

CSDL – BÀI TẬP LUYỆN TẬP

Phụ thuộc hàm và Dạng chuẩn

Phụ thuộc hàm: $X \rightarrow Y$

(X, Y: Hai tập thuộc tính trên cùng lược đồ quan hệ R

Tập thuộc tính X xác định tập thuộc tính Y trên R

Tập Y phụ thuộc vào X trên R

Một giá trị tập X cho ra một giá trị duy nhất trên tập Y)

Ví dụ trên lược đồ QLDT: GiaoVien có các PTH:

mãgv \rightarrow họten, gvqlcm

gvqlcm \rightarrow mãgv: có phải PTH, nếu có quy tắc nghiệp vụ là mỗi người chỉ quản lý 1 người khác

Khoá: Siêu khoá (superkey), Khoá (ứng viên, key/candidate key), Khoá chính (primary key)

- Siêu khoá: tập thuộc tính bất kỳ xác định duy nhất 1 dòng dữ liệu (xác định tất cả thuộc tính của quan hệ),

R(ABCDE) cho f: $AB \rightarrow ABCDE = R$ thì AB là một siêu khoá

- Khoá ứng viên (khoá): là siêu khoá nhỏ nhất

R(ABCDE), $F = \{f_1: AB \rightarrow CD, f_2: B \rightarrow E, f_3: C \rightarrow D\}$

SK: ABCDE, ABC ($\rightarrow ABCED = R$), ABE, ...

K (sẽ chứa AB): $AB \rightarrow ABCDE = R$: AB là tập thuộc tính nhỏ nhất suy ra toàn bộ R (AB là siêu khoá nhỏ nhất \rightarrow AB là khoá ứng viên của quan hệ)

- Cách tìm khoá ứng viên:

- B1: Tìm tập thuộc tính chỉ nằm bên vế trái ($LHS = AB$) hoặc không xuất hiện trong tập PTH (\rightarrow Bắt buộc phải có trong khoá vì không có thuộc tính nào xác định được LHS ngoại trừ chính nó)
- B2: Tìm tập thuộc tính chỉ nằm bên vế phải ($RHS = DE$) (\rightarrow Hiển nhiên không phải là thuộc tính khoá vì không xác định được thuộc tính nào cả)
- B3: Kiểm tra thử LHS đủ là khoá chưa? OK: LHS là khoá duy nhất.
- B4: Tìm tập thuộc tính trung gian ($I = C$) là tập vừa nằm bên trái vừa nằm bên phải của bất kỳ PTH nào trong tập F.

Thuộc tính Khoá và không khoá:

- Thuộc tính khoá: thuộc ít nhất 1 khoá ứng viên
- Thuộc tính không khoá: không thuộc bất kỳ khoá ứng viên nào cả
- VD1: $R(\underline{ABCDE})$, $F = \{f1: AB \rightarrow CD, f2: B \rightarrow E, f3: C \rightarrow D\}$
 - TT khoá: A, B
 - TT không khoá: CDE
- VD2: $R(ABCDEF)$, $K = \{AB, BC, BD\}$
 - TT khoá: A, B, C, D
 - TT không khoá: E, F

Dạng chuẩn: các tiêu chí để xác định mức độ chuẩn hoá (mức độ tổ chức dữ liệu) của một lược đồ CSDL, mức cao: ít trùng lặp thông tin; mức thấp: trùng lặp nhiều (\rightarrow các bất thường trong việc thêm/xoá/sửa).

Mức độ dạng chuẩn	Tiêu chí	Vi phạm <i>Các tập thuộc tính gây vi phạm chính là tập làm trùng lặp thông tin.</i>	Nguyên tắc chuẩn hoá
DC1 (1NF)	Không đa trị, không kết hợp.	Có ít nhất một hoặc một tập thuộc tính bị đa trị hoặc kết hợp.	Đa trị: <ul style="list-style-type: none">• C1: Đem tập trùng lặp đa trị tách ra, kéo theo khoá của quan hệ cũ.• C2: Đem các dòng dữ liệu vi phạm đa trị tách ra thành nhiều dòng. Kết hợp: <ul style="list-style-type: none">• Tách thành phần con thành cột khác nhau.

