuẫn như vậy có thể phát sinh khi
  - Có một bộ đặc biệt được lưu trữ tại nhiều địa điểm (các bản sao);
  - Nhưng không phải tất cả các bản sao đều được cập nhật.

### Dị thường dữ liệu - ví dụ

- Dị thường do cập nhật
  - Thay đổi địa chỉ cho "Huỳnh Huy" → Phải thay đổi đ/c tất cả các bộ mô tả "Huỳnh Huy"
- Dị thường do thêm
  - Thêm bộ mới mà không có thông tin (hoặc NULL) về sothich =>
     không thể thêm vì sothich là một phần của Khóa
- Dị thường do xoá :
  - Giả sử rằng "Lê Văn Tám" không thích leo núi nữa và ta muốn xoá sở thích này khỏi quan hệ:
    - Không có cách nào để chỉ xoá sở thích
    - Hoặc xóa tất cả thông tin mô tả "Lê Văn Tám" → Mất thông tin do xóa
    - Hoặc thay thế sothich bởi NULL → vấn đề NULL trong khóa chính

### Dị thường dữ liệu

- Nhận xét: Nếu chỉ có một bộ có thể mô tả một cá nhân
  - → sothich sẽ không là một phần của khoá
  - → Các vấn đề nêu trong các ví dụ sẽ không xảy ra

### Các tiêu chí đánh giá thiết kế LĐQH

- Đảm bảo rằng ngữ nghĩa của các thuộc tính là rõ ràng trong lược đồ
- Giảm thông tin dư thừa trong các bộ. Dư thừa dữ liệu gây:
  - Dị thường dữ liệu khi thêm hoặc sửa
  - Mất thông tin khi xoá
- Giảm giá trị NULL trong các bộ. Các giá trị NULL làm:
  - Lãng phí không gian lưu trữ
  - Khó thực hiện việc chọn, các hàm kết tập và nối kết
- Không chấp nhận khả năng tạo ra các bộ giả (spurious tuples)
  - Sinh ra do kết nối các quan hệ không dựa trên khoá chính và khoá ngoài.

TAIKHOAN(stk, sodu, ngaymo, MAKH)

### Các tiêu chí đánh giá thiết kế LĐQH

- Giảm giá trị NULL trong các bộ. Các giá trị NULL làm:
  - TAIKHOAN(<u>STK</u>, ngay, sodu, MAKH, MADN)
  - → TAIKHOAN(<u>STK</u>, ngay, sodu, MAKH, loai, ghichu)

### Nội dung

- Giới thiệu vấn đề
- Phụ thuộc hàm

### Giới thiệu PTH

- Khái niệm quan trọng nhất trong lý thuyết thiết kế lược đồ quan hệ là phụ thuộc hàm (PTH)
- PTH là công cụ hình thức để phân tích các lược đồ quan hệ:
  - Cho phép phát hiện và
  - Mô tả một số các vấn đề vừa nêu trên
- Một PTH là một ràng buộc giữa hai tập thuộc tính từ một cơ sở dữ liệu.
- PTH được sử dụng để xác định các dạng chuẩn (Normal Form).

### Định nghĩa

- Cho lược đồ quan hệ R(U) với:
  - U={ A1, A2, ..., An},
  - $-X\neq\varnothing$ ,  $Y\neq\varnothing$ ,  $X\subseteq U$ ,  $Y\subseteq U$

#### Định nghĩa

X xác định Y hay Y phụ thuộc hàm vào X, nếu và chỉ nếu với mỗi giá trị của X xác định duy nhất một giá trị của Y, hay:

 $\forall r \in R$ ,  $\forall t1, t2 \in r$ , t1[X]=t2[X] thì t1[Y]=t2[Y]

- Ký hiệu X → Y
- X là vế trái và Y là vế phải của PTH

R(U)	Α	В
	1	4
	1	5
	3	7

- A → B?
- B → A?
- MSSV → hoten
- Hoten → MSSV

• Cho quan hệ R với tập phụ thuộc hàm F:

$$F= \left\{ \begin{array}{ccc} A & \rightarrow & B \\ B,C & \rightarrow & D \\ D & \rightarrow & E \\ A,C & \rightarrow & D \\ A,C & \rightarrow & E \end{array} \right\}$$

