

PHẦN I – NHẬP MÔN CƠ SỞ DỮ LIỆU QUAN HỆ

(Tổng số: 15 tiết, Lý thuyết: 15 tiết)

CHƯƠNG 1 – TỔNG QUAN VỀ CƠ SỞ DỮ LIỆU

(Tổng số: 2 tiết, Lý thuyết: 2 tiết)

§ 1. NHỮNG KHÁI NIỆM CƠ BẢN VỀ CƠ SỞ DỮ LIỆU**I. SỰ CẦN THIẾT CỦA CƠ SỞ DỮ LIỆU**

Trong những năm gần đây, thuật ngữ “Cơ sở dữ liệu” (CSDL - Database) đã trở nên khá quen thuộc không chỉ riêng với những người làm Tin học mà còn đối với cả những người làm trong nhiều lĩnh vực khác như Thống kê, Kinh tế, Quản lý Doanh nghiệp v.v... Các ứng dụng của Tin học vào công tác quản lý ngày càng nhiều hơn và càng đa dạng hơn. Có thể nói hầu hết các lĩnh vực kinh tế, xã hội, giáo dục, y tế v.v... 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 lĩnh vực thiết kế và xây dựng các CSDL.

Mục đích của Chương 1 chỉ đơn giản là cung cấp các khái niệm cơ bản về CSDL để các học viên có một cái nhìn ban đầu về một CSDL và một hệ quản trị CSDL. Trước hết chúng ta sẽ tìm hiểu lý do tại sao cần phải có một CSDL.

Hệ thống các tập tin cổ điển (File System):

Cho đến nay vẫn còn một số đơn vị kinh tế, hành chính sự nghiệp v.v... sử dụng mô hình hệ thống các tập 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. Chẳng hạn, ta hãy xét ví dụ sau:

Ví dụ: Tại một công ty người ta trang bị máy vi tính cho tất cả các phòng, ban nghiệp vụ. Bộ phận Văn phòng sử dụng máy vi tính để soạn thảo văn bản bằng Microsoft Word do thủ trưởng yêu cầu về tình hình hoạt động của đơn vị, trong đó có chỉ tiêu về tổng số công nhân viên chức chia theo trình độ chuyên môn được đào tạo. Phòng Kế toán sử dụng máy vi tính để tính lương và in danh sách lương của từng bộ phận trong đơn vị dựa trên danh sách cán bộ viên chức cùng hệ số lương và các hệ số phụ cấp của họ do phòng Tổ chức cung cấp. Thông tin mà phòng Kế toán quản lý và khai thác là: Họ và Tên, Hệ số lương, Hệ số phụ cấp, Phụ cấp khác của các công nhân viên chức (CNVC) xếp theo từng phòng ban và sử dụng công cụ văn phòng là Microsoft Excel. Phòng Tổ chức quản lý thông tin lý lịch của CNVC chi tiết hơn gồm: Họ, Tên (để riêng thành một cột “Tên” để tiện sắp xếp Alphabet), Giới tính, Ngày sinh, Ngày tuyển dụng, Hoàn cảnh gia đình, Quá trình đào tạo, Hệ số lương, Hệ số phụ cấp, Ngày xếp lương trên... nhưng thiếu thông tin về Phụ cấp khác của CNVC. Phần mềm được sử dụng để quản lý là FoxPro for Windows.

Trong khi đó, tại Tổng công ty của họ, các phòng ban nghiệp vụ cũng được trang bị máy vi tính. Phòng Tổ chức cán bộ tại Tổng công ty sử dụng phần mềm Microsoft Access để quản lý CNVC gồm các cán bộ chủ chốt từ trường phó phòng, quản đốc và phó quản đốc xí nghiệp trở lên của các công ty con trực thuộc. Thông tin quản lý tại đây cũng giống như thông tin quản lý tại phòng tổ chức của công ty con.