Mức độ dạng chuẩn	Tiêu chí	Vi phạm <i>Các tập thuộc tính gây vi phạm chính là tập làm trùng lặp thông tin.</i>	Nguyên tắc chuẩn hoá
DC2 (2NF)	Không tồn tại phụ thuộc riêng phần giữa thuộc tính không khoá và bất kỳ khoá ứng viên nào.	Có ít nhất một thuộc tính không khoá phụ thuộc một phần vào khoá hay: Tìm thấy ít nhất một $X \rightarrow Y$ trong đó: X là phần con của khoá và Y là không khoá.	<ul style="list-style-type: none"> Tách tất cả các trường hợp $X \rightarrow Y$ vi phạm (tức là $X \rightarrow Y$ có X là con của khoá và Y là không khoá) ra (giữ lại X ở bảng cũ để kết bảng).
DC3 (3NF)	Không tồn tại phụ thuộc bắc cầu giữa không khoá và bất kỳ khoá ứng viên nào. → Các tập thuộc tính không khoá suy ra lẫn nhau.	<p>Có ít nhất một thuộc tính không khoá phụ thuộc vào một hay một tập thuộc tính không khoá khác: Tìm thấy ít nhất một $X \rightarrow Y$ trong đó: X không chứa khoá và Y là không phải thuộc tính không khoá.</p> <p>$R(\underline{AB}/\underline{CDE})$ $F = \{$ $f1: AB \rightarrow C;$ $f2: AB \rightarrow D;$ $f3: AB \rightarrow E;$ $f4: C \rightarrow A;$ $f5: C \rightarrow B;$ $f6: E \rightarrow A \}$</p> <p>Đạt DC2 vì không có PTH có vế trái là con khoá và vế phải là thuộc tính không khoá.</p> <p>Đạt DC3 vì không có PTH vế trái không chứa khoá và vế phải là thuộc tính không khoá.</p>	<ul style="list-style-type: none"> Tách tất cả các trường hợp $X \rightarrow Y$ vi phạm ra (giữ lại X ở bảng cũ để kết bảng).

Mức độ dạng chuẩn	Tiêu chí	Vi phạm <i>Các tập thuộc tính gây vi phạm chính là tập làm trùng lặp thông tin.</i>	Nguyên tắc chuẩn hoá
DCBC (BCNF)	Mọi phụ thuộc hàm về trái phải là siêu khoá (chứa bất kỳ khoá ứng viên nào).	Tìm thấy ít nhất một $X \rightarrow Y$ trong đó: X không chứa khoá. Ví dụ trên vi phạm BC vì f6: $E \rightarrow A$ có E không phải là siêu khoá.	Đem $E \rightarrow A$ tách ra bảng mới dữ lại E để kết bảng và xác định lại tập PTH cho từng quan hệ mới.

1. Cho bảng dữ liệu SINHVIÊN(MãSV, ThôngTin, KếtQuả). Trong đó thuộc tính ThôngTin bao gồm {TênSV, Ngày Sinh, Email, (Điện thoại1, Điện thoại 2, ...)}, thuộc tính KếtQuả bao gồm {Môn học, Lần học, Điểm} lưu trữ môn học sinh viên đã tham gia học và điểm của môn học đó cho từng lần khác nhau.

- Giải thích tại sao danh sách SINHVIÊN không đạt dạng chuẩn 1. Hãy thiết kế lại để được quan hệ SINHVIÊN đạt dạng chuẩn 1 và vẽ lược đồ quan hệ.
- Từ kết quả câu a), chỉ ra các điểm trùng lặp thông tin? Quan hệ có đạt dạng chuẩn 2 hay 3 không? Giải thích. Hãy thiết kế lại để được bảng SINHVIÊN đạt dạng chuẩn 3 và vẽ lược đồ quan hệ.