A			222
/\	H		ノノノ
$\boldsymbol{\vdash}$			

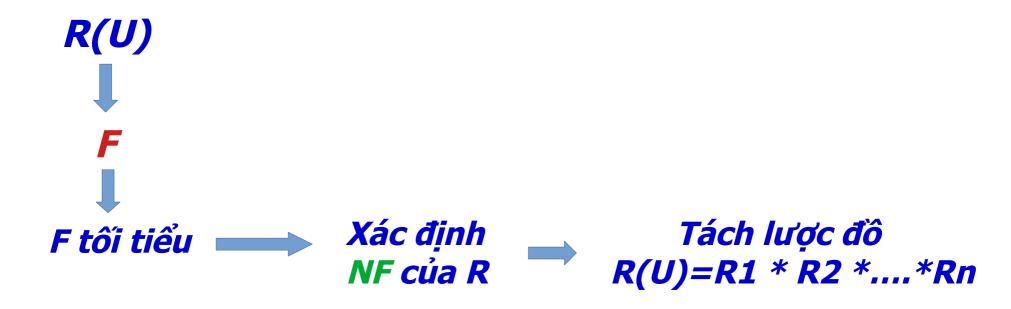
R	Α	В	С	D	E
	a1	b1	c1	d3	e2
	a1	b1	сЗ	d4	e3
	a2	b2	с4	d2	e1
	аЗ	b1	c1	d3	e2
	a2	b2	c4	d2	e1

PTH

```
MASV → hoten
MASV → diachi
MASV, MM, hk, nk → diem
```

- Chú ý: không tồn tại các PTH sau
  - MASV → MAMON
  - MCB → ngaybay
- Chú ý: Sự tồn tại của một số PTH trong một LĐQH có thể dẫn đến dị thường (khi thêm/sửa/xóa).
  - Ví dụ: id, sothich → hoten, diachi
    - => Chuẩn hoá và tách LĐQH

# Các bước thiết kế CSDL theo dạng chuẩn



## Luật suy diễn - Hệ tiên đề Armstrong

- Cho lược đồ quan hệ R(U), U={ A1, A2, ..., An}, X ≠ Ø,
   Y ≠ Ø, X, Y, Z, W ⊆ U
- Hệ tiên đề Armstrong gồm các luật sau:

```
Phản xạ: Nếu Y ⊆ X Thì X → Y
Tăng trưởng: Nếu X → Y Thì XZ → YZ
Bắc cầu: Nếu X → Y và Y → Z Thì X → Z
```

• 3 luật trên có thể suy diễn ra các luật sau:

```
Hợp:

Giả bắc cầu:
Phân rã:

Nếu X → Y và X → Z, Thì X → YZ
Nếu X → Y và YZ → W, Thì XZ → W
Nếu X → YZ, Thì X → Y và X → Z
```

$$X \rightarrow Y \quad va \quad X \rightarrow Z <=> \quad X \rightarrow YZ$$

### Sử dụng hệ tiên đề Armstrong

- Sử dụng hệ tiên đề Armstrong để suy diễn một phụ thuộc hàm mới từ một tập các phụ thuộc hàm cho trước
- Ví dụ: Cho quan hệ R với tập PTH F như sau:

$$F= \{ A \rightarrow C \\ B \rightarrow H \\ C,G \rightarrow H,I \\ A,B \rightarrow I \}$$

Chứng minh rằng AG → I được suy diễn từ F

```
- Ta có: A → C

C,G → H,I => AG → H, I (tựa bắc cầu)

=> AG → I (phân rã)

=> Vậy AG → I được suy diễn từ F
```

#### Các tính chất của PTH

```
Nếu Y \subseteq X Thì X \to Y
F1:
            Phản xạ:
F2:
            Bắc cầu:
                                   Nếu X \rightarrow Y và Y \rightarrow Z Thì X \rightarrow Z
                                   Nếu X \rightarrow Y Thì XZ \rightarrow YZ
F3:
            Tăng trưởng:
            Giả bắc cầu: Nếu X \rightarrow Y và YZ \rightarrow W, Thì XZ \rightarrow W
F4:
F5:
            Phản xa chắt: X \rightarrow X
            Mở rộng vế trái, thu hẹp vế phải
F6:
                               Nếu X → Y thì XZ → Y\W với Z, W \subseteq U
                                   Nếu X \rightarrow Y và Z \rightarrow W thì XZ \rightarrow YW
F7: Cộng tính đầy đủ:
      Mở rộng vế trái:
                                  Nếu X \rightarrow Y thì XZ \rightarrow Y
F9:
                                   Nếu X \rightarrow Y và X \rightarrow Z, Thì X \rightarrow YZ
      Hợp:
F10 : Phân rã :
                                   Nếu X \rightarrow YZ, Thì X \rightarrow Y và X \rightarrow Z
F11 : Tích lũy
                                   Nếu X \rightarrow YZ, Z \rightarrow AW thì X \rightarrow YAW
```

