

CƠ SỞ DỮ LIỆU

Đề cương tóm tắt



CHƯƠNG I

CƠ SỞ DỮ LIỆU VÀ MỘT SỐ

MÔ HÌNH CƠ SỞ DỮ LIỆU

LỊCH SỬ CƠ SỞ DỮ LIỆU

- Những năm **60**: Mô hình **dữ liệu phân cấp** + mô hình dạng **mạng**
- Năm **1970**: Mô hình **quan hệ**
- Năm **1976**: Mô hình **thực thể liên kết**
- Những năm **80**: Mô hình **hướng đối tượng**
- Những năm **90**: Oracle Database, SQL Server, MySQL, PostgreSQL, SQLite
- Đầu thế kỷ **21**: NoSQL (MongoDB)

KHÁI NIỆM

- **Cơ sở dữ liệu** là tập hợp có tổ chức các dữ liệu có quan hệ logic với nhau
- Các loại dữ liệu:
 - + Dữ liệu **có cấu trúc**: bảng dữ liệu, hàng, cột,...
 - + Dữ liệu **không có cấu trúc**: hình ảnh, video, âm thanh,...
 - + Dữ liệu **bán cấu trúc**: email,...
- Hệ quản trị cơ sở dữ liệu là **hệ thống phần mềm** cho phép: tao lập cơ sở dữ liệu và điều khiển mọi truy nhập đối với cơ sở dữ liệu đó
- Chức năng của Hệ quản trị cơ sở dữ liệu:
 - + Định nghĩa
 - + Xây dựng
 - + Xử lý
 - + Chia sẻ
 - + Bảo vệ

Đề cương tóm tắt Cơ sở dữ liệu - SITDE PK

- Hệ cơ sở dữ liệu gồm 4 thành phần:
 - + Cơ sở dữ liệu hợp nhất
 - + Người sử dụng
 - + Hệ quản trị cơ sở dữ liệu
 - + Phần cứng

CÁC NGÔN NGỮ GIAO TIẾP GIỮA NGƯỜI SỬ DỤNG VÀ CSDL

- Ngôn ngữ định nghĩa dữ liệu (DDL)
- Ngôn ngữ thao tác dữ liệu (DML)
- Ngôn ngữ truy vấn có cấu trúc (SQL)
- Ngôn ngữ quản lý dữ liệu (DCL)

PHÂN LOẠI HỆ CƠ SỞ DỮ LIỆU

- Theo mô hình dữ liệu: mạng – phân cấp – quan hệ - hướng đối tượng
- Theo số người sử dụng: một người dùng – nhiều người dùng
- Theo tính phân tán: tập trung (cá nhân, trung tâm, khách/chủ) – phân tán
- Theo tính thống nhất của dữ liệu: đồng nhất – không đồng nhất

CÁC MỨC BIỂU DIỄN CỦA MỘT CƠ SỞ DỮ LIỆU

- Mức vật lý:
 - + Mức **thấp nhất của sự trừu tượng**
 - + Tồn tại trong **thiết bị lưu trữ** (lưu trữ ở **mức trong**)
 - + Dành cho **người quản trị** và **người sử dụng chuyên môn**
- Mức quan niệm:
 - + Định nghĩa cấu trúc logic của dữ liệu, dữ liệu nào được lưu trữ và mối quan hệ giữa chúng
 - + Là **biểu diễn trừu tượng** của mức vật lý
 - + Dành cho chuyên viên tin học khảo sát và phân tích, quản trị CSDL
- Mức khung nhìn (mức ngoài, lược đồ con):
 - + Mô tả cách người sử dụng nhìn thấy dữ liệu
 - + Mức **người sử dụng** và **chương trình ứng dụng**
 - + Nhìn thấy toàn bộ hoặc một phần CSDL theo góc độ khác nhau
 - + Chỉ có thể làm việc với **một phần** CSDL

CÁC MÔ HÌNH DỮ LIỆU

- Mô hình mức cao:
 - + Sử dụng các **khái niệm gần gũi, tự nhiên, giàu ngữ nghĩa**
 - + Gồm: Mô hình ERD, mô hình đối tượng
- Mô hình cài đặt: Gồm mô hình **quan hệ, mạng, phân cấp**
- Mô hình mức thấp

MÔ HÌNH DỮ LIỆU QUAN HỆ

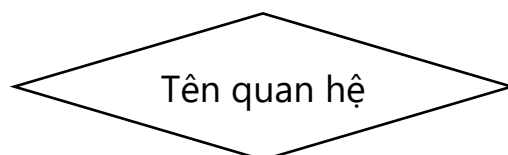
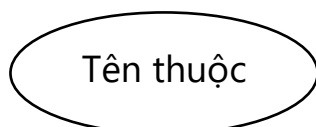
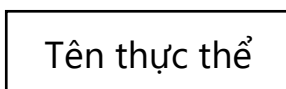
- Ra đời: Codd, năm **1970**
- Biểu diễn dưới dạng bảng
- Dựa trên lý thuyết toán học, gần với cấu trúc tệp, cấu trúc dữ liệu
- Thuộc tính: Một tính chất riêng biệt của một đối tượng
- Phân loại thuộc tính:
 - + Đơn trị
 - + Đa trị
 - + Kết hợp
 - + Suy diễn
- Miền: Là tập con của một kiểu dữ liệu. Ký hiệu $dom(A)$
- Quan hệ: Là một tập hữu hạn các thuộc tính
- Bộ: Là thông tin của một đối tượng thuộc quan hệ

Khi biểu diễn dưới dạng bảng: Bộ chính là các dòng, cột chính là các thuộc tính, quan hệ là các bảng

- Khóa: 2 loại chính: Khóa chính và khóa ngoại
- + Khóa chính: Đại diện đặc trưng cho một bộ dữ liệu, các giá trị là duy nhất
- + Khóa ngoại: Một hay nhiều thuộc tính của quan hệ này là khóa của quan hệ khác

MÔ HÌNH DỮ LIỆU LIÊN KẾT THỰC THỂ

- Ký hiệu:



- Loại liên kết:

+ 1 – 1

+ 1 – nhiều

+ nhiều – 1

+ nhiều – nhiều

CHƯƠNG II

MỘT SỐ NGÔN NGỮ KHAI THÁC CƠ SỞ DỮ LIỆU

ĐẠI SỐ QUAN HỆ

- Phép hợp: $r \cup s$:

Tập tất cả các bộ thuộc r hoặc thuộc s hoặc thuộc cả 2

Hiểu đơn giản: Gom hết các dòng của 2 bảng lại thành 1 bảng duy nhất, các dòng nào có dữ liệu giống nhau thì chỉ xuất hiện 1 lần.

r	A	B
	α	1
	α	2
	β	1

s	A	B
	α	2
	β	3

$r \cup s$	A	B
	α	1
	α	2
	β	1
	α	2
	β	3

- Phép giao: $r \cap s$:

Tập tất cả các bộ thuộc cả 2 quan hệ r và s

Hiểu đơn giản: Các dòng nào xuất hiện ở cả 2 bảng thì lấy.

Đề cương tóm tắt Cơ sở dữ liệu - SITDE PK

r	A	B
	α	1
	α	2
	β	1

s	A	B
	α	2
	β	3

$r \cap s$	A	B
	α	2

- Phép trừ: $r - s$:

Tập tất cả các bộ thuộc r nhng không thuộc s

Hiểu đơn giản: Lấy bảng r và loại bỏ đi những dòng của r mà cũng xuất hiện ở s.

Lưu ý: $r - s \neq s - r$

r	A	B
	α	1
	α	2
	β	1

s	A	B
	α	2
	β	3

$r - s$	A	B
	α	1
	β	1

- Tích Descartes (Đề-các): $r \times s$

Tập tất cả các $n + m$ bộ có n thành phần đầu tiên thuộc r và m thành phần sau là một bộ thuộc s

Hiểu đơn giản: Lấy từng dòng của r ghép với tất cả các dòng còn lại của s.

Đề cương tóm tắt Cơ sở dữ liệu - SITDE PK

r	A	B
α		1
β		2

s	X	C	D
α		10	+
β		10	+
β		20	-
γ		10	-

r x s	A	R.B	X	C	D
α		1	α	10	+
α		1	β	10	+
α		1	β	20	-
α		1	γ	10	-
β		2	α	10	+
β		2	β	10	+
β		2	β	20	-
β		2	γ	10	-

- Phép chiếu: $\pi_X(r)$

Tập các bộ của r xác định trên tập thuộc tính X

Hiểu đơn giản: Lấy ra các cột X từ bảng r

r	A	B	C
α		10	1
α		20	1
β		30	1
β		40	2

 $\pi_{A,C}(r)$

$\pi_{A,C}(r)$	A	C
α		1
β		1
β		2

- Phép chọn: $\sigma_F(r)$

Tập tất cả các bộ thuộc r thỏa mãn điều kiện F

Hiểu đơn giản: Lấy ra các dòng mà có dữ liệu khiến điều kiện F đúng

$$\sigma_{(A=B) \wedge (D > 5)}(r)$$

r	A	B	C	D
α	α		1	7
α	β		5	7
β	β		12	3
β	β		23	10

A	B	C	D
α	α	1	7
β	β	23	10

- Phép kết nối: $r \bowtie s$

Hiểu đơn giản: Ghép hai bảng lại với nhau dựa trên 2 cột có giá trị giống nhau hoặc thỏa mãn điều kiện nào đó

r	A	B	C
1		2	3
4		5	6
7		8	9

s	D	E
	3	1
	6	2

$$r \bowtie_{C=D} s$$

A	B	C	D	E
1	2	3	3	1
4	5	6	6	2

r	A	B	C	s	D	E
1		2	3		3	1
4		5	6		6	2
7		8	9			

$$r \bowtie_{B < D} s$$

A	B	C	D	E
1	2	3	3	1
1	2	3	6	2
4	5	6	6	2

- Phép chia: $r \div s$

Tập tất cả các bộ t sao cho với mọi bộ v $\in s$ thì t ghép với v thuộc r

r	A	B	C	D	E
	α	a	α	a	1
	α	a	γ	a	1
	α	a	γ	b	1
	β	a	γ	a	1
	β	a	γ	b	3
	γ	a	γ	a	1
	γ	a	γ	b	1
	γ	a	β	b	1

s	D	E
	a	1
	b	1

$r \div s$	A	B	C
	α	a	γ
	γ	a	γ

Ta thấy ở đây chỉ có hai bộ (A, B, C) thuộc r mà có đủ 2 dòng của s là:

+ Bộ (α, a, γ): ($\alpha, a, \gamma, a, 1$), ($\alpha, a, \gamma, b, 1$)

+ Bộ (γ, a, γ): ($\gamma, a, \gamma, a, 1$), ($\gamma, a, \gamma, b, 1$)

Vậy nên kết quả của phép $r \div s$ chính là hai bộ đó.

- Một số các hàm kết hợp:

Đề cương tóm tắt Cơ sở dữ liệu - SITDE PK

- + SUM: tính tổng
- + AVG: tính trung bình
- + MAX: tính giá trị lớn nhất
- + MIN: tính giá trị nhỏ nhất
- + COUNT: đếm số giá trị

PHÉP VỊ TỪ BIẾN BỘ, BIẾN MIỀN, NGÔN NGỮ ĐẠI SỐ QUAN HỆ THUẦN TÚY ISBL

Nhóm vị từ biến bộ

- $t[X]$: lấy ra một tập X các thuộc tính từ bảng t

Ví dụ: $HANGHOA[MAHANG, SOLUONG]$

- $t[X] \theta c$: lấy ra một tập X các thuộc tính từ bảng t thỏa mãn điều kiện " θc "

Ví dụ: $HANGHOA[MAU] = "Đỏ"$

Trong trường hợp này: θ là phép "="; c là "Đỏ"

\Rightarrow Lấy ra các hàng hóa có thuộc tính MAU là "Đỏ"

- $t[X] \theta s[Y]$: lấy ra một tập X các thuộc tính từ bảng t thỏa mãn điều kiện " θ " trong so sánh với tập Y các thuộc tính lấy ra từ bảng s

Ví dụ: $HANGHOA[TENHANG] = (HOADON[TENHANG] \wedge HOADON[MAU] = "Đỏ")$

\Rightarrow Lấy ra từ bảng $HANGHOA$ các hàng hóa có tên hàng trong bảng $HOADON$ và màu của hàng hóa đó trong bảng $HOADON$ là màu "Đỏ"

Nhóm vị từ biến miền

Là biểu thức có dạng $\{ \langle X_1, X_2, \dots, X_n \rangle \mid P(X_1, X_2, \dots, X_n) \}$

Trong đó $\langle X_1, X_2, \dots, X_n \rangle \in t$ là một tập các thuộc tính sao cho thỏa mãn điều kiện $P(X_1, X_2, \dots, X_n)$

Ví dụ: $\{ \langle MAHANG, TENHANG, MAU, KHOILUONG \rangle \mid \langle MAHANG, TENHANG,$

$MAU, KHOILUONG \rangle \in HANGHOA \wedge MAU = "Đỏ" \}$

\Rightarrow Lấy ra các bộ gồm các thuộc tính MAHANG, TENHANG, MAU, KHOILUONG sao cho các bộ này đều thuộc bảng $HANGHOA$ và MAU của bộ đó là "Đỏ"

Nhóm ngôn ngữ đại số quan hệ thuần túy ISBL

- Là cách viết khác của đại số quan hệ

Đại số quan hệ	Đại số quan hệ thuần túy (ISBL)
$R \cup S$	$R + S$
$R - S$	$R - S$
$R \cap S$	$R . S$
$\sigma_F(R)$	$R : F$
$\pi_{A_1, A_2, \dots, A_n}(R)$	$R \% A_1 . A_2, \dots A_N$
$R \bowtie S$	$R * S$

- Cú pháp: LIST <tên quan hệ>: <biểu thức đại số quan hệ thuần túy>

Ví dụ: LIST HANGHOA: MAU="Đỏ"

⇒ Liệt kê các bộ trong bảng HANGHOA có MAU là "Đỏ"

NGÔN NGỮ TRUY VẤN DỮ LIỆU CÓ CẤU TRÚC (SQL)

Riêng phần này, đi thi sẽ chỉ xoay quanh câu lệnh sau:

SELECT (DISTINCT) <X> FROM <Y> WHERE <P> GROUP BY <C> (HAVING <F>) ORDER BY <Z> ASC|DESC

Trong đó:

- DISTINCT: Loại bỏ các dòng trùng lặp
- X: Các cột cần hiển thị. Nếu muốn hiển thị tất cả các cột thì ghi *
- Y: Tên bảng để lấy dữ liệu
- P: Các biểu thức điều kiện trên các thuộc tính (Ví dụ MAU="Đỏ")
- GROUP BY: Nhóm theo C (C có thể là một hay nhiều cột), có thể kèm theo điều kiện F bằng cách sử dụng từ khóa HAVING
- ORDER BY: Sắp xếp các dòng trong kết quả truy vấn theo các cột/biểu thức Z. Mặc định (nếu không ghi gì) sẽ sắp xếp ASC (tăng dần). Sắp xếp giảm dần thì phải ghi DESC

Một số biểu thức khác có thể gặp:

- **LIKE "S%"**: Đây là biểu thức điều kiện, nghĩa là kiểm tra xem giá trị đó có bắt đầu bằng xâu S hay không. Ví dụ: LIKE "ABC%" thì các giá trị thỏa mãn có thể là ABCD, ABCABC, ABCDE (bắt đầu bằng ABC là thỏa mãn). Dấu % có thể thay bằng dấu *

Đề cương tóm tắt Cơ sở dữ liệu - SITDE PK

- **NOT X:** Biểu thức điều kiện. Nghĩa là “Phủ định của X” hay “Không phải X”
- **BETWEEN x AND y:** Biểu thức điều kiện. Kiểm tra xem giá trị có nằm trong đoạn $[x, y]$ không
- **X AS Y:** Thường gặp trong phần SELECT, nghĩa là lấy ra cột X nhưng không để tên là X nữa mà đổi tên cột thành Y

Ngoài ra, đi thi còn có thể vào thêm các câu lệnh sau:

- INSERT: Chèn một bộ dữ liệu vào quan hệ
- DELETE: Xóa một hoặc nhiều bộ dữ liệu khỏi quan hệ
- ALTER TABLE: Thay đổi cấu trúc bảng
- CREATE VIEW: Tạo khung nhìn

CHƯƠNG III

LÝ THUYẾT THIẾT KẾ CƠ SỞ DỮ LIỆU QUAN HỆ

KHÁI NIỆM

- Tập thuộc tính U: Tập hữu hạn khác rỗng bao gồm các thuộc tính

- Phụ thuộc hàm: Cho hai tập thuộc tính X, Y. Ta nói **X xác định hàm cho Y** hay **Y phụ thuộc hàm (vào) X** (ký hiệu $X \rightarrow Y$) nếu với mọi quan hệ r xác định trên tập thuộc tính chứa X, Y mà $u[X] = v[X]$ thì $u[Y] = v[Y]$

Ví dụ: MASINHVIENTENSINHVIENT

(Với mỗi MASINHVIENT phân biệt (lúc này X là MASINHVIENT) thì luôn chỉ có một TENSINHVIENT tương ứng (lúc này Y là TENSINHVIENT) dù X và Y có nằm ở bảng u hay v bất kỳ

- Lược đồ quan hệ R: Là cặp (U, F) trong đó:

U là tập thuộc tính

F là tập các phụ thuộc hàm xác định trên U

PHỤ THUỘC HÀM

- Suy diễn logic: $X \rightarrow Y$ được gọi là suy diễn logic từ F nếu với mỗi quan hệ R bất kỳ xác định trên $R(U)$ thỏa mãn các phụ thuộc hàm trong tập F đã cho thì cũng thỏa mãn $X \rightarrow Y$.

Ví dụ: Cho $R=(U, F)$. $U = \{A, B, C\}$; $F = \{A \rightarrow B; B \rightarrow C\}$

Khi đó $A \rightarrow C$ được suy diễn logic từ F

- Bao đóng: Là tập tất cả các phụ thuộc hàm được suy diễn logic từ F . Ký hiệu: F^+

Hệ tiên đề Amstrong

- Phản xạ: Nếu $Y \subseteq X \subseteq U$ thì $X \rightarrow Y$
- Tăng trưởng: Nếu $X \rightarrow Y$; và có một $Z \subseteq U$ thì $XZ \rightarrow YZ$
- bắc cầu: Nếu $X \rightarrow Y$; $Y \rightarrow Z$ thì $X \rightarrow Z$

Một số tính chất khác

Tính tựa bắc cầu: Nếu $X \rightarrow Y, YZ \rightarrow W$ thì $XZ \rightarrow W$

Tính phản xạ chặt: $\forall X \subseteq U$ thì $X \rightarrow X$

Tính mở rộng trái: $\forall X \subseteq U$, nếu $X \rightarrow Y$, $\forall Z \subseteq U$ thì $XZ \rightarrow Y$

Tính thu hẹp phải: $\forall X, Y, Z \subseteq U$, nếu $X \rightarrow YZ$ thì $X \rightarrow Y, X \rightarrow Z$

Tính cộng: $\forall X, Y, Z \subseteq U$

Nếu $X \rightarrow Y$ và $X \rightarrow Z$ thì $X \rightarrow YZ$,

Nếu $X \rightarrow Y$ và $Z \rightarrow W$ thì $XZ \rightarrow YW$

TÍNH TOÁN BAO ĐÓNG

Phương pháp tìm bao đóng

Cho một lược đồ $R=(U,F)$, tập thuộc tính U và tập phụ thuộc hàm F . Tìm tập bao đóng $\{A\}^+$ với $A \subseteq U$

Giải

Đề cương tóm tắt Cơ sở dữ liệu - SITDE PK

Trước tiên, ta quy ước với nhau một phụ thuộc hàm $X \rightarrow Y$ bất kỳ sẽ có phần X là **phần vế trái**, phần Y là **phần vế phải**

Bước 1: Tìm trong tập phụ thuộc hàm F một phụ thuộc hàm đầu tiên mà A **chứa** phần vế trái. Nếu tìm thấy thì sang bước 2, không thấy thì sang bước 3

Bước 2: Ta gom **phần vế phải** của phụ thuộc hàm vừa tìm được vào A. Lúc này ta có một A mới và quay lại bước 1

Bước 3: Tập A lúc này chính là tập bao đóng $\{A\}^+$ cần tìm.

Ví dụ: Cho quan hệ $R=(U, F)$; trong đó $U = \{A, B, C, D\}$; $F = \{A \rightarrow B, AB \rightarrow C, BC \rightarrow D\}$. Tìm $\{AB\}^+$

Giải

Ở đây ta cần tìm bao đóng của AB. Ta làm theo phương pháp trên

1. Tìm trong tập F phụ thuộc hàm đầu tiên mà AB chứa phần vế trái: Ở đây ta phát hiện $A \rightarrow B$ thỏa mãn điều kiện.
2. Lấy phần vế phải gộp vào AB: Ở đây vế phải bằng B, AB đã có B rồi, nên ta bỏ qua
3. Tiếp tục tìm trong tập F: Tiếp tục thấy $AB \rightarrow C$ thỏa mãn
4. Lấy phần vế phải gộp vào AB: Ở đây vế phải bằng C, AB chưa có C. Ta thu được tập mới là ABC
5. Tìm trong tập F phụ thuộc hàm đầu tiên mà ABC chứa phần vế trái: Lúc này trong tập F chỉ còn $BC \rightarrow D$, và ta thấy thỏa mãn vì ABC có chứa BC.
6. Lấy phần vế phải gộp vào ABC: Ta được ABCD
7. Lúc này tập F đã được xét hết. Ta dừng lại.

Vậy bao đóng $\{AB\}^+ = \{ABCD\}$

HAI TẬP PHỤ THUỘC HÀM TƯƠNG ĐƯƠNG

Cho hai tập phụ thuộc hàm F và G. Hai tập này gọi là tương đương nếu tập $X \rightarrow Y$ bất kỳ thuộc F thì đều có thể suy diễn logic ra từ G và ngược lại. Ký hiệu: $F \approx G$

Ví dụ: Cho tập $F = \{A \rightarrow BC; A \rightarrow D; CD \rightarrow E\}$ và $G = \{A \rightarrow BCD; A \rightarrow ABD; CD \rightarrow E\}$. Kiểm tra xem F có tương đương với G không?

Giải

- Xét F:

Đề cương tóm tắt Cơ sở dữ liệu - SITDE PK

+ $A \rightarrow BC$: Ta thấy có thể suy ra được điều này từ $A \rightarrow BCD$ thuộc G

+ $A \rightarrow D$: Ta thấy có thể suy ra được điều này từ $A \rightarrow ABD$ thuộc G

+ $CD \rightarrow E$: Phụ thuộc hàm này cũng thuộc G

- Tương tự kiểm tra G

=> Ta thấy rằng thỏa mãn định nghĩa hai tập phụ thuộc hàm tương đương.

Vậy F tương đương G

PHỦ TỐI THIỂU CỦA TẬP CÁC PHỤ THUỘC HÀM

- Phụ thuộc hàm đầy đủ: Cho phụ thuộc hàm $X \rightarrow Y$. Phụ thuộc hàm này gọi là đầy đủ nếu loại bỏ bất kỳ thuộc tính A nào đó ra khỏi X thì phụ thuộc hàm đó không còn đúng nữa và không có thuộc tính dư thừa.

Phương pháp tìm phủ tối thiểu

Cho $R = (U, F)$. Yêu cầu tìm phủ tối thiểu của tập F.

Giải

Trước tiên, ta quy ước với nhau một phụ thuộc hàm $X \rightarrow Y$ bất kỳ sẽ có phần X là **phần vế trái**, phần Y là **phần vế phải**

Bước 1: Chắc chắn rằng tất cả các phụ thuộc hàm trong F đều là **phụ thuộc hàm đầy đủ**. Nếu tồn tại một phụ thuộc hàm bất kỳ chưa phải là đầy đủ thì phải biến đổi nó về dạng đầy đủ.

Bước 2: Tách tất cả các phụ thuộc hàm thành dạng phụ thuộc hàm đơn (vế trái và vế phải chỉ có **duy nhất một** thuộc tính)

Bước 3: Loại bỏ tất cả các phụ thuộc hàm **có thể suy diễn logic ra** từ các phụ thuộc hàm khác và các phụ thuộc hàm bị trùng lặp dư thừa.

Bước 4: Tập phụ thuộc hàm còn lại chính là phủ tối thiểu cần tìm.

Ví dụ: Cho $R = (U, F)$ với $F = \{AB \rightarrow CD, B \rightarrow C, C \rightarrow D\}$

Giải

- Kiểm tra tính phụ thuộc hàm đầy đủ:

Ở đây ta thấy phụ thuộc hàm $AB \rightarrow CD$ chưa phải là đầy đủ do A là thuộc tính dư thừa. Do:

$B \rightarrow C; C \rightarrow D \Rightarrow B \rightarrow D$. Như vậy khi $B \rightarrow C$ và $B \rightarrow D$ thì $B \rightarrow CD$. A là không cần thiết.

Ta thay $AB \rightarrow CD$ bằng $B \rightarrow CD$. Khi đó $F = \{B \rightarrow CD; B \rightarrow C; C \rightarrow D\}$

- Tách các phụ thuộc hàm thành dạng đơn:

Ta được $F = \{B \rightarrow C; B \rightarrow D; B \rightarrow C; C \rightarrow D\}$

Đề cương tóm tắt Cơ sở dữ liệu - SITDE PK

- Loại bỏ các phụ thuộc hàm có thể suy diễn logic và bị trùng lặp:

Ta thấy rằng có thể loại bỏ $B \rightarrow D$ vì $B \rightarrow C$; $C \rightarrow D$ thì có thể suy ra $B \rightarrow D$.

Loại bỏ bớt 1 $B \rightarrow C$ vì bị trùng.

Vậy phủ tối thiểu của F là $F = \{B \rightarrow C; B \rightarrow D\}$

KHÓA CỦA MỘT LƯỢC ĐỒ QUAN HỆ

Phương pháp xác định MỘT khóa của một lược đồ quan hệ

Cho $R = (U, F)$. $U = \{X_1, X_2, \dots, X_n\}$ Tìm MỘT khóa của lược đồ quan hệ R?

Giải

Bước 1: Gán $K = U$

Bước 2: Lần lượt tính $(K \setminus X_i)^+$ với $i = 1, 2, \dots, n$. Nếu bao đóng thu được đúng bằng U^+ thì ta loại X_i đang xét ra khỏi K.

Bước 3: K còn lại chính là một khóa của lược đồ quan hệ R

Ví dụ: Cho $R = (U, F)$ với $U = \{ABCDE\}$; $F = \{AB \rightarrow C, AC \rightarrow B, BC \rightarrow DE\}$. Tìm một khóa của R xác định trên U và F?

Giải

Bước 1: Gán $K=U$ tức là $K=ABCDE$

Bước 2: Tính $(K \setminus A)^+ = (BCDE)^+ = BCDE \neq U^+$ nên $K=ABCDE$

Bước 3: Tính $(K \setminus B)^+ = (ACDE)^+ = ABCDE = U^+$ nên loại B ra tập K ban đầu
 $K=ACDE$

Bước 4: Tính $(K \setminus C)^+ = (ADE)^+ = ADE \neq U^+$ nên không bỏ C ra tập K ta có
 $K=ACDE$

Bước 5: Tính $(K \setminus D)^+ = (ACE)^+ = ACEBD = U^+$ nên bỏ D ra tập K ta có
 $K=ACE$

Bước 6: Tính $(K \setminus E)^+ = (AC)^+ = ACBDE = U^+$ nên bỏ E ra tập K ta có $K=AC$
 \Rightarrow Kết quả Khóa là $K=AC$

Phương pháp xác định TẤT CẢ các khóa của một lược đồ quan hệ

Cho $R = (U, F)$. $U = \{X_1, X_2, \dots, X_n\}$ Tìm TẤT CẢ khóa của lược đồ quan hệ R?

Giải

Bước 1: Xác định tất cả các tổ hợp (tất cả tập con) của U.

Bước 2: Lần lượt xét từng tổ hợp. Giả sử tổ hợp đang xét là X_i .

Ta tính $(X_i)^+$. Nếu $(X_i)^+ = U$ thì ta thêm X_i vào tập S

Bước 3: Xét tập S thu được sau khi xét hết các tổ hợp. Nếu chọn ra một S_i mà có một S_j là **con** của S_i (hay S_i chứa S_j) thì loại S_i khỏi S .

Bước 4: Các phần tử còn lại trong S chính là tập khóa của R .

Ví dụ: Cho $R = (U, F)$ với $U = \{ABC\}$, $F = \{AB \rightarrow C, C \rightarrow A\}$. Tìm tất cả các khóa của R .

Giải

1. Các tổ hợp của U là: A, B, C, AB, AC, BC, ABC

2. Tính toán:

- $(A)^+ = \text{Rỗng}$
- $(B)^+ = \text{Rỗng}$
- $(C)^+ = \{CA\}$ khác U
- $(AB)^+ = \{ABC\} = U \Rightarrow S = \{AB\}$
- $(AC)^+ = \{AC\}$ khác U
- $(BC)^+ = \{ABC\} = U \Rightarrow S = \{AB, BC\}$
- $(ABC)^+ = \{ABC\} = U \Rightarrow S = \{AB, BC, ABC\}$

3. Xét S :

Ta thấy rằng, $AB \subset ABC$ và $BC \subset ABC$ nên loại ABC khỏi S . Lúc này $S = \{AB, BC\}$
 AB không phải con của BC và ngược lại. Do đó, không cần loại phần tử nào nữa.

4. Kết luận: Vậy tập khóa của lược đồ R là $K = \{AB, BC\}$

CHUẨN HÓA LƯỢC ĐỒ QUAN HỆ

Chuẩn hóa 1NF

- Các thuộc tính đều là nguyên tố (không thể chia nhỏ thành các phần/ý nghĩa nhỏ hơn)
- Không chứa thuộc tính đa trị (nhiều giá trị)
- Không có thuộc tính nào có thể tính toán được từ một hay một số thuộc tính khác

Chuẩn hóa 2NF

- Là 1NF
- Các thuộc tính không phải khóa phải **phụ thuộc hàm đầy đủ** vào khóa chính

Phương pháp kiểm tra xem $R = (U, F)$ có phải 2NF không?

Bước 1: Tìm tập khóa K của R .

Bước 2: Tính $(X_i)^+$ với X_i là tập con thực sự của các khóa K . (Tập con thực sự của U là tập hợp con của U mà không bao gồm U)

Bước 3: Nếu $(X_i)+$ chứa thuộc tính không khóa của R thì không phải 2NF, ngược lại thì R đã thỏa mãn 2NF.

Chuẩn hóa 3NF

- Là 2NF
- Các thuộc tính không phải khóa phải **phụ thuộc trực tiếp** vào khóa chính.

Phương pháp kiểm tra xem $R = (U, F)$ có phải 3NF không?

Bước 1: Tìm tập khóa K của R

Bước 2: Nếu có các phụ thuộc hàm $X \rightarrow Y$ trong đó Y chứa thuộc tính không khóa mà X là siêu khóa thì R đạt chuẩn 3NF, ngược lại thì R chưa thỏa mãn 3NF

NF	Cách nhận biết chưa thỏa mãn chuẩn	Cách chuẩn hóa
1	Quan hệ có thuộc tính đa trị hoặc quan hệ lặp	Loại bỏ các thuộc tính vi phạm 1NF và đặt chúng vào 1 bảng riêng cùng với khóa chính của quan hệ ban đầu. Khóa chính của bảng này là một tổ hợp của khóa chính ban đầu và thuộc tính đa trị hoặc khóa bộ phận của nhóm lặp. Các thuộc tính còn lại lập thành một quan hệ với khóa chính là khóa chính ban đầu
2	Phụ thuộc một phần vào thuộc tính khóa	Loại bỏ các thuộc tính không khóa phụ thuộc vào một phần của khóa chính và tách thành bảng riêng với khóa chính của bảng mới là một phần của khóa mà chúng phụ thuộc vào. Các thuộc tính còn lại lập thành một quan hệ với khóa chính là khóa chính ban đầu.
3	Phụ thuộc ẩn, tồn tại phụ thuộc hàm giữa các thuộc tính không phải là khóa	Loại bỏ các thuộc tính bắc cầu ra khỏi quan hệ và tách chúng thành một quan hệ riêng có khóa chính là thuộc tính bắc cầu. Các thuộc tính còn lại lập thành một quan hệ với khóa chính là khóa chính ban đầu.

CHƯƠNG IV

TỔ CHỨC DỮ LIỆU VẬT LÝ

MÔ HÌNH TỔ CHỨC BỘ NHỚ NGOÀI

- Bộ nhớ ngoài (bộ nhớ thứ cấp) có thể là đĩa từ, băng từ, ...
- Được chia thành các khối vật lý có kích thước khác nhau; mỗi tệp dữ liệu chiếm một khối; mỗi khối chứa nhiều bản ghi
- Các bản ghi đều có địa chỉ

TỔ CHỨC TẬP ĐỒNG

- Là tổ chức lưu trữ **đơn giản nhất** trong đó các bản ghi được lưu trữ kế tiếp nhau trong các khối **không tuân theo một thứ tự đặc biệt nào** và **không có tổ chức** nào áp dụng với các khối
- Các thao tác:
 - + Tìm kiếm: Quét toàn bộ hoặc ít nhất một nửa tập
 - + Thêm: Tìm bản ghi cần thêm trước để tránh trùng lặp
 - + Xóa: Tìm -> Đánh dấu xóa -> Xóa -> Tổ chức lại tập
 - + Sửa: Tìm -> Sửa
- **Luôn luôn tốn 4 bytes để móc nối dữ liệu. Cái này cần lưu ý kỹ để làm các bài tập hỏi về số khối cần để lưu bản ghi.**

Ví dụ: Cho quan hệ SINHVIEN(MASV, HODEM, TENS, NGAYSINH, QUEQUAN) với MASV chiếm 4 bytes, HODEM chiếm 20 bytes, TENS chiếm 10 bytes, NGAYSINH chiếm 16 bytes, QUEQUAN chiếm 25 bytes. Nếu mỗi khối có dung lượng 256 bytes thì cần bao nhiêu khối để lưu được 16 bản ghi?

Giải

Ta có tổng dung lượng lưu trữ của một bản ghi là: $4 + 20 + 10 + 16 + 25 + 4$ (**4 bytes móc nối dữ liệu**) = 79 bytes

Như vậy mỗi khối sẽ lưu được: $256 / 79 = 3$ (Là 3.24 nhưng phải **lấy phần nguyên**)

Vậy số khối cần để lưu được 16 bản ghi là: $16 / 3 = 5.333 = 6$ (**lấy hàm trần** - tức là chỉ cần có phần lẻ thì lấy phần nguyên rồi cộng 1)

TỔ CHỨC TẬP BĂM

- Phân chia tập hợp các bản ghi của tập dữ liệu thành các cụm (buckets)

+ Mỗi cụm: Một hoặc nhiều khối

+ Mỗi khối: Số lượng cố định các bản ghi

+ Tổ chức lưu trữ dữ liệu: Tổ chức đồng

+ Giá trị khóa: $[0; B-1]$ (B là số cụm, có thể tiến đến cực đại)

- Giá trị khóa: Sử dụng hàm băm

Mỗi bản ghi có một khóa là x , thì hàm băm $h(x)$ nhận một giá trị trong khoảng $[0, k]$ với k là một giá trị nguyên dương bất kỳ (thường k là số nguyên tố). Nên chọn một số k là số nguyên tố nhỏ nhất thỏa mãn $k \geq$ số cụm (B).

$$h(x) = x \bmod k \text{ (x chia cho k và lấy phần dư)}$$

- **Lưu ý: Luôn tốn 4 bytes cho con trỏ/bản ghi**

- Các thao tác:

+ Tìm kiếm:

- Tính hàm băm $h(x)$
- Tìm địa chỉ băm của cụm $h(x)$
- Thực hiện tìm kiếm của tổ chức đồng

+ Thêm:

- Tìm bản ghi X
- Thêm vào khối đầu tiên trong cụm còn chỗ trống (nếu không có thì tạo khối mới)

+ Xóa:

- Tìm bản ghi
- Loại bỏ bản ghi khỏi cụm

+ Sửa:

- Tìm bản ghi X
- Loại bỏ bản ghi này và thêm vào bản ghi mới

Ví dụ: Cho quan hệ NHACC(MACC, TENCC, DIACHI, SDT, SOKHACH) với MACC chiếm 4 bytes, TENCC chiếm 20 bytes, DIACHI chiếm 20 bytes, SDT chiếm 16 bytes và SOKHACH chiếm 18 bytes. Biết rằng mỗi khối có thể chứa được 256 bytes, NHACC có 16 bản ghi và quan hệ này được biểu diễn dưới dạng tệp băm có 5 cụm. Thực hiện thêm bản ghi có giá trị khóa 32 vào quan hệ này. Cho biết rằng bản ghi này sẽ nằm ở khối nào, thuộc cụm thứ bao nhiêu trong cấu trúc tệp băm?

Giải

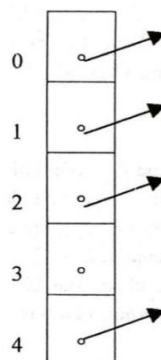
Trước tiên, ta tính số bytes một bản ghi cần dùng:

$$T = 4 + 20 + 20 + 16 + 18 + \mathbf{4 \text{ (4 bytes dành cho móc nối dữ liệu)}} = 82$$

Mỗi khối lưu được 256 bytes, do đó, mỗi khối lưu được $256 / 82 = 3$ bản ghi

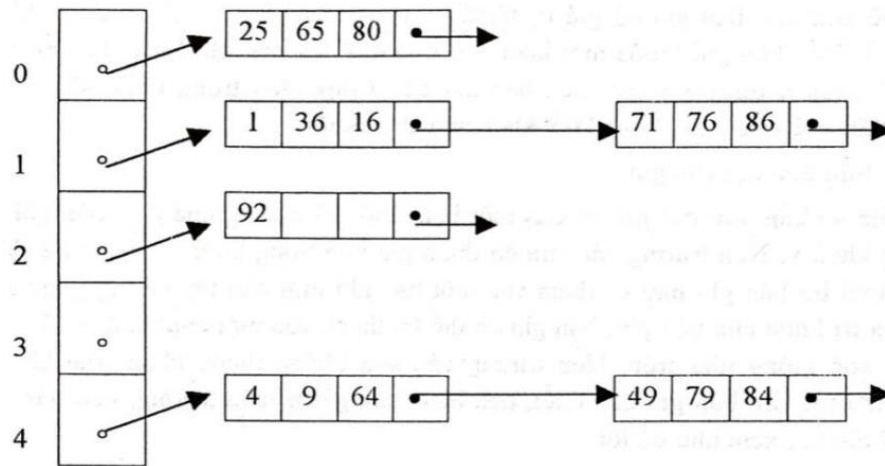
Mà NHACC có 16 bản ghi, do đó cần: $16 / 3 = 6$ khối

Tổ chức tệp băm này sẽ gồm 5 cụm, biểu diễn như sau:



Đề cương tóm tắt Cơ sở dữ liệu - SITDE PK

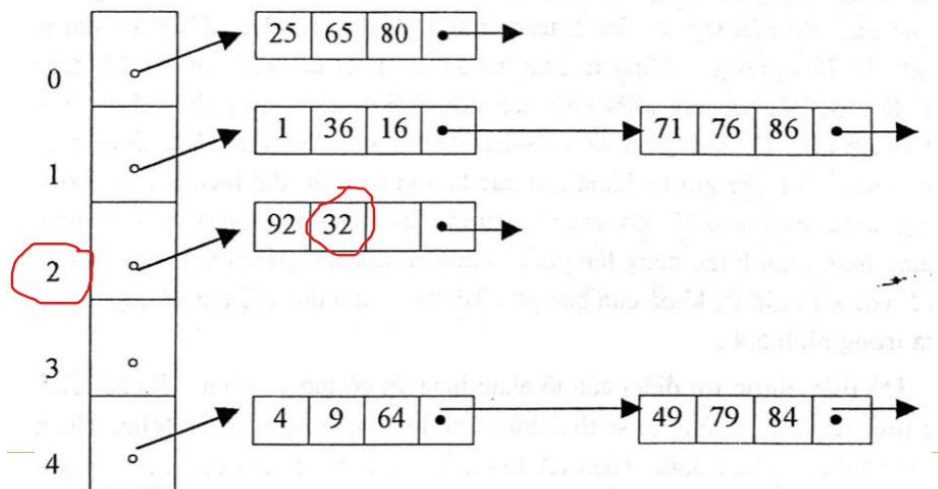
Giả sử khi ta thêm các bản ghi vào, ta sẽ được:



Bây giờ, ta chọn $k = 5$ (do 5 là số nguyên tố và $\geq B$). Khi đó hàm $h(x) = x \bmod 5$

Thực hiện thêm bản ghi có khóa là 32

Khi đó, $h(32) = 32 \% 5 = 2$. Nghĩa là ta sẽ đẩy bản ghi này vào cụm 2. Ở đây, ngay khối đầu tiên ta thấy vẫn còn ô trống, nên ta đẩy bản ghi này vào ô trống đó.



TỔ CHỨC TẬP CHỈ DẪN

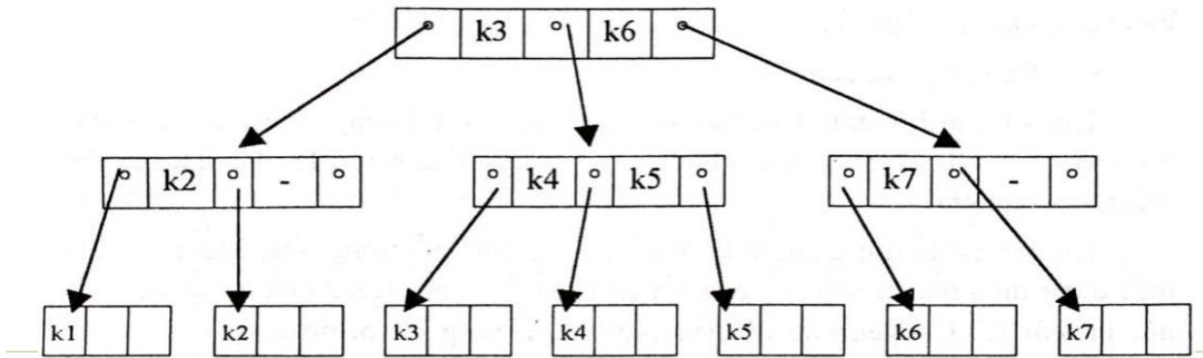
- Nhằm tạo ra tập mới từ tập dữ liệu ban đầu (tập chính) sao cho tập mới hỗ trợ tập ban đầu trong thực hiện một số thao tác cơ bản như thêm, cập nhật, sửa, xóa, ...
- Tập chính: Luôn được sắp dẫn
- Một chỉ dẫn (chỉ mục) của một tập có ý nghĩa như một danh mục để cải thiện tốc độ cho việc tìm kiếm và tra cứu

TỔ CHỨC B-CÂY (B-TREE)

- Tổ chức theo cấp m , thỏa mãn đồng thời các tính chất sau:
 - + Nút gốc của cây: Là nút lá hoặc ít nhất có 2 con
 - + Mỗi nút nhánh (trừ nút lá và nút gốc) có từ $\lceil m/2 \rceil$ đến m con

Đề cương tóm tắt Cơ sở dữ liệu - SITDE PK

+ Mỗi đường đi (từ gốc -> lá) có độ dài bằng nhau



CHƯƠNG V

TỐI ƯU HÓA TRUY VẤN

TỔNG QUAN

- 03 bước thực hiện xử lý truy vấn:
- + Phân tích và dịch
- + Tối ưu hóa
- + Đánh giá

PHÂN TÍCH VÀ DỊCH

- Kiểm tra xem câu truy vấn có đúng cú pháp hay không: Kết quả cho ra là một cây phân tích
- Kiểm tra ngữ nghĩa:
 - + Kiểm tra xem tên quan hệ xuất hiện trong truy vấn có phải tên của các quan hệ trong CSDL hay không
 - + Kiểm tra ngữ nghĩa giữa các quan hệ trong mệnh đề FROM với các thuộc tính trong các mệnh đề khác
 - + Kiểm tra kiểu dữ liệu có phù hợp với thuộc tính hay không, tên thuộc tính có nhập hàng hay không
- Đưa về dạng biểu diễn trong (biểu thức quan hệ): Câu truy vấn được phân rã thành các query block (khối truy vấn - QB)
 - + Một QB chứa một biểu thức (SELECT - FROM - WHERE - GROUP BY - HAVING)
 - + Các câu truy vấn lồng trong 1 câu truy vấn là các QB độc lập

TỐI ƯU HÓA TRUY VẤN

- Là tiến trình lựa chọn kế hoạch thực thi câu hỏi một cách hiệu quả nhất sao cho
- + Tốc độ thực thi câu truy vấn nhanh nhất có thể
- + Việc xử lý câu truy vấn chiếm dụng bộ nhớ ít nhất có thể

ĐÁNH GIÁ

- Để đặc tả đánh giá một câu hỏi:
- + Cung cấp biểu thức đại số quan hệ
- + Chú thích các chỉ thị đặc tả đánh giá mỗi phép toán

CÁC CHIẾN LƯỢC TỐI ƯU HÓA CÁC BIỂU THỨC ĐẠI SỐ QUAN HỆ

- Ưu tiên thực hiện các phép chọn và phép chiếu đầu tiên
- Tổ hợp các phép chọn xác định với tích Đề-các thành phép kết nối
- Xác định các biểu thức con chung trong một biểu thức
- Xử lý các tệp trước khi tính toán một phép kết nối
- Ước lượng chi phí và lựa chọn thứ tự thực hiện thích hợp

CHƯƠNG VI

AN TOÀN VÀ TOÀN VỆN DỮ LIỆU

AN TOÀN DỮ LIỆU

- Xác minh người sử dụng: Đảm bảo không cho phép người sử dụng thực hiện bất cứ thao tác nào trên CSDL nếu không có quyền làm điều đó
- Kiểm tra truy nhập
- Sử dụng các khung nhìn
- Một số lệnh an toàn dữ liệu trong SQL:
 - + CREATE VIEW <tên> [ds cột] AS <câu truy vấn>: Tạo khung nhìn cho người sử dụng

Đề cương tóm tắt Cơ sở dữ liệu - SITDE PK

- + GRANT <ds thao tác> ON <đối tượng> TO <ds người sử dụng>: Cấp quyền để thực hiện <ds các thao tác>
- + REVOKE <ds thao tác> ON <đối tượng> FROM <ds người sử dụng>: Tước quyền thực hiện <ds các thao tác>

TOÀN VỆN DỮ LIỆU

- Ràng buộc miền: CONSTRAINT <tên ràng buộc> CHECK <điều kiện>
- Ràng buộc khóa: CONSTRAINT <tên ràng buộc> [PRIMARY KEY <ds khóa chính>] / [UNIQUE KEY <ds khóa duy nhất>] / [FOREIGN KEY <ds khóa ngoại>]
- Các khẳng định: CREATE ASSERTION <tên khẳng định> CHECK <vị từ>
- Các kích hoạt: Một số nguyên tắc:
 - + Chỉ được kiểm tra khi các sự kiện, đặc tả xuất hiện
 - + Ngăn chặn lập tức sự kiện vừa phát sinh bằng một kích hoạt kiểm tra điều kiện
 - + Khi điều kiện kích hoạt thỏa mãn, hành động kết hợp với kích hoạt được thực hiện bởi DBMS