Gợi ý đáp án:

- Không đạt DC1 vì trường thuộc tính ThôngTin và KếtQuả gây trùng lặp trên tập đa trị và tập thuộc tính kết hợp dẫn đến ngữ nghĩa cột thuộc tính không rõ ràng.

MãSV	ThôngTin (Đa trị, Kết hợp)	KếtQuả (Đa trị, Kết hợp)
001	{Lan, 12/03/2024, vl@gmail, (02433512, 0923231321)}	{CSDL, 1, 10} {CTDL, 1, 8}
002	{Phong, 12/03/2024, np@gmail, (0923332321, 0923231321)}	{CSDL, 1, 7} {CTDL, 2, 8} {HDH, 1, 8}

Đề xuất quan hệ đạt DC1-Giải pháp 1: Phân rã đa trị thành nhiều dòng và kết hợp thành nhiều cột

MãSV	TênSV	NgàySinh	Email	ĐiệnThoại	MônHọc	LầnHọc	Điểm
001	Lan	12/03/2024	vl@gmail	02433512	CSDL	1	10
001	Lan	12/03/2024	vl@gmail	02433512	CTDL	1	8
001	Lan	12/03/2024	vl@gmail	0923332321	CSDL	1	10
001	Lan	12/03/2024	vl@gmail	0923231321	CTDL	1	8
...							

SinhVien (MãSV, TênSV, NgàySinh, Email, ĐiệnThoai, MônHoc, LầnHoc, Điểm)

$F = \{ f1: \text{MãSV} \rightarrow \text{TênSV, NgàySinh, Email};$
 $f2: \text{MãSV, MônHoc, LầnHoc} \rightarrow \text{Điểm};$
 $f3: \text{ĐiệnThoai} \rightarrow \text{MãSV} \}$

Lược đồ SinhVien đạt DC1 vì tất cả cột là giá trị nguyên tố (không đa trị, không kết hợp)

b) Các điểm trùng lặp thông tin: tìm phụ thuộc hàm gây vi phạm dạng chuẩn cao \rightarrow tập thuộc tính liên quan là chỗ bị trùng lặp thông tin.

- DC2: Vi phạm DC2 vì cả f1 và f2 đều có vế trái là con của khoá và vế phải là thuộc tính không khoá; VD: TênSV là thuộc tính không khoá phụ thuộc vào một phần khoá là MãSV. \rightarrow Trùng lặp thông tin trên tập thuộc tính (MãSV, TênSV, NgàySinh, Email) và (MãSV, MônHoc, LầnHoc, Điểm).

- Làm sao chuẩn hoá lược đồ đạt DC2? \rightarrow Tách thuộc riêng phần giữa không khoá và khoá (f1, f2). Nếu lược đồ CSDL có tập lược đồ quan hệ R1, R2, R3 thì lược đồ CSDL được coi là đạt DC2 khi có ít nhất 1 lược đồ R1, R2 và R3 đạt DC2; VÀ các lược đồ còn lại đạt từ chuẩn 2 trở lên.

- B1: Tách theo f1

SinhVien ---tách theo f1---

SinhVien1(MãSV, TênSV, NgàySinh, Email)

DCBC vì các PTH vế trái là khoá

$F1 = \{ f1: \text{MãSV} \rightarrow \text{TênSV, NgàySinh, Email} \}$

SinhVien2(MãSV, ĐiệnThoai, MônHoc, LầnHoc, Điểm),

DC1 vì f2 có Điểm là không khoá phụ thuộc một phần khoá (MãSV, MônHoc, LầnHoc)

$F2 = \{ f2: \text{MãSV, MônHoc, LầnHoc} \rightarrow \text{Điểm}; f3: \text{ĐiệnThoai} \rightarrow \text{MãSV} \}$