### Bao đóng (Closure)

- Bao đóng của tập các PTH
- Bao đóng của tập các thuộc tính

### Bao đóng của tập các PTH

- Cho lược đồ quan hệ R(U), U={ A1, A2, ..., An}
- F là tập PTH trên R
- Bao đóng của F, ký hiệu F+ bao gồm:
  - F
  - Và các PTH được suy diễn từ F
- F gọi là đầy đủ nếu F = F<sup>+</sup>
- Trên thực tế, việc tính F+ khó thực hiện vì có thể dẫn đến sự bùng nổ tổ hợp
  - => Thay vào đó, ta sẽ xét xem một PTH dạng X → Y có thuộc F+ hay không, nghĩa là X → Y được suy diễn từ F không ?

### Bao đóng của tập các PTH

- Ví dụ:
- Cho F= { AB → C
   BC → D
   D → E, G }
  - Chứng minh rằng AB → E ∈ F<sup>+</sup>
  - Ta có: AB → C BC → D => AB → D (tựa bắc cầu) và ta có D → E,G => D → E (phân rã) => AB → E (bắc cầu)

=> Vậy AB → E ∈ F<sup>+</sup> hay AB → E được suy diễn từ F

### Bao đóng của tập thuộc tính

- Cho lược đồ quan hệ R(U), U = {A1, A2, ..., An}
- X là tập các thuộc tính trên U, F là tập các PTH trên R
- Bao đóng của tập thuộc tính X đối với F, ký hiệu X+ bao gồm tập các thuộc tính PTH vào X, nghĩa là:
  - $-X^{+}=\{A\in U\mid X\rightarrow A\in F^{+}\}$
- Nhận xét:
  - Làm thế nào biết được một PTH X → Y có được suy diễn từ F không ?
    - $X \rightarrow Y \in F^+ <=> X^+ \supseteq Y$
  - Nếu X+ = U thì X là siêu khóa của R

### Thuật toán tìm X+

- Dữ liệu vào: lược đô quan hệ R với tập thuộc tính U, tập PTH F và X ⊆ U
- Dữ liêu ra: X+
- Giải thuật:
  - Bước 1: X⁺ = X
  - Bước 2: Nếu tồn tại  $A \rightarrow B \in F$  và  $A \subseteq X^+$  thì :

$$X^+ = X^+ \cup B$$

- Lặp lại bước 2 cho đến khi không thể thêm thuộc tính cho X<sup>+</sup> hoặc tất cả các PTH đã được xét.
- Bước 3 : Kết quả là X<sup>+</sup>

### Thuật toán tìm X+

Ví dụ: cho tập PTH

$$F= \{ A \rightarrow C \\ B \rightarrow H \\ C,G \rightarrow H,I \\ A,B \rightarrow I \}$$

- Tim (A)+ (AB)+
- PTH A → H có suy diễn được từ F không ?
   Hay A → H có thuộc F+ không ?
- PTH AB → CH có suy diễn được từ F không ?
   Hay AB → CH có thuộc F<sup>+</sup> không ?