Nhận xét:

Ưu điểm:

- Việc xây dựng hệ thống các tập 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 cho mục đích hẹp nên khả năng đáp ứng nhanh chóng, kịp thời.

Nhược điểm:

- Do thông tin được tổ chức ở mỗi phòng ban mỗi khác, cũng như phần mềm công cụ để triển khai mỗi nơi cũng rất khác nhau nên sự phối hợp tổ chức và khai thác ở các phòng ban là khó khăn. Thông tin ở phòng ban này không sử dụng được cho phòng ban khác, tại đơn vị con với đơn vị cấp trên. Cùng một thông tin được nhập vào máy tại nhiều nơi khác nhau gây ra lãng phí công sức nhập tin và không gian lưu trữ trên các vật mang tin. Sự trùng lặp thông tin có thể dẫn đến tình trạng không nhất quán dữ liệu. Chẳng hạn, nhân viên Nguyễn Văn Quang được ghi đầy đủ ở phòng Tổ chức, nhưng tại phòng Kế toán chỉ ghi tắt là Nguyễn V Quang.

- Thông tin được tổ chức ở nhiều nơi nên việc cập nhật cũng dễ làm mất tính nhất quán dữ liệu. Một cán bộ chủ chốt của công ty có thay đổi về hoàn cảnh gia đình (mới cưới vợ / lấy chồng, sinh thêm con...) có thể được cập nhật ngay tại đơn vị nhưng sau một thời gian mới được cập nhật tại Tổng công ty.

- Do hệ thống được tổ chức thành các hệ thống tập tin riêng lẻ nên thiếu sự chia sẻ thông tin giữa các nơi. Việc kết nối các hệ thống này hay việc nâng cấp ứng dụng sẽ là rất khó khăn.

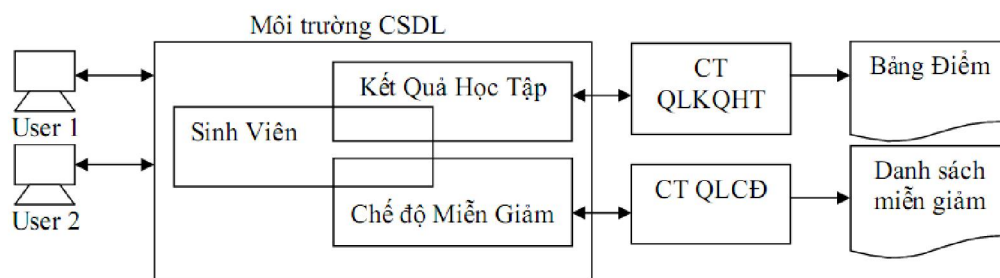
Qua phân tích trên chúng ta nhận thấy việc tổ chức dữ liệu theo hệ thống các tập tin có nhiều nhược điểm. Việc xây dựng một hệ thống tin đảm bảo được tính chất nhất quán dữ liệu, không trùng lặp thông tin mà vẫn đáp ứng được nhu cầu khai thác đồng thời của tất cả các phòng ban ở công ty và tổng công ty là thực sự cần thiết.

II. NHỮNG KHÁI NIỆM CƠ BẢN VỀ CƠ SỞ DỮ LIỆU

1. Khái niệm cơ sở dữ liệu

Cơ sở dữ liệu (CSDL) là sự tập hợp có tổ chức (có cấu trúc) các dữ liệu có liên quan luận lý với nhau, được lưu trữ trên các thiết bị lưu trữ thông tin (băng từ, đĩa từ...) để có thể thỏa mãn yêu cầu khai thác thông tin đồng thời của nhiều người sử dụng hay nhiều chương trình ứng dụng với nhiều mục đích khác nhau.

