

MỤC LỤC

1	Chương 1. CÁC KHÁI NIỆM CƠ BẢN	5
1.1	Tại sao phải có một cơ sở dữ liệu	5
1.2	Định nghĩa một cơ sở dữ liệu.....	5
1.2.1	Khái niệm.....	5
1.2.2	Ưu điểm.....	6
1.2.3	Vấn đề cần giải quyết.....	6
1.3	Hệ quản trị cơ sở dữ liệu (DataBase Management System_DBMS)	7
1.3.1	Ví dụ.....	7
1.3.2	Khái niệm.....	7
1.4	Hệ thống cơ sở dữ liệu (Database System)	8
1.5	Các đối tượng sử dụng CSDL.....	8
1.5.1	Đối tượng trực tiếp.....	8
1.5.2	Đối tượng gián tiếp	9
1.6	Lợi ích của việc sử dụng HQTCSDL	9
2	Chương 2. NHỮNG KHÁI NIỆM VÀ KIẾN TRÚC CỦA HỆ THỐNG CƠ SỞ DỮ LIỆU	11
2.1	Mô hình dữ liệu, lược đồ và trường hợp (Data Models, Schemas, Instances) 11	
2.1.1	Phân loại mô hình dữ liệu	11
2.1.2	Lược đồ(Schema) , minh hoạ (instances), và trạng thái (State).....	14
2.2	Lược đồ kiến trúc của hệ quản trị cơ sở dữ liệu (DBMS Architecture) và sự độc lập dữ liệu (Data Independence)	15
2.2.1	Lược đồ kiến trúc 3 mức của HQTCSDL.....	16
2.2.2	Độc lập dữ liệu.....	17
2.3	Ngôn ngữ của HQTCSDL.....	17
2.4	Các tính năng của HQTCSDL	17
2.5	Phân loại HQTCSDL	17
3	Chương 3. MÔ HÌNH QUAN HỆ - THỰC THỂ (Entity – Relationship Model)	19
3.1	Sử dụng mô hình dữ liệu khái niệm mức cao để thiết kế cơ sở dữ liệu...19	
3.2	Mục đích của mô hình khái niệm ER(Entity – Relationship Model)	20

3.3	Ví dụ về một cơ sở dữ liệu ứng dụng	20
3.4	Kiểu thực thể(Entity Type), Thuộc tính (Attributes), Khoá (Keys)	22
3.4.1	Thực thể (Entities) và thuộc tính (Attributes).....	22
3.4.2	Kiểu thực thể, Khoá và tập giá trị	25
3.5	Liên kết, Kiểu liên kết và các Ràng buộc liên kết.....	25
3.5.1	Định nghĩa liên kết và kiểu liên kết	25
3.5.2	Bậc của kiểu liên kết	26
3.5.3	Ràng buộc liên kết.....	27
3.6	Kiểu thực thể yếu(Weak Entity)	29
3.7	Tổng quát hóa và chuyên biệt hóa	29
3.7.1	Thực thể con và thực thể chính	30
3.7.2	Các thực thể con loại trừ	30
3.8	Các ký hiệu và quy ước đặt tên trong mô hình ER.....	31
3.8.1	Các ký hiệu.....	31
3.8.2	Quy tắc đặt tên	31
3.9	Xây dựng một mô hình ER.....	32
3.9.1	Các bước xây dựng sơ đồ ER.....	32
3.9.2	Mô hình ER cho cơ sở dữ liệu COMPANY	33
3.9.3	Bài tập	34
4	Chương 4. MÔ HÌNH CƠ SỞ DỮ LIỆU QUAN HỆ.....	37
4.1	Khái niệm mô hình quan hệ	37
4.2	Các thành phần cơ bản của mô hình	37
4.2.1	Một số khái niệm của mô hình quan hệ	37
4.2.2	Quan hệ:	37
4.2.3	Các tính chất của một quan hệ	38
4.2.4	Các ràng buộc toàn vẹn trên quan hệ	38
4.2.5	Các phép toán trên CSDL quan hệ.....	41
5	Chương 5. CHUYỂN TỪ MÔ HÌNH ER SANG MÔ HÌNH QUAN HỆ	48
6	Chương 6. PHỤ THUỘC HÀM VÀ CHUẨN HOÁ QUAN HỆ	55
6.1	Một số hướng dẫn khi thiết kế cơ sở dữ liệu quan hệ	55
6.2	Phụ thuộc hàm(Functional Dependencies).....	56
6.2.1	Định nghĩa phụ thuộc hàm.....	56

6.2.2	Hệ tiên đề Armstrong.....	57
6.2.3	Bao đóng của tập phụ thuộc hàm.....	57
6.2.4	Bao đóng của tập thuộc tính X trên F.....	57
6.2.5	Khoá của quan hệ.....	58
6.2.6	Tập phụ thuộc hàm tương đương.....	59
6.2.7	Tập phụ thuộc hàm tối thiểu.....	59
6.3	Các dạng chuẩn của quan hệ.....	60
6.3.1	Định nghĩa các dạng chuẩn.....	60
6.3.2	Phép phân rã các lược đồ quan hệ.....	66
6.4	Chuẩn hoá quan hệ.....	70
6.4.1	Thuật toán phân rã lược đồ quan hệ thành các lược đồ quan hệ con ở BCNF.....	70
6.4.2	Thuật toán phân rã một lược đồ quan hệ thành các lược đồ con ở 3NF.....	72
7	Chương 7. THIẾT KẾ CƠ SỞ DỮ LIỆU VẬT LÝ (Tham khảo).....	75
7.1	Nội dung thiết kế file vật lý và cơ sở dữ liệu vật lý.....	75
7.1.1	Quá trình thiết kế.....	75
7.1.2	Sản phẩm thiết kế.....	76
7.2	Thiết kế các trường.....	77
7.2.1	Yêu cầu thiết kế trường.....	77
7.2.2	Chọn kiểu và cách biểu diễn dữ liệu.....	78
7.3	Thiết kế các bản ghi vật lý.....	80
7.3.1	Phi chuẩn.....	80
7.3.2	Quản lý trường có độ dài cố định.....	81
7.3.3	Quản lý trường có độ dài biến đổi.....	81
7.4	Thiết kế file vật lý.....	82
7.4.1	Các loại file.....	82
7.4.2	Các phương pháp truy cập.....	82
7.4.3	Tổ chức file.....	83
7.4.4	Ví dụ về thiết kế file.....	87

1 Chương 1. CÁC KHÁI NIỆM CƠ BẢN

Trong nhiều năm gần đây, thuật ngữ Cơ sở dữ liệu - Database đã trở nên quen thuộc trong nhiều lĩnh vực. Các ứng dụng tin học vào quản lý ngày càng nhiều và đa dạng, hầu hết các lĩnh vực kinh tế, xã hội... đều đã ứng dụng các thành tựu mới của tin học vào phục vụ công tác chuyên môn của mình. Chính vì lẽ đó mà ngày càng nhiều người quan tâm đến thiết kế, xây dựng và ứng dụng cơ sở dữ liệu (CSDL).