Tìm (A)<sup>+</sup>

 Bước 1: khởi tao
 A<sup>+</sup>= A

 Bước 2: lặp
 A<sup>+</sup>= A + C = AC (A → C)
 A<sup>+</sup>= AC + Ø

Ta có A<sup>+</sup>= AC không chứa H
 => A → H ∉ F<sup>+</sup>

- B3:  $A^{+} = AC$ 

### Thuật toán tìm X+

Ví dụ: cho tập PTH

$$F= \{ A \rightarrow C \\ B \rightarrow H \\ C,G \rightarrow H,I \\ A,B \rightarrow I \}$$

- Tim (A)+ (AB)+
- PTH A → H có suy diễn được từ F không ? Hay A → H có thuộc F<sup>+</sup> không ?
- PTH AB → CH có suy diễn được từ F không ?
   Hay AB → CH có thuộc F<sup>+</sup> không ?

```
Tìm (AB)^+
- Bước 1

(AB)^+= AB
- Bước 2 lặp
• (AB)^+= AB + C + H + I
= ABCHI (A \rightarrow C, B \rightarrow H, AB \rightarrow I)
• AB<sup>+</sup>= ABCHI + \varnothing
- B3: AB<sup>+</sup>= ABCHI
```

Tương tự ta có: (AB)<sup>+</sup> = ABCHI

Ta có:  $(AB)^+$  = ABCHI  $\supseteq$  CH => AB  $\rightarrow$  CH  $\in$  F<sup>+</sup>

- Tập PTH tương đương
  - Hai tập PTH F và G trên cùng lược đồ quan hệ là tương đương nếu và chỉ nếu F+= G+

- Tập PTH tối tiểu: F được gọi là tối tiểu nếu F thỏa các điều kiện sau:
  - Mọi PTH trong F chỉ có một thuộc tính ở vế phải
  - Không tồn tại PTH thừa
  - Không tồn tại PTH mà vế trái của nó có thuộc tính thừa
     => Phủ tối tiểu của tập PTH F là tập F' tương đương
     với F
    - => Mọi tập PTH đều có ít nhất một tập PTH tối tiểu

- Phụ thuộc hàm thừa trong một tập các PTH
  - Một PTH X → Y được gọi là PTH thừa trong tập PTH F nếu và chỉ nếu
    - F tương đương  $F\setminus\{X \to Y\}$
    - Hay  $\{X \to Y\}$  được suy diễn từ  $F\setminus\{X \to Y\}$
  - Ví dụ:  $F=\{A \rightarrow B, B \rightarrow C, D \rightarrow BE, A \rightarrow C\}$ 
    - PTH thừa ???

=> A → C vì PTH này được suy diễn từ hai PTH A→ B và B → C

- Thuôc tính thừa ở vế trái
  - Xét PTH có dạng X<sub>i</sub>X<sub>j</sub> → Y
  - Thuộc tính X<sub>i</sub> ở vế trái của PTH X<sub>i</sub>X<sub>j</sub> → Y được gọi là thuộc tính thừa nếu trong F thay X<sub>i</sub>X<sub>j</sub> → Y bằng X<sub>j</sub> → Y thì F<sup>+</sup> vẫn không thay đổi
    - Hay  $F^+ = (F \setminus \{X_i X_j \to Y\} \cup \{X_j \to Y\})^+$
  - Ví dụ:  $F=\{A \rightarrow B, AB \rightarrow C, D \rightarrow BE\}$ 
    - Thuộc tính thừa ???

=> B trong PTH AB  $\rightarrow$  C vì A $\rightarrow$  B

### PTH rút gọn tự nhiên

- Tập PTH F được gọi là rút gọn tự nhiên nếu
  - Vế phải và vế trái của mọi PTH không có thuộc tính chung
    - Nếu có thuộc tính chung thì bỏ thuộc tính chung đó ở vế phải
  - Hai PTH khác nhau có vế trái khác nhau
    - Nếu có hai PTH cùng vế trái dạng L₁ → R₁ và L₁ → R₂ thì gom hai PTH đó lại thành L₁ → R₁R₂
- Ví dụ: Tìm tập PTH rút gọn tự nhiên của F

$$F = \{AD \rightarrow CD \\ B \rightarrow H \\ C,G \rightarrow H \\ C,G \rightarrow I \\ A,B \rightarrow I \}$$

$$=> F_{TN} = \{AD \rightarrow C \\ B \rightarrow H \\ C,G \rightarrow H,I \\ A,B \rightarrow I \}$$

### Bài toán tìm phủ tối tiểu

- Các bước tìm tập PTH tối tiểu của F:
  - Tách các PTH sao cho VP có 1 thuộc tính (dùng phân rã)
  - Loại bỏ các PTH thừa
  - Loai bỏ các thuộc tính thừa ở VT
  - Tìm tập PTH rút gọn tự nhiên của F