- B2: Tách SinhVien2 theo f2:

KếtQuả(MãSV, MônHoc, LầnHoc, Điểm) DCBC vì các PTH vế trái là khoá

$F3 = \{ f2: \text{MãSV, MônHoc, LầnHoc} \rightarrow \text{Điểm} \}$

SinhVien2(MãSV, ĐiệnThoai, MônHoc, LầnHoc)

SV_DT(MãSV, ĐiệnThoai) DCBC vì các PTH vế trái là khoá

$F4 = \{ f3: \text{ĐiệnThoai} \rightarrow \text{MãSV} \}$

Kết luận: từ lược đồ kết quả của **câu a** như sau:

SinhVien (MãSV, TênSV, NgàySinh, Email, ĐiệnThoai, MônHoc, LầnHoc, Điểm)

$F = \{ f1: \text{MãSV} \rightarrow \text{TênSV, NgàySinh, Email};$
 $f2: \text{MãSV, MônHoc, LầnHoc} \rightarrow \text{Điểm};$
 $f3: \text{ĐiệnThoai} \rightarrow \text{MãSV} \}$

Lược đồ đạt DC1 sau sau khi tách theo các phụ thuộc hàm vi phạm (f1, f2, f3) ta có lược đồ CSDL mới bao gồm 3 lược đồ quan hệ con:

SinhVien(MãSV, TênSV, NgàySinh, Email); $F1 = \{ f1: \text{MãSV} \rightarrow \text{TênSV, NgàySinh, Email} \}$

KếtQuả(MãSV, MônHoc, LầnHoc, Điểm) $F3 = \{ f2: \text{MãSV, MônHoc, LầnHoc} \rightarrow \text{Điểm} \}$

SV_DT (MãSV, ĐiệnThoai) $F4 = \{ f3: \text{ĐiệnThoai} \rightarrow \text{MãSV} \}$

Cả 3 lược đồ con **SinhVien**, **KếtQuả** và **SV_DT** đều đạt DCBC \rightarrow Lược đồ CSDL mới đạt DCBC.

2. Xét bảng NHÂNVIÊN sau, với TÊN NV là khóa chính:

TÊN NV	DỰ ÁN	PHÒNG BAN
Anh	Phân công việc	Phòng thí nghiệm mạng
Bình	Định tuyến mạng	Phòng thí nghiệm mạng
Duy	Thiết kế khảo sát	Phòng quản lý nhân sự
Hạ	Phân công việc	Phòng thí nghiệm mạng
Tâm	Thiết kế khảo sát	Phòng quản lý nhân sự
Minh	Định tuyến mạng	Phòng thí nghiệm mạng
Hùng	Dịch vụ phân tích tin tức	Phòng quản lý nhân sự

- Chỉ ra các điểm trùng lặp thông tin và giải thích vì sao NHÂNVIÊN không đạt DC3?
- Hãy chuẩn hóa NHÂNVIÊN thành dạng chuẩn 3.

3. Một hệ thống thông tin quản lý sinh viên bao gồm các quan hệ sau đây:

SINHVIÊN (MSSV, HỌTÊN, SỐCMND, ĐỊACHỈ, NGÀY SINH, TÊNNGÀNH)

F1 = {

f1: MSSV → HỌTÊN, SỐCMND, ĐỊACHỈ, NGÀY SINH, TÊNNGÀNH;

f2: SỐCMND → MSSV }

NGÀNH (TÊNNGÀNH, NĂM, SLTĐ, SLSV, MÃPHG, TÊNPHG, TÒANHÀ)

F2 = {

f3: TÊNNGÀNH, NĂM → SLTĐ;

f4: TÊNNGÀNH → SLSV, MÃPHG;

f5: MÃPHG → TÊNPHG, TÒANHÀ }

KQ (MÃMH, TÊN MH, MÔTẢ, SỐTC, HK, NĂM, TÊN GV, ĐIỀMMÔN, MSSV)

F3 = {

f6: MÃMH → TÊN MH, MÔTẢ, SỐTC;

f7: MÃMH, HK, NĂM → TÊN GV;

f8: MÃMH, HK, NĂM, MSSV → ĐIỀMMÔN }

Gợi ý đáp án:

a) Xác định khoá chính cho từng quan hệ.