Ví dụ: Để quản lý hoạt động học tập trong một trường Đại học, chúng ta sẽ quản lý các dữ liệu là các thông tin về sinh viên, về các môn học, điểm thi... Do đó, chúng ta tổ chức các dữ liệu thành các bảng và lưu giữ chúng vào sổ sách hoặc sử dụng một phần mềm máy tính để lưu giữ chúng trên máy tính. Ta có một tập hợp các dữ liệu có liên quan đến nhau và mang nhiều ý nghĩa, đó là một CSDL.



Hình: CSDL quản lý sinh viên

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

2. Các tính chất của một cơ sở dữ liệu

Một CSDL biểu thị một khía cạnh nào đó của thế giới thực như hoạt động của một công ty, một nhà trường, một ngân hàng... Những thay đổi của thế giới thực phải được phản ánh

một cách trung thực vào trong CSDL. Những thông tin được đưa vào trong CSDL tạo thành một không gian CSDL hoặc là một “thế giới nhỏ”.

Một CSDL là một tập hợp dữ liệu liên kết với nhau một cách logic và mang một ý nghĩa cố hữu nào đó. Một CSDL không phải là một tập hợp tùy tiện.

Một CSDL được thiết kế và được phổ biến cho một mục đích riêng. Nó có một nhóm người sử dụng có chủ định và có một số ứng dụng được xác định phù hợp với mối quan tâm của người sử dụng. Nói cách khác, một CSDL có một nguồn cung cấp dữ liệu, một mức độ tương tác với các sự kiện trong thế giới thực và một nhóm người quan tâm tích cực đến các nội dung của nó.

Một CSDL có thể có cỡ tùy ý và có độ phức tạp thay đổi. Có những CSDL chỉ gồm vài trăm bản ghi (như CSDL phục vụ việc quản lý lương ở một cơ quan nhỏ), và có những CSDL có dung lượng rất lớn (như các CSDL phục vụ cho việc tính cước điện thoại, quản lý nhân sự trên phạm vi của một tổng công ty lớn). Các CSDL phải được tổ chức quản lý sao cho những người sử dụng có thể tìm kiếm dữ liệu, cập nhật dữ liệu và lấy dữ liệu ra khi cần thiết. Một CSDL có thể được tạo ra và duy trì một cách thủ công và cũng có thể được Tin học hóa.

CSDL là tài nguyên thông tin dùng chung cho nhiều người. Bất kỳ người sử dụng nào trên mạng máy tính, tại các thiết bị đầu cuối, về nguyên tắc có quyền truy nhập khai thác toàn bộ hay một phần dữ liệu theo chế độ trực tuyến hay tương tác mà không phụ thuộc vào vị trí địa lý của người sử dụng với các tài nguyên đó.

CSDL được các hệ ứng dụng khai thác bằng ngôn ngữ con dữ liệu hoặc bằng các chương trình ứng dụng để xử lý, tìm kiếm, tra cứu, sửa đổi, bổ sung hay loại bỏ dữ liệu. Tìm kiếm và tra cứu thông tin là một trong những chức năng qua trọng và phổ biến nhất của các dịch vụ CSDL. Hệ quản trị CSDL (Database Management System - DBMS) là phần mềm điều khiển các hoạt động truy nhập CSDL. Khi người sử dụng đưa ra yêu cầu truy nhập bằng một ngôn ngữ con dữ liệu nào đó, HQTCSDL tiếp nhận và thực hiện các thao tác trên CSDL đã lưu trữ.

3. Ưu điểm của cơ sở dữ liệu

Tổ chức lưu trữ dữ liệu theo lý thuyết CSDL có những ưu điểm sau:

- Giảm bớt dư thừa dữ liệu trong lưu trữ: Trong các ứng dụng lập trình truyền thống, phương pháp tổ chức lưu trữ dữ liệu vừa tốn kém, lãng phí bộ nhớ và các thiết bị lưu trữ, vừa dư thừa thông tin lưu trữ. Nhiều chương trình ứng dụng khác nhau cùng xử lý trên các dữ liệu như nhau, dẫn đến sự dư thừa đáng kể về dữ liệu.