### Ví dụ Phủ tối tiểu

Cho F={ A → B, B → C, D → BE, A → C}, tìm F tối tiểu ?
 1) Tách các PTH sao cho VP có 1 thuộc tính

- 2) Loại bỏ PHT thừa
  - Ta có: A → B và B → C => A → C (thừa => loại PTH này)
- 3) Không có thuộc tính thừa ở VT do các VT chỉ có 1 tt
- 4) F rút gọn tự nhiên

- F={ 
$$A \rightarrow B$$
,  $B \rightarrow C$ ,  $D \rightarrow BE$  }

• Tìm tập PTH tối tiểu của:

$$F=\{A\rightarrow BC,$$

$$B\rightarrow CE,$$

$$A\rightarrow E,$$

$$AC\rightarrow H,$$

$$D\rightarrow B\}$$

Tìm tập PTH tối tiểu của:

$$F=\{A \rightarrow BC, B \rightarrow CE, A \rightarrow E, AC \rightarrow H, D \rightarrow B\}$$

• B1: 
$$F = \{A \rightarrow B, A \rightarrow C, B \rightarrow C, B \rightarrow C, B \rightarrow E, A \rightarrow E, A \rightarrow E, A \rightarrow B, B \rightarrow C \Rightarrow A \rightarrow C \text{ (thừa)} A \rightarrow B, B \rightarrow E \Rightarrow A \rightarrow E \text{ (thừa)}$$

• 
$$F=\{A \rightarrow BC, B \rightarrow CE, A \rightarrow E, AC \rightarrow H, D \rightarrow B\}$$

• B3: 
$$F = \{A \rightarrow B, B \rightarrow C, B \rightarrow C, B \rightarrow E, A \rightarrow H, D \rightarrow B\}$$

F= $\{A \rightarrow B, B \rightarrow C, B \rightarrow C, B \rightarrow C, B \rightarrow E, D \rightarrow B\}$ 

Ià tập tối tiểu

Ta có AC → H có vế trái có 2 thuộc tính

Do 
$$A \rightarrow B$$
 và  $B \rightarrow C => A \rightarrow C$   
=> thuộc tính C trong AC → H thừa

- Bài 1: Cho quan hệ sau KHACHHANG(id, ten, tpho) như sau
  - Các phát biểu sau đúng hay sai:

001	Albert	Bruxelles
002	Francois	Liege
003	Brabo	Anvers
004	Albert	Anvers
005	Leon	Liege
006	Philippe	Bruxelles
007	Brabo	Anvers

Phụ thuộc hàm

- **Bài 2**: Cho R(A, B, C, D, E) và  $F = \{AB \rightarrow C, B \rightarrow D, CD \rightarrow E\}$ 
  - Tính (BC)<sup>+</sup>
  - Chứng minh rằng F⁺ chứa AB → E
  - Chứng minh rằng AB là khóa

- **Bài 2**: Cho R(A, B, C, D, E) và  $F = \{AB \rightarrow C, B \rightarrow D, CD \rightarrow E\}$ 
  - Tính (BC) $^+$  = BC
    - $(BC)^+ = BC + D = BCD (B \rightarrow D)$
    - $(BC)^+ = BCD + E = BCDE (CD \rightarrow E)$
    - $(BC)^+ = BCDE + \emptyset = BCDE$
  - Chứng minh rằng F⁺ chứa AB → E
    - (AB)+= ABCDE  $\supseteq$  E => AB  $\rightarrow$  E  $\in$  F<sup>+</sup>
  - Chứng minh rằng AB là khóa

Bài 3 :Cho quan hệ

**KhamBenh**(idBN, tenBN, idBsi, tenBsi, ngaykham, loaibenh) Một bác sĩ có thể khám nhiều bệnh liên quan nhiều bệnh nhân

 Xác định tất cả các PTH của quan hệ KhamBenh biét rằng mỗi bệnh nhân có thể khám nhiều lần trong ngày nhưng không quá 1 lần với cùng bác sĩ.

#### • Bài 4:

Cho F = {AB  $\rightarrow$  C, A  $\rightarrow$  D, D  $\rightarrow$  E, AC  $\rightarrow$  B} Hai PTH AB  $\rightarrow$  E và D  $\rightarrow$  C có được suy diễn từ F hay không?