- SinhViên có khoá: (MSSV), (SỐCMND) do MSSV suy ra toàn bộ thuộc tính của quan hệ và số cmnd suy ra mssv nên cũng suy ra toàn bộ thuộc tính của quan hệ. Hai khoá (ứng viên) là MSSV và SốCMND, MSSV được chọn là khoá chính.
- Ngành có khoá: (TÊNNGÀNH, NĂM) do TÊNNGÀNH, NĂM suy ra tất cả các thuộc tính khác của quan hệ. TÊNNGÀNH, NĂM → TÊNNGÀNH, NĂM, SLTĐ, SLSV, MÃPHG, TÊNPHG, TÒANHÀ.
- KếtQuả có khoá: (MÃMH, HK, NĂM, MSSV) do tập này suy ra toàn bộ thuộc tính của quan hệ KếtQuả.

- c) Chỉ ra các trùng lặp thông tin và cho ví dụ về các dị thường dữ liệu có thể gặp phải.
- SinhViên không có trùng lặp vì về trái các PTH đều là khoá nên không có thuộc tính không khoá phụ thuộc bắc cầu riêng phần vào khoá.
 - Ngành trùng lặp: SLSV và MÃPHG là không khoá phụ thuộc vào một phần khoá (f4); và TÊNPHG, TÒANHÀ phụ thuộc bắc cầu vào Khoá (f5) → Trùng lặp thông tin trên các tập thuộc tính (TÊNNGÀNH, SLSV, MÃPHG); và (MÃPHG, TÊNPHG, TÒANHÀ).
 - KếtQuả trùng lặp trên các tập thuộc tính của f6 và f7 vì về trái là phần con của khoá còn về phải là thuộc tính không khoá → thuộc tính không khoá phụ thuộc riêng phần vào khoá.
- d) Xác định dạng chuẩn của lược đồ CSDL bên trên. Giải thích.
- SinhVien: DCBC vì về trái các PTH là khoá.
 - Ngành đạt DC1 vi phạm DC2 do PTH f4 có thuộc tính không khoá phụ thuộc không đầy đủ (riêng phần) vào khoá.
 - KếtQuả đạt DC1 vi phạm DC2 do PTH f6 và f7 có thuộc tính không khoá phụ thuộc không đầy đủ (riêng phần) vào khoá.
- Cả lược đồ CSDL đạt DC1.
- e) Đề xuất lược đồ mới đạt tối thiểu dạng chuẩn 3. Giải thích.
- **SINHVIÊN** (MSSV, HỌTÊN, SỐCMND, ĐỊACHỈ, NGÀY SINH, TÊNNGÀNH) **giữ nguyên vì đã đạt DCBC.**
 $F1 = \{f1: MSSV \rightarrow HỌTÊN, SỐCMND, ĐỊACHỈ, NGÀY SINH, TÊNNGÀNH;$
 $f2: SỐCMND \rightarrow MSSV \}$
 - **NGÀNH:**
B1: tách theo f4 để nâng lên chuẩn 2 được quan hệ còn NGÀNH1 và NGÀNH_CHỈTIÊU.
NGÀNH1 (TÊNNGÀNH, SLSV, MÃPHG, TÊNPHG, TÒANHÀ) **DC2, vi phạm DC3 vì f5 có tt không khoá phụ thuộc bắc cầu vào khoá.**
 $F2 = \{ f4: TÊNNGÀNH \rightarrow SLSV, MÃPHG;$
 $f5: MÃPHG \rightarrow TÊNPHG, TÒANHÀ \}$

NGÀNH_CHỈTIÊU(TÊNNGÀNH, NĂM, SLTĐ) **DCBC** vì các PTH đều có về trái là khoá.

$$F3 = \{ f3: TÊNNGÀNH, NĂM \rightarrow SLTĐ \}$$

B2: tách NGÀNH1 theo f5 để nâng lên chuẩn 3 được hai con hệ con NGÀNH2 và PHÒNG.