- Tổ chức lưu trữ dữ liệu theo lý thuyết CSDL sẽ tránh được sự không nhất quán trong lưu trữ dữ liệu và bảo đảm được tính toàn vẹn của dữ liệu: Nếu một thuộc tính được mô tả trong nhiều tệp dữ liệu khác nhau và lặp lại nhiều lần thì khi thực hiện việc cập nhật, sửa đổi, bổ sung dữ liệu sẽ không sửa hết nội dung các mục đó. Nếu dữ liệu càng nhiều thì sự sai sót khi cập nhật, bổ sung càng lớn. Khả năng xuất hiện mâu thuẫn, không nhất quán thông tin càng nhiều, dẫn đến không nhất quán dữ liệu trong lưu trữ. Tất yếu kéo theo sự dị thường thông tin, thừa, thiếu và mâu thuẫn thông tin.

- Tổ chức lưu trữ dữ liệu theo lý thuyết CSDL có thể triển khai đồng thời nhiều ứng dụng trên cùng một CSDL. Điều này có nghĩa là các ứng dụng không chỉ chia sẻ chung tài nguyên dữ liệu mà còn trên cùng một CSDL có thể triển khai đồng thời nhiều ứng dụng khác nhau tại các thiết bị đầu cuối khác nhau.

- Tổ chức dữ liệu theo lý thuyết CSDL sẽ thống nhất các tiêu chuẩn, thủ tục và các biện pháp bảo vệ, an toàn dữ liệu. Các hệ CSDL sẽ được quản lý tập trung bởi một người hay một nhóm người quản trị CSDL, bằng các hệ quản trị CSDL. Người quản trị CSDL có thể áp dụng thống nhất các tiêu chuẩn, quy định, thủ tục chung như quy định thống nhất về mẫu biểu báo cáo, thời gian bổ sung, cập nhật dữ liệu. Điều này làm dễ dàng cho công việc bảo trì dữ liệu.

Người quản trị CSDL có thể bảo đảm việc truy nhập tới CSDL, có thể kiểm tra, kiểm soát các quyền truy nhập của người sử dụng. Ngăn chặn các truy nhập trái phép, sai quy định từ trong ra hoặc từ ngoài vào...

4. Những vấn đề cần giải quyết

Để đạt được hiệu quả khi sử dụng CSDL, ta cần giải quyết những vấn đề đặt ra sau đây:

- Tính chủ quyền của dữ liệu: Do tính chia sẻ của CSDL nên tính chủ quyền của dữ liệu có thể bị lu mờ và làm mờ nhạt tinh thần trách nhiệm, được thể hiện trên vấn đề an toàn dữ liệu, khả năng biểu diễn các mối liên hệ ngữ nghĩa của dữ liệu, và tính chính xác của dữ liệu. Điều này có nghĩa là người khai thác CSDL phải có nghĩa vụ cập nhật các thông tin mới nhất của CSDL.

- Tính bảo mật và quyền khai thác thông tin của người sử dụng: Do có nhiều người được phép khai thác CSDL một cách đồng thời nên cần phải có một cơ chế bảo mật và phân quyền hạn khai thác CSDL. Các hệ điều hành nhiều người sử dụng hay hệ điều hành mạng cục bộ (Novell Netware, Windows for Workgroup, WinNT...) đều có cung cấp cơ chế này.

- Tranh chấp dữ liệu: Nhiều người được phép truy nhập vào cùng một tài nguyên dữ liệu của CSDL với những mục đích khác nhau như xem, thêm, xóa hoặc sửa dữ liệu, do đó cần phải có một cơ chế ưu tiên khi truy nhập dữ liệu. Cơ chế ưu tiên có thể được thực hiện bằng việc cấp quyền ưu tiên cho từng người khai thác - người nào được cấp quyền hạn ưu tiên cao hơn thì được ưu tiên truy nhập dữ liệu trước; có thể dựa trên thời điểm truy nhập - ai có yêu cầu truy xuất trước thì có quyền truy nhập dữ liệu trước...

