

Chương 2

Thiết kế cơ sở dữ liệu logic

Giáo viên: Nguyễn Mậu Uyên

Bộ môn: Hệ thống thông tin

Biên soạn: Đỗ Thị Mai Hương

Nội dung

- ❖ Các khái niệm cơ bản
- ❖ Các phương pháp thiết kế cơ sở dữ liệu logic

Các khái niệm cơ bản

- ❖ Mô hình CSDL quan hệ hay ngắn gọn là mô hình quan hệ được E.F.Codd phát triển vào đầu những năm 1970.
- ❖ Các thành phần trong mô hình quan hệ gồm: Các quan hệ - các bộ - các thuộc tính.

Các khái niệm cơ bản

❖ Một quan hệ:

Là một bảng dữ liệu 2 chiều (cột và dòng), mô tả thực thể. Mỗi cột tương ứng với một thuộc tính của thực thể. Mỗi dòng chứa các giá trị dữ liệu của một đối tượng cụ thể thuộc thực thể.

❖ Mô hình quan hệ:

Là cách thức biểu diễn dữ liệu dưới dạng các quan hệ (các bảng).

❖ Lược đồ quan hệ:

$R(A_1, \dots, A_n)$, trong đó R là tên quan hệ, A_i là các thuộc tính, mỗi A_i có miền giá trị tương ứng $\text{Dom}(A_i)$.

Lược đồ quan hệ được sử dụng để mô tả một quan hệ, bao gồm: Tên quan hệ, các thuộc tính và bậc của quan hệ (số lượng các thuộc tính)

Các khái niệm cơ bản

❖ Một quan hệ r của $R(A_1, \dots, A_n)$, ký hiệu $r(R)$ là một tập hợp m -bộ $r = \{ t_1, \dots, t_m \}$

Mỗi $t_i = \langle v_1, \dots, v_n \rangle$, $v_i \in \text{dom}(A_i)$. $r(R) \subseteq \text{dom}(A_1) \times \dots \times \text{dom}(A_n)$

$r = \{ (v_{i1}, v_{i2}, \dots, v_{in}) / i=1, \dots, m \}$

❖ Ta có A_i là các thuộc tính

❖ Miền giá trị của A_i là: $D_1 = \text{dom}(A_1)$, $D_2 = \text{dom}(A_2)$, ..., $D_n = \text{dom}(A_n)$.

❖ Các tập (D_1, D_2, \dots, D_n) là tập các miền trị của R

❖ n được gọi là bậc của quan hệ r , m được gọi là lực lượng của r .

Các khái niệm cơ bản

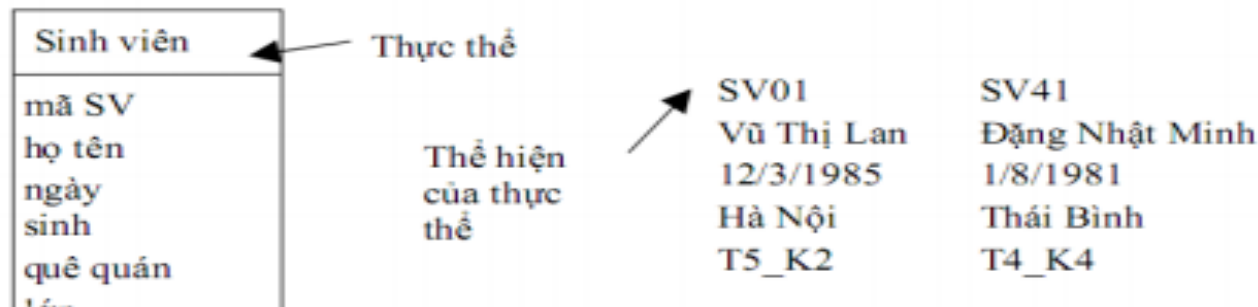
❖ Sự tương ứng giữa mô hình liên kết thực thể và mô hình quan hệ

Mô hình thực thể liên kết	Mô hình quan hệ	Các bảng trong hệ QTCSDL
Thực thể	Quan hệ	Bảng
Thể hiện của thực thể	Bộ	Dòng hay bản ghi
Thuộc tính	Thuộc tính	Cột hay trường

Các khái niệm cơ bản

Ví dụ:

- Mô hình thực thể liên kết:



- Biểu diễn quan hệ dưới dạng mô hình quan hệ :

Quan hệ **SINHVIÊN** (mã SV, họ tên, ngày sinh, quê quán, lớp)

Các bộ: (SV01, Vũ Thị Lan, 12/3/1985, Hà Nội, T5_K2)

(SV41, Đặng Nhật Minh, 1/8/1981, Thái Bình, T4_K4)

- Biểu diễn quan hệ dưới dạng bảng:

SINH VIÊN	mã SV	họ tên	ngày sinh	quê quán	lớp
	SV01	Vũ Thị Lan	12/3/1985	Hà Nội	T5_K2
	SV41	Đặng Nhật Minh	1/8/1981	Thái Bình	T4_K4

Các khái niệm cơ bản

- ❖ Các tính chất của một quan hệ
 - Giá trị đưa vào cột là đơn nhất
 - Các giá trị trong cùng một cột phải thuộc cùng một miền giá trị (cùng kiểu)

Các khái niệm cơ bản

❖ Mỗi quan hệ:

- Thể hiện quan hệ giữa các bảng với nhau.

❖ Các loại mỗi quan hệ:

- Trong mô hình quan hệ chỉ có 2 loại mỗi quan hệ là 1-1 và 1-n.

❖ Ví dụ:

- Sinhvien và Thethuvien
- Lop và Sinhvien

Các khái niệm cơ bản

❖ Khoá chính của một quan hệ (Primary Key-PK):

- Là một hoặc một nhóm thuộc tính xác định duy nhất một bộ trong quan hệ. Khoá chính của quan hệ là định danh của thực thể tương ứng.

❖ Khoá ngoài:

- Một khoá ngoài được sử dụng để thiết lập một mối quan hệ. Đó là thuộc tính mô tả của quan hệ này nhưng đồng thời lại là thuộc tính khoá trong quan hệ khác.

❖ Ví dụ:

Lop(Malop, Tenlop, Siso)

Sinhvien(MaSV, Hoten, NS, GT, *Malop*)

Các khái niệm cơ bản

❖ Các ràng buộc trong mô hình quan hệ:

- Ràng buộc là những quy tắc được áp đặt lên trên dữ liệu đảm bảo tính tin cậy và độ chính xác của dữ liệu. Các luật toàn vẹn được thiết kế để giữ cho dữ liệu phù hợp và đúng đắn.

❖ Các loại ràng buộc:

- Ràng buộc thực thể: là một ràng buộc trên khoá chính. Nó yêu cầu khoá chính phải tối thiểu, xác định duy nhất và không null

Các khái niệm cơ bản

❖ Các loại ràng buộc:

- Ràng buộc tham chiếu (ràng buộc khoá ngoài): liên quan đến tính toàn vẹn của mỗi quan hệ tức là liên quan đến tính toàn vẹn của khoá ngoài.
- Một ràng buộc tham chiếu yêu cầu một giá trị khoá ngoài trong một quan hệ cần phải tồn tại là một giá trị khoá chính trong một quan hệ khác hoặc là giá trị null.

SINH VIÊN	mã SV	tên SV	tên lớp
	SV01	Đỗ thị Cúc	T1-K3
	SV52	Vũ Thu Hà	
	SV34	Hoàng Anh	<u>T4-K7</u>

Vi phạm ràng buộc tham chiếu

LỚP	tên lớp	phòng học
	T1_K3	302
	T3_K4	414

Các khái niệm cơ bản

❖ Ràng buộc miền giá trị:

- Là một hợp các kiểu dữ liệu và những giá trị giới hạn mà thuộc tính có thể nhận được.
- Thông thường việc xác định miền giá trị của các thuộc tính bao gồm một số các yêu cầu sau: Kiểu dữ liệu, Độ dài dữ liệu, khuôn dạng của dữ liệu, các giá trị giới hạn cho phép, có duy nhất hay không, có cho phép giá trị rỗng hay không.

Các phương pháp thiết kế cơ sở dữ liệu logic

❖ Phương pháp Top-down

- Tiếp cận theo hướng mô hình liên kết thực thể, sau đó áp dụng các quy tắc chuyển đổi mô hình liên kết thực thể sang mô hình quan hệ.

