

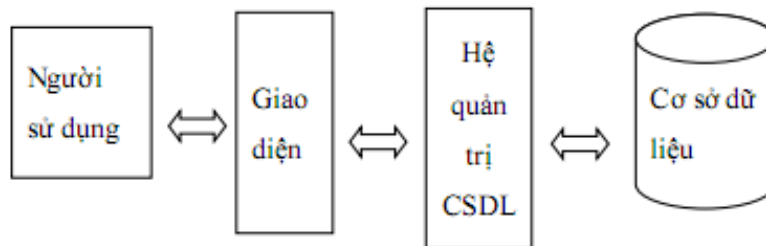
CHƯƠNG 4 : PHÂN TÍCH VÀ THIẾT KẾ DỮ LIỆU

4.1. TỔNG QUAN

Tất cả mọi hệ thống đều phải sử dụng một cơ sở dữ liệu của mình, đó có thể là một cơ sở dữ liệu đã có hoặc một cơ sở dữ liệu được xây dựng mới. Cũng có những hệ thống sử dụng cả cơ sở dữ liệu cũ và mới. Việc phân tích và thiết kế cơ sở dữ liệu cho một hệ thống có thể tiến hành đồng thời với việc phân tích và thiết kế hệ thống hoặc có thể tiến hành riêng. Vấn đề đặt ra là cần xây dựng một cơ sở dữ liệu giảm được tối đa sự dư thừa dữ liệu đồng thời phải dễ khôi phục và bảo trì.

4.1.1. Các khái niệm

- Cơ sở dữ liệu (CSDL): CSDL máy tính là một kho chứa một bộ sưu tập có tổ chức các file dữ liệu, các bản ghi và các trường.
- Hệ quản trị cơ sở dữ liệu (HQTCSDL): là một phần mềm điều khiển mọi truy nhập đối với CSDL.
- Các HQTCSDL được phân loại theo mô hình dữ liệu như sau:



- Các HQTCSDL phân cấp ứng với mô hình phân cấp (ví dụ: IMS của IBM)
- Các HQTCSDL mạng ứng với mô hình mạng (ví dụ: IDMS của Cullinet Software)
- Các HQTCSDL quan hệ ứng với mô hình quan hệ (ví dụ: ORACLE của Oracle, DB2 của IBM, Access và SQL server của Microsoft)
- Các HQTCSDL hướng đối tượng ứng với mô hình hướng đối tượng (ví dụ: Jasmine)

4.1.2. Các bước tiến hành phân tích và thiết kế CSDL

a. Phân tích CSDL

Bước phân tích CSDL độc lập với các hệ quản trị CSDL, bước này thực hiện các công việc sau:

- + Xác định các yêu cầu về dữ liệu: Phân tích các yêu cầu dữ liệu của hệ thống để xác định các yêu cầu về dữ liệu.
- + Mô hình hoá dữ liệu: Xây dựng mô hình thực thể liên kết biểu diễn các yêu cầu về dữ liệu.

b. Thiết kế CSDL quan hệ

Thiết kế CSDL bao gồm:

- + Thiết kế logic CSDL: độc lập với một hệ quản trị CSDL.
 - Xác định các quan hệ: Chuyển từ mô hình thực thể liên kết sang mô hình quan hệ.
 - Chuẩn hoá các quan hệ: chuẩn hoá các quan hệ về dạng chuẩn ít nhất là chuẩn 3 (3NF)
- + Thiết kế vật lý CSDL: dựa trên một hệ quản trị CSDL cụ thể.
 - Xây dựng các bảng trong CSDL quan hệ: quyết định cấu trúc thực tế của các bảng lưu trữ trong mô hình quan hệ.
 - Hỗ trợ các cài đặt vật lý trong CSDL: cài đặt chi tiết trong HQTCSDL lựa chọn.

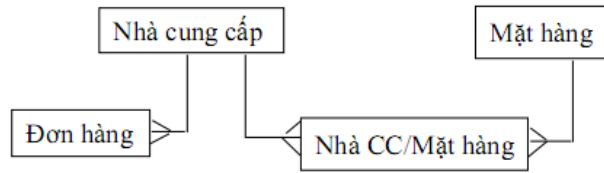
4.2. MÔ HÌNH THỰC THỂ LIÊN KẾT

4.2.1. Mục đích của việc xây dựng mô hình

Xây dựng mô hình thực thể liên kết nhằm các mục đích sau;

- Mô tả thế giới thực gần với quan niệm, suy nghĩ của ta. Đây là mô hình tốt với lượng thông tin ít nhất, mô tả thế giới dữ liệu đầy đủ nhất.
- Việc xây dựng mô hình nhằm thành lập một biểu đồ cấu trúc dữ liệu bao gồm dữ liệu cần xử lý và cấu trúc nội tại của nó.

Ví dụ một mô hình thực thể liên kết



Mô hình liên kết thực thể của việc bán hàng

4.2.2. Các thành phần của mô hình thực thể liên kết

a. Thực thể

Khái niệm về thực thể

Thực thể là khái niệm để chỉ một lớp các đối tượng có cùng đặc tính chung mà người ta muốn quản lý thông tin về nó. Ví dụ, sinh viên, hàng hóa, vật tư

Một đối tượng cụ thể trong thực thể được gọi là một cá thể (còn gọi là một thể hiện của thực thể). Ví dụ Lê Văn Bình là một cá thể của thực thể Sinh viên, Xi măng là cá thể của thực thể Hàng hóa

Thuộc tính của thực thể

Để mô tả thông tin về một thực thể người ta thường dựa vào các đặc trưng riêng của thực thể đó. Các đặc trưng đó được gọi là thuộc tính của thực thể. Ví dụ thực thể Sinh viên có các thuộc tính Mã sinh viên, Họ tên sinh viên, Ngày sinh, Địa chỉ, Trường, Khoa, Khóa sinh viên, Lớp,... Đối với thực thể Hàng hóa thì giá trị của các thuộc tính Mã hàng hóa, Tên hàng hóa, Đơn vị tính, Đơn giá,... là những thông tin mà nhà quản lý cần quan tâm để quản lý hàng hóa.

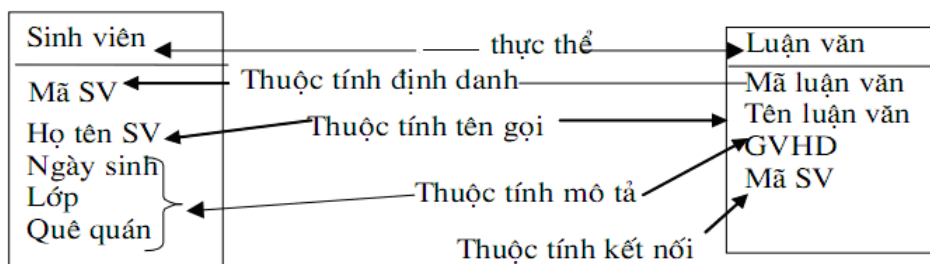
Thuộc tính của thực thể bao gồm các loại sau

Thuộc tính định danh (còn gọi là định danh thực thể, đôi khi còn gọi là thuộc tính khoá): Là một hoặc một số thuộc tính mà giá trị của nó cho phép phân biệt các thực thể khác nhau. Một thực thể bao giờ cũng được xác định một thuộc tính định danh làm cơ sở để phân biệt các thể hiện cụ thể của nó. Ví dụ : Số hiệu khách hàng, Mã mặt hàng, Mã sinh viên,...

Thuộc tính mô tả: Là các thuộc tính mà giá trị của chúng chỉ có tính mô tả cho thực thể hay liên kết mà thôi. Hầu hết các thuộc tính trong một kiểu thực thể đều là mô tả. Có một số thuộc tính mô tả đặc biệt như sau:

- Thuộc tính tên gọi: là thuộc tính mô tả để chỉ tên các đối tượng thuộc thực thể. Thuộc tính tên gọi để phân biệt các thực thể (tách các thực thể).
- Thuộc tính kết nối (thuộc tính khoá ngoài): là thuộc tính chỉ ra mối quan hệ giữa một thực thể đã có và một thực thể trong bảng khác. Thuộc tính kết nối giống thuộc tính mô tả thông thường trong thực thể chứa nó nhưng nó lại là thuộc tính khoá của một thực thể trong bảng khác.

Ví dụ:



b. Liên kết và các kiểu liên kết

Liên kết (còn gọi là quan hệ) là sự kết hợp giữa hai hay nhiều thực thể phản ánh sự ràng buộc trong quản lý. Đặc biệt: Một thực thể có thể liên kết với chính nó ta thường gọi là tự liên kết. Giữa hai thực thể có thể có nhiều hơn một liên kết.

Có ba kiểu liên kết: một - một, một - nhiều, nhiều - nhiều.

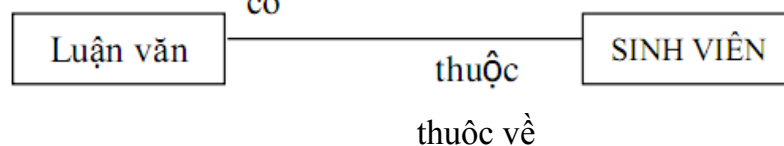
Liên kết một – một (1-1):

Mỗi thể hiện của thực thể A quan hệ với một thể hiện của thực thể B và ngược lại.

Kí hiệu:



Ví dụ: Một sinh viên có một luận văn. Một luận văn thuộc về một sinh viên



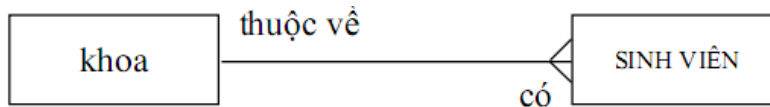
Liên kết một – nhiều (1-N)



Mỗi thể hiện của thực thể A quan hệ với nhiều thể hiện của thực thể B. Ngược lại mỗi thể hiện của thực thể B quan hệ với chỉ một thể hiện của thực thể A.

Kí hiệu:

Ví dụ: Một khoa có nhiều sinh viên. Một sinh viên thuộc về một khoa.



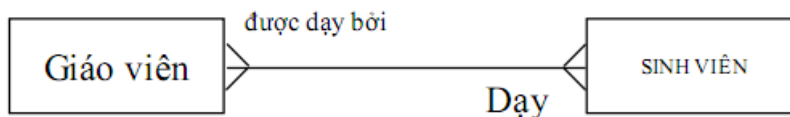
Liên kết nhiều – nhiều (N-N)

Mỗi thể hiện của thực thể A quan hệ với nhiều thể hiện của thực thể B. Ngược lại mỗi thể hiện của thực thể B quan hệ với nhiều thể hiện của thực thể A.

Kí hiệu



Ví dụ: Một giáo viên dạy nhiều sinh viên. Một sinh viên được dạy bởi nhiều giáo viên.



Điều kiện để một cá thể của thực thể tham gia vào liên kết với một thực thể khác gọi là loại thành viên. Nó có thể là bắt buộc hay tùy chọn trong quan hệ. Các loại thành viên cho biết số thể hiện nhỏ nhất của mỗi thực thể tham gia vào liên kết với một thể hiện của một thực thể khác.

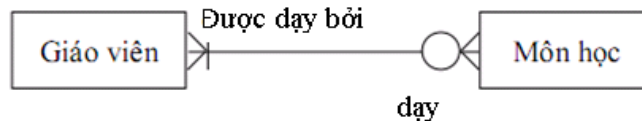
Kí hiệu:



Ví dụ:

Tùy chọn (ít nhất 0) – “một giáo viên có thể dạy không, một hoặc nhiều môn học.”

Bắt buộc (ít nhất 1) – “một môn học cần phải được một hoặc nhiều giáo viên dạy.”



Chú ý:

- Mô hình dữ liệu không chỉ là công cụ phân tích thiết kế mà còn như một phương pháp kiểm tra chặt chẽ các yêu cầu nghiệp vụ của người sử dụng.
- Nếu hai thực thể có quan hệ một - một thường có ít lý do để coi chúng như hai bảng tách biệt nên người ta thường gộp hai thực thể làm một
- Nếu hai thực thể có quan hệ nhiều - nhiều thì không có sự khác biệt về bản chất giữa các chiều vì vậy ít khi được sử dụng.

Tóm lại trong ba kiểu liên kết trên, liên kết một nhiều là quan trọng hơn cả và hầu như các mối quan hệ trong mô hình thực thể liên kết đều là một nhiều.

Bài tập: Xác định các liên kết của các kiểu thực thể trong hệ thống quản lý bến xe

Khách hàng - Vé (1-n); Vé - Phương tiện (n-1); Đơn hàng- Nhà cc(n-1); Đơn hàng-Phương tiện (1-n).

4.2.3. Xây dựng mô hình thực thể liên kết của hệ thống

a. Các bước tiến hành

Bước 1: Xác định và định danh thực thể

Xác định các thực thể là các mục thông tin cần thiết cho hệ thống và hệ thống cần lưu giữ. Tìm các thực thể từ ba nguồn :

- *Thông tin tài nguyên:* con người, kho bãi, tài sản (VD: nhà cung cấp, mặt hàng, kho...)

- *Thông tin giao dịch*: là các luồng thông tin đến từ môi trường và kích hoạt một *chuỗi hoạt động của hệ thống* (VD: đơn hàng (mua, bán), dự trữ, phiếu yêu cầu,...)
- *Thông tin tổng hợp*: thường ở dưới dạng thống kê liên quan đến các kế hoạch hoặc kiểm soát (VD: dự toán chi tiêu, tính lương...)

Việc xác định và định danh thực thể phải thoả mãn:

- Tên gọi là danh từ.
- Có nhiều thể hiện.
- Có duy nhất một định danh.
- Có ít nhất một thuộc tính mô tả.
- Có quan hệ với ít nhất một thực thể khác.

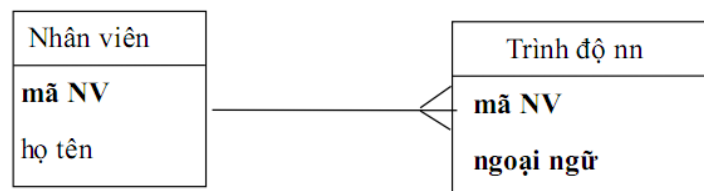
Bước 2: Xác định các thuộc tính mô tả cho các thực thể

Khi xác định các thuộc tính mô tả cho các thực thể cần chú ý rằng:

- Mỗi thuộc tính chỉ xuất hiện một lần trong thực thể tương ứng.
- Nếu không chắc chắn là thuộc tính hay thực thể cần tiếp tục nghiên cứu và phân tích nó.

Chú ý: Khi một thuộc tính của thực thể A có nhiều giá trị ta sẽ mô hình hoá thuộc tính đó là một thực thể B có quan hệ phụ thuộc với thực thể A. Định danh của thực thể B sẽ bao gồm các thuộc tính định danh của thực thể A và một số thuộc tính khác của thực thể B. Liên kết giữa thực thể A và thực thể B được gọi là liên kết phụ thuộc.

Ví dụ: Một nhân viên có thể có nhiều trình độ ngoại ngữ với các ngôn ngữ khác nhau. Khi đó trình độ ngoại ngữ của nhân viên không được mô hình hoá là một thuộc tính mà được mô hình hoá là một thực thể như sau:



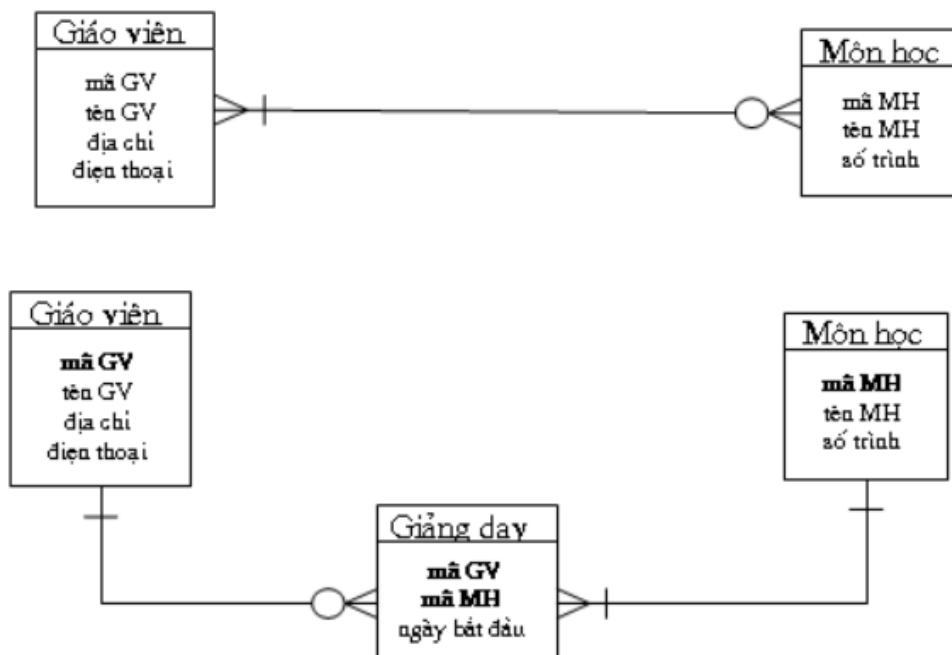
Như vậy định danh của thực thể TRÌNH ĐỘ NN gồm hai thuộc tính mã NV và ngoại ngữ.

Bước 3: Xác định liên kết giữa các thực thể

Xác định liên kết giữa các thực thể theo trình tự sau:

- Thiết lập sự tồn tại của liên kết (Vẽ đường thẳng và đặt tên quan hệ tại hai đầu)
- Xác định loại liên kết (1-1, 1-N, N-N) và loại thành viên (tùy chọn hay bắt buộc).
- Tách liên kết N-N thành hai liên kết 1-N với một thực thể kết hợp. Khi đó thực thể kết hợp sẽ có định danh được tạo thành từ hai thuộc tính định danh của các thực thể ban đầu.

Ví dụ: Tách liên kết giữa 2 thực thể giáo viên và môn học như sau:



Ví dụ về xây dựng mô hình thực thể liên kết:

Một công ty thương mại Y chuyên kinh doanh các mặt hàng điện tử. Công ty nhập các mặt hàng từ các nhà cung cấp khác nhau. Chi tiết về các mặt hàng gồm có: mã hàng (duy nhất), tên hàng và các mô tả mặt hàng.

Công ty cũng cần lưu giữ thông tin về các nhà cung cấp như tên, địa chỉ, điện thoại, fax. Mỗi nhà cung cấp có một mã duy nhất. Mỗi nhà cung cấp có

thể cung cấp nhiều mặt hàng nhưng mỗi mặt hàng chỉ được cung cấp từ một nhà cung cấp.

Các mặt hàng được lưu giữ trong các kho. Mỗi kho hàng có một diện tích khác nhau và chỉ chứa một loại mặt hàng.

Công ty có nhiều cửa hàng đại lý để bán các mặt hàng. Hàng được cung cấp cho các cửa hàng thông qua các phiếu xuất. Thông tin trên mỗi phiếu xuất cần có mã số cửa hàng nhận hàng, ngày xuất, thông tin về các mặt hàng được xuất như tên hàng, số lượng, đơn giá, thành tiền.

Yêu cầu: Vẽ mô hình thực thể liên kết của hệ thống.

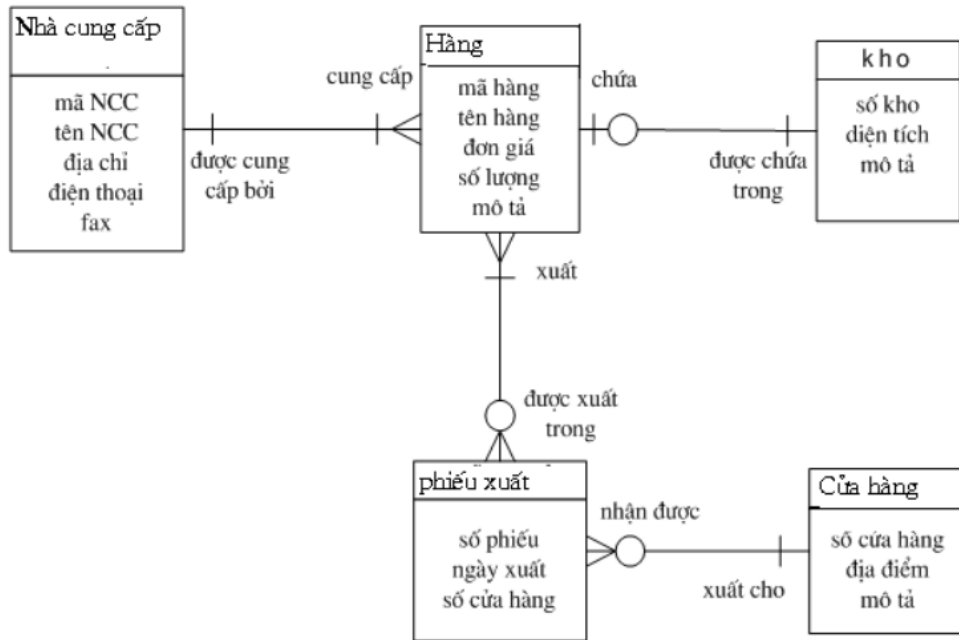
Xác định các thực thể, định danh thực thể và các thuộc tính mô tả:

Tên thực thể	Định danh	Thuộc tính mô tả
HÀNG	mã hàng	tên hàng, đơn giá, số lượng
NHÀ CUNG CẤP	mã NCC	tên NCC, địa chỉ, điện thoại, fax
KHO	số kho	diện tích
PHIẾU XUẤT	số phiếu	ngày xuất, số cửa hàng
CỬA HÀNG	số cửa hàng	địa điểm

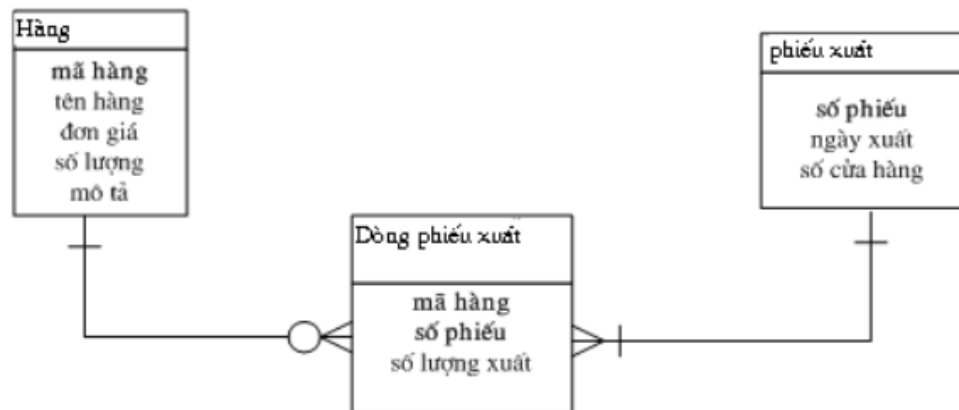
Xác định liên kết giữa các thực thể

- + Một mặt hàng cần được cung cấp bởi một nhà cung cấp. Một nhà cung cấp cần cung cấp một hoặc nhiều mặt hàng.
- + Một mặt hàng cần được lưu giữ trong một kho. Mỗi kho lưu giữ 0 hoặc một loại hàng.
- + Một mặt hàng được xuất trong 0, 1 hoặc nhiều phiếu xuất. Một phiếu xuất có thể xuất 1 hoặc nhiều mặt hàng.
- + Một cửa hàng nhận được 0, 1 hoặc nhiều phiếu xuất. Mỗi phiếu xuất cần được xuất cho chỉ một cửa hàng.

Vẽ mô hình thực thể liên kết



- Quan hệ n-n giữa Hàng và Phiếu xuất có thể được tách thành 2 quan hệ 1- n với thực thể kết hợp Dòng phiếu xuất như sau:



4.3. MÔ HÌNH QUAN HỆ

4.3.1. Các khái niệm

a. Khái niệm về mô hình quan hệ

Mô hình CSDL quan hệ hay ngắn gọn là mô hình quan hệ được E.F.Codd phát triển vào đầu những năm 1970. Mô hình này được thiết lập trên cơ sở lý thuyết tập hợp nên nó rất dễ hiểu và được sử dụng rất rộng rãi trong việc tổ chức dữ liệu cho các hệ thống.

- Các thành phần trong mô hình quan hệ gồm: Các quan hệ - các bộ - các thuộc tính.

Tương ứng với các thành phần trong mô hình thực thể liên kết là: Các thực thể - các thể hiện của thực thể – các thuộc tính.

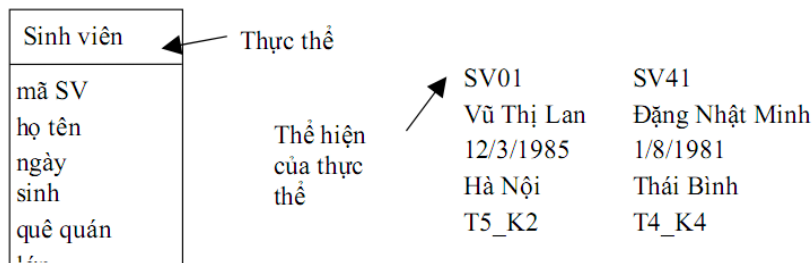
Mô hình thực thể liên kết => Mô hình quan hệ => Các bảng trong hệ quản trị cơ sở dữ liệu.

So sánh tương ứng giữa mô hình thực thể liên kết, mô hình quan hệ và hệ quản trị CSDL chúng ta có bảng sau:

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

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

Chú ý:

- Các bộ trong một quan hệ không được trùng nhau.

- Thuộc tính được xác định bởi tên, thứ tự của các thuộc tính trong quan hệ là không quan trọng. Trong một quan hệ, tên các thuộc tính phải khác nhau (các miền giá trị của các thuộc tính không nhất thiết khác nhau)

b. Khái niệm về khoá

Khoá chính:

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. Trong quan hệ các thuộc tính thuộc khoá chính được gạch chân và được gọi là các thuộc tính khoá.

Ví dụ: SINH VIÊN (mã SV, họ tên, ngày sinh, quê quán, lớp) – mã SV là khoá chính của quan hệ SINH VIÊN.

Khi chọn khoá chính cần phải xem xét các tiêu chuẩn sau: khoá chính phải xác định được duy nhất một bộ trong quan hệ, phải có số thuộc tính ít nhất, phải không thay đổi theo thời gian.

Khoá ghép:

Khoá ghép là khoá có từ hai thuộc tính trở lên.

Ví dụ: GIẢNG DẠY(mã GV, mã MH, ngày bắt đầu, ngày kết thúc) – khoá chính của quan hệ GIẢNG DẠY là một khoá ghép gồm hai thuộc tính mã GV và mã MH.

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. Trong quan hệ các thuộc tính khoá ngoài được in nghiêng hoặc gạch chân bằng nét đứt.

Ví dụ: LỚP (tên lớp, khoa, phòng học)

SINH VIÊN (mã SV, họ tên, ngày sinh, quê quán, *tên lớp*) – *tên lớp* là khoá ngoài của quan hệ SINH VIÊN

Khoá giả:

Là thuộc tính do con người đặt ra để làm khoá chính. Thuộc tính này không mô tả đặc điểm của các đối tượng quan tâm mà chỉ có tác dụng để xác định duy nhất đối tượng đó.

Ví dụ: mã SV, số hoá đơn... Thông thường khi khoá chính có từ 3 thuộc tính trở lên người ta thường đặt ra một khoá giả làm khoá chính để tiện lợi hơn trong việc truy vấn dữ liệu.

Các ràng buộc trong mô hình quan hệ là:

- 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. (Giá trị null tức là không có giá trị. Nó khác với giá trị 0 hay dấu cách.)
- 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.

Ví dụ: trong quan hệ sau:

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>

Bản ghi mờ còi

LỚP	tên lớp	phòng học
	T1_K3	302
	T3_K4	414

Tên lớp là T4-K7 trong quan hệ SINH VIÊN không có trong quan hệ LỚP vì vậy nó vi phạm ràng buộc tham chiếu.

- Các ràng buộc được định nghĩa bởi người dùng: đây là các ràng buộc liên quan đến miền giá trị của dữ liệu thực tế.

4.3.2. Các dạng chuẩn

a. Phụ thuộc hàm

Khái niệm

Trong một quan hệ R, thuộc tính B phụ thuộc hàm vào thuộc tính A (hay thuộc tính A xác định hàm thuộc tính B) ký hiệu $A \rightarrow B$ nếu với mỗi giá trị của thuộc tính A xác định một giá trị duy nhất của thuộc tính B.

Ví dụ: Mã SV \rightarrow Ngày sinh

Mã SV	Ngày sinh
1234	1/2/78
1235	15/5/81
1237	30/12/80
1236	1/2/78

Phụ thuộc hàm giữa nhiều thuộc tính: thuộc tính B phụ thuộc hàm vào các thuộc tính A1 và A2 ký hiệu $\{A1, A2\} \rightarrow B$ nếu với mỗi cặp giá trị của A1 và A2 xác định duy nhất một giá trị của B.

Ví dụ: {Số hoá đơn, Mã hàng} \rightarrow Số lượng

Số hoá đơn	Mã hàng	Số lượng
1234	P1	50*
1234	P2	70
1235	P1	50*
1235	P2	30
1236	P1	80

Chú ý: $A1 \rightarrow \{A2, A3\} \equiv A1 \rightarrow A2$ và $A1 \rightarrow A3$

$\{A1, A2\} \rightarrow A3 \neq A1 \rightarrow A3$ và $A2 \rightarrow A3$

Các loại phụ thuộc hàm

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

Phụ thuộc hàm bắc cầu: Nếu có $A1 \rightarrow A2$ và $A2 \rightarrow A3$ thì $A1 \rightarrow A3$. Khi đó A3 được gọi là phụ thuộc bắc cầu vào A1.

Theo quan niệm phụ thuộc hàm thì định nghĩa khoá như sau: Trong quan hệ R, tập các thuộc tính K là khoá của quan hệ nếu có $K \rightarrow B_i$ với B_i là tất cả các thuộc tính còn lại.

b. Các dạng chuẩn

Dạng chuẩn 1 (1NF):

Một quan hệ là ở dạng chuẩn 1 nếu toàn bộ các miền thuộc tính đều là các miền đơn và không tồn tại nhóm 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: Khoá chính là Mã SV. Nhóm thuộc tính lặp là Môn học và Điểm.

Mã SV	Tên SV	Môn học	Điểm	Môn học	Điểm
135	Anh	SA1	5	SA2	7
136	Bình	SA2	6	SD	5
140	Lan	SD	8		

Dạng chuẩn 2 (2NF):

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: Cho một quan hệ: R (A, B, C, D, E)

Khoá chính là A,B

Các phụ thuộc hàm:

{A,B} \rightarrow D;

A \rightarrow C; (Phụ thuộc hàm bộ phận vào khoá)

D \rightarrow E

Chú ý: Quan hệ có khoá chính là một thuộc tính luôn ở dạng chuẩn 2

Dạng chuẩn 3 (3NF):

Một quan hệ ở dạng chuẩn 3 nếu nó đã ở dạng chuẩn 2 và không tồn tại phụ thuộc hàm bắc cầu vào khoá (hay phụ thuộc hàm giữa các thuộc tính không khoá).

Ví dụ: Cho một quan hệ: R (A, B, D, E)

Khoá chính là A,B

Các phụ thuộc hàm:

{A,B} \rightarrow D;

D \rightarrow E (Phụ thuộc hàm giữa các thuộc tính không khoá)

4.4. THIẾT KẾ LOGIC CSDL

4.4.1. Các cách tiếp cận để mô hình hoá dữ liệu:

Có hai cách tiếp cận để mô hình hoá dữ liệu như sau:

a. Cách 1: Vẽ mô hình thực thể liên kết – Cách tiếp cận từ trên xuống (Top - down):

Để xây dựng mô hình thực thể liên kết cần xác định một cách trực giác các đối tượng quan trọng mà một hệ thống cần phải lưu trữ như dữ liệu (đó là các thực thể) và xác định các thuộc tính mô tả cho các thực thể đó cùng với quan hệ giữa các thực thể. Nếu áp dụng đúng các luật trong mô hình thực thể liên kết thì ta sẽ có các quan hệ đã được chuẩn hoá.

b. Cách 2: Chuẩn hoá – Các tiếp cận từ dưới lên (Bottom - up):

Để chuẩn hoá cần 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.

Mỗi cách tiếp cận đều có ưu điểm và nhược điểm riêng vì vậy trong thực tế người ta thường phối hợp cả hai cách tiếp cận này để có được một mô hình dữ liệu chính xác nhất.

Kết quả cuối cùng của phần thiết kế logic CSDL là tập các bản ghi logic biểu diễn các quan hệ trong CSDL.

4.4.2. Chuyển đổi từ mô hình thực thể liên kết thành các bản ghi logic

a. Các luật chuyển đổi

- Mỗi thực thể trong mô hình thực thể liên kết trở thành một quan hệ.
- Mỗi thuộc tính trong mô hình thực thể liên kết trở thành một thuộc tính trong quan hệ tương ứng.
- Định danh thực thể trong mô hình thực thể liên kết trở thành khoá chính trong quan hệ tương ứng. Nếu khoá chính không đáp ứng được các tiêu chuẩn về tính tối thiểu, tính xác định duy nhất và tính ổn định thì đưa vào một khoá giả làm khoá chính.
- Thi hành các liên kết thông qua việc đặt khóa ngoài vào quan hệ

- Với liên kết 1-1: đặt khoá chính của một trong hai quan hệ vào quan hệ kia làm khoá ngoài.

Ví dụ: Với mô hình thực thể liên kết



Ta có các quan hệ:

CÔNG NHÂN (mã CN, họ tên, ngày sinh, số máy)

MÁY (số máy, mác, công suất)

Hay: CÔNG NHÂN (mã CN, họ tên, ngày sinh)

MÁY (số máy, mác, công suất, mã CN)

Với liên kết 1-N: đặt khoá chính của quan hệ đầu một vào quan hệ đầu nhiều làm khoá ngoài.

Ví dụ: Với mô hình thực thể liên kết



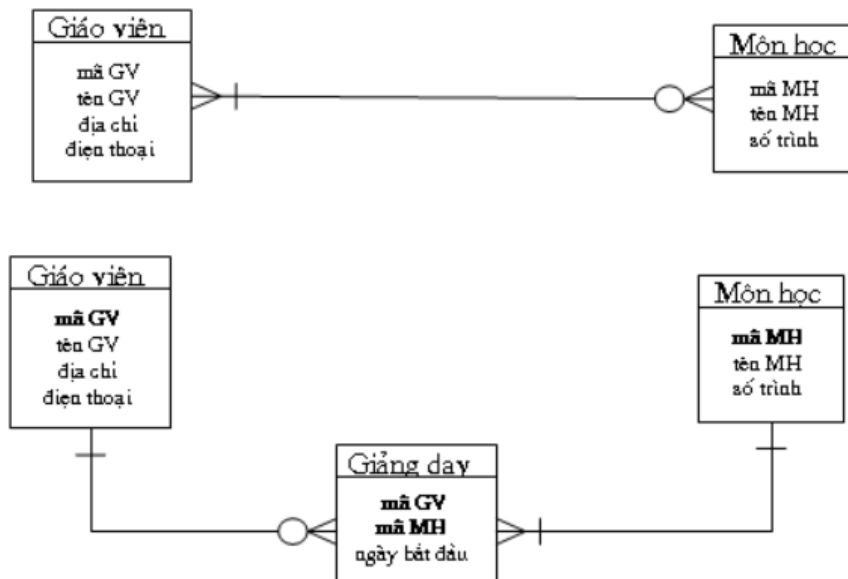
Ta có các quan hệ:

SINH VIÊN (mã SV, họ tên, ngày sinh, tên lớp)

LỚP (tên lớp, phòng học)

- Với liên kết N-N: Cần tách thành quan hệ 1-N rồi mới chuyển thành quan hệ. Khi đó quan hệ tương ứng với thực thể kết hợp sẽ có khoá ghép và mỗi thuộc tính khoá của quan hệ này là một khoá ngoài.

Ví dụ: Với mô hình thực thể liên kết



Ta có các quan hệ:

GIÁO VIÊN (mã SV, tên GV, đại chỉ, điện thoại)

MÔN HỌC (mã MH, tên MH, số trình)

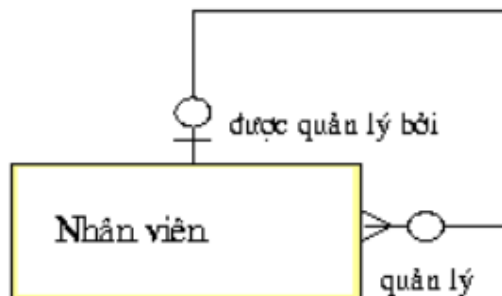
GIẢNG DẠY (mã GV, mã MH, ngày bắt đầu)

- Một số liên kết đặc biệt:

+ Tự liên kết:

Ví dụ: Mỗi nhân viên có không hoặc một người quản lý (giám đốc không có người quản lý nào). Mỗi nhân viên có thể quản lý 0, 1 hay nhiều nhân viên.

Ta có một quan hệ:



NHÂN VIÊN (mã NV, họ tên, ngày sinh, giới tính, địa chỉ, mã NV quản lý)

Trong đó Mã NV quản lý là một mã NV nào đó.

+ Liên kết phụ thuộc

Ví dụ: Một mặt hàng có nhiều giá khác nhau trong những giai đoạn khác nhau.

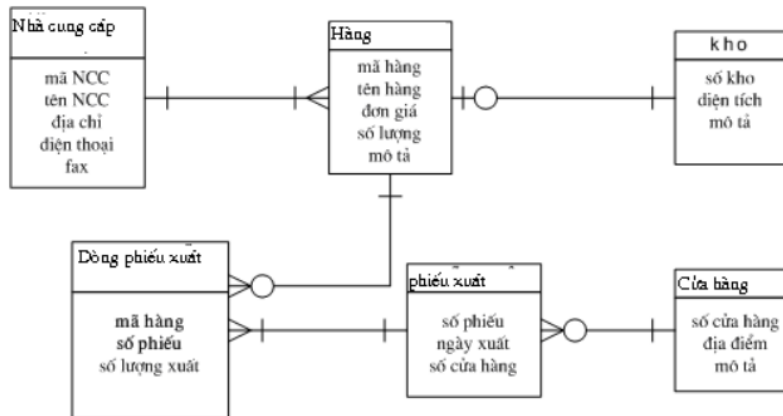
Ta có các quan hệ:

MẶT HÀNG (Mã hàng, Tên hàng)

GIÁ (Mã hàng, Ngày bắt đầu, đơn giá)

b. Ví dụ

Chuyển mô hình thực thể liên kết sau thành các bản ghi logic biểu diễn các quan hệ:



Các bản ghi logic mô tả các quan hệ trong công ty thương mại Y:

NHÀ CUNG CẤP (mã NCC, tên NCC, địa chỉ, điện thoại, fax)

HÀNG (mã hàng, tên hàng, đơn giá, số lượng, mô tả, mã NCC, số kho)

KHO (số kho, diện tích, mô tả)

PHIẾU XUẤT (số phiếu, ngày xuất, số cửa hàng)

DÒNG PHIẾU XUẤT (mã hàng, số phiếu, số lượng xuất)

CỬA HÀNG (số cửa hàng, địa điểm, mô tả)

4.4.3. Chuẩn hoá quan hệ

a. Các bước tiến hành:

Bước 1: Từ một biểu mẫu (tài liệu xuất: hoá đơn, chứng từ,...) 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 (như thành tiền = đơn giá x số lượng) và các thuộc tính trình bày như 'số thứ tự' nếu có.
- 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.

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

- Tách các thuộc tính không nằm trong nhóm lặp thành một quan hệ (R1). Xác định khoá chính của quan hệ này.
- Các thuộc tính của nhóm lặp và khoá chính của quan hệ trên (R1) tạo thành một quan hệ (R2). Xác định khóa chính cho quan hệ R2 (khóa chính của R2 sẽ là một khóa ghép giữa khóa của R1 và một thuộc tính khác trong R2)
- Ví dụ: $R\{A, B, C, D, E\}$ và khóa là $\{A\}$. Tồn tại nhóm thuộc tính lặp $\{C,D\}$ thì tách thành 2 quan hệ $R1\{A, B, E\}$ và $R2\{A,C,D\}$.

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

- Tách các thuộc tính tham gia vào phụ thuộc hàm được xác định bởi một phần của khóa vào một quan hệ mới (R3). Khóa chính của quan hệ là thuộc tính xác định hàm.
- Phần còn lại với khóa chính của quan hệ trên (R3) là một quan hệ giữ nguyên khóa chính như quan hệ ban đầu.

- Ví dụ: $R\{A, B, C, D\}$ và khoá là $\{A, B\}$. Tồn tại phụ thuộc hàm $A \rightarrow C$ thì tách thành 2 quan hệ $R_1\{A, C\}$ và $R_2\{A, B, D\}$.

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

- Tách các thuộc tính tham gia vào phụ thuộc hàm giữa các thuộc tính không khoá vào một quan hệ mới (R_4). Khoá chính của quan hệ là thuộc tính xác định hàm.
- Phần còn lại và khoá chính của quan hệ trên (R_4) là một quan hệ giữ nguyên khoá chính của quan hệ ban đầu.
- Ví dụ: $R\{A, B, C, D\}$ mà khoá là $\{A, B\}$. Tồn tại phụ thuộc hàm $C \rightarrow D$ thì tách thành 2 quan hệ $R_1\{C, D\}$ và $R_2\{A, B, C\}$

Ví dụ:

Hãy xây dựng mô hình dữ liệu cho hệ thống từ mẫu phiếu xuất của công ty thương mại Y dưới đây.

- Danh sách thuộc tính: số PX, ngày, số đại lý, địa chỉ, tên hàng, đơn vị tính, đơn giá, số lượng.
- Bổ sung thêm thuộc tính mã hàng.
- Phụ thuộc hàm: (quy định giá bán của mặt hàng không thay đổi) số PX \rightarrow ngày, mã KH
số đại lý \rightarrow địa chỉ
mã hàng \rightarrow tên hàng, đơn vị tính, đơn giá
 $\{\text{số PX, mã hàng}\} \rightarrow$ số lượng
- Khoá chính: số PX.
Nhóm lặp (*): mã hàng, tên hàng, đơn vị tính, đơn giá, số lượng.

Chuẩn hoá quan hệ:

0NF	1NF	2NF	3NF	Tên gọi
số PX ngày số đại lý địa chỉ * <i>mã hàng</i> * <i>tên hàng</i> * <i>đơn vị tính</i> * <i>đơn giá</i> * <i>số lượng</i>	<u>số PX</u> ngày số đại lý địa chỉ	<u>số PX</u> ngày số đại lý địa chỉ	<u>số đại lý</u> địa chỉ	ĐẠI LÝ
			<u>số PX</u> ngày <i>số đại lý</i>	PHIẾU XUẤT
	<u>số PX</u> <u>mã hàng</u> tên hàng đơn vị tính đơn giá số lượng	<u>mã hàng</u> tên hàng đơn vị tính	<u>mã hàng</u> tên hàng đơn vị tính	HÀNG
		<u>số PX</u> <u>mã hàng</u> số lượng	<u>số PX</u> <u>mã hàng</u> số lượng	HÀNG XUẤT

Các bản ghi logic biểu diễn các quan hệ:

ĐẠI LÝ (số đại lý, địa chỉ)

PHIẾU XUẤT (số PX, ngày, số đại lý)

HÀNG (mã hàng, tên hàng, đơn vị tính, đơn giá)

HÀNG XUẤT (số PX, mã hàng, số lượng xuất)

4.4.4. Hoàn thiện mô hình CSDL logic

a. Một số nguyên tắc

Sau khi tiến hành theo hai hướng khác nhau: xây dựng mô hình thực thể liên kết và chuẩn hoá dữ liệu chúng ta sẽ có hai tập bản ghi logic khác nhau của cùng một hệ thống. Khi đó cần phải kết hợp lại để có một mô hình CSDL logic thống nhất cho hệ thống.

Một số nguyên tắc kết hợp:

- Kiểm tra sự thống nhất về tên gọi của các quan hệ và các thuộc tính trong hai kết quả. Nếu cùng tên những 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.

b. Ví dụ:

Trong ví dụ phần 1 ta thu được tập các bản ghi logic mô tả các quan hệ trong hệ thống công ty thương mại Y là:

Nhóm 1:

NHÀ CUNG CẤP (mã NCC, tên NCC, địa chỉ, điện thoại, fax)

HÀNG (mã hàng, tên hàng, đơn giá, số lượng, mô tả, mã NCC, số kho)

KHO (số kho, diện tích, mô tả)

PHIẾU XUẤT (số phiếu, ngày xuất, số cửa hàng)

DÒNG PHIẾU XUẤT (mã hàng, số phiếu, số lượng xuất)

CỬA HÀNG (số cửa hàng, địa điểm, mô tả)

Trong ví dụ phần 2 ta thu được tập các bản ghi logic mô tả các quan hệ trong hệ thống công ty thương mại Y là:

Nhóm 2:

ĐẠI LÝ (số đại lý, địa chỉ)

PHIẾU XUẤT (số PX, ngày, số đại lý)

HÀNG (mã hàng, tên hàng, đơn vị tính, đơn giá)

DÒNG PHIẾU XUẤT (số PX, mã hàng, số lượng xuất)

Ta thấy

- + Trong nhóm 1 có thực thể CỬA HÀNG và nhóm 2 có thực thể ĐẠI LÝ là hai thực thể khác tên nhưng cùng nghĩa.
- + Thuộc tính số phiếu và số PX, thuộc tính ngày xuất và ngày là cùng nghĩa.

Kết quả thiết kế logic CSDL của hệ thống công ty thương mại Y là:

NHÀ CUNG CẤP (mã NCC, tên NCC, địa chỉ, điện thoại, fax)

HÀNG (mã hàng, tên hàng, đơn vị tính, đơn giá, số lượng, mô tả, mã NCC, số kho)

KHO (số kho, diện tích, mô tả)

PHIẾU XUẤT (số phiếu, ngày xuất, số đại lý)

DÒNG PHIẾU XUẤT (mã hàng ,số phiếu, số lượng xuất)

ĐẠI LÝ (số đại lý, địa chỉ đại lý, mô tả)

4.5. THIẾT KẾ VẬT LÝ CSDL

Phần thiết kế vật lý CSDL sẽ phụ thuộc vào một HQTCSDDL mà chúng ta sẽ sử dụng để cài đặt CSDL. Chúng ta cần chuyển từ các bản ghi logic với các thuộc tính thành các quan hệ được biểu diễn dưới dạng bảng với các trường hay các cột được cài đặt trong một HQTCSDDL cụ thể.

4.5.1. Lập các bảng

- Với mỗi bảng cần chỉ rõ:
 - + Khoá chính (có thể gồm một hay nhiều thuộc tính).
 - + Mô tả của tất cả các cột (trường).
 - Với mỗi cột (trường) cần phải có:
 - + Một tên duy nhất (trong bảng lưu giữ nó).
 - + Một mô tả ngắn gọn.
 - + Một kiểu dữ liệu (ví dụ: integer, char, date, logical,... phụ thuộc vào HQTCSDDL cụ thể cài đặt CSDL)
 - + Một kích thước (mặc định hay chỉ rõ tùy từng kiểu dữ liệu)
 - + Chú ý: tên của bảng, cột không nên quá dài, cần đủ nghĩa và thường không có dấu.
 - Các cột (trường) tùy chọn và các cột (trường) bắt buộc:
 - + Nếu cột là bắt buộc thì người sử dụng cần phải cung cấp một giá trị cho cột này trong mỗi dòng (bản ghi) thêm vào bảng khi cập nhật CSDL. Cột sẽ cần phải được chỉ rõ là NOT NULL
 - + Nếu cột là tùy chọn thì nó có thể nhận giá trị null.
- Chú ý: có thể lưu trữ một giá trị 0 cho một cột tùy chọn khi nó không có giá trị nhưng cách này không phải là cách hay.

4.5.2. Xác định khóa

- Khoá chính và các khoá ngoài

- + Hầu hết các HQTCSDL hiện đại cho phép chỉ rõ khoá chính, khoá ngoài khi định nghĩa các bảng.
- + Nếu các HQTCSDL không cho phép định nghĩa tự động thì cần phải chỉ rõ. Khoá chính cần phải duy nhất và not null. Khoá ngoài cần phải liên quan tới một thể hiện của khoá chính đã có hoặc là nhận giá trị null.
- Giá trị hợp lệ và giá trị mặc định

Ví dụ: Một mã KH được định nghĩa nằm trong khoảng từ 1 đến 1000 (giá trị hợp lệ). Giá trị mặc định cho số lượng bán là 1 (giá trị mặc định).

- + Thuận lợi của việc chỉ rõ giá trị hợp lệ là nó sẽ được gắn vào trong tất cả các chương trình khi lập trình. Đảm bảo sự thống nhất trong việc kiểm tra tính đúng đắn của dữ liệu.
- + Giá trị mặc định có thể được sử dụng để giảm bớt việc gõ máy cho người sử dụng. Nó giúp cho việc nhập dữ liệu nhanh hơn và giảm bớt lỗi.
- Khoá giả
 - + Để tăng tốc độ tìm kiếm, tiết kiệm không gian nhớ và giảm lỗi các khoá cần phải ngắn. Nếu một vài khoá quá dài cần phải thay chúng bằng một khoá giả.
 - + Có hai vấn đề khi đưa ra một khoá giả:
 - * Cần phải thiết lập một kỹ thuật để sinh tự động các giá trị duy nhất cho khoá (sinh khoá tự động)
 - * Cần phải cung cấp một ràng buộc duy nhất cho các định danh tự nhiên để ngăn chặn việc lưu trữ một vài bản ghi cho cùng một thực thể có cùng một dữ liệu nhưng chỉ khác nhau về giá trị khoá.

4.5.3. Thiết lập quan hệ

- Các quan hệ
 - + Các quan hệ giữa các bảng sẽ được thiết lập bởi các khoá ngoài.
 - + Cần phải đảm bảo là cái ràng buộc khoá ngoài sẽ được cài đặt.

4.6. Xem xét hiệu suất thực thi CSDL

Cần phải đánh giá việc thực thi CSDL, đó là tính hợp lý của thời gian đáp ứng các câu hỏi truy vấn. Khi xem xét hiệu suất thực thi CSDL cần phải:

- Chạy thử CSDL với một tập đủ lớn các bản ghi trong môi trường thực tế.
- Mô phỏng một số người dùng cùng thực hiện một truy vấn trong CSDL đa người sử dụng.
- Chạy trộn lẫn một số chức năng với thời gian quy định.

4.7. Điều chỉnh CSDL

Sau khi Xem xét hiệu suất thực thi CSDL ta cần thực hiện tiếp các công việc sau :

- Tạo các thủ tục sao lưu và phục hồi CSDL.
- Tổ chức bảo trì và bảo mật cho CSDL.
- Thiết lập các ràng buộc toàn vẹn cho CSDL.
- Xác định không gian nhớ cho các bảng CSDL và các chỉ số.
- Thiết lập vùng nhớ đệm cho CSDL
- Phân cụm CSDL, tổ chức lưu trữ vật lý tối ưu cho CSDL trên đĩa.
- Đảm bảo điều khiển tương tranh trong CSDL đa người sử dụng.
- Tạo ra các khung nhìn của CSDL cho từng đối tượng người dùng.
- Xác định sự phân tán dữ liệu và xử lý giữa các máy trạm (client) và máy chủ (server) trong một hệ thống phân tán.

CÂU HỎI ÔN TẬP CHƯƠNG 4

1. Trình bày các bước tiến hành phân tích và thiết kế cơ sở dữ liệu
2. Các thành phần của mô hình thực thể liên kết và cách xây dựng mô hình thực thể liên kết của hệ thống
3. Các dạng chuẩn trong mô hình quan hệ
4. Phương pháp chuẩn hóa quan hệ

Thiết lập quan hệ trong cơ sở dữ liệu