Trong chương này, chúng ta sẽ tìm hiểu thế nào là cơ sở dữ liệu và các khái niệm liên quan đến nó. Trước hết, chúng ta sẽ tìm hiểu lý do tại sao cần phải quản lý dữ liệu bằng CSDL?

1.1 Tại sao phải có một cơ sở dữ liệu

Hệ thống các tệp tin cổ điển

Cho đến nay vẫn còn một số đơn vị kinh tế, hành chính sự nghiệp... sử dụng mô hình hệ thống các tệp tin cổ điển: chúng được tổ chức riêng rẽ, phục vụ cho một mục đích của một đơn vị hay một đơn vị con trực thuộc cụ thể.

- Ưu điểm:

Việc xây dựng hệ thống các tệp tin riêng tại từng đơn vị quản lý ít tốn thời gian bởi khối lượng thông tin cần quản lý và khai thác là nhỏ, không đòi hỏi đầu tư vật chất và chất xám nhiều, do đó triển khai ứng dụng nhanh.

Thông tin được khai thác chỉ phục vụ mục đích hẹp nên khả năng đáp ứng nhanh chóng, kịp thời.

- Nhược điểm:

Thông tin được tổ chức riêng rẽ ở nhiều nơi nên việc cập nhật dễ làm mất tính nhất quán dữ liệu.

Hệ thống thông tin được tổ chức thành các hệ thống file riêng lẻ nên thiếu sự chia sẻ thông tin giữa các nơi.

Qua phân tích trên, chúng ta nhận thấy việc tổ chức dữ liệu theo hệ thống tệp tin hoàn toàn không phù hợp với những hệ thống thông tin lớn. Việc xây dựng một hệ thống thông tin đảm bảo được tính nhất quán dữ liệu, đáp ứng được nhu cầu khai thác đồng thời của nhiều người là thực sự cần thiết.

1.2 Định nghĩa một cơ sở dữ liệu

1.2.1 Khái niệm

CSDL và công nghệ CSDL đã có những tác động to lớn trong việc phát triển sử dụng máy tính. Có thể nói rằng CSDL ảnh hưởng đến tất cả các nơi có sử dụng máy tính:

Kinh doanh (thông tin về sản phẩm, khách hàng, ...)

Giáo dục (thông tin về sinh viên, điểm, ..)

Thư viện (thông tin về tài liệu, tác giả, độc giả...)

Y tế (thông tin về bệnh nhân, thuốc....)...

Như vậy, cơ sở dữ liệu là gì?

CSDL là tập hợp các dữ liệu có cấu trúc và liên quan với nhau được lưu trữ trên máy tính, được nhiều người sử dụng và được tổ chức theo một mô hình.

Ví dụ:

Danh bạ điện thoại là một ví dụ về CSDL.

- Là các thông tin có ý nghĩa
- Là tập hợp các thông tin có cấu trúc.
- Các thông tin này có liên quan với nhau và có thể hệ thống được.

Trong khái niệm này, chúng ta cần nhấn mạnh, CSDL là tập hợp các thông tin có tính chất hệ thống, không phải là các thông tin rời rạc, không có liên quan với nhau. Các thông tin này phải có cấu trúc và tập hợp các thông tin này phải có khả năng đáp ứng nhu cầu khai thác của nhiều người sử dụng một cách đồng thời. Đó cũng chính là đặc trưng của CSDL.

1.2.2 Ưu điểm

Từ khái niệm trên, ta thấy rõ ưu điểm nổi bật của CSDL là:

Giảm sự trùng lặp thông tin xuống mức thấp nhất và do đó đảm bảo được tính nhất quán và toàn vẹn dữ liệu (Cấu trúc của cơ sở dữ liệu được định nghĩa một lần. Phần định nghĩa cấu trúc này gọi là meta-data, và được Catalog của HQTCSDB lưu trữ).

Đảm bảo sự độc lập giữa dữ liệu và chương trình ứng dụng (Insulation between programs and data): Cho phép thay đổi cấu trúc, dữ liệu trong cơ sở dữ liệu mà không cần thay đổi chương trình ứng dụng.

Trừu tượng hoá dữ liệu (Data Abstraction): Mô hình dữ liệu được sử dụng để làm ẩn lưu trữ vật lý chi tiết của dữ liệu, chỉ biểu diễn cho người sử dụng mức khái niệm của cơ sở dữ liệu.

Nhiều khung nhìn (multi-view) cho các đối người dùng khác nhau: Đảm bảo dữ liệu có thể được truy xuất theo nhiều cách khác nhau. Vì yêu cầu của mỗi đối tượng sử dụng CSDL là khác nhau nên tạo ra nhiều khung nhìn vào dữ liệu là cần thiết.

Đa người dùng (multi-user): Khả năng chia sẻ thông tin cho nhiều người sử dụng và nhiều ứng dụng khác nhau.

1.2.3 Vấn đề cần giải quyết

Để đạt được các ưu điểm trên, CSDL đặt ra những vấn đề cần giải quyết. Đó là:

Tính chủ quyền của dữ liệu: Do tính chia sẻ của CSDL nên chủ quyền của CSDL dễ bị xâm phạm.

Tính bảo mật và quyền khai thác thông tin của người sử dụng: Do có nhiều người được phép khai thác CSDL nên cần thiết phải có một cơ chế bảo mật và phân quyền hạn khai thác CSDL.

Tranh chấp dữ liệu: Nhiều người được phép cùng truy cập vào CSDL với những mục đích khác nhau: Xem, thêm, xóa hoặc sửa dữ liệu. Cần phải có cơ chế ưu

tiên truy cập dữ liệu hoặc giải quyết tình trạng xung đột trong quá trình khai thác cạnh tranh. Cơ chế ưu tiên có thể được thực hiện bằng việc cấp quyền (hay mức độ) ưu tiên cho từng người khai thác.

Đảm bảo dữ liệu khi có sự cố: Việc quản lý dữ liệu tập trung có thể làm tăng nguy cơ mất mát hoặc sai lệch thông tin khi có sự cố mất điện đột xuất hoặc đĩa lưu trữ bị hỏng. Một số hệ điều hành mạng có cung cấp dịch vụ sao lưu ảnh đĩa cứng (cơ chế sử dụng đĩa cứng dự phòng - RAID), tự động kiểm tra và khắc phục lỗi khi có sự cố. Tuy nhiên, bên cạnh dịch vụ của hệ điều hành, để đảm bảo an toàn cho CSDL, nhất thiết phải có một cơ chế khôi phục dữ liệu khi có sự cố xảy ra.

1.3 Hệ quản trị cơ sở dữ liệu (DataBase Management System_DBMS)

1.3.1 Ví dụ

Như chúng ta đã biết, kích thước và độ phức tạp của CSDL rất khác nhau.

Ví dụ:

Danh bạ điện thoại của một quốc gia, một thành phố.. chứa tới hàng triệu số và những thông tin cần thiết về khách hàng.

Trong trường đại học có tới hàng ngàn sinh viên. Nhà trường phải quản lý tất cả những thông tin liên quan đến sinh viên như: tên, ngày sinh, quê quán, địa chỉ, kết quả học tập...

Xét một Ví dụ về CSDL quản lý tài liệu và độc giả trong thư viện quốc gia. Giả sử rằng có 100 triệu cuốn sách, mỗi cuốn sách cần lưu 10 thông tin liên quan, mỗi thông tin chứa tối đa 400 kí tự thì CSDL sẽ phải có tối thiểu $100 * 10^6 * 400 * 10$ kí tự (bytes). Như vậy, dung lượng bộ nhớ cần dùng là: $100 * 10^6 * 400 * 10 = 400$ GB.

Ta thấy, bộ nhớ cũng là vấn đề cần phải được giải quyết. Tuy nhiên, vấn đề quan trọng hơn ở đây lại là cách thức tổ chức dữ liệu trong một cơ sở dữ liệu để phục vụ cho việc truy cập, tìm kiếm, cập nhật,... nhanh chóng và an toàn hơn.

Việc tổ chức dữ liệu như thế nào được thực hiện thông qua Hệ quản trị cơ sở dữ liệu(HQTCSDL).

Vậy hệ quản trị cơ sở dữ liệu (HQTCSDL) là gì?

1.3.2 Khái niệm.

HQTCSDL là tập hợp các phần mềm cho phép định nghĩa các cấu trúc để lưu trữ thông tin trên máy, nhập dữ liệu, thao tác trên các dữ liệu đảm bảo sự an toàn và bí mật của dữ liệu.

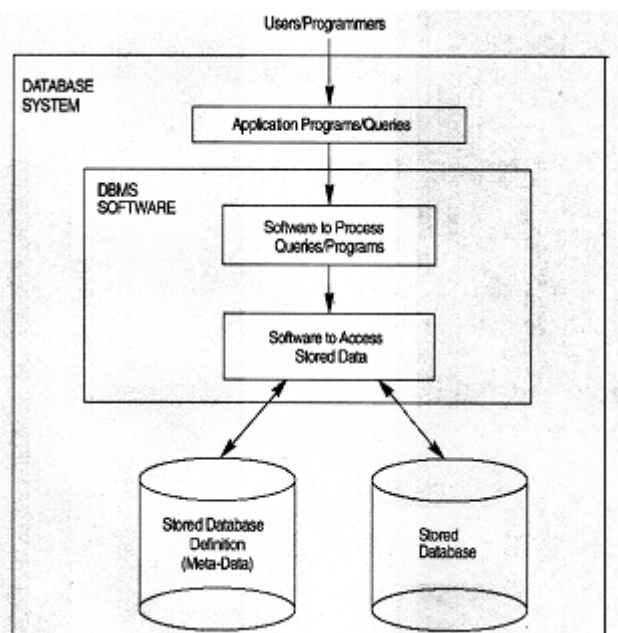
Định nghĩa cấu trúc: Định nghĩa cấu trúc CSDL bao gồm việc xác định kiểu dữ liệu, cấu trúc và những ràng buộc cho dữ liệu được lưu trữ trong CSDL.

Nhập dữ liệu: Là việc lưu trữ dữ liệu vào các thiết bị lưu trữ trung gian được điều khiển bằng HQTCSDL.

Thao tác dữ liệu: thao tác trên CSDL bao gồm những chức năng như truy xuất cơ sở dữ liệu để tìm kiếm thông tin cần thiết, cập nhật cơ sở dữ liệu và tổng hợp những báo cáo từ dữ liệu.

1.4 Hệ thống cơ sở dữ liệu (Database System)

Là phần mềm HQTCSDL cùng với dữ liệu của bản thân cơ sở dữ liệu đó.



Hình 1.1. Môi trường hệ thống cơ sở dữ liệu đơn giản

1.5 Các đối tượng sử dụng CSDL

Đối với các cơ sở dữ liệu nhỏ, mang tính cá nhân như lịch làm việc, danh bạ điện thoại cá nhân... thì chỉ cần một người để tạo ra và thao tác trên nó. Tuy nhiên, đối với các CSDL lớn như: quản lý tài chính của ngân hàng nhà nước, điều hành các chuyến bay cho các sân bay quốc tế... cần phải có rất nhiều người tham gia thiết kế, xây dựng, bảo trì CSDL và hàng trăm người sử dụng. Trong phần này, chúng ta tìm hiểu xem ai là người thao tác với CSDL hàng ngày. Và trong phần sau, chúng ta xem xét những người không trực tiếp tham gia một CSDL cụ thể, họ là người duy trì môi trường hệ thống CSDL.

1.5.1 Đối tượng trực tiếp

1.5.1.1 Quản trị cơ sở dữ liệu

Trong những tổ chức có nhiều người cùng sử dụng chung một nguồn dữ liệu thì nhất thiết phải có một người đứng đầu quản lý, chịu trách nhiệm đối với nguồn dữ liệu này. Đó chính là người quản trị cơ sở dữ liệu (Database Administrators _ DBA).

DBA có nhiệm vụ tổ chức nội dung của cơ sở dữ liệu, tạo và phân quyền cho người sử dụng, đưa ra yêu cầu về phần cứng và phần mềm... nếu cần thiết. DAB chịu trách nhiệm bảo vệ an toàn, Backup thông tin... khi có sự cố.

1.5.1.2 Thiết kế cơ sở dữ liệu

Người thiết kế CSDL chịu trách nhiệm:

- Xác định những dữ liệu nào cần lưu trữ trong CSDL
- Lựa chọn những cấu trúc thích hợp để biểu diễn và lưu trữ những dữ liệu này.

- Phỏng vấn tất cả những người sử dụng CSDL sau này để hiểu được những yêu cầu của họ đối với CSDL
- Tiến hành phân tích thiết kế hệ thống sau khi thống nhất được tất cả các yêu cầu của người sử dụng

1.5.1.3 Người sử dụng cuối

Người sử dụng cuối là những người truy cập CSDL để:

- Truy vấn
- Cập nhật
- Thống kê, báo cáo

1.5.1.4 Phân tích hệ thống và Lập trình ứng dụng

Phân tích hệ thống để định rõ những yêu cầu của người sử dụng cuối cùng, thống nhất để đưa ra khung nhìn cho từng đối tượng người sử dụng, quản lý các giao tác (transactions)...

Lập trình ứng dụng:

- Thực hiện các yêu cầu thông qua lập trình bằng những ngôn ngữ phù hợp
- Chạy thử chương trình (test)
- Chữa lỗi và gỡ rối chương trình (debug)
- Viết tài liệu, hướng dẫn sử dụng.
- Bảo trì hệ thống

1.5.2 Đối tượng gián tiếp

Ngoài những đối tượng trực tiếp tham gia vào một CSDL cụ thể như đã nói ở trên, còn có một đội ngũ những người phân tích, phát triển, và thực hiện tạo ra môi trường hệ thống và phần mềm của hệ quản trị cơ sở dữ liệu. Những người này không trực tiếp thao tác trên một hệ quản trị CSDL nào cụ thể. Họ là:

- Người phân tích và thực hiện tạo ra hệ thống của HQTCSDL
- Những nhà phát triển hệ công cụ (Tool developers)
- Người kiểm thử và bảo trì hệ thống

1.6 Lợi ích của việc sử dụng HQTCSDL

- Hạn chế dư thừa dữ liệu.
- Ngăn cản truy cập dữ liệu bất hợp pháp (bảo mật và phân quyền sử dụng).
- Cung cấp khả năng lưu trữ lâu dài cho các đối tượng và cấu trúc dữ liệu.
- Cho phép suy dẫn dữ liệu (từ dữ liệu này suy ra dữ liệu khác) sử dụng Rules.
- Cung cấp giao diện đa người dùng.
- Cho phép biểu diễn mối quan hệ phức tạp giữa các dữ liệu.

- Đảm bảo ràng buộc toàn vẹn dữ liệu (Enforcing Integrity Constraints).
- Cung cấp thủ tục sao lưu và phục hồi (backup và recovery)

TaiLieu.vn

2 Chương 2. NHỮNG KHÁI NIỆM VÀ KIẾN TRÚC CỦA HỆ THỐNG CƠ SỞ DỮ LIỆU

2.1 Mô hình dữ liệu, lược đồ và trường hợp (Data Models, Schemas, Instances)

Một trong những đặc điểm cơ bản của cơ sở dữ liệu là cung cấp một số mức độ trừu tượng hoá dữ liệu bằng cách làm ẩn đi cách thức tổ chức dữ liệu- cái mà hầu hết người dùng không cần biết đến.

Mô hình dữ liệu (Data Model): Là một tập những khái niệm dùng để biểu diễn cấu trúc của cơ sở dữ liệu-cung cấp những điều kiện cần thiết để đạt được mức độ trừu tượng dữ liệu. Cấu trúc dữ liệu bao gồm kiểu dữ liệu (data types) và mối quan hệ giữa các dữ liệu (relationships) và những ràng buộc (constraints) mà cơ sở dữ liệu phải tuân theo.

Hầu hết mô hình dữ liệu đều có một tập hợp các thao tác cơ bản (basic operations) để truy vấn và cập nhật dữ liệu.

- Thao tác chung (generic operations): Thêm (insert), Xoá (delete), Sửa (modify), Truy cập (retrieve)
- Thao tác do người dùng định nghĩa (user-defined operations)

2.1.1 Phân loại mô hình dữ liệu

Có rất nhiều mô hình dữ liệu đã được đưa ra, chúng ta có thể phân loại chúng theo những kiểu khái niệm mà họ đã dùng để biểu diễn cấu trúc cơ sở dữ liệu. Mô hình dữ liệu được chia làm 3 loại sau:

a. Mô hình khái niệm (Conceptual (high-level, semantic) data models):

Cung cấp những khái niệm gần gũi với đa số người sử dụng, mô hình này chỉ ra cái gì được đưa vào để quản lý. Mô hình này là phương tiện để những người phân tích thiết kế giao tiếp với người sử dụng, nhằm thu thập thông tin, xác định đúng đắn và đầy đủ yêu cầu của hệ thống.

Mô hình này sử dụng cấu trúc dữ liệu là: thực thể (entity), thuộc tính (attribute) và mối liên kết (relationship)

b. Mô hình dữ liệu vật lý (Physical (low-level, internal) data models):

Cung cấp những khái niệm để biểu diễn chi tiết cách thức dữ liệu được lưu trữ trong máy tính. Mô hình này chỉ ra định dạng bản ghi (record formats), thứ tự sắp xếp các bản ghi (record ordering) và đường dẫn để truy cập dữ liệu (access paths).

c. Mô hình dữ liệu thể hiện (Implementation (record-oriented) data models):

Mô tả các dữ liệu bằng cách sử dụng những ký pháp tương ứng với mô hình dữ liệu mà một hệ quản trị cơ sở dữ liệu sử dụng.

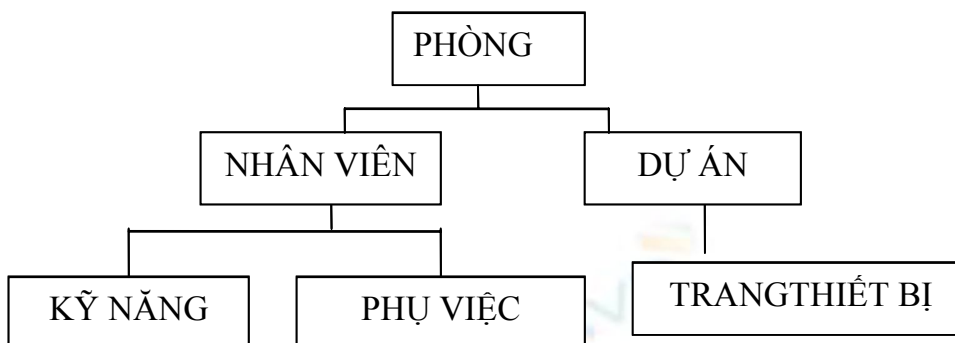
Các loại mô hình cơ sở dữ liệu thể hiện:

c1. Mô hình phân cấp

Mô hình CSDL phân cấp được biểu diễn dưới dạng cây và các đỉnh của cây là các bản ghi. Các bản ghi liên kết với nhau theo mối quan hệ cha-con.

- Một cha có nhiều con
- Một con chỉ có một cha

Ví dụ:



Hình 2.1. Minh họa mô hình cơ sở dữ liệu phân cấp

Ưu điểm:

- Thể hiện dễ dàng quan hệ 1-N.
- Việc phân chia dữ liệu dễ thể hiện, đảm bảo an toàn dữ liệu
- Tính độc lập của chương trình và các dữ liệu được đảm bảo

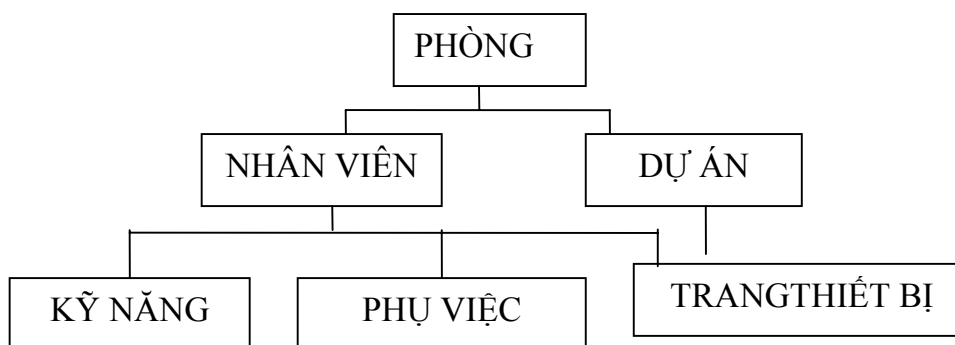
Nhược điểm:

- Không thể hiện được mối quan hệ M-N
- Trong một hệ thống phân cấp, dữ liệu được tổ chức như trên dẫn đến khó sửa đổi dữ liệu.

c2. Mô hình mạng

Cấu trúc cơ bản trong mô hình mạng là những tập hợp và mỗi tập hợp có bản ghi là bản ghi chủ và một số bản ghi thành viên. Mỗi thành viên có thể thuộc về nhiều tập hợp.

Ví dụ:



Hình 2.2. Minh họa mô hình cơ sở dữ liệu mạng

Ưu điểm:

- Dễ thể hiện mối liên kết M-N
- Kiểu truy cập dữ liệu mềm dẻo hơn kiểu phân cấp

Nhược điểm:

- Việc sửa đổi số liệu khó khăn.
- Với những lập trình viên, việc thiết kế CSDL khó.

c3. Mô hình quan hệ

Trong mô hình quan hệ, các dữ liệu được biểu diễn ở dạng các bảng với các dòng và các cột.

Trong mô hình quan hệ không có một cấu trúc vật lý nào của dữ liệu mô tả sự kết nối giữa các bảng. Thay vào đó, sự kết nối giữa các bảng được mô tả logic bằng các giá trị được lưu trữ trong các dòng của bảng. Chẳng hạn trong hình dưới đây, thuộc tính *ProCode*(Mã tỉnh) được lưu trong cả 2 bảng PROVINCE và bảng STUDENT, giá trị chung này cho phép người dùng liên kết được 2 bảng.

PROVINCE	
ProCode	ProName
04	Hà Nội
08	Tp Hồ Chí Minh
...	...

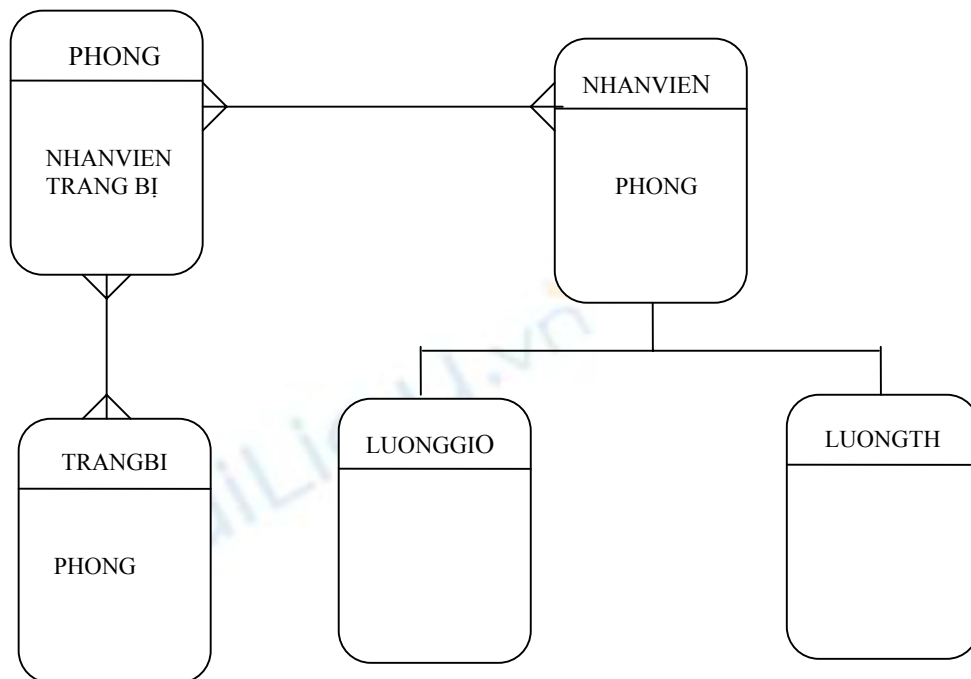
STUDENT			
StdNo	StdName	StdBird	ProCode
TD001	AA	9/16/1979	04
TD002	BB	6/19/1979	08
...

Hình 2.3. Minh họa mô hình cơ sở dữ liệu quan hệ

c4. Mô hình hướng đối tượng

Trong mô hình hướng đối tượng, các thuộc tính dữ liệu và các thao tác trên các dữ liệu này được bao gói trong một cấu trúc gọi là đối tượng.

Đối tượng có thể chứa các dữ liệu phức hợp như văn bản, hình ảnh, tiếng nói và hình ảnh động. Một đối tượng có thể yêu cầu hoặc xử lý dữ liệu từ một đối tượng khác bằng việc gửi đi một thông báo đến đối tượng đó. Mô hình hướng đối tượng biểu diễn một sơ đồ mới để lưu trữ và thao tác dữ liệu. Từ một đối tượng có thể sinh ra một đối tượng khác.



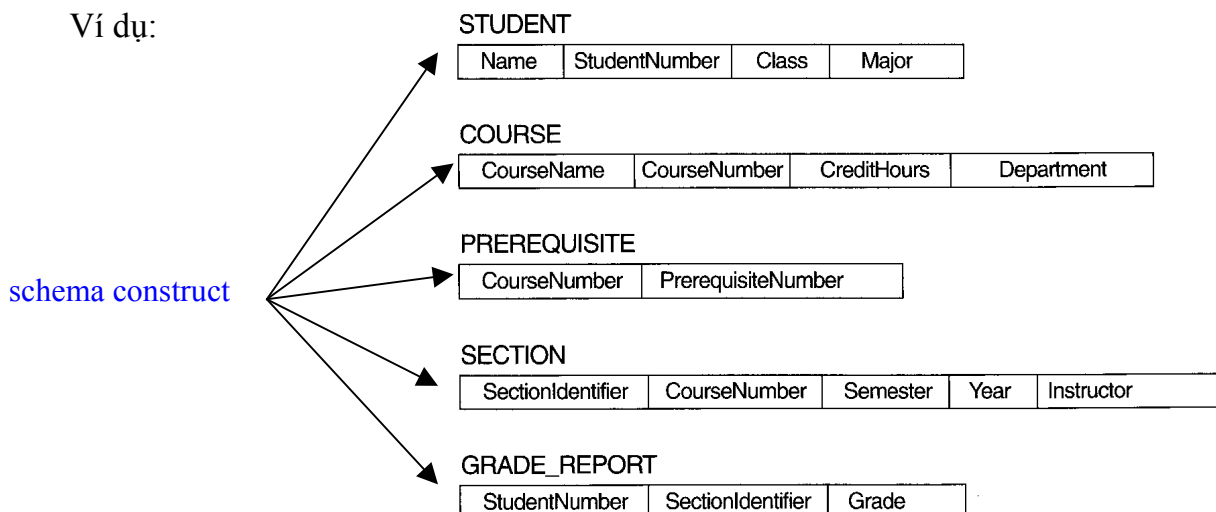
Hình 2.4. Minh họa mô hình cơ sở dữ liệu hướng đối tượng

2.1.2 Lược đồ (Schema) , minh họa (instances), và trạng thái (State)

Lược đồ cơ sở dữ liệu (Database Schema): là biểu diễn của cơ sở dữ liệu, bao gồm cấu trúc cơ sở dữ liệu và những ràng buộc trên dữ liệu.

Sơ đồ của lược đồ cơ sở dữ liệu (Schema Diagram): Là lược đồ cơ sở dữ liệu được biểu diễn thông qua sơ đồ.

Ví dụ:



Hình 2.5. Lược đồ cơ sở dữ liệu UNIVERSITY

Minh học cơ sở dữ liệu (Database Instance): Là dữ liệu thực sự được lưu trữ trong cơ sở dữ liệu ở thời điểm hiện tại. Database Instance cũng được gọi là trạng thái của cơ sở dữ liệu (**database state**)

Ví dụ:

STUDENT	Name	StudentNumber	Class	Major
	Smith	17	1	CS
	Brown	8	2	CS

COURSE	CourseName	CourseNumber	CreditHours	Department
	Intro to Computer Science	CS1310	4	CS
	Data Structures	CS3320	4	CS
	Discrete Mathematics	MATH2410	3	MATH
	Database	CS3380	3	CS

SECTION	SectionIdentifier	CourseNumber	Semester	Year	Instructor
	85	MATH2410	Fall	98	King
	92	CS1310	Fall	98	Anderson
	102	CS3320	Spring	99	Knuth
	112	MATH2410	Fall	99	Chang
	119	CS1310	Fall	99	Anderson
	135	CS3380	Fall	99	Stone

GRADE_REPORT	StudentNumber	SectionIdentifier	Grade
	17	112	B
	17	119	C
	8	85	A
	8	92	A
	8	102	B
	8	135	A

PREREQUISITE	CourseNumber	PrerequisiteNumber
	CS3380	CS3320
	CS3380	MATH2410
	CS3320	CS1310

Hình 2.5. Cơ sở dữ liệu UNIVERSITY

Như vậy, **Database Schema** rất hiếm khi thay đổi, còn **Database State** thay đổi bất kỳ khi nào có sự cập nhập dữ liệu.

2.2 Lược đồ kiến trúc của hệ quản trị cơ sở dữ liệu (DBMS Architecture) và sự độc lập dữ liệu (Data Independence)

Như chúng ta đã biết, các tính chất quan trọng nhất của cơ sở dữ liệu là: (1) Đảm bảo sự độc lập giữa chương trình ứng dụng và dữ liệu. (2) Hỗ trợ nhiều khung nhìn cho các đối tượng người dùng khác nhau. (3) Sử dụng danh mục để lưu trữ biểu diễn dữ liệu (schema). Trong phần này, chúng ta sẽ tìm hiểu kiến trúc của hệ quản trị

cơ sở dữ liệu, gọi là **Lược đồ kiến trúc 3 mức** (three –schema architecture). Sau đó chúng ta sẽ tìm hiểu về khái niệm độc lập dữ liệu.

2.2.1 Lược đồ kiến trúc 3 mức của HQTCSDL

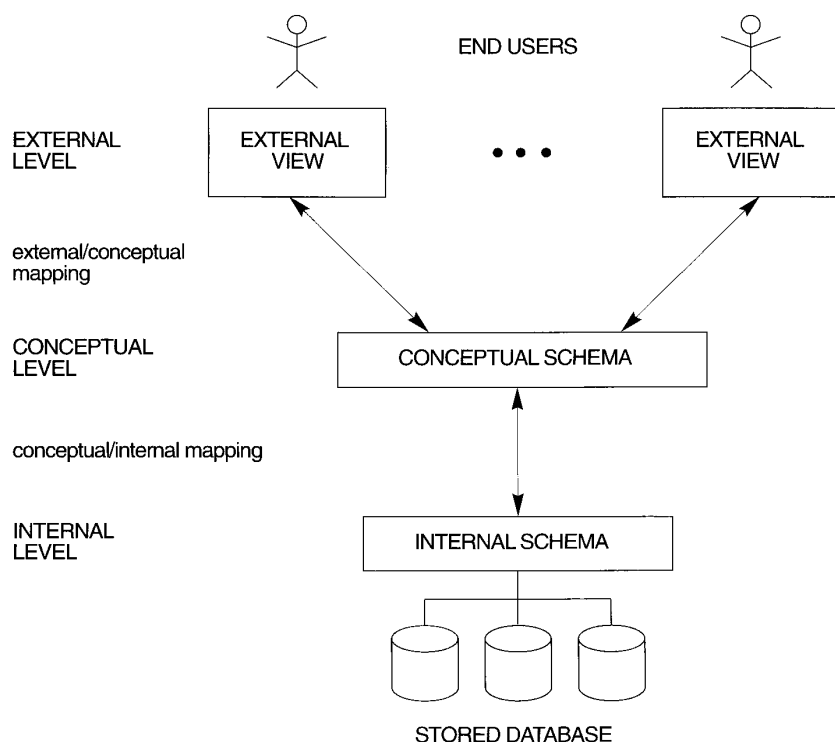
Mục đích của việc phân thành 3 mức trong kiến trúc của mô hình HQTCSDL là để tách biệt các ứng dụng của người sử dụng với cơ sở dữ liệu vật lý. Trong kiến trúc này, lược đồ có thể được định nghĩa ở 3 mức sau:

Lược đồ mức trong (Internal schema) ở Mức trong (Internal level) để biểu diễn chi tiết cấu trúc lưu trữ dữ liệu và cách thức truy cập dữ liệu. Lược đồ mức trong sử dụng mô hình dữ liệu vật lý (physical data model).

Lược đồ khái niệm (Conceptual schema) ở Mức khái niệm (Conceptual level) để biểu diễn cấu trúc và các ràng buộc trong toàn bộ cơ sở dữ liệu phục vụ cho việc giao tiếp với người sử dụng. Lược đồ khái niệm ẩn đi cách thức tổ chức vật lý của dữ liệu, chỉ tập trung vào việc biểu diễn các thực thể, các kiểu dữ liệu, mối quan hệ giữa các thực thể, các thao tác của người sử dụng và các ràng buộc giữa các dữ liệu. Mô hình dữ liệu mức khái niệm (Conceptual data model) hoặc Mô hình dữ liệu thể hiện (Implementation data model) có thể được sử dụng ở mức này.

Lược đồ mức ngoài (External Level) ở Mức ngoài (External level hoặc View level) để biểu diễn hàng loạt những khung nhìn của người sử dụng (user views). Mô hình dữ liệu mức cao (High-level data model) hoặc Mô hình dữ liệu thể hiện (Implementation data model) có thể được sử dụng ở mức này.

Ánh xạ giữa các mức này là cần thiết. Những chương trình làm việc với dữ liệu ở mức ngoài và được HQTCSDL ánh xạ tới dữ liệu vật lý ở mức trong để thực hiện.



Hình 2.6. Lược đồ kiến trúc 3 mức của HQTCSDL

2.2.2 Độc lập dữ liệu

Kiến trúc 3 mức của HQTCSDL có thể được sử dụng để giải thích khái niệm về độc lập dữ liệu. Độc lập dữ liệu là khả năng thay đổi lược đồ ở một mức nào đó của hệ thống cơ sở dữ liệu mà không cần phải thay đổi lược đồ ở mức cao hơn. Chúng ta có thể định nghĩa 2 kiểu của độc lập dữ liệu:

Độc lập dữ liệu logic (Logical data independence): cho phép thay đổi lược đồ khái niệm mà không cần phải thay đổi lược đồ mức ngoài hoặc những chương trình ứng dụng. Chúng ta có thể thay đổi lược đồ khái niệm để mở rộng (thêm các trường dữ liệu, các bản ghi) hoặc thu nhỏ cơ sở dữ liệu (xóa các trường dữ liệu, các bản ghi).

Độc lập dữ liệu vật lý (Physical data independence): cho phép thay đổi lược đồ mức trong mà không cần thay đổi lược đồ khái niệm. Có khi chúng ta cần thay đổi lược đồ mức trong vì các file vật lý đôi khi cần tổ chức lại để tăng hiệu quả thực hiện. Nếu kiểu dữ liệu không thay đổi thì chúng ta không cần thay đổi lại lược đồ khái niệm.

2.3 Ngôn ngữ của HQTCSDL

Vì HQTCSDL phục vụ có nhiều đối tượng người sử dụng khác nhau nên nó phải hỗ trợ ngôn ngữ để người sử dụng tương tác với nó. Trong phần này chúng ta sẽ tìm hiểu một số những ngôn ngữ được HQTCSDL hỗ trợ.

Ngôn ngữ định nghĩa dữ liệu (Data Definition Language - DDL): Là ngôn ngữ được các nhà quản trị cơ sở dữ liệu (DBA) và các nhà thiết kế cơ sở dữ liệu (database designers) dùng để xây dựng lược đồ khái niệm của cơ sở dữ liệu. Trong nhiều HQTCSDL, DDL cũng được sử dụng để định nghĩa lược đồ mức trong và mức ngoài (các khung nhìn). Một số HQTCSDL chia thành 2 ngôn ngữ: Ngôn ngữ định nghĩa lưu trữ (storage definition language – SDL) và Ngôn ngữ định nghĩa khung nhìn (view definition language -VDL) được dùng để định nghĩa lược đồ mức trong và mức ngoài.

Ngôn ngữ thực hiện dữ liệu (Data Manipulation Language -DML): Là ngôn ngữ được sử dụng để thao tác với dữ liệu bao gồm việc truy cập đến bản ghi và cập nhật dữ liệu cho bản ghi (thêm, sửa, xóa). Các lệnh DML có thể được nhúng trong ngôn ngữ lập trình hoặc thực hiện độc lập (ngôn ngữ truy vấn).

2.4 Các tính năng của HQTCSDL

- Nạp dữ liệu đang lưu trữ ở các tệp tin vào cơ sở dữ liệu (Conversion Tool).
- Cung cấp các thao tác truy xuất
- Đảm bảo tính độc lập dữ liệu
- Cung cấp thủ tục sao lưu và phục hồi (backup và recovery)
- Cung cấp các thủ tục điều khiển đồng thời (Do tính truy xuất đồng thời và cạnh tranh)
- Cung cấp các thủ tục kiểm soát bản quyền, kiểm tra tính đúng đắn của dữ liệu (để đảm bảo tính an toàn và toàn vẹn dữ liệu)

2.5 Phân loại HQTCSDL

Người ta phân loại HQTCSDL dựa trên một số tiêu chí:

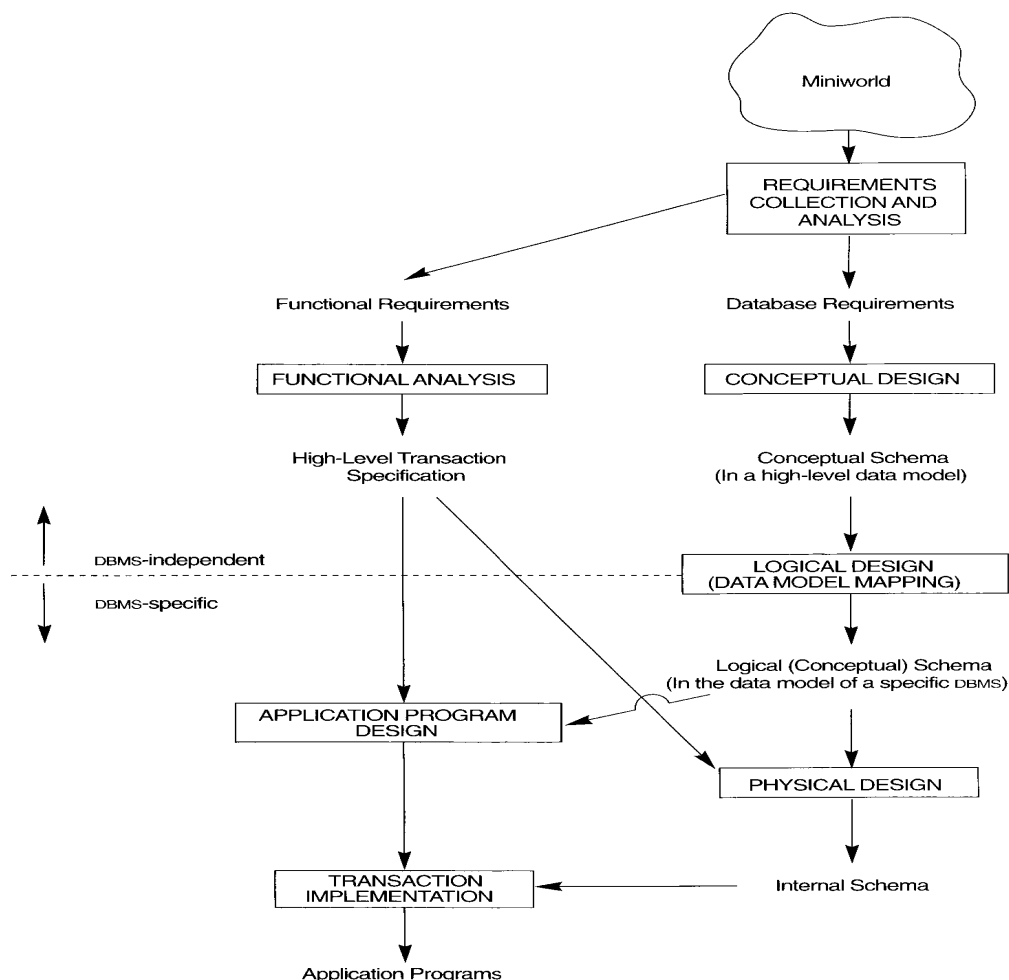
Dựa trên mô hình dữ liệu được HQTCSDL sử dụng: Mô hình quan hệ (Relational Data Model), Mô hình hướng đối tượng (Object Data Model), Mô hình mạng (Network Data Model), mô hình phân cấp (Hierarchical Data Model), mô hình quan hệ thực thể (Entity – Relationship Data Model)...

Tiêu chí khác:

- Số người sử dụng HQTCSDL: Đơn người dùng (Single-user), Đa người dùng (multi-user). Hầu hết các HQTCSDL hiện nay đều là HQTCSDL đa người dùng.
- Vị trí HQTCSDL: tập trung (sử dụng một máy đơn) hay phân tán (sử dụng nhiều máy tính).
- Giá của HQTCSDL: 100\$- 300\$; 10000\$- 100000\$

3 Chương 3. MÔ HÌNH QUAN HỆ - THỰC THỂ (Entity – Relationship Model)

3.1 Sử dụng mô hình dữ liệu khái niệm mức cao để thiết kế cơ sở dữ liệu



Hình 3.1. Các giai đoạn thiết kế cơ sở dữ liệu

Hình trên chỉ ra tiến trình thiết kế cơ sở dữ liệu một cách đơn giản. Bước đầu tiên là *tập hợp và phân tích yêu cầu hệ thống*. Trong bước này, người thiết kế cơ sở dữ liệu phải tiến hành thu thập yêu cầu của người sử dụng, sau đó viết tài liệu *những yêu cầu dữ liệu*. Kết quả của bước này là viết ra được tập hợp những yêu cầu tất cả các đối tượng người dùng một cách xúc tích. Từ đó, ta xác định được yêu cầu chức năng (*Functional Requirements*) của hệ thống.

Sau khi tất cả các yêu cầu đã được tập hợp và phân tích, bước tiếp theo là xây dựng *lược đồ khái niệm (conceptual schema)* cho cơ sở dữ liệu. Lược đồ khái niệm là nơi biểu diễn xúc tích những yêu cầu của người sử dụng và biểu diễn chi tiết những kiểu thực thể (*entity types*), quan hệ (*relationships*) và những ràng buộc (*constraints*) của dữ liệu, phần này sử dụng những khái niệm được cung cấp trong mô hình dữ liệu mức cao (*High level data model*). Mô hình dữ liệu mức cao là mô hình dữ liệu được dùng để giao tiếp với người sử dụng bình thường vì thế nó rất dễ hiểu, nó chỉ ra *cái gì*

cần được lưu trong cơ sở dữ liệu chứ không chỉ ra cụ thể dữ liệu được thực hiện *như thế nào*.

Bước tiếp theo trong quá trình thiết kế là cài đặt cơ sở dữ liệu trên một *mô hình dữ liệu thực hiện*, sử dụng các Hệ quản trị cơ sở dữ liệu nào đó (Hầu hết các hệ quản trị cơ sở dữ liệu hiện nay sử dụng mô hình dữ liệu quan hoặc hướng đối tượng). Vì thế, chúng ta cần thiết phải chuyển từ mô hình dữ liệu mức cao sang những mô hình dữ liệu thực hiện. Bước này được gọi là thiết kế logic (*Logical design*) hoặc ánh xạ mô hình dữ liệu (*Data model mapping*) và kết quả của bước này là *lược đồ cơ sở dữ liệu* trong mô hình cơ sở dữ liệu thực hiện.

Bước cuối cùng là thiết kế vật lý cho cơ sở dữ liệu (*physical design*), bao gồm việc thiết kế những cấu trúc lưu trữ dữ liệu bên trong, đường dẫn truy cập, tổ chức file của các file dữ liệu.

Trong chương này, chúng ta sẽ sử dụng *mô hình khái niệm ER* cho thiết kế lược đồ khái niệm (*Conceptual Schema*).

3.2 Mục đích của mô hình khái niệm ER(Entity – Relationship Model)

Qua bước xem xét tổng quát trên, ta thấy rằng, mô hình E-R là một mô tả logic chi tiết dữ liệu của một tổ chức hay một lĩnh vực nghiệp vụ. Nó giúp người thiết kế cơ sở dữ liệu mô tả thế giới thực gắn gũi với quan niệm và cách nhận nhìn nhận bình thường của con người. Nó là công cụ để phân tích thông tin nghiệp vụ.

Mục đích của mô hình E – R:

- Làm thống nhất quan điểm về dữ liệu của những người tham gia hệ thống: Người quản lý, người dùng cuối, người thiết kế hệ thống
- Xác định các xử lý về dữ liệu cũng như các ràng buộc trên các dữ liệu.
- Giúp đỡ việc thể hiện cơ sở dữ liệu về mặt cấu trúc: Sử dụng thực thể và các mối liên kết giữa các thực thể. Biểu diễn mô hình quan hệ thực thể bằng một sơ đồ.

3.3 Ví dụ về một cơ sở dữ liệu ứng dụng

Trong phần này, chúng ta sẽ tìm hiểu một ví dụ về cơ sở dữ liệu ứng dụng - gọi là *COMPANY* để minh họa các khái niệm trong mô hình ER và sử dụng nó trong khi thiết kế lược đồ khái niệm.

Cơ sở dữ liệu *COMPANY* cần lưu trữ thông tin về nhân viên (*employees*), phòng/ban (*departments*), và các dự án (*projects*) trong công ty. Sau khi tập hợp được tất cả các yêu cầu của hệ thống, người thiết kế cơ sở dữ liệu tiến hành mô tả lại *miniworld* bằng mô hình ER.

Sau đây là các quy tắc nghiệp vụ của hệ thống cơ sở dữ liệu *COMPANY*, chúng ta sẽ xây dựng mô hình ER từng bước để giới thiệu các khái niệm của mô hình ER.

1. Công ty (*COMPANY*) có nhiều phòng/ban (*DEPARTMENTS*). Mỗi phòng/ban có tên (*name*), mã số (*number*) duy nhất và có một nhân viên (*employee*) làm quản lý (*manages*) phòng/ban. Chúng ta lưu lại ngày bắt đầu (*start date*) làm quản lý phòng/ban của nhân viên đó.