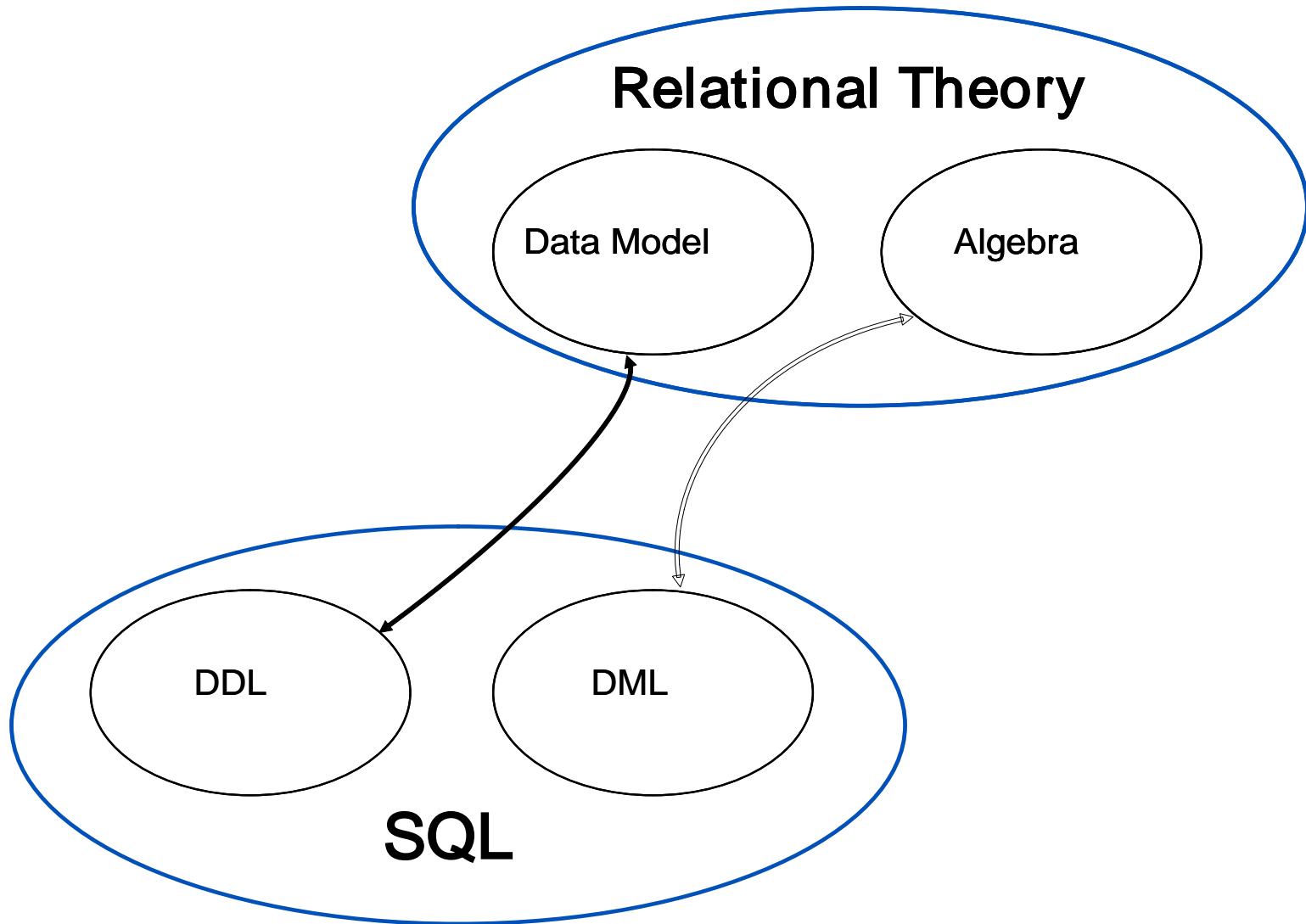


## **Chương 2**

# **CƠ SỞ DỮ LIỆU QUAN HỆ**

# CƠ SỞ DỮ LIỆU QUAN HỆ



# NỘI DUNG

## ■ Mô hình dữ liệu quan hệ

- ◆ Khái niệm
- ◆ Khoá của quan hệ
- ◆ Ràng buộc quan hệ

## ■ Đại số quan hệ

## ■ Ngôn ngữ vấn tin SQL

# MÔ HÌNH DỮ LIỆU QUAN HỆ

## ■ *Relational Database [Codd, 1970]*

- *Dữ liệu lưu trữ trong bảng hai chiều*
- *Mỗi bảng có tên duy nhất*
- *Mỗi cột có tên duy nhất trong một bảng*
- *Mỗi giá trị trên một cột thuộc cùng một miền*
- *Thứ tự các hàng trong bảng không quan trọng*
- *Thứ tự các cột trong bảng không quan trọng*
- *Mỗi giá trị trên cột là nguyên tố*

◆ Ví dụ: Quan hệ SV

MSV	HOTEN	NS
20051000	Nguyễn Hải Nam	1/1/1985
20051001	Trần Mai Ly	2/2/1986
20051002	Lê Phương Bắc	3/3/1985