NGÀNH2 (TÊNNGÀNH, SLSV, MÃPHG) **DCBC** vì các PTH đều có về trái là khoá.

$$F4 = \{ f4: TÊNNGÀNH \rightarrow SLSV, MÃPHG \}$$

PHÒNG (MÃPHG, TÊNPHG, TOÀNHÀ) **DCBC** vì các PTH đều có về trái là khoá.

$$F5 = \{ f5: MÃPHG \rightarrow TÊNPHG, TÒANHÀ \}$$

- **KẾTQUẢ**: tương tự tách theo các PTH làm vi phạm dạng chuẩn cao (f6, f7) sẽ ra được các phụ thuộc hàm con:

MÔN HỌC (MÃMH, TÊNMH, MÔTẢ, SỐTC), $F6 = \{f6\}$ **DCBC**

PHÂN CÔNG (MÃMH, HK, NĂM, TÊN GV), $F7 = \{f7\}$ **DCBC**

KẾTQUẢ (MÃMH, HK, NĂM, MSSV, ĐIỀMMÔN), $F8 = \{f8\}$ **DCBC**

Tóm tắt: khi tách quan hệ R(XYZK) theo phụ thuộc hàm f1: $X \rightarrow Y$ thì hai quan hệ mới bao gồm: $R1(XYZ)$ nếu X dẫn đến Y và Y dẫn được đến Z và $R2(XK)$.

Lược đồ CSDL mới:

- **SINHVIÊN** (MSSV, HỌTÊN, SỐCMND, ĐỊACHỈ, NGÀY SINH, TÊNNGÀNH), $F1 = \{f1, f2\}$ **DCBC**
 - **NGÀNH_CHỈTIÊU**(TÊNNGÀNH, NĂM, SLTĐ), $F3 = \{f3\}$ **DCBC**
 - **NGÀNH2** (TÊNNGÀNH, SLSV, MÃPHG), $F2 = \{f4\}$ **DCBC**
 - **PHÒNG** (MÃPHG, TÊNPHG, TOÀNHÀ), $F5 = \{f5\}$ **DCBC**
 - **MÔN HỌC** (MÃMH, TÊNMH, MÔTẢ, SỐTC), $F6 = \{f6\}$ **DCBC**
 - **PHÂN CÔNG** (MÃMH, HK, NĂM, TÊN GV), $F7 = \{f7\}$ **DCBC**
 - **KẾTQUẢ** (MÃMH, HK, NĂM, MSSV, ĐIỀMMÔN), $F8 = \{f8\}$ **DCBC**
- ➔ Lược đồ CSDL mới đạt DCBC do tất cả lược đồ con đều đạt DCBC.

4. Hãy xác định các khoá và dạng chuẩn của từng Q_i, F_i , được cho như sau:

a) $Q_1(ABCDEFGH)$, $F_1 = \{A \rightarrow H; AB \rightarrow C; BC \rightarrow D; G \rightarrow B\}$

b) $Q_2(ABCSXYZ)$, $F_2 = \{S \rightarrow A; AX \rightarrow B; S \rightarrow B; BY \rightarrow C; CZ \rightarrow X\}$

c) $Q_3(ABCD)$, $F_3 = \{A \rightarrow B; BC \rightarrow D; D \rightarrow A\}$

Chuẩn hoá các lược đồ bên trên về DC3 (3NF) hoặc DCBC (BCNF).