- Đảm bảo dữ liệu khi có sự cố: Việc quản lý dữ liệu tập trung có thể làm tăng khả năng mất mát hoặc sai lệch thông tin khi có sự cố như mất điện đột xuất, một phần đĩa lưu trữ CSDL bị hư hỏng v.v... 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 CSDL luôn luôn ổn định, một CSDL nhất thiết phải có một cơ chế khôi phục dữ liệu khi các sự cố bất ngờ xảy ra.

Giá phải trả cho ứng dụng CSDL

Như với bất kỳ quyết định kinh doanh nào, ứng dụng theo hướng CSDL cần trả giá và chịu một vài rủi ro như cần đến nhân sự mới và chuyên sâu, chi phí quản lý và thực hiện, giải quyết va chạm tâm lý...

- Nhu cầu về nhân sự mới, chuyên môn: Cần phải thuê hay tự đào tạo các chuyên gia công nghệ thông tin để phân tích, thiết kế, thực hiện CSDL trong tổ chức. Mặt khác, công nghệ ngày càng phát triển nhanh đã khiến các hệ thống CSDL thường xuyên cập nhật tri thức cho các cán bộ.

- Chi phí về quản lý, thực hiện: Hệ thống quản trị CSDL nhiều người dùng thường lớn và phức tạp, đòi hỏi nhiều người tham gia và bảo trì. Người ta cần tính đến chi phí thiết bị máy tính, chi phí phần mềm, chi phí truyền thông.

- Chi phí chuyển đổi hệ thống: Trước khi áp dụng CSDL, tổ chức đã xử lý dữ liệu theo phương pháp thủ công, hay sử dụng hệ thống xử lý tập tin. Để chuyển công nghệ xử lý dữ liệu sang cách tiếp cận mới, tổ chức cần có kinh phí chuyển các mô hình cũ sang mô hình mới, chuyển đổi dữ liệu cho phù hợp với môi trường mới.

- Va chạm về quan điểm sử dụng CSDL: Trước khi CSDL được áp dụng trong tổ chức, người ta cần thống nhất về mô tả dữ liệu, các khuôn dạng thể hiện dữ liệu... Sau khi hình thành CSDL, việc sử dụng dữ liệu của CSDL cần đạt hiệu quả. Cần khuyến khích các bộ phận phát huy tốt hệ thống CSDL vì rất có thể một số bộ phận nào đó muốn độc quyền thông tin hay e ngại dùng hệ thống tự động hoá, cho nên có xu hướng giảm hiệu quả của CSDL.

5. Các đối tượng sử dụng CSDL

- Người sử dụng CSDL không chuyên về lĩnh vực tin học và CSDL, do đó CSDL cần có các công cụ để cho những người sử dụng không chuyên có thể sử dụng để khai thác CSDL khi cần thiết.

- Chuyên viên Tin học biết khai thác CSDL: Những người này có thể xây dựng các ứng dụng khác nhau phục vụ cho các mục đích khác nhau trên CSDL.

- Người quản trị CSDL: Đây là những người hiểu biết về tin học, về các hệ quản trị CSDL và hệ thống máy tính. Họ là người tổ chức CSDL (khai báo cấu trúc CSDL, ghi nhận các yêu cầu bảo mật cho các dữ liệu cần bảo vệ...), do đó họ phải nắm rõ các vấn đề kỹ thuật về CSDL để có thể phục hồi dữ liệu khi có sự cố. Họ là những người cấp quyền hạn khai thác CSDL, do vậy họ có thể giải quyết được các vấn đề tranh chấp dữ liệu, nếu có.

6. Các ứng dụng của cơ sở dữ liệu

Hiện nay, hầu như CSDL gắn liền với mọi ứng dụng của Tin học; chẳng hạn như việc quản lý hệ thống thông tin trong các cơ quan nhà nước, việc lưu trữ và xử lý thông tin trong các doanh nghiệp, trong các lĩnh vực nghiên cứu khoa học, trong công tác giảng dạy, cũng như trong việc tổ chức thông tin đa phương tiện...

§ 2. KIẾN THỨC TỔNG QUÁT CỦA HỆ QUẢN TRỊ CƠ SỞ DỮ LIỆU**I. HỆ QUẢN TRỊ CƠ SỞ DỮ LIỆU – HỆ CƠ SỞ DỮ LIỆU****1. Khái niệm hệ quản trị cơ sở dữ liệu**

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...

Ta thấy, bộ nhớ 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 CSDL để phục vụ cho việc truy cập, tìm kiếm, cập nhật... nhanh chóng và an toàn hơn.

Để giải quyết tốt tất cả các vấn đề đặt ra cho một CSDL như đã nêu trên, cần phải có một hệ thống các phần mềm chuyên dụng được gọi là hệ quản trị CSDL (Database Management System - DBMS). Đó là các công cụ hỗ trợ tích cực cho các nhà phân tích và thiết kế CSDL và những người khai thác CSDL.

Các khái niệm:

Một hệ quản trị CSDL là một tập hợp các chương trình giúp cho người sử dụng quản lý cấu trúc và dữ liệu của CSDL, điều khiển truy xuất dữ liệu trong CSDL, duy trì và khai thác một CSDL.

Định nghĩa một CSDL: Tạo cấu trúc dữ liệu tương ứng với mô hình dữ liệu được chọn, bao gồm việc đặc tả các kiểu dữ liệu, các cấu trúc và các ràng buộc cho các dữ liệu sẽ được lưu trữ trong CSDL.

Xây dựng một CSDL: Là quá trình lưu trữ các dữ liệu trên các phương tiện lưu trữ được hệ quản trị CSDL kiểm soát.

Thao tác một CSDL: Bao gồm các chức năng như truy vấn CSDL để lấy ra các dữ liệu cụ thể, cập nhật CSDL để phản ánh các thay đổi trong thế giới nhỏ và tạo ra các báo cáo từ các dữ liệu.

Cho đến nay có khá nhiều hệ quản trị CSDL mạnh được đưa ra thị trường như: Visual FoxPro, Microsoft Access, SQL Server, DB2, Sybase, Paradox, Informix, Oracle... với các chất lượng khác nhau.

2. Các chức năng của một hệ quản trị cơ sở dữ liệu

Một hệ quản trị CSDL thực hiện các chức năng sau:

a. Cung cấp môi trường tạo lập CSDL

Một hệ quản trị CSDL phải cung cấp một môi trường cho người dùng dễ dàng khai báo kiểu dữ liệu, các cấu trúc dữ liệu thể hiện thông tin và các ràng buộc trên dữ liệu. Với các hệ quản trị CSDL hiện nay, người dùng có thể tạo lập CSDL thông qua các giao diện đồ họa. Để thực hiện được chức năng này, mỗi hệ quản trị CSDL cung cấp cho người dùng một ngôn ngữ định nghĩa dữ liệu. Ta có thể hiểu ngôn ngữ định nghĩa dữ liệu thực chất là hệ thống các ký hiệu để mô tả CSDL.

b. Cung cấp môi trường cập nhật và khai thác dữ liệu

Ngôn ngữ để người dùng diễn tả yêu cầu cập nhật hay tìm kiếm, kết xuất thông tin được gọi là ngôn ngữ thao tác dữ liệu.

Thao tác dữ liệu gồm:

- Cập nhật: Nhập thêm, sửa, xoá dữ liệu.
- Khai thác: Tìm kiếm, kết xuất dữ liệu.

Trong thực tế, ngôn ngữ định nghĩa dữ liệu và ngôn ngữ thao tác dữ liệu là 2 thành phần của một ngôn ngữ CSDL duy nhất. Ngôn ngữ CSDL được sử dụng phổ biến hiện nay là SQL (Structured Query Language).

c. Cung cấp công cụ kiểm soát, điều khiển truy cập vào CSDL

Để góp phần đảm bảo được các yêu cầu đặt ra cho một hệ CSDL, hệ quản trị CSDL phải có các bộ chương trình thực hiện những nhiệm vụ sau:

- Phát hiện và ngăn chặn sự truy cập không được phép: Chức năng này góp phần đáp ứng yêu cầu an toàn và bảo mật thông tin.
- Duy trì tính nhất quán của dữ liệu.
- Tổ chức và điều khiển các truy cập đồng thời để bảo vệ các ràng buộc toàn vẹn và tính nhất quán.
- Khôi phục CSDL khi có sự cố ở phần cứng hay phần mềm.
- Quản lý các mô tả dữ liệu.

Nói chung, mọi hệ quản trị CSDL đều có thể cung cấp các chương trình nêu trên, nhưng các hệ quản trị CSDL khác nhau có chất lượng và khả năng khác nhau khi đáp ứng các nhu cầu thực tế. Chẳng hạn, một hệ quản trị CSDL dùng trên một máy tính cá nhân chỉ cung cấp những phương tiện bảo vệ dữ liệu, duy trì tính nhất quán dữ liệu, khôi phục dữ liệu một cách hạn chế. Trong khi đó, những hệ quản trị CSDL lớn, phục vụ cùng lúc cho nhiều người dùng, ngoài các phương tiện kể trên có thể có thêm những chức năng khác như xử lý các truy cập đồng thời...

Các hệ quản trị CSDL luôn phát triển theo hướng đáp ứng các đòi hỏi ngày càng cao hơn của người dùng, bởi vậy các chức năng của hệ quản trị CSDL ngày càng được mở rộng hơn.

3. Các thành phần của một hệ quản trị cơ sở dữ liệu

Các hệ CSDL đều có các thành phần sau:

- Ngôn ngữ giao tiếp giữa người sử dụng và CSDL, bao gồm:

- + Ngôn ngữ mô tả dữ liệu (Data Definition Language - DDL): Cho phép khai báo cấu trúc CSDL, khai báo các mối liên hệ của dữ liệu (Data Relationship) và các quy tắc (Rules, Constraint) quản lý áp đặt lên các dữ liệu đó.

- + Ngôn ngữ thao tác dữ liệu (Data Manipulation Language - DML): Cho phép người sử dụng có thể thêm (Insert), xóa (Delete), sửa (Update) dữ liệu trong CSDL.

- + Ngôn ngữ truy vấn dữ liệu (Data Query Language - DQL): Cho phép những người khai thác (chuyên nghiệp hoặc không chuyên) sử dụng để tìm kiếm các thông tin cần thiết trong CSDL.

- + Ngôn ngữ quản lý dữ liệu (Data Control Language - DCL): Cho phép những người quản trị hệ thống khai báo bảo mật thông tin và cấp quyền hạn khai thác CSDL cho người sử dụng.

- Từ điển dữ liệu (Data Dictionary): Dùng để ghi nhận các thành phần cấu trúc của CSDL, mô tả các ánh xạ liên kết, mật mã, quyền hạn sử dụng, các chương trình ứng dụng...

- Có biện pháp phân quyền và bảo mật tốt khi có yêu cầu bảo mật.
- Cơ chế giải quyết vấn đề tranh chấp dữ liệu. Mỗi hệ quản trị CSDL cũng có thể cài đặt một cơ chế riêng để giải quyết các vấn đề này. Một số biện pháp sau đây được sử dụng:
 - + Cấp quyền ưu tiên cho từng người sử dụng (người quản trị CSDL thực hiện).
 - + Đánh dấu yêu cầu truy xuất dữ liệu, