

Chương 6

Dạng chuẩn và chuẩn hóa

Dạng chuẩn và chuẩn hóa

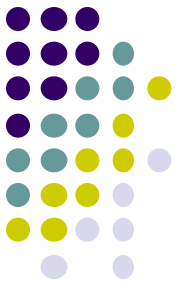


- 6.1. Sự cần thiết phải chuẩn hóa
- 6.2 Các dạng chuẩn của quan hệ
- 6.3. Chuẩn hóa quan hệ
- 6.4 Chuẩn hóa trong thực tế



Dạng chuẩn

Sự cần thiết phải chuẩn hóa



- **Do thiết kế kém sẽ gây nguy hiểm cho CSDL.**
- Trùng lặp thông tin: không có khả năng trình bày thông tin một cách chắc chắn.
- VD: Cho một lược đồ quan hệ dùng để ghi nhận giáo viên và lớp giảng dạy của giáo viên
- GIANGDAY(MONHOC, SOTIET, LOP, GV, HV, DC)
- Các phụ thuộc hàm: MONHOC \rightarrow SOTIET; MONHOC, LOP \rightarrow GV; GV \rightarrow HOCVI, DC. Có tình trạng dl như sau:

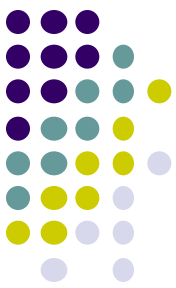
MONHOC	SOTIET	LOP	GV	HOCVI	DC
CSDL	60	CNTT1	D.T.Hien	TS	HN
CSDL	60	CNTT2	D.T.Hien	TS	HN
CTDL	45	CNTT1	H.V.A	ThS	HP
CTDL	45	CNTT2	H.V.A	ThS	HP

Sự cần thiết phải chuẩn hóa...



- Do có phụ thuộc hàm MONHOC \rightarrow SOTIET nên số tiết của dòng thứ 2 và dòng thứ 4 gây nên trùng lặp thông tin.
- Do phụ thuộc hàm GV \rightarrow HOCVI, DC nên học vị và địa chỉ của dòng thứ 2 và dòng thứ 4 gây nên trùng lặp thông tin.
- Các dl gây trùng lặp thông tin là các dl có thể suy đoán được một cách chắc chắn và duy nhất từ phụ thuộc hàm.

Phân rã



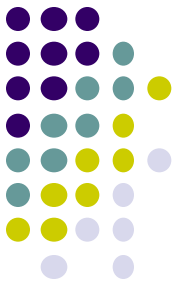
- Từ một lược đồ quan hệ kém phân rã thành những lược đồ quan hệ tốt hơn.
- **Ví dụ:** Phân rã lược đồ quan hệ GIANGDAY thành hai lược đồ TKB và GV
- TKB(MONHOC, SOTIET, LOP)
- GV(LOP, GV, HOCVI, DC)
- Tình trạng dữ liệu của hai lược đồ trên như sau:

$$\text{TKB} = \Pi_{\text{MONHOC, SOTIET, LOP}}(\text{GIANGDAY}) \quad \text{GV} = \Pi_{\text{LOP, GV, HOCVI, DC}}(\text{GIANGDAY})$$

MONHOC	SOTIET	LOP
CSDL	60	CNTT1
CSDL	60	CNTT2
CTDL	45	CNTT1
CTDL	45	CNTT2

LOP	GV	HOCVI	DC
CNTT1	D.T.Hien	TS	HN
CNTT2	D.T.Hien	TS	HN
CNTT1	H.V.A	ThS	HP
CNTT2	H.V.A	ThS	HP

Phân rã...



- Những rắc rối xảy ra
- Để trả lời câu hỏi “Cho biết thông tin của giáo viên dạy CSDL của CNTT1” ta phải kết nối tự nhiên hai quan hệ TKB và GV.

MONHOC	SOTIET	LOP	GV	HOCVI	DC
CSDL	60	CNTT1	D.T.Hien	TS	HN
CSDL	60	CNTT1	H.V.A	ThS	HP
CSDL	60	CNTT2	D.T.Hien	TS	HN
CSDL	60	CNTT2	H.V.A	ThS	HP
CTDL	45	CNTT1	H.V.A	ThS	HP
CTDL	45	CNTT1	D.T.Hien	TS	HN
CTDL	45	CNTT2	H.V.A	ThS	HP
CTDL	45	CNTT2	D.T.Hien	TS	HN

- Ta thấy hai giáo viên dạy môn CSDL của lớp CNTT1 trong khi thông tin ban đầu chỉ có N.V.A
- → **Vấn đề này gọi là phân rã không bảo toàn thông tin.**

Phân rã...



- Xét phụ thuộc hàm trên lược đồ phân rã:
- $TKB(MONHOC, SOTIET, LOP) \quad MONHOC \rightarrow SOTIET$
- $GV(LOP, GV, HOCVI, DC) \quad GV \rightarrow HOCVI, DC$
- Từ hai phụ thuộc hàm trên ta không thể suy ra được phụ thuộc hàm $MONHOC, LOP \rightarrow GV$.
- Như vậy, hai phụ thuộc hàm trên không đảm bảo kiểm tra các ràng buộc toàn vẹn do 3 phụ thuộc hàm ban đầu gây ra.
- **→ Vấn đề này gọi là phân rã không bảo toàn phụ thuộc hàm.**
- *Phải có quy tắc phân rã để không vi phạm hai vấn đề trên.*

Phân rã bảo toàn thông tin

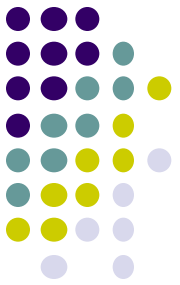


- Cho lược đồ quan hệ Q . Ta có định nghĩa sau:
- Tập $\{Q_1, Q_2, \dots, Q_n\}$ là một phân rã của Q nếu:
 - $Q = Q_1 \cup Q_2 \cup \dots \cup Q_n$
- Một cách tổng quát TQ là một quan hệ của Q thì:
 - $TQ \subseteq \Pi_{R_1}(TQ) \bowtie \Pi_{R_2}(TQ) \bowtie \dots \bowtie \Pi_{R_n}(TQ)$
- Phân rã thông tin trên bảo toàn thông tin nếu:
 - $TQ = \Pi_{R_1}(TQ) \bowtie \Pi_{R_2}(TQ) \bowtie \dots \bowtie \Pi_{R_n}(TQ)$

Phân rã bảo toàn thông tin...

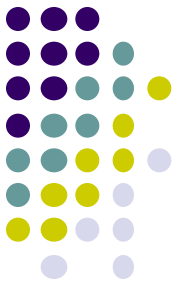


- **Điều kiện để phân rã bảo toàn thông tin**
- Cho Q và F là tập phụ thuộc hàm, Q_1 và Q_2 là một phân rã bảo toàn thông tin trên Q nếu thoả một trong hai phụ thuộc hàm sau:
 - $Q_1 \cap Q_2 \rightarrow Q_1 \setminus Q_2$ hoặc $Q_1 \cap Q_2 \rightarrow Q_2 \setminus Q_1$
- Vì vậy nếu $X \rightarrow Y \in F^+$ thì phân rã sau sẽ bảo toàn thông tin
- $Q_1(XY), Q_2(Q-Y)$
- Thật vậy, vì Q_1 có $X \rightarrow Y$ và $Q_1 \cap Q_2 = X$, $Q_1 \setminus Q_2 = Y$ do đó $Q_1 \cap Q_2 \rightarrow Q_1 \setminus Q_2$



Phân rã bảo toàn thông tin...

- **VD:** Cho $R(ABCDE)$, $F=\{AB \rightarrow C, C \rightarrow D, D \rightarrow AE\}$
- Kiểm tra xem các phép tách có bảo toàn thông tin không?
 - $R1(ABD), R2(ACE)$
 - $R1(ABC), R2(ABDE)$
 - $R1(ADE), R2(DEBC)$
- **VD:** Lược đồ GIANGDAY nếu phân rã thành hai lược đồ sau thì bảo toàn thông tin.
- $Q1(MONHOC, SOTIET, LOP, GV), Q2(GV, HOCVI, DC)$
- vì $Q1 \cap Q2 = GV$, $Q2 - Q1 = HOCVI, DC$ mà $GV \rightarrow HOCVI, DC$



Phân rã bảo toàn thông tin...

- Phương tiện để kiểm tra phân rã bảo toàn thông tin:
- Dùng kỹ thuật Tableau: là một bảng T như sau:

	1	2	...	m
Q1				
Q2				
...				
Q _n				

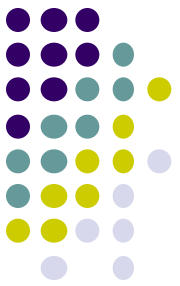
- m cột cho m thuộc tính của Q
- n dòng cho n quan hệ phân rã
- $(i,j) = a_j$ nếu Q_i có chứa thuộc tính thứ j của Q
- $= b_k$ nếu ngược lại, k bắt đầu bằng 1 và tăng dần.



Phân rã bảo toàn thông tin...

- Áp dụng luật phụ thuộc hàm để biến đổi bảng T thành T^* theo thuật toán sau:
- **While ($X \rightarrow A \in F$)**
- { Chọn dòng W1 và W2 sao cho $W1.X = W2.X$
- If ($W1.A \neq W2.A$)
- { Nếu $W1.A = a_j$ và $W2.A = b_k$ thay $W2.A$ bằng $W1.A$
- Nếu $W1.A = b_k$ và $W2.A = a_j$ thay $W1.A$ bằng $W2.A$
- Nếu $W1.A = b_j$ và $W2.A = b_k$ thay $W2.A$ bằng $W1.A$
- }
- }
- Cuối cùng xem bảng kết quả nếu trong bảng xuất hiện hàng gồm các kí hiệu $a_1, a_2, a_3, \dots, a_m$ thì phân rã bảo toàn thông tin.

Phân rã bảo toàn thông tin...



- Ví dụ: Q1(MONHOC, SOTIET, LOP, GV), Q2(GV, HOCVI, DC)
- $F = \{MONHOC \rightarrow SOTIET; MONHOC, LOP \rightarrow GV; GV \rightarrow HOCVI, DC\}$

T

	MONHOC	SOTIET	LOP	GV	HOCVI	DC
Q1	a ₁	a ₂	a ₃	a ₄	b ₁	b ₂
Q2	b ₃	b ₄	b ₅	a ₄	a ₅	a ₆

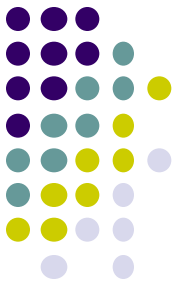
- Từ $GV \rightarrow HOCVI, DC$ ta thay thế b₁ thành a₅ và b₂ thành a₆

T*

	MONHOC	SOTIET	LOP	GV	HOCVI	DC
Q1	a ₁	a ₂	a ₃	a ₄	a ₅	a ₆
Q2	b ₃	b ₄	b ₅	a ₄	a ₅	a ₆

- Ta được dòng thứ nhất toàn a_j nên phân rã trên **bảo toàn thông tin**
- Ví dụ: cho R(ABCDE), $F = \{A \rightarrow BC, B \rightarrow C, C \rightarrow D, DE \rightarrow C, CE \rightarrow A\}$
- Kiểm tra tách có bảo toàn thông tin R1(AD), R2(AB), R3(BD), R4(CDE)

Phân rã bảo toàn phụ thuộc hàm



- Cho LĐQH Q và tập PTH F
- Phân rã Q thành $\{Q_1, Q_2 \dots Q_n\}$ thì mỗi Q sẽ xác định một tập PTH F_i :
 - $F_i = \{X \rightarrow Y : XY \subseteq Q_i \text{ và } X \rightarrow Y \in F^+\}$
 - F_i được gọi là tham chiếu của F^+ lên Q_i
- Phân rã trên bảo toàn phụ thuộc hàm nếu:
 - Đặt $F' = F_1 \cup F_2 \cup \dots \cup F_n$, thì $F' \equiv F$ (nghĩa là $F'^+ = F^+$)
- Để kiểm tra phân rã bảo toàn PTH ta đi kiểm tra $F_1 \cup F_2 \cup \dots \cup F_n \equiv F$
- *Lưu ý: Khi tính các F_i thường hay thiếu sót các phụ thuộc hàm vì F_i là chiếu của F^+ lên Q_i chứ không phải F lên Q_i . Như vậy để tính đầy đủ F_i của Q_i ta tính bao đóng của tất cả các tập con thực sự của Q_i .*
- $X \subset Q_i$. Nếu $X^+ \cap Q_i \neq X$ thì $(X \rightarrow (X^+ \cap (Q_i - X))) \in F_i$.

Phân rã bảo toàn phụ thuộc hàm...



- **Ví dụ:** Cho $Q(ABCD)$, $F = \{A \rightarrow B, B \rightarrow C, C \rightarrow D, D \rightarrow A\}$
- Phân rã Q thành $\{Q1(AB), Q2(BC), Q3(CD)\}$ sẽ dễ dàng nhầm lẫn:
 - $Q1(AB)$, $F1 = \{A \rightarrow B\}$
 - $Q2(BC)$, $F2 = \{B \rightarrow C\}$
 - $Q3(CD)$, $F3 = \{C \rightarrow D\}$
- Lúc này $F' = F1 \cup F2 \cup F3 = \{A \rightarrow B, B \rightarrow C, C \rightarrow D\}$
- Rõ ràng là F' không tương đương với F vì $F'^+ \neq F^+$ do $D \rightarrow A \notin F'^+$

=> Nếu vội vã kết luận phân rã trên không bảo toàn phụ thuộc hàm là sai.

Phân rã bảo toàn phụ thuộc hàm...



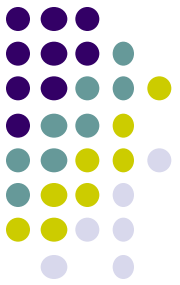
- Thực ra:
- Q1(AB) $A_f^+ = ABCD \Rightarrow A \rightarrow B \in F1$ $B_f^+ = BCDA \Rightarrow B \rightarrow A \in F1$
 - Vậy $F1 = \{A \rightarrow B, B \rightarrow A\}$
- Q2(BC) $B_f^+ = BCDA \Rightarrow B \rightarrow C \in F2$ $C_f^+ = CDAB \Rightarrow C \rightarrow B \in F2$
 - Vậy $F2 = \{B \rightarrow C, C \rightarrow B\}$
- Q3(CD) $C_f^+ = CDAB \Rightarrow C \rightarrow D \in F3$ $D_f^+ = DABC \Rightarrow D \rightarrow C \in F3$
 - Vậy $F3 = \{C \rightarrow D, D \rightarrow C\}$
- Vậy $F' = F1 \cup F2 \cup F3 = \{A \rightarrow B, B \rightarrow A, B \rightarrow C, C \rightarrow B, C \rightarrow D, D \rightarrow C\}$
- Ta tính được $F'^+ = F^+ \Rightarrow F' \equiv F$
- Kết luận: Phân rã trên **bảo toàn phụ thuộc hàm**

Dạng chuẩn (Normal Form-NF)



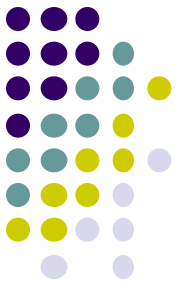
- Xét dạng chuẩn dựa trên phụ thuộc hàm
- **Thuộc tính khoá:** Thuộc tính tham gia vào bất kỳ khoá nào đó của quan hệ chứa nó. Ngược lại gọi là **thuộc tính không khoá**.
- **Ví dụ:** Q(ABCDEF) A, B, D, E là các thuộc tính khoá. C, F là các thuộc tính không khoá.
- **Thuộc tính đơn:** Miền giá trị của nó không phải là tích hợp của các miền giá trị khác.
- $X \rightarrow A$ là **PTH nguyên tố** nếu: Không $\exists Y$ là tập con thực sự của X , $Y \rightarrow A \in F^+$

Dạng chuẩn...



- Ví dụ: Cho $F = \{AB \rightarrow C, B \rightarrow C\}$ thì:
 - $AB \rightarrow C$: không là phụ thuộc hàm nguyên tố vì có $B \rightarrow C$.
 - $B \rightarrow C$: là phụ thuộc hàm nguyên tố
- A là thuộc tính **phụ thuộc đầy đủ** vào X nếu $X \rightarrow A$ là phụ thuộc hàm nguyên tố
- Ví dụ trên thì C là thuộc tính phụ thuộc đầy đủ vào B chứ không phụ thuộc đầy đủ vào AB.

Dạng chuẩn 1(1NF)



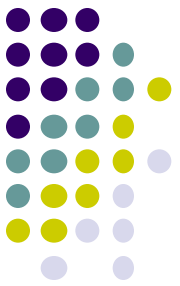
- **ĐN:** Lược đồ quan hệ Q ở dạng 1NF nếu tất cả các thuộc tính của Q đều là thuộc tính đơn/nguyên tố.
- Lược đồ CSDL C ở 1NF nếu tất cả các Qi của C đều ở 1NF.

DEPARTMENT			
DNAME	<u>DNUMBER</u>	DMGRSSN	DLOCATIONS
Research	5	333445555	{Bellaire, Sugarland, Houston}
Administration	4	987654321	{Stafford}
Headquarters	1	888665555	{Houston}

Vi phạm dạng chuẩn 1

- ☞ Đây là dạng chuẩn đơn giản nhất, nó không chú ý đến các phụ thuộc hàm do đó có rất nhiều trùng lặp thông tin do các phụ thuộc hàm trên gây ra.

Dạng chuẩn 1(1NF)...



- Chuyển quan hệ trên thành dạng chuẩn 1 (bằng cách xác định tập thuộc tính {DNumber, DLocation} là khoá chính)

DEPARTMENT

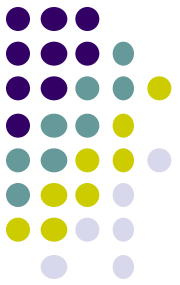
DNAME	<u>DNUMBER</u>	DMGRSSN	<u>DLOCATION</u>
Research	5	333445555	Bellaire
Research	5	333445555	Sugarland
Research	5	333445555	Houston
Administration	4	987654321	Stafford
Headquarters	1	888665555	Houston

Dư
thừa



- ☞ Đây là dạng chuẩn đơn giản nhất, nó không chú ý đến các phụ thuộc hàm do đó có rất nhiều trùng lặp thông tin do các phụ thuộc hàm trên gây ra.

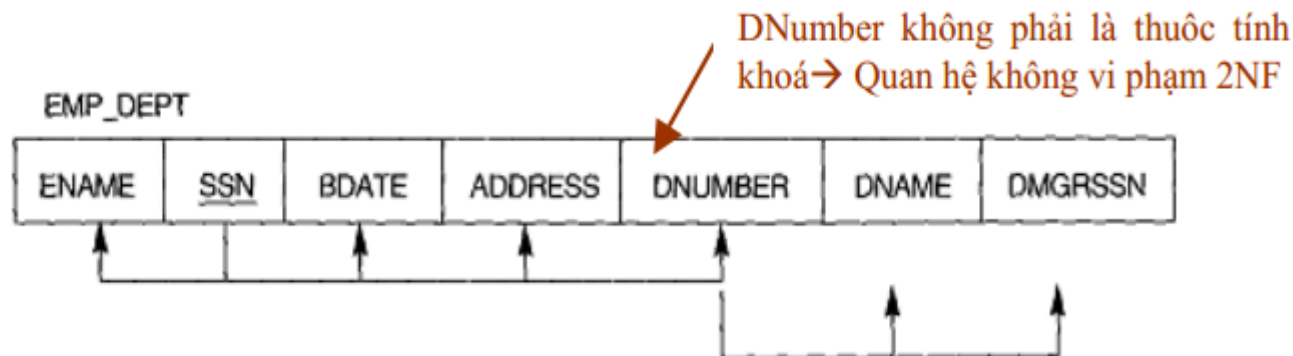
Dạng chuẩn 2 (2NF)



- **ĐN**: Lược đồ quan hệ Q ở dạng 2NF nếu ở 1NF và tất cả thuộc tính không khoá đều phụ thuộc đầy đủ vào khoá.
- Lược đồ CSDL C ở dạng 2NF nếu tất cả các Q_i của C đều ở dạng 2NF
- **VD**: Cho $Q(\underline{AB}CD)$, $F = \{A \rightarrow C, B \rightarrow D\}$, không đạt 2NF vì:
 - Khoá chính là AB; C,D là hai thuộc tính không khoá.
 - $AB \rightarrow C$ không là phụ thuộc hàm nguyên tố vì có $A \rightarrow C$
 - $AB \rightarrow D$ không là phụ thuộc hàm nguyên tố vì có $B \rightarrow D$
 - C và D không phụ thuộc đầy đủ vào khoá.
- Xét một tình trạng Q có sự trùng lặp thông tin (các giá trị trong ngoặc là trùng lặp)
- | Q | <u>A</u> | <u>B</u> | C | D |
|---|----------|----------|------|------|
| ● | (a1) | b1 | (c1) | d1 |
| ● | (a1) | b2 | (?) | d2 |
| ● | a2 | (b3) | c2 | (d3) |
| ● | a3 | (b3) | c3 | (?) |

Dạng chuẩn 2 (2NF)...

- EMP_DEPT đạt 2NF



EMP_DEPT

ENAME	<u>SSN</u>	BDATE	ADDRESS	DNUMBER	DNAME	DMGRSSN
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5	Research	333445555
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5	Research	333445555
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4	Administration	987654321
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire, TX	4	Administration	987654321
Narayan, Ramesh K.	666884444	1962-09-15	975 FireOak, Humble, TX	5	Research	333445555
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5	Research	333445555
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4	Administration	987654321
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1	Headquarters	888665555

Dư thừa dữ liệu

Dạng chuẩn 2 (2NF)...



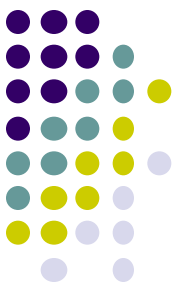
- **VD:** Cho $Q(ABCD)$, $F = \{AB \rightarrow C, C \rightarrow D\}$ ở 2NF vì:
 - Khoá chính là AB; C,D là hai thuộc tính không khoá. $AB \rightarrow C$ và $AB \rightarrow D$ đều là các phụ thuộc hàm nguyên tố.
 - $\Rightarrow C$ và D đều là phụ thuộc đầy đủ vào khoá.
- Xét một tình trạng Q có sự trùng lặp thông tin:
- | Q | A | B | C | D |
|---|----|----|------|------|
| | a1 | b1 | (c1) | (d1) |
| | a2 | b2 | (c1) | (?) |
- Ta thấy rằng ở VD trên, $C \rightarrow D$ gây ra trùng lặp thông tin vì thuộc tính không khoá D phụ thuộc bắc cầu vào khoá (nghĩa là phụ thuộc hàm **khoá $(AB) \rightarrow D$** suy diễn nhờ qui tắc bắc cầu Armstrong).
- **Nhận xét dạng chuẩn 2NF**
- ☞ Vẫn bị trùng lặp thông tin do phụ thuộc hàm bắc cầu.

Dạng chuẩn 3 (3NF)



- **ĐN1**: Lược đồ quan hệ Q ở dạng 3NF nếu ở 2NF và tất cả các thuộc tính không khoá không phụ thuộc bắc cầu vào khoá.
- **ĐN2**: Lược đồ quan hệ Q ở dạng 3NF nếu ở 1NF và tất cả phụ thuộc hàm không hiển nhiên $X \rightarrow Y$ của F^+ thoả một trong hai điều kiện sau:
 - (i) X là một siêu khoá (X chứa một khoá nào đó)
 - (ii) Mỗi thuộc tính trong tập (Y - X) nằm trong một khoá nào đó.
- Lược đồ CSDL C ở dạng 3NF nếu tất cả các Qi của C đều ở dạng 3NF

Dạng chuẩn 3 (3NF)...



- Tách thành EMPLOYEE, DEPARTMENT đạt 3NF

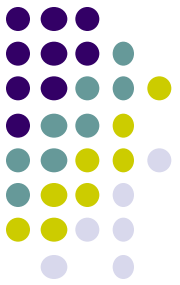
EMPLOYEE

ENAME	<u>SSN</u>	BDATE	ADDRESS	DNUMBER
Smith, John B.	123456789	1965-01-09	731 Fondren, Houston, TX	5
Wong, Franklin T.	333445555	1955-12-08	638 Voss, Houston, TX	5
Zelaya, Alicia J.	999887777	1968-07-19	3321 Castle, Spring, TX	4
Wallace, Jennifer S.	987654321	1941-06-20	291 Berry, Bellaire, TX	4
Narayan, Remesh K.	666884444	1962-09-15	975 Fire Oak, Humble, TX	5
English, Joyce A.	453453453	1972-07-31	5631 Rice, Houston, TX	5
Jabbar, Ahmad V.	987987987	1969-03-29	980 Dallas, Houston, TX	4
Borg, James E.	888665555	1937-11-10	450 Stone, Houston, TX	1

DEPARTMENT

DNAME	<u>DNUMBER</u>	DMGRSSN
Research	5	333445555
Administration	4	987654321
Headquarters	1	888665555


Dạng chuẩn 3 (3NF)...



- **VD:** Cho $Q(\underline{ABCD})$, $F = \{AB \rightarrow CD\}$ ở dạng 3NF
 - Vì $AB \rightarrow CD$ có vế trái là một siêu khoá.
- **VD:** Cho $Q(\underline{ABCDE})$, $F = \{AB \rightarrow CDE, B \rightarrow D, DE \rightarrow ABC\}$ ở dạng 3NF vì:
 - $AB \rightarrow CDE$ có vế trái là một siêu khoá.
 - $B \rightarrow D$ có $(VP) - (VT) = D$ chứa trong khoá DE .
 - $DE \rightarrow ABC$ có vế trái là một siêu khoá.
- Xét một tình trạng Q có sự trùng lặp thông tin:
- | Q | <u>A</u> | B | C | D | <u>E</u> |
|---|----------|------|----|------|----------|
| | a1 | (b1) | c1 | (d1) | e1 |
| | a2 | (b1) | c2 | (?) | e1 |
- 🖐 Ở dạng 3NF nếu có nhiều khoá sẽ có khả năng trùng lặp thông tin do các phụ thuộc hàm loại (ii) trong ĐN2.

Dạng chuẩn Boyce –codd (BCNF)

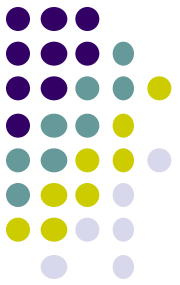


- **ĐN**: Lược đồ quan hệ Q ở BCNF nếu ở dạng 1NF và tất cả phụ thuộc hàm không hiển nhiên $X \rightarrow Y$ của F^+ thì X là một siêu khoá (X chứa một khoá nào đó).
- Lược đồ CSDL C ở dạng BCNF nếu tất cả các Q_i của C đều ở dạng BCNF.
- **VD**: Cho $Q(\underline{ABCD})$, $F\{AB \rightarrow CD, D \rightarrow AB\}$ ở dạng BCNF vì:
 - $AB \rightarrow CD$ có vế trái là một siêu khoá
 - $D \rightarrow AB$ có vế trái là một siêu khoá
 - $D \rightarrow C \in F^+$ vế trái là một siêu khoá
-  Ở dạng chuẩn BCNF không có sự trùng lặp thông tin do phụ thuộc hàm gây ra. Tuy nhiên dạng BCNF có điều kiện quá khắt khe, đôi khi người ta chỉ cần chấp nhận ở dạng 3NF.



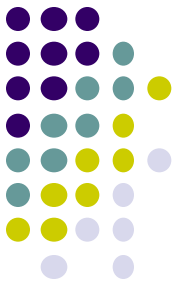
Chuẩn hoá lược đồ CSDL quan hệ

Chuẩn hoá lược đồ CSDL



- Thuật toán phân rã
- Thuật toán tổng hợp
- Cách thức chuẩn hoá trong thực tế

Phương pháp phân rã (Decomposition)



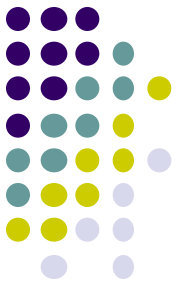
- Dựa vào điều kiện phân rã bảo toàn thông tin: Q thành Q_1 và Q_2 thoả $Q_1 \cap Q_2 \rightarrow Q_1 \setminus Q_2$ hay $Q_1 \cap Q_2 \rightarrow Q_2 \setminus Q_1$.
- Thuật toán phân rã thành các lược đồ ở dạng chuẩn BCNF như sau:
- Cho Q và tập phụ thuộc hàm F xác định trên Q
- $\text{Phân_rã} = \{Q\}$;
- $\text{done} = \text{false}$;
- Tính F^+ ;
- while (not done)
- if (có một Q_i trong Phân_rã không ở dạng BCNF)
- { $X \rightarrow Y$ là phụ thuộc hàm không hiển nhiên trên Q_i thoả:
- $X \rightarrow Q_i \notin F^+$ và $X \cap Y = \emptyset$ thì
- $\text{Phân_rã} = (\text{Phân_rã} - Q_i) \cup (XY) \cup (Q_i - Y)$
- }
- else $\text{done} = \text{true}$;
- Kết quả ta được tập Phân_rã gồm các lược đồ ở dạng **BCNF**.

Phương pháp phân rã (Decomposition)...



- **VD:** Cho $Q(ABCD)$, $F = \{AB \rightarrow C, C \rightarrow A, B \rightarrow D\}$
- $Q(ABCD)$ không ở dạng **BCNF**, chọn $C \rightarrow A$. Phân rã thành Q_1 và Q_2
- $Q_1(\underline{C}A)$, $F_1 = \{C \rightarrow A\}$, **BCNF**
- $Q_2(\underline{B}CD)$, $F_2 = \{B \rightarrow D\}$, không **BCNF**
 - $Q_2(\underline{B}CD)$ không ở **BCNF**, chọn $B \rightarrow D$. Phân rã thành Q_{21} , Q_{22}
 - ◆ $Q_{21}(\underline{B}D)$, $F_{21} = \{B \rightarrow D\}$, **BCNF**
 - ◆ $Q_{22}(\underline{B}C)$, $F_{22} = \emptyset$, **BCNF**
- *Cuối cùng ta phân rã thành :*
- $Q_1(\underline{A}C)$, $F_1 = \{C \rightarrow A\}$, **BCNF**
- $Q_2(\underline{B}D)$, $F_2 = \{B \rightarrow D\}$, **BCNF**
- $Q_3(\underline{B}C)$, $F_3 = \emptyset$, **BCNF**

Phương pháp phân rã (Decomposition)...



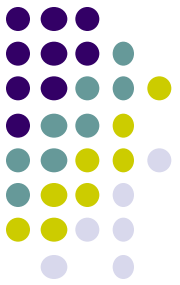
- **Nhận xét:**
- ☞ Cho dạng chuẩn BCNF .
- ☞ Bảo toàn thông tin
- ☞ Không phải lúc nào cũng bảo toàn PTH (VD trên 3 tập PTH cuối cùng F_1, F_2, F_3 không tương đương với tập phụ thuộc hàm F ban đầu).

Phương pháp tổng hợp (Synthesis)



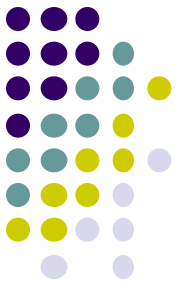
- ***Thuật toán sau cho phân rã đạt tối thiểu dạng 3NF.***
- Cho Q và tập PTH F xác định trên Q.
- Tính Fc là một phủ tối thiểu của F;
- **Xác định các khoá của Q ;**
- **$i = 0$;**
- **for (Mỗi phụ thuộc hàm $X \rightarrow Y$ trong Fc)**
- **if (không có Qj, $j = 1, 2, \dots, i$ chứa XY)**
- **{ $i++$;**
- **$Q_i = XY$;**
- **}**
- **if (Không có Qj, $j=1, 2, \dots, i$ chứa khoá của Q)**
- **{ $i++$;**
- **$Q_i =$ bất kỳ khoá nào của Q ;**
- **}**
- **return (Q1, Q2, ..., Qi) ;**

Phương pháp tổng hợp (Synthesis)...



- **VD:** Phân rã $Q(\underline{A}BCDE)$, $F = \{A \rightarrow CD, C \rightarrow D, B \rightarrow E\}$
- Tính được $F_c = \{A \rightarrow C, C \rightarrow D, B \rightarrow E\}$; khoá là AB
- Suy ra phân rã thành:
 - $Q1 = \underline{A}C$
 - $Q2 = \underline{C}D$
 - $Q3 = \underline{B}E$
- Không có Q_i nào chứa khoá AB nên ta có: $Q4 = \underline{A}B$
- **Kết quả đạt được:**
 - $Q1(\underline{A}C)$, $F1 = \{A \rightarrow C\}$, đạt 3NF (đạt luôn BCNF)
 - $Q2(\underline{C}D)$, $F2 = \{C \rightarrow D\}$, đạt 3NF (đạt luôn BCNF)
 - $Q3(\underline{B}E)$, $F3 = \{B \rightarrow E\}$, đạt 3NF (đạt luôn BCNF)
 - $Q4(\underline{A}B)$, $F4 = \emptyset$, đạt 3NF (đạt luôn BCNF)

Phương pháp tổng hợp (Synthesis)...



- **Nhận xét:**
- ☞ Tối thiểu đạt dạng chuẩn 3NF .
- ☞ Bảo toàn PTH vì mỗi PTH trong F_c cho một quan hệ và quan hệ này xác định luôn PTH đó.
Vậy $F' = \cup F_i$, $F' \equiv F_c \equiv F$.
- ☞ Bảo toàn thông tin vì có ít nhất một lược đồ trong phân rã chứa khoá.

Cách thức chuẩn hoá thực tế



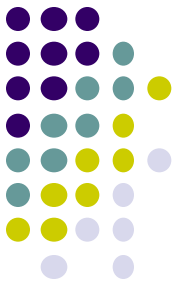
- Trong thực tế khi chuẩn hoá lược đồ CSDL thường thực hiện theo các bước:
- **Bước 1**: Kiểm tra xem quan hệ đã đạt dạng chuẩn 1NF chưa?. Nếu chưa ở 1NF có nghĩa là có các thuộc tính chưa nguyên tố/lặp. Tiến hành tách các thuộc tính đó.
- **Bước 2**: Kiểm tra xem chúng có ở dạng 2NF không?, nghĩa là kiểm tra xem các thuộc tính không khoá có *phụ thuộc hoàn toàn* vào khoá chính không?. Tiến hành tách những PTH bộ phận đó thành các bảng con để giảm sự trùng lặp thông tin.

Cách thức chuẩn hoá thực tế...

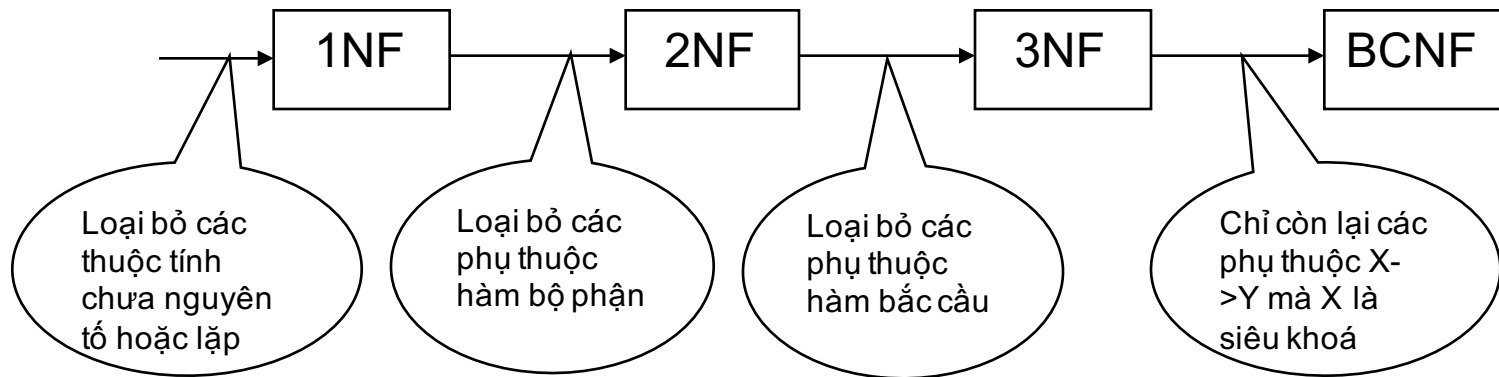


- **Bước 3**: Kiểm tra xem chúng đã đạt dạng chuẩn 3NF chưa?, nghĩa là các thuộc tính không khoá thì *phụ thuộc trực tiếp* vào khoá chính. Tiến hành tách những PTH bắc cầu thành bảng con.
- **Bước 4**: kiểm tra xem chúng đã đạt dạng chuẩn BCNF chưa?, nghĩa là tất cả các phụ thuộc hàm đều có vế trái là siêu khoá. Tiến hành tách PTH có vế trái chưa phải là siêu khoá.

Cách thức chuẩn hoá thực tế...



- ***Ta có thể tóm tắt lại quá trình chuẩn hoá như sau:***



- ***Trong thực tế ta chấp nhận dạng chuẩn thấp để đổi lại đơn giản trong cài đặt và hệ thống chạy nhanh hơn.***

Cách thức chuẩn hoá thực tế...



- **Ví dụ minh họa**
- Giả sử khi khảo sát thực tế về việc mua bán vật tư, ta có được lược đồ quan hệ sau:
- **PHIEUNHAP**(sophieu, ngaynhap, makhach, tenkh, diachikh, dienthoai, makho,
- | | | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| 7 | | | | | | |
- diachikho, hinhthucthanhtoan, loaitien, mavattu*, tenvattu*, soluong*,
- | | | | | | |
|---|---|----|----|----|----|
| 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|----|----|----|----|
- donvitinh*, dongia*, tyleVAT*)
- | | | |
|----|----|----|
| 14 | 15 | 16 |
|----|----|----|
- **Có các phụ thuộc hàm sau:**
- $F = \{1 \rightarrow (2,3,4,5,6,7,8,9,10); 3 \rightarrow (4,5,6); 7 \rightarrow 8; 11 \rightarrow (12,14,15,16); (1,11) \rightarrow 13\}$

Cách thức chuẩn hoá thực tế...



- **Bước 1:** Khoá chính của quan hệ PHIEUNHAP là $K=(1,11)=(\text{sophieu}, \text{mavattu})$.
- Chưa ở dạng 1NF vì có các thuộc tính (có dấu *) là thuộc tính lặp. Ta tiến hành phân rã thành hai quan hệ:
- Quan hệ 1: Các thuộc tính lặp và phần khoá chính xác định chúng.
- Quan hệ 2: các thuộc tính còn lại và phần khoá chính xác định phần này (các thuộc tính không lặp)
- Quan hệ 1 gồm các thuộc tính lặp (11,12,13,14,15,16) và khoá chính là (1)
- Tức là **VATTU**(Sophieu, mavattu, tenvattu, soluong, donvitinh,
1 11 12 13 14
dongia, tyleVAT)
- 15 16
- Phụ thuộc hàm của quan hệ này là $F1=\{ 11 \rightarrow (12,14,15,16); (1,11) \rightarrow 13 \}$
- Khoá chính của quan hệ này là $\{1,11\}$

Cách thức chuẩn hoá thực tế...



- Quan hệ 2 gồm các thuộc tính (2,3,4,5,6,7,8,9,10) và khoá (1)
- **PHIEUNHAP**(sophieu, ngaynhap, makhach, tenkh,
1 2 3 4
- diachikh, dienthoai, makho,
5 6 7
- diachikho, hinhthucthanhtoan, loaitien)
8 9 10
- Phụ thuộc hàm của quan hệ này là $F2 = \{1 \rightarrow (2,3,4,5,6,7,8,9,10); 3 \rightarrow (4,5,6); 7 \rightarrow 8\}$
- Khoá chính là $\{1\}$.

- Quan hệ 1_1: **VATTU**(mavattu, tenvattu, donvitinh, dongia, tyleVAT)

11
12
14
15

16

- - Xét quan hệ 2: đã đạt dạng chuẩn 2NF vì thuộc tính khoá của nó là sophieu và các thuộc tính đều phụ thuộc hoàn toàn vào khoá.
- **Bước 3:** Xem chúng đã đạt dạng 3NF chưa?. Nếu chưa ta tiến hành tách đôi thành quan hệ 1: gồm các thuộc tính phụ thuộc bậc cầu và thuộc tính cầu; quan hệ 2: gồm các thuộc tính còn lại và thuộc tính cầu.
- - Xét quan hệ 1_1:
- Quan hệ 1_1: **VATTU**(mavattu, tenvattu, donvitinh, dongia, tyleVAT)

11
12
14
15
16
- Đã là dạng chuẩn 3NF vì không có phụ thuộc hàm bậc cầu.
- - Xét quan hệ 1_2:
- Quan hệ 1_2: **DONGVATTU**(sophieu, mavattu, soluong)

1
11
13
- Đã là dạng chuẩn 3NF vì không có phụ thuộc hàm bậc cầu.

Cách thức chuẩn hoá thực tế...



- - Xét quan hệ 2:
- **PHIEUNHAP**(sophieu, ngaynhap, makhach, tenkh, diachikh, dienthoai, makho,
- 1 2 3 4 5 6 7
- diachikho, hinhthucthanhtoan, loaitien
- 8 9 10
- $F2 = \{1 \rightarrow (2, 3, 4, 5, 6, 7, 8, 9, 10); 3 \rightarrow (4, 5, 6); 7 \rightarrow 8\}$, Khoá chính là $\{1\}$
- Chưa đạt dạng 3NF vì có PTH bắc cầu: Các thuộc tính (4, 5, 6) PTH bắc cầu vào khoá chính qua cầu (3), còn thuộc tính (8) phụ thuộc bắc cầu vào khoá chính qua cầu (7).
- **Để có dạng chuẩn 3NF tách thành những quan hệ sau:**
- Quan hệ 2_1: **KHACH**(makhach, tenkh, diachikh, dienthoai)
- 3 4 5 6
- $F2_1 = \{3 \rightarrow 4, 5, 6\}$, khoá chính là (3)

Cách thức chuẩn hoá thực tế...



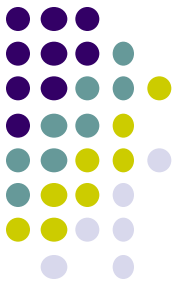
- Quan hệ 2_2: **KHO**(makho, diachikho)
7 8
- F2_2={7->8}; khoá chính là (7)
- Quan hệ 2_3: **PHIEUNHAP**(sophieu, ngaynhap, makhach,
1 2 3
makho, hinhthucthanhtoan, loaitien)
7 9 10
- F2_3={1-(2,3,7,9,10)}; khoá chính là (1)
- **Bước 4**: Kiểm tra xem chúng đạt dạng BCNF chưa?
- Ta thấy tất cả các quan hệ trên đều đã đạt ở dạng BCNF vì tất cả các vế trái của các phụ thuộc hàm đều là siêu khoá.

Cách thức chuẩn hoá thực tế...



- Tóm lại: Ta có các quan hệ sau khi chuẩn hoá đạt dạng chuẩn BCNF như sau:
- 1. **VATTU**(mavattu, tenvattu, donvitinh, dongia, tyleVAT)
- 2. **DONGVATTU**(sophieu, mavattu, soluong)
- 3. **KHACH**(makhach, tenkh, diachikh, dienthoai)
- 4. **KHO**(makho, diachikho)
- 5. **PHIEUNHAP**(sophieu, ngaynhap, makhach, makho, hinhthucthanhtoan, loaitien)

Bài tập chương 6



6.1: Xác định khoá và xét các phân rã sau đây theo hai tiêu chuẩn bảo toàn thông tin và bảo toàn phụ thuộc hàm.

- a. $Q(ABCD), F = \{A \rightarrow BC, C \rightarrow D\}$
- $Q1(AB), F1 = \{A \rightarrow B\}$
- $Q2(CD), F2 = \{C \rightarrow D\}$
- b. $Q(ABCD), F = \{A \rightarrow B, AC \rightarrow D\}$
- $Q1(AB), F1 = \{A \rightarrow B\}$
- $Q2(ADC), F2 = \{AC \rightarrow D\}$
- c. $Q(ABDCE), F = \{A \rightarrow C, B \rightarrow C, C \rightarrow D, DE \rightarrow C, CE \rightarrow A\}$
- $Q1(AB), F1 = \{A \rightarrow D\}$
- $Q2(CD), F2 = \emptyset$
- $Q3(AB), F3 = \emptyset$
- $Q4(CD), F4 = \{C \rightarrow D, DE \rightarrow C, CE \rightarrow D\}$
- $Q5(AB), F5 = \emptyset$
- d. $Q(ABCD), F = \{A \rightarrow B, C \rightarrow D\}$
- $Q1(AB), F1 = \{A \rightarrow B\}$
- $Q2(CD), F2 = \{C \rightarrow D\}$



Bài tập chương 6...

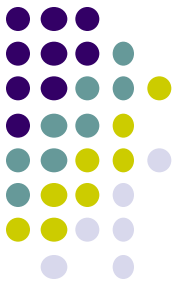
6.2: Xét dạng chuẩn của các lược đồ quan hệ sau:

- $Q(ABCD)$, $F=\{A \rightarrow B, C \rightarrow D\}$
- $Q(ABCD)$, $F=\{AB \rightarrow C, C \rightarrow D\}$
- $Q(ABCD)$, $F=\{AB \rightarrow CD, CD \rightarrow AB, C \rightarrow B\}$
- $Q(ABCDE)$, $F=\{AB \rightarrow CD, D \rightarrow E, DE \rightarrow ABC\}$
- $Q(ABCDEF)$, $F=\{AB \rightarrow E, AC \rightarrow F, AD \rightarrow B, B \rightarrow C, C \rightarrow D\}$

6.3: Cho lược đồ quan hệ $Q(ABCDEF)$, $F=\{C \rightarrow F, E \rightarrow A, CE \rightarrow D, A \rightarrow B\}$

- Xác định khoá của Q
- Phân rã thành dạng chuẩn Boyce-Codd bảo toàn thông tin
- Phân rã thành dạng chuẩn 3 bảo toàn thông tin và bảo toàn phụ thuộc hàm.

6.4: Giả sử bài 3 được phân rã thành $Q1(CDEF)$ và $Q2(ABE)$, hãy xác định $F1$, $F2$ và đánh giá phân rã trên.

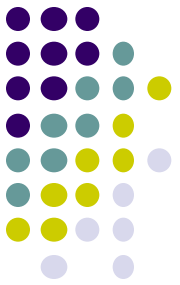


Bài tập chương 6...

6.5: Nếu bài 3 được phân rã thành Q1(CF), Q2(AE), Q3(CDE), Q4(AB).
Hãy xác định F1, F2, F3, F4 và đánh giá chúng.

6.6: Cho lược đồ quan hệ

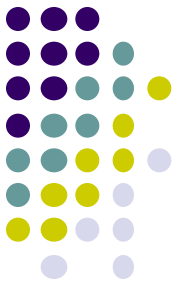
- VẠCHUYỀN(TÀU, LOAITÀU, CHUYẾN, HÀNG, CẢNG, NGÀY)
- Mỗi tàu (TÀU) thuộc duy nhất một loại tàu nào đó (LOAITÀU), mỗi chuyến có một mã số riêng biệt (CHUYẾN) dùng để xác định một chuyến tàu (TÀU) chở một khối lượng hàng hoá nào đó (HÀNG), mỗi chiếc tàu trong một ngày(NGÀY) chỉ cập vào một cảng duy nhất (CẢNG) của một chuyến vận chuyển nào đó (CHUYẾN)
- Xác định tập các phụ thuộc hàm trên.
- Xác định dạng chuẩn của VẠCHUYỀN
- Nếu VẠCHUYỀN chưa tốt hãy tìm một phân rã tốt cho nó



Bài tập chương 6...

6.7. Cho quan hệ **PHIEUNHAP**(số phiếu, ngày, mã NCC, tên NCC, địa chỉ, Mã vật tư, tên vật tư, số lượng, đơn vị, đơn giá)

- Có các phụ thuộc hàm
- $F = \{ \text{Số phiếu} \rightarrow \text{ngày, mã NCC};$
- $\text{mã NCC} \rightarrow \text{tên NCC, địa chỉ};$
- $\text{mã vật tư} \rightarrow \text{tên vật tư, đơn vị, đơn giá};$
- $\text{số phiếu, mã vật tư} \rightarrow \text{số lượng}$
- Giả sử tách thành hai quan hệ:
- **PHIEUNHAP**(số phiếu, ngày, mã NCC, Mã vật tư, tên vật tư, số lượng, đơn vị, đơn giá)
- **NHACUNGCAP**(mã NCC, tên NCC, địa chỉ)
- Kiểm tra xem chúng đạt dạng chuẩn nào? vì sao?



Bài tập chương 6...

- Giả sử tách thành các quan hệ:
- PHIEUNHAP(số phiếu, ngày, mã NCC)
- DONGPHIEU(số phiếu, Mã vật tư, số lượng)
- NHACUNGCAP(mã NCC, tên NCC, địa chỉ)
- VATTU(Mã vật tư, tên vật tư, đơn vị, đơn giá)
- Kiểm tra xem chúng đạt dạng chuẩn nào? vì sao?

6.8. Giả sử có quan hệ BANHANG(Ngày tháng, mã hàng, tên hàng, đơn giá, số lượng, tổng tiền theo ngày, thanh toán)

- Có các phụ thuộc hàm sau:
- {mã hàng} \rightarrow {tên hàng, đơn giá}
- {ngày tháng} \rightarrow { tổng tiền theo ngày, thanh toán}
- {ngày tháng, mã hàng} \rightarrow {số lượng}
- Hãy chuẩn hoá quan hệ BANHANG thành dạng chuẩn BCNF.