❖ Các bước xây dựng mô hình liên kết thực thể

- Xác định các thực thể.
- Xác định các thuộc tính và gắn thuộc tính vào mỗi liên kết.
- Xác định mỗi liên kết và loại mỗi liên kết giữa các thực thể. Gắn thuộc tính vào mỗi liên kết (nếu có).
- Vẽ mô hình liên kết thực thể.

Các phương pháp thiết kế cơ sở dữ liệu logic

❖ Phương pháp Top-down

- Chuyển đổi từ mô hình liên kết thực thể sang mô hình quan hệ
 - Tên thực thể -> Tên lược đồ quan hệ (Tên bảng)
 - Thuộc tính -> Thuộc tính
 - Thuộc tính khóa -> Khóa chính
 - Thuộc tính đa trị -> chuyển thành các thuộc tính đơn trị.
 - Thực thể yếu -> Lược đồ quan hệ (có khóa là thuộc tính khóa của bảng cha và một thuộc tính định danh của nó)

Các phương pháp thiết kế cơ sở dữ liệu logic

❖ Phương pháp Top-down

❖ Chuyển đổi từ mô hình liên kết thực thể sang mô hình quan hệ

■ Đối với mỗi liên kết:

- 1-1: Có 2 cách chuyển đổi:
 - C1: Lấy khóa chính của bảng này sang làm khóa ngoại bảng kia hoặc ngược lại
 - C2: Gộp 2 bảng làm 1, xác định lại thuộc tính khóa
- 1-n: Lấy thuộc tính khóa chính của bảng cha sang làm thuộc tính khóa ngoại của bảng con.
- n-n: Chuyển mỗi liên kết thành lược đồ quan hệ mới, thuộc tính là thuộc tính của mỗi liên kết, khóa chính là tổ hợp khóa của hai bảng liên quan

Các phương pháp thiết kế cơ sở dữ liệu logic



❖ Phương pháp Bottom - up

- Nhóm tất cả các thuộc tính liên quan của hệ thống vào trong một quan hệ.
- Áp dụng các luật chuẩn hoá để tách quan hệ đó thành các quan hệ có cấu trúc tốt hơn, giảm bớt dư thừa dữ liệu.

Các phương pháp thiết kế cơ sở dữ liệu logic

❖ Phương pháp Bottom – up

❖ Phụ thuộc hàm

- Trong một quan hệ R , thuộc tính Y phụ thuộc hàm vào thuộc tính X (hay thuộc tính X xác định hàm thuộc tính Y) ký hiệu $X \rightarrow Y$ nếu với mỗi giá trị của thuộc tính X xác định một giá trị duy nhất của thuộc tính Y .
- $X \rightarrow Y$ trong r nếu với mọi t và t' của r mà t, t' bằng nhau trên tập X thì chúng cũng bằng nhau trên tập Y , tức là $\forall t, t' \in r$ nếu $t.X = t'.X \Rightarrow t.Y = t'.Y$

Các phương pháp thiết kế cơ sở dữ liệu logic

❖ Các loại phụ thuộc hàm

■ Phụ thuộc hàm đầy đủ:

- Thuộc tính Y gọi là phụ thuộc đầy đủ vào tập thuộc tính X (có từ 2 thuộc tính trở lên) nếu nó chỉ phụ thuộc hàm vào X và không phụ thuộc hàm vào bất cứ tập con nào của X. Ngược lại Y gọi là phụ thuộc hàm bộ phận vào tập thuộc tính X.

■ Phụ thuộc hàm bắc cầu:

- Nếu có $A_1 \rightarrow A_2$ và $A_2 \rightarrow A_3$ thì $A_1 \rightarrow A_3$. Khi đó A_3 được gọi là phụ thuộc bắc cầu vào A_1 .

Các phương pháp thiết kế cơ sở dữ liệu logic

❖ Hệ tiên đề Amstrong

- A1. Tính phản xạ

$X \rightarrow X$, hay tổng quát hơn nếu $Y \subset X$ thì $X \rightarrow Y$

- A2. Tính mở rộng hai vế

$X \rightarrow Y$ thì $XZ \rightarrow YZ$. (Mở rộng hai vế Z)

- A3. Tính bắc cầu: $X \rightarrow Y$ và $Y \rightarrow Z$ thì $X \rightarrow Z$.

Các phương pháp thiết kế cơ sở dữ liệu logic

❖ Bao đóng và thuật toán tìm bao đóng

- $X^+ = \{A: A \in R \text{ và } X \rightarrow A \in F^+\}$

hoặc $X^+ = \{X \cup A: A \in R, A \notin X \text{ và } X \rightarrow A \in F\}$

❖ Bổ đề:

- Giả sử $X \subseteq R$, nếu gọi X^+ là tập các thuộc tính A của R mà $F \models X \rightarrow A$ thì với mọi tập $Y \subseteq R$, $F \models X \rightarrow Y \Leftrightarrow Y \subseteq X^+$.

Các phương pháp thiết kế cơ sở dữ liệu logic

❖ Sơ đồ quan hệ:

- Là một lược đồ quan hệ R và tập phụ thuộc hàm F

❖ Định nghĩa khóa

- Trong quan hệ R , tập các thuộc tính K là khoá của quan hệ nếu có $K \rightarrow A_i$ với A_i là tất cả các thuộc tính còn lại.

Các phương pháp thiết kế cơ sở dữ liệu logic

❖ Các tính chất của khóa:

1. Các thuộc tính không xuất hiện trong cả hai vế trái hoặc phải của tập F phải có trong mọi khóa k .
2. Các thuộc tính chỉ xuất hiện bên trái của các PTH trong F cũng phải thuộc mọi khóa k .
3. Những thuộc tính xuất hiện, và chỉ xuất hiện bên vế phải của tập PTH sẽ không thuộc bất kỳ khóa nào.

Các phương pháp thiết kế cơ sở dữ liệu logic

❖ Thuật toán tìm một khóa

- Bước 1: Đặt $k = R$
- Bước 2: Lặp quá trình loại khỏi k phần tử A mà $(k-A)^+ = R$

Ví dụ:

Hoadon1(SoHD, NgayHD, MaKH, TenKH, DC)

Hoadon2(SoHD, NgayHD, MaKH, MaMH, Soluong, Dongia)

Các phương pháp thiết kế cơ sở dữ liệu logic

❖ Thuật toán tìm mọi khóa

- Bước 1: Tìm tất cả tập con khác rỗng của R
- Bước 2: Loại tập con có bao đóng khác R
- Bước 3: Loại tập con bao tập con khác
- Bước 4: Những tập còn lại là khóa

Ví dụ:

Docgia(MaSV,Hoten,NS,Sothe,Ngaycap)

MaSV->Hoten,NS,Sothe,Ngaycap;MaSV->Hoten,NS;

Sothe->MaSV,Hoten,NS,Ngaycap;Sothe->Ngaycap

Các phương pháp thiết kế cơ sở dữ liệu logic

❖ Các dạng chuẩn

❖ Dạng chuẩn 1:

■ Định nghĩa:

- Một quan hệ là ở dạng chuẩn 1 nếu toàn bộ các miền thuộc tính đều là các miền đơn và không tồn tại thuộc tính lặp.
- Một thuộc tính A là thuộc tính lặp nếu với một giá trị cụ thể của khoá chính có nhiều giá trị của thuộc tính A kết hợp với khoá chính này.

■ Ví dụ:

Sinhvien(Masv,hoten,ns,dc,trinhdongoaingu,lop)

Hoadon(SoHD,NgayHD,MaKH,MaMH,Soluong,Dongia,
HotenKH,TenMH,DVT)

Các phương pháp thiết kế cơ sở dữ liệu logic

- Chuẩn hóa đưa về dạng chuẩn 1:
 - Loại bỏ các thuộc tính vi phạm dạng chuẩn 1 và đặt chúng vào một bảng riêng cùng với khoá chính của quan hệ ban đầu. Khoá chính của bảng này là một tổ hợp của khoá chính của quan hệ ban đầu và thuộc tính đa trị hoặc khoá bộ phận của nhóm lặp.
 - Các thuộc tính còn lại lập thành một quan hệ với khóa chính là khóa chính ban đầu.
- Bài tập áp dụng:
Sinhvien(Masv,hoten,ns,dc,trinhdongoaingu,lop)
Hoadon(SoHD,NgayHD,MaKH,MaMH,Soluong,Dongia,
HotenKH,TenMH,DVT)

Các phương pháp thiết kế cơ sở dữ liệu logic



❖ Các dạng chuẩn

❖ Dạng chuẩn 2:

■ Định nghĩa:

- Một quan hệ ở dạng chuẩn 2 nếu nó đã ở dạng chuẩn 1 và không tồn tại phụ thuộc hàm bộ phận vào khoá.

■ Ví dụ:

- Sinhvien(MaSV,MaMH, TenMH,sodvht,diem)
- Hoadon(SoHD,MaMH,Soluong,Dongia,TenMH,DVT)

Các phương pháp thiết kế cơ sở dữ liệu logic

LOGO

❖ Các dạng chuẩn

❖ Dạng chuẩn 2:

■ Chuẩn hóa đưa về dạng chuẩn 2:

- Tách các thuộc tính phụ thuộc vào bộ phận của khóa ra thành một lược đồ quan hệ mới, bổ sung thêm thuộc tính mà nó phụ thuộc vào làm khóa chính.
- Các thuộc tính còn lại tạo thành một lược đồ quan hệ, khóa là khóa ban đầu.

■ Bài tập áp dụng:

- Sinhvien(Masv,mamh, tenmh,sodvht,diem)
- Hoadon(SoHD,MaMH,Soluong,Dongia,TenMH,DVT)

Các phương pháp thiết kế cơ sở dữ liệu logic



❖ Các dạng chuẩn

- Dạng chuẩn 3:
- Định nghĩa:
 - Một quan hệ ở dạng chuẩn 3 nếu nó ở dạng chuẩn 2 và không tồn tại phụ thuộc hàm ngoài khóa (phụ thuộc hàm bắc cầu).
- Ví dụ:
 - Sinhvien(Masv,Hoten,NS,GT,Malop,Tenlop)
 - Hoadon(SoHD,NgayHD,MaKH,HotenKH)

Các phương pháp thiết kế cơ sở dữ liệu logic



❖ Các dạng chuẩn

❖ Dạng chuẩn 3:

■ Chuẩn hóa đưa về dạng chuẩn 3:

- Tách các thuộc tính phụ thuộc vào thuộc tính không khóa ra thành một lược đồ quan hệ mới, bổ sung thêm thuộc tính mà nó phụ thuộc vào làm khóa chính.
- Các thuộc tính còn lại tạo thành một lược đồ quan hệ, khóa là khóa ban đầu.

■ Bài tập áp dụng:

- Sinhvien(Masv,Hoten,NS,GT,Malop,Tenlop)
- Hoadon(SoHD,NgayHD,MaKH,HotenKH)

Các phương pháp thiết kế cơ sở dữ liệu logic



❖ Các dạng chuẩn

❖ Dạng chuẩn BCNF:

- Định nghĩa: Một quan hệ ở dạng chuẩn BCNF nếu:
 - Nó ở dạng chuẩn 3 và không tồn tại phụ thuộc hàm có nguồn là thuộc tính không khóa, đích là thuộc tính khóa.

Ví dụ:

- Sinhvien(Malop, MaMH, diem, MaSV)

Các phương pháp thiết kế cơ sở dữ liệu logic

❖ Các dạng chuẩn

❖ Dạng chuẩn BCNF:

■ Chuẩn hóa đưa về chuẩn BCNF :

- Tách các thuộc tính phụ thuộc vào thuộc tính không khóa ra thành một lược đồ quan hệ, bổ sung thêm thuộc tính gây ra sự phụ thuộc vào làm khóa của quan hệ này.
- Các thuộc tính còn lại tạo thành 1 LĐQH, khóa là một phần của khóa ban đầu và một phần là thuộc tính gây ra sự phụ thuộc.

■ Bài tập áp dụng:

- Sinhvien(Malop,MaMH, diem, MaSV)

Các phương pháp thiết kế cơ sở dữ liệu logic

❖ Bài tập vận dụng:

- Cho mẫu phiếu yêu cầu mua sách của khách hàng và mẫu hóa đơn bán sách.
- Yêu cầu: Sử dụng 2 phương pháp thiết kế dữ liệu logic top-down và bottom-up để xây dựng mô hình dữ liệu logic

Các phương pháp thiết kế cơ sở dữ liệu logic

❖ *Phương pháp Topdown:*

- Liệt kê các thuộc tính từ mẫu biểu đã cho.
- Xác định các thực thể
- Gắn các thuộc tính vào thực thể
- Xác định mối quan hệ giữa các thực thể.
- Gắn các thuộc tính vào mối quan hệ (nếu có)
- Xây dựng mô hình ER
- Chuyển đổi từ mô hình ER sang mô hình quan hệ.
- Vẽ mô hình quan hệ.

Các phương pháp thiết kế cơ sở dữ liệu logic

❖ *Phương pháp Bottom-up:*

- Bước 1: Từ một biểu mẫu lấy ra một danh sách các thuộc tính cho quan hệ chưa được chuẩn hoá (còn gọi là dạng chuẩn 0).
 - Mỗi tiêu đề trong biểu mẫu là một thuộc tính.
 - Bỏ qua phần đầu đề và phần dưới cùng (một số ghi chú, chữ ký ...) của biểu mẫu.
 - Không lấy các thuộc tính được suy diễn từ những thuộc tính khác.
 - Bổ sung thêm một số thuộc tính định danh tương ứng với một số thuộc tính tên gọi chưa có định danh nếu cần thiết.
 - Xác định nhóm thuộc tính lặp, các phụ thuộc hàm giữa các thuộc tính.

Các phương pháp thiết kế cơ sở dữ liệu logic

❖ *Phương pháp Bottom-up:*

- **Bước 2:** Chuẩn hoá về dạng chuẩn 1 (1NF): Tách nhóm thuộc tính lặp.

Các phương pháp thiết kế cơ sở dữ liệu logic

❖ *Phương pháp Bottom-up:*

- **Bước 3:** Chuẩn hoán về dạng chuẩn 2 (2NF): Loại bỏ phụ thuộc bộ phận vào khoá (chỉ áp dụng với các quan hệ có khoá ghép)

Các phương pháp thiết kế cơ sở dữ liệu logic

❖ *Phương pháp Bottom-up:*

- **Bước 4:** Chuẩn hoá về dạng chuẩn 3 (3NF): Loại bỏ phụ thuộc hàm giữa các thuộc tính không khoá.

❖ Hoàn thiện mô hình CSDL logic

- *Một số nguyên tắc kết hợp kết quả từ 2 cách tiến hành:*
 - Kiểm tra sự thống nhất về tên gọi của các quan hệ và các thuộc tính trong hai kết quả. Nếu cùng tên nhưng khác nghĩa thì phải đặt lại tên cho khác nhau. Nếu cùng nghĩa nhưng khác tên thì phải đặt lại tên cho giống nhau.
 - Lấy tất cả các quan hệ khác nhau từ hai kết quả. Với hai quan hệ trùng nhau thì lấy tất cả các thuộc tính có trong hai quan hệ từ hai kết quả.

Các phương pháp thiết kế cơ sở dữ liệu logic

❖ *Hướng dẫn bài tập vận dụng:*

- Từ mẫu phiếu yêu cầu mua sách của khách hàng xác định một lược đồ quan hệ duy nhất. Áp dụng các dạng chuẩn và quy tắc chuẩn hóa để tách lược đồ quan hệ về dạng chuẩn 3NF hoặc BCNF.
- Tương tự áp dụng với mẫu hóa đơn bán sách.

❖ Mô hình quan hệ

Nhaxuatban(MaNXB, TenNXB, Diachi, Dienthoai)

Tacgia(MaTG, Hoten, Diachi, Dienthoai)

Chude(MaCD, TenCD)

Sach(Masach, Tensach, soluong, dongia, MaNXB, MaCD)

Tacgia_Sach(MaTG, Masach, sotrang, ghichu)

Khachhang(MaKH, Hoten, Diachi, Dienthoai)

Donhang(MaDH, MaKH, NgayDH)

Chitietdonhang(MaDH, MaSach, Soluong)

Bài tập về nhà

A decorative header image featuring a close-up of a fountain pen nib and a small logo with the word "LOGO" in green capital letters.

- ❖ Bài tập 1: Thiết kế cơ sở dữ liệu logic cho “Hệ thống bán sách online”
- ❖ Bài 2: Thiết kế cơ sở dữ liệu logic cho “Hệ thống quản lý đào tạo LQDUNI”
- ❖ Bài 3: Tìm hiểu những nhu cầu báo cáo đặt biệt trong các hệ thống
- ❖ Bài 4: Tìm hiểu những vấn đề thiết kế phi chuẩn trong các hệ thống thực