# MÔ HÌNH DỮ LIỆU QUAN HỆ

## ■ Thuật ngữ

Người dùng

Lý thuyết

**Bảng**

**Quan hệ**

**Hàng hoặc Bản ghi**

**Bộ**

**Cột hay Trường**

**Thuộc tính**

**Tập giá trị hợp lệ**

**Miền**

**Số cột trong bảng**

**Cấp**

**Số hàng trong bảng**

**Lực lượng**

# MÔ HÌNH DỮ LIỆU QUAN HỆ

■ Miền (Domain): *Tập các giá trị nguyên tố*

■ Thuộc tính (Attribute): *Một thuộc tính A gồm:*

◆ *tên thuộc tính*

◆ *miền trị,  $dom(A)$ , hữu hạn hoặc vô hạn,  $||dom(A)|| \geq 2$*

■ Tích Đề-các:

$$D1 \times D2 \times \dots \times Dn = \{(v1, v2, \dots, vn) | vi \in Di\}$$

Ví dụ:

$$D1 = \{0,1\}; D2 = \{a,b,c\}$$

$$D1 \times D2 = \{(0,a), (0,b), (0,c), (1,a), (1,b), (1,c)\}$$

$$D1 = \{'Lan', 'Mai', 'Cúc'\}; D2 = \{'nam', 'nu'\}; D3 = \{19, 20\}$$

$$D1 \times D2 \times D3 = \{('Lan', 'nam', 19), ('Lan', 'nam', 20), ('Lan', 'nu', 19), \dots\}$$

# MÔ HÌNH DỮ LIỆU QUAN HỆ

## ■ Quan hệ:

$U = \{A1, A2, \dots, An\}$  - tập hữu hạn các thuộc tính,

Quan hệ  $R$  trên tập thuộc tính  $U$  là:

$$R \subseteq \text{dom}(A1) \times \text{dom}(A2) \times \dots \times \text{dom}(An)$$

◆ Kí hiệu:  $R(U)$  hay  $R(A1, A2, \dots, An)$  hay  $R$  trong trường hợp ngầm định.

◆ Ví dụ: Quan hệ  $SV(\text{MSV}, \text{HOTEN}, \text{NS})$

- Tên quan hệ:  $SV$

- Tập thuộc tính:  $\{\text{MSV}, \text{HOTEN}, \text{NS}\}$

- Miền trị:  $\text{dom}(\text{MSV}): \text{string}(7)$

$\text{dom}(\text{HOTEN}): \text{string}(30)$

$\text{dom}(\text{NS}): \text{date}$

# MÔ HÌNH DỮ LIỆU QUAN HỆ

## ■ Thể hiện (trạng thái) của Quan hệ

các bộ của  
quan hệ SV

MSV	HOTEN	NS
20051000	Nguyễn Hải Nam	1/1/1985
20051001	Trần Mai Ly	2/2/1986
20051002	Lê Phương Bắc	3/3/1986

## ■ Một bộ của quan hệ:

$t = \langle '20051000', 'Nguyễn Hải Nam', '1/1/1985' \rangle$ , và  $t \in SV$ .

## ■ Cơ sở dữ liệu quan hệ là tập các quan hệ được biến thiên theo thời gian.



# MÔ HÌNH DỮ LIỆU QUAN HỆ

■ Ví dụ 1 CSDL quan hệ (đặc tả):

**SV(MASV: string, HOTEN: string, NS: date)**

**HP(MAHP: string, TENHP: string, SOTC: integer)**

**KQHT(MAHP: string, MASV: string, DIEM: digital)**

# MÔ HÌNH DỮ LIỆU QUAN HỆ

## ■ Quy ước

- ◆ Tên quan hệ là các chữ cái in hoa, ở cuối bảng, như R, S, T,...
- ◆ Các bộ được kí hiệu bởi chữ thường, ở cuối bảng, như u, v, t
- ◆ Các thuộc tính được kí hiệu bởi chữ in hoa, ở đầu bảng, như A, B, C,..và viết ABC thay cho {A, B, C}
- ◆ Tập các thuộc tính được kí hiệu bởi chữ in hoa, ở cuối bảng, như U, V, X, Y, Z,..và viết XY thay cho  $X \cup Y$
- ◆ Kí hiệu t.X là hạn chế của bộ t trên tập thuộc tính X

t = <'200510000', 'Nguyễn Hải Nam', '1/1/1985'>, và  $t \in SV$

X = {MSV, HOTEN}

t.X = <'200510000', 'Nguyễn Hải Nam'>

# KHOÁ CỦA QUAN HỆ

## ■ Khoá (key)

*Cho quan hệ  $R(U)$ ,  $K \subseteq U$ ,  $K$  là khoá của quan hệ  $R$  nếu:*

*(i) với  $\forall t, s \in R, t \neq s: t.K \neq s.K$*

*(ii)  $\nexists K' \subset K$  mà  $K'$  thoả (i).*

**Nếu  $K$  chỉ thoả (i),  $K$  được gọi là siêu khoá (super key).**

◆ Một quan hệ có thể có một hay nhiều khoá.

◆ Khoá chính (primary key), Khoá dự tuyển (candidate key).

Quan hệ SV

Khoá của  
SV??

Có thể bổ sung  
bộ này vào SV →  
???

MASV	HOTEN	NS
20051000	Nguyễn Hải Nam	1/1/1985
20051001	Trần Mai Ly	2/2/1986
20051002	Lê Phương Bắc	3/3/1985
20051003	Lê Phương Bắc	12/12/1985
20051002	Lê Ngọc Lam	12/12/1985

# KHOÁ CỦA QUAN HỆ

## ■ Ví dụ, cho quan hệ SVHT:

MASV	HOTEN	NS	MAHP	TENHP	HK	SOTC	DIEM
20051000	Nguyễn Hải Nam	1-1-85	m1	Đại số	1	4	5
20051000	Nguyễn Hải Nam	1-1-85	m2	Giải tích 1	1	3	7
20051000	Nguyễn Hải Nam	1-1-85	m3	Giải tích 2	2	4	5
20051001	Trần Mai Ly	2-2-85	m1	Đại số	1	4	8
20051001	Trần Mai Ly	2-2-85	m2	Giải tích 1	1	3	4

■ Có thể bổ sung vào SVHT bộ  $t = \langle "20051000", \dots, "m1", \dots, 2 \rangle$  ?

## ■ Xác định khoá của quan hệ SVHT

- ◆ Nếu giả thiết, các môn khác nhau phải có tên khác nhau ?
- ◆ Nếu giả thiết có thể có hai môn giống tên nhau ?
- ◆ Nếu giả thiết một môn học có thể học trong nhiều kì khác nhau?

# RÀNG BUỘC QUAN HỆ

- Ràng buộc miền
- Ràng buộc NULL
- Ràng buộc khoá
  - ◆ Ràng buộc khoá chính
  - ◆ Ràng buộc duy nhất
- Ràng buộc toàn vẹn tham chiếu
  - ◆ dựa trên khoá ngoại lai (foreign key)

# Ví dụ cơ sở dữ liệu quan hệ

Quan hệ SV

MASV	HOTEN	NS
20051000	Nguyễn Hải Nam	1/1/1985
20051001	Trần Mai Ly	2/2/1986
20051002	Lê Phương Bắc	3/3/1985

Quan hệ KQHT

MASV	MAHP	DIEM
20051000	M1	4
20051000	M2	6
20051001	M2	7
20051001	M1	5

Quan hệ HP

MAHP	TENHP	SOTC
M1	Anh văn	4
M2	Toán rời rạc	3
M3	Cơ sở dữ liệu	4

# Ràng buộc dữ liệu

## ■ Tình huống 1

Ràng buộc miền

MAHP	TENHP	SOTC
M1	Anh văn	aa
M2	Toán rời rạc	3
M3	Cơ sở dữ liệu	4

## ■ Tình huống 2

Ràng buộc NULL

MASV	HOTEN	NS
20051000	Nguyễn Hải Nam	
20051001		2/2/1986
20051002	Lê Phương Bắc	3/3/1985

Giá trị trên thuộc tính có thể là NULL hay không?

# Ràng buộc dữ liệu

## ■ Tình huống 3

Ràng buộc khoá chính

Không chấp nhận !

MAHP	TENHP	SOTC
M1	Anh văn	4
M1	Toán rời rạc	3
	Cơ sở dữ liệu	4

## ■ Tình huống 2

Ràng buộc khoá duy nhất

MAHP	TENHP	SOTC
M1	Anh văn	4
M2	Anh văn	3
M3	Cơ sở dữ liệu	4



# Ràng buộc dữ liệu

- Tình huống 4. Ràng buộc toàn vẹn tham chiếu

Khoá ngoại lai

SV

MASV	HOTEN	NS
20051000	Nguyễn Hải Nam	1/1/1985
20051001	Trần Mai Ly	2/2/1986
20051002	Lê Phương Bắc	3/3/1985

KQHT

MASV	MAHP	DIEM
20051000	M1	4
20051000	M2	6
20051001	M2	7
20051001	M1	6
20051004	M4	5

HP

MAHP	TENHP	SOTC
M1	Anh văn	4
M2	Toán rời rạc	3
M3	Cơ sở dữ liệu	4

???

## ■ Khóa chính (primary key) là một tập thuộc tính:

- ◆ Hai bộ bất kỳ phải khác nhau trên khóa chính
- ◆ Không chấp nhận giá trị NULL trên khóa chính

## ■ Khóa duy nhất (unique key) là một (tập) thuộc tính:

- ◆ Hai bộ bất kỳ phải khác nhau trên khóa duy nhất
- ◆ Chấp nhận giá trị NULL trên khóa duy nhất

## ■ Khóa ngoại lai (foreign key) :

- ◆ Thuộc tính không là khóa của quan hệ nhưng là khóa của quan hệ khác trong CSDL

# ĐẠI SỐ QUAN HỆ

## ■ Thế nào là “Đại số” ?

◆ Hệ toán học, bao gồm:

- *Toán tử / Phép toán (Operators)* --- kí pháp mô tả thủ tục để cho ra kết quả từ các giá trị vào cho trước
- *Toán hạng (Operands)* --- biến/giá trị mà toán tử thao tác và cho ra kết quả.

## ■ Đại số quan hệ

- ◆ Hệ đại số, toán hạng là các quan hệ
- ◆ Toán tử được thiết kế để thực hiện những thao tác thường gặp nhất đối với quan hệ trong CSDL.
  - Kết quả là một hệ đại số mà có thể được sử dụng như một ngôn ngữ văn tin trên các quan hệ.

# ĐẠI SỐ QUAN HỆ

## ■ Các phép toán quan hệ cơ bản

- ◆ Phép hợp
- ◆ Phép giao
- ◆ Phép hiệu
- ◆ Tích đề các
- ◆ Phép chọn
- ◆ Phép chiếu
- ◆ Phép kết nối
- ◆ Phép chia
- ◆ \*\*\*Phép đổi tên quan hệ/tên thuộc tính

# ĐẠI SỐ QUAN HỆ

- **Các phép tính quan hệ:** Hợp, Giao, Hiệu, Tích Đề các, Chọn, Chiếu, Kết nối, Chia
- **Các toán hạng:** Các quan hệ
- **Biểu thức quan hệ**
  - ◆ Toán tử: là các phép toán quan hệ
  - ◆ Toán hạng: là các quan hệ
  - ◆ Kết quả: là một quan hệ

# ĐẠI SỐ QUAN HỆ

## ■ Phép hợp (Union)

- ◆ Cho hai quan hệ  $R(U)$ ,  $S(U)$  ( $R$  và  $S$  tương thích)
- ◆ Phép hợp hai quan hệ  $R$  và  $S$  là một quan hệ trên  $U$ :

$$R \cup S = \{t \mid t \in R \vee t \in S\}$$

## ■ Phép giao (Intersection)

- ◆ Cho hai quan hệ  $R(U)$ ,  $S(U)$  ( $R$  và  $S$  tương thích)
- ◆ Phép giao hai quan hệ  $R$  và  $S$  là một quan hệ trên  $U$ :

$$R \cap S = \{t \mid t \in R \wedge t \in S\}$$

# ĐẠI SỐ QUAN HỆ

## ■ Phép trừ (Set difference)

- ◆ Cho hai quan hệ  $R(U)$ ,  $S(U)$  ( $R$  và  $S$  tương thích)
- ◆ Phép trừ quan hệ  $R$  cho quan hệ  $S$  là một quan hệ  $U$ :

$$R - S = \{t \mid t \in R \wedge t \notin S\}$$

## ■ Tích đề các (Catersian Product)

- ◆ Cho  $U = \{A_1, A_2, \dots, A_n\}$ ,  $V = \{B_1, B_2, \dots, B_m\}$ ; Quan hệ  $R(U)$ ,  $S(V)$ .
- ◆ Tích đề các hai quan hệ  $R$  và  $S$  là:

$$R \times S = \{t = (a_1, a_2, \dots, a_n, b_1, b_2, \dots, b_n) \mid (a_1, a_2, \dots, a_n) \in R, (b_1, b_2, \dots, b_n) \in S\}$$

# CÁC PHÉP TOÁN QUAN HỆ

## ■ Phép chọn (Selection)

*Cho quan hệ  $R(U)$ ,  $E$  là biểu thức chọn trên  $U$ . Phép chọn quan hệ  $R$  theo điều kiện  $E$  cho một quan hệ trên  $U$ :*

$$\sigma_E(R) = \{t \mid t \in R \wedge t(E)\} \quad \text{hoặc } R(E)$$

$t(E)$ : bộ  $t$  thoả  $E$

## ■ Phép chiếu (Projection)

*Cho quan hệ  $R(U)$ ,  $X \subseteq U$ . Phép chiếu quan hệ  $R$  trên tập thuộc tính  $X$  là một quan hệ trên  $X$ :*

$$\pi_X(R) = \{t.X \mid t \in R\} \quad \text{hoặc } R[X]$$



# CÁC PHÉP TOÁN QUAN HỆ

■  $R1 = \sigma_{MASV='20051000'}(SV)$

Quan hệ R2

■  $R2 = \Pi_{MASV, HOTEN}(SV)$

MASV	HOTEN	NS
20051000	Nguyễn Hải Nam	1/1/1985
20051001	Trần Mai Ly	2/2/1986
20051002	Lê Phương Bắc	3/3/1985

MASV	HOTEN
20051000	Nguyễn Hải Nam
20051001	Trần Mai Ly
20051002	Lê Phương Bắc

Quan hệ R1

Quan hệ SV

MASV	HOTEN	NS
20051000	Nguyễn Hải Nam	1/1/1985

# CÁC PHÉP TOÁN QUAN HỆ

$$R3 = \prod_{MASV, HOTEN} (\sigma_{MASV='20051000'}(SV))$$

MASV	HOTEN	NS
20051000	Nguyễn Hải Nam	1/1/1985
20051001	Trần Mai Ly	2/2/1986
20051002	Lê Phương Bắc	3/3/1985

Quan hệ SV

Quan hệ R3

MASV	HOTEN
20051000	Nguyễn Hải Nam

# CÁC PHÉP TOÁN QUAN HỆ

## ■ PHÉP KẾT NỐI TỰ NHIÊN

$$\blacksquare R_4 = SV \bowtie KQHT$$

MASV	HOTEN	NS
20051000	Nguyễn Hải Nam	1/1/1985
20051001	Trần Mai Ly	2/2/1986
20051002	Lê Phương Bắc	3/3/1985

MASV	MAHP	DIEM
20051000	M1	4
20051000	M2	6
20051001	M2	7

Quan hệ R4

MASV	HOTEN	NS	MAHP	DIEM
20051000	Nguyễn Hải Nam	1/1/1985	M1	4
20051000	Nguyễn Hải Nam	1/1/1985	M2	6
20051001	Trần Mai Ly	2/2/1986	M2	7

# CÁC PHÉP TOÁN QUAN HỆ

## ■ Phép kết nối $\theta$ (Theta-Joins)

*Cho  $R(U)$ ,  $S(V)$ .  $A \in U$ ,  $B \in V$ ,*

*$\text{dom}(A)$  và  $\text{dom}(B)$  có cùng kiểu so sánh được.*

*$\theta$  - phép tính so sánh.*

*Phép kết nối  $\theta$  hai quan hệ  $R$  và  $S$ :*

$$R \bowtie_{A \theta B} S = \{ \langle u, v \rangle \mid u \in R, v \in S, u.A \theta v.B \}$$

◆ Kí hiệu  $\langle u, v \rangle$ : một bộ được tạo ra nhờ sự xếp bộ  $v$  cạnh bộ  $u$

## ■ Kết nối tự nhiên 2 quan hệ $R$ và $S$ , kí hiệu $R \bowtie S$ hoặc $R * S$ , là kết nối bằng trên hai thuộc tính cùng tên, sau đó loại bớt đi những thuộc tính cùng tên, chỉ giữ lại một đại diện.

◆  $R \bowtie S$  là quan hệ trên tập thuộc tính  $U \cup V$

◆ Một bộ thuộc  $R \bowtie S$  tạo nên bởi bộ  $u \in R$ ,  $v \in S$ , được gọi là “bộ nối được”  
(joined tuple)

# CÁC PHÉP TOÁN QUAN HỆ

■  $R5 = \sigma_{MAHP='M2'}(SV \bowtie KQHT)$

MASV	HOTEN	NS
20051000	Nguyễn Hải Nam	1/1/1985
20051001	Trần Mai Ly	2/2/1986
20051002	Lê Phương Bắc	3/3/1985

MASV	MAHP	DIEM
20051000	M1	4
20051000	M2	6
20051001	M2	7

Quan hệ R5

MASV	HOTEN	NS	MAHP	DIEM
20051000	Nguyễn Hải Nam	1/1/1985	M2	6
20051001	Trần Mai Ly	2/2/1986	M2	7

# CÁC PHÉP TOÁN QUAN HỆ

$$R6 = \Pi_{MASV, HOTEN, DIEM} (\sigma_{MAHP='M2'}(SV \bowtie KQHT))$$

MASV	HOTEN	NS
20051000	Nguyễn Hải Nam	1/1/1985
20051001	Trần Mai Ly	2/2/1986
20051002	Lê Phương Bắc	3/3/1985

MASV	MAHP	DIEM
20051000	M1	4
20051000	M2	6
20051001	M2	7

MASV	HOTEN	DIEM
20051000	Nguyễn Hải Nam	6
20051001	Trần Mai Ly	7

Quan hệ R6

# CÁC PHÉP TOÁN QUAN HỆ

■  $R7 = \prod_{NS} (\sigma_{MASV='20051002'}(SV))$

■  $R8 = SV \bowtie_{NS > NS} R7$

MASV	HOTEN	NS
20051000	Nguyễn Hải Nam	1/1/1985
20051001	Trần Mai Ly	2/2/1986
20051002	Lê Phương Bắc	3/3/1985

Quan hệ R7

NS
3/3/1985

Quan hệ R8

MASV	HOTEN	SV.NS	R7.NS
20051001	Trần Mai Ly	2/2/1986	3/3/1985

# CÁC PHÉP TOÁN QUAN HỆ

## ■ Phép chia (Quotient)

Cho  $R(U)$  và  $S(V)$ ,  $\emptyset \neq V \subset U$ ,  $X = U - V$ .

Phép chia quan hệ  $R$  cho  $S$ , kí hiệu  $R \div S$ , là một quan hệ trên  $X$ :

$$R \div S = \{t.X \mid t \in R: \forall v \in S, \langle t.X, v \rangle \in R\}$$

## ■ Thực hiện phép toán:

◆ Tính  $\prod_X(R)$

◆ Với mỗi  $t \in \prod_X(R)$ , nếu  $\forall v \in S$  đều thỏa mãn  $\langle t, v \rangle \in R$  thì  $t \in R \div S$



# CÁC PHÉP TOÁN QUAN HỆ

■  $T = R \div S$

R ( HOTEN, NGOAINGU)

Lan Pháp

Mai Anh

Lan Anh

Lan Nga

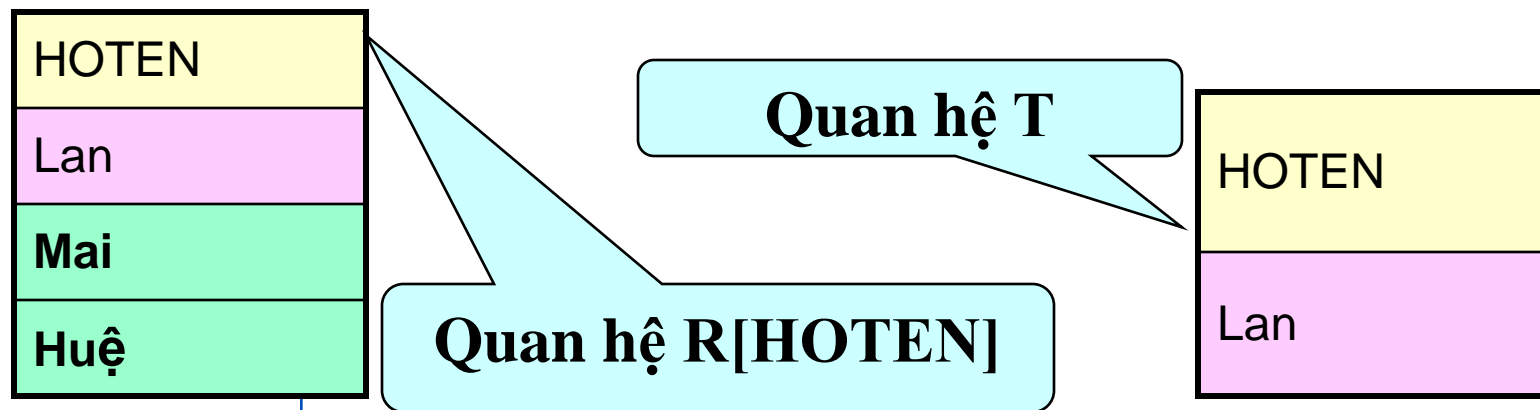
Hue Nga

S (NGOAINGU)

Anh

Pháp

Nga



## ■ Phép đổi tên quan hệ

- ◆ Đổi tên quan hệ R thành quan hệ S(A1, A2, ..., An)

$$\rho_S(A_1, A_2, \dots, A_n)(R)$$

- ◆ Quan hệ kết quả có các bộ giống các bộ của R, nhưng tên quan hệ là S
- ◆ Có thể đổi tên thuộc tính
- ◆ Nếu chỉ đổi tên quan hệ:

$$\rho_S(R)$$

# Tìm kiếm bằng đại số quan hệ

## ■ Biểu thức quan hệ

- ◆ Toán hạng là các quan hệ
- ◆ Toán tử là các phép toán quan hệ
- ◆ Kết quả là một quan hệ

## ■ Thứ tự ưu tiên

- ◆ Phép toán trong (..) có thứ tự ưu tiên cao hơn
- ◆ Đối với phép toán cùng ngôi, thứ tự ưu tiên là từ trái sang phải
- ◆ Phép toán một ngôi có thứ tự ưu tiên cao hơn phép toán hai ngôi.

# Tìm kiếm bằng đại số quan hệ

## ■ Ví dụ

**Cho CSDL gồm 3 quan hệ chứa dữ liệu về sinh viên của một khoa**

- ◆ SV (MASV, HOTEN, NS, GT, QUE, LOP)
- ◆ HP (MAHP, TENHP, SOTC)
- ◆ KQHT (MASV, MAHP, DIEM)

# TÌM KIẾM BẰNG ĐẠI SỐ QUAN HỆ

## ■ Ví dụ một trạng thái của CSDL

### Quan hệ SV

MASV	HOTEN	NS	GT	QUE	LOP
20051000	Nguyễn Hải Nam	1/1/1985	Nam	Hà Nội	Toán 1
20051001	Trần Mai Ly	2/2/1986	Nữ	Hà Tây	Toán 1
20051002	Lê Phương Bắc	3/3/1985	Nam	Nam Hà	Toán 2

### Quan hệ HP

MAHP	TENHP	SOTC
M1	Cơ sở dữ liệu	4
M2	Toán rời rạc	3
M3	Mạng MT	4

### Quan hệ KQHT

MASV	MAHP	DIEM
20051000	M1	4
20051001	M1	6
20051001	M2	7
20051001	M3	5

# TÌM KIẾM BẰNG ĐẠI SỐ QUAN HỆ

- Cho biết mã số và họ tên những sinh viên lớp 'Toán 1'

$$\Pi_{MASV, HOTEN}(\sigma_{LOP='Toan1'}(SV))$$

- Cho biết mã số và họ tên những sinh viên lớp 'Toán 1', quê ở Hà Nội

- Cho biết mã số và họ tên những sinh viên lớp 'Toán 1' đã có điểm học phần 'M1'

$$R = \prod_{MASV, HOTEN} (\sigma_{MAHP='M1' \wedge LOP='Toan1'}(SV \bowtie KQHT))$$

- Cho biết mã số và họ tên những sinh viên lớp 'Toán 1' không có điểm học phần 'M1'

$$\prod_{MASV, HOTEN} (\sigma_{LOP='Toan1'}(SV)) - R$$

- hoặc viết:

$$R1 = \prod_{MASV} (\sigma_{LOP='Toan1'}(SV)) - \prod_{MASV} (\sigma_{MAHP='M1'}(KQHT))$$

$$R2 = \prod_{MASV, HOTEN} (\sigma_{LOP='Toan1'}(SV))$$

$$ANSWER = R1 \bowtie R2$$

# VÍ DỤ VỀ TÌM KIẾM BẰNG ĐẠI SỐ QUAN HỆ

- Cho biết mã số và họ tên những sinh viên lớp 'Toán 1' có điểm học phần 'M1' lớn hơn hoặc bằng 4.

$$R1 = \prod_{MASV, HOTEN} (\sigma_{LOP='Toan1'}(SV))$$

$$R2 = \prod_{MASV} (\sigma_{MAHP='M1' \wedge DIEM \geq 4}(KQHT))$$

$$ANSWER = R1 \bowtie R2$$

- *Các cách biểu diễn khác?*



# TÌM KIẾM BẰNG ĐẠI SỐ QUAN HỆ

- Cho biết mã số và họ tên những sinh viên lớp 'Toán 1' có điểm học phần 'Cơ sở dữ liệu' lớn hơn hoặc bằng 4.

$$R1 = SV \bowtie_{KQHT} HP$$

$$ANSWER = \prod_{MASV, HOTEN} (\sigma_{LOP='Toan1' \wedge TENHP='Co so du lieu' \wedge DIEM \geq 4}(R1))$$

- *Thay đổi thứ tự phép kết nối?*
- *Các cách biểu diễn khác?*

# TÌM KIẾM BẰNG ĐẠI SỐ QUAN HỆ

- Cho biết mã số sinh viên có điểm < 4 ở cả hai học phần mã số M1 và M2

? Cách biểu diễn sau đúng hay sai?

$$ANSWER = \prod_{MASV} (\sigma_{DIEM < 4 \wedge MAHP = 'M1' \wedge MAHP = 'M2'}(KQHT))$$

$$R1 = \prod_{MASV} (\sigma_{DIEM < 4 \wedge MAHP = 'M1'}(KQHT))$$

$$R2 = \prod_{MASV} (\sigma_{DIEM < 4 \wedge MAHP = 'M2'}(KQHT))$$

$$ANSWER = R1 \bowtie R2$$

$$\text{Hoặc: } ANSWER = R1 \cap R2$$

# TÌM KIẾM BẰNG ĐẠI SỐ QUAN HỆ

- Cho biết danh sách mã số và họ tên những sinh viên lớp Toán 1 và Toán 2 có điểm học phần CSDL < 4

$$R1 = SV \triangleright \triangleleft KQHT \triangleright \triangleleft HP$$

$$ANS = \prod_{MASV, HOTEN} (\sigma_{(LOP='Toan1' \vee LOP='Toan2') \wedge DIEM < 4 \wedge TENHP='CSDL'}(R1))$$

Cách biểu diễn khác:

$$R1 = \prod_{MASV, HOTEN} (\sigma_{LOP='Toan1' \vee LOP='Toan2'}(SV))$$

$$R2 = \prod_{MASV} (\sigma_{DIEM < 4 \wedge TENHP='CSDL'}(HP \triangleright \triangleleft KQHT))$$

$$ANSWER = R1 \triangleright \triangleleft R2$$

Cách biểu diễn khác? .....

# TỐI ƯU HOÁ BIỂU THỨC QUAN HỆ

- **Tối ưu hóa là phương pháp thay một biểu thức quan hệ bằng một biểu thức tương đương đơn giản hơn, thời gian thực hiện nhanh hơn mà kết quả không đổi.**

## ■ **Nhận xét**

- ◆ Đối với quan hệ nhỏ, việc tính toán nhanh hơn, sử dụng ít bộ nhớ hơn
- ◆ Phép toán làm quan hệ nhỏ đi: chọn, trừ, giao, chiếu, chia
- ◆ Phép toán làm quan hệ lớn hơn: kết nối, hợp, tích Đề các

# TỐI ƯU HOÁ BIỂU THỨC QUAN HỆ

## ■ Chiến lược tối ưu hoá

- ◆ Thực hiện các phép toán làm quan hệ nhỏ đi càng sớm càng tốt
- ◆ Thực hiện các phép toán làm quan hệ lớn hơn càng về sau càng tốt

## ■ Ví dụ

- ◆ Cho biết danh sách mã số và họ tên những sinh viên lớp Toán 1 và Toán 2 thi không đạt (điểm  $< 4$ ) học phần CSDL

# TỐI ƯU HOÁ BIỂU THỨC QUAN HỆ

- Biểu thức quan hệ biểu diễn câu hỏi trên là:

$$\prod_{MASV, HOTEN} (\sigma_E (SV \triangleright \triangleleft KQHT \triangleright \triangleleft HP))$$

$$E = ( (LOP = \text{'Toán 1'} \vee LOP = \text{'Toán 2'}) \wedge DIEM < 4 \\ \wedge TENHP = \text{'CSDL'})$$

$$= E1 \wedge E2 \wedge E3$$

Trong đó:

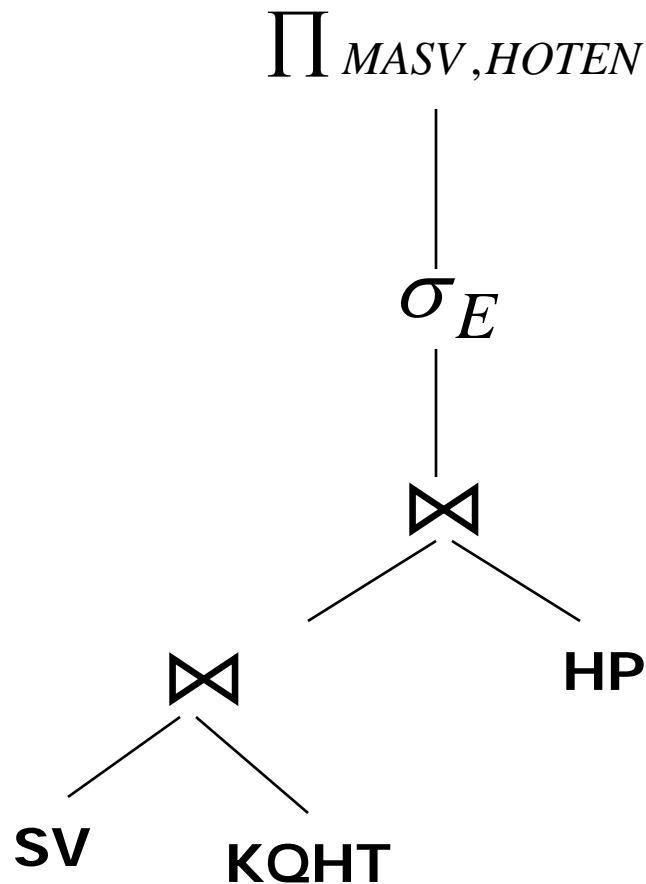
- ◆  $E1 = (LOP = \text{'Toán 1'} \vee LOP = \text{'Toán 2'})$
- ◆  $E2 = (DIEM < 4)$
- ◆  $E3 = (TENHP = \text{'CSDL'})$

# TỐI ƯU HOÁ BIỂU THỨC QUAN HỆ

## ■ Cây biểu thức quan hệ

- ◆ Các nút trong là các toán tử
- ◆ Các nút lá là các quan hệ
- ◆ Cây con là một biểu thức con

- $\Pi_{MASV, HOTEN}(\sigma_E(SV \triangleright \triangleleft KQHT \triangleright \triangleleft HP))$



*Biểu thức quan hệ tối ưu?*



# Phép toán quan hệ mở rộng

## ■ Phép toán tổng gộp (Aggregation Operators)

- ◆ SUM
- ◆ AVG
- ◆ MIN và MAX
- ◆ COUNT

## ■ Phép nhóm (Grouping Operator)

## ■ Phép sắp xếp (Sorting Operator)

## ■ Phép kết nối ngoài (Outer Join)

## ■ Phép nhóm (Grouping Operator)

$$\gamma_{A_1, A_2, \dots, A_k}(R)$$

- ◆ Phân hoạch các bộ của R thành các nhóm
- ◆ Tất cả các bộ trong một nhóm có giá trị giống nhau trên danh sách các thuộc tính  $L = \{A_1, A_2, \dots, A_k\}$  (các thuộc tính này gọi là “thuộc tính nhóm” – grouping attribute)
- ◆ Đối với mỗi nhóm, tạo ra một bộ bởi:
  - Giá trị trên thuộc tính nhóm của các bộ trong nhóm
  - Hàm tổng gộp, được tính toán trên các thuộc tính tổng gộp – aggregated attribute (thuộc danh sách L) của tất cả các bộ thuộc nhóm đó.

## ■ Phép nhóm (Grouping Operator)

$$\gamma_{MASV, \text{MAX}(\text{DIEM}) \rightarrow \text{MaxDiem}} (KQHT)$$

$$\gamma_{MASV, \text{COUNT}(\text{MAHP}) \rightarrow \text{SLHP}} (KQHT)$$

$$\gamma_{LOP, \text{COUNT}(MASV) \rightarrow SSSVLOP, \text{MIN}(NS) \rightarrow \text{MinNS}} (SV)$$

## ■ Phép sắp xếp (Sorting Operator)

$$\tau_L(R)$$

R: quan hệ

L: danh sách các thuộc tính

## ■ Phép kết nối ngoài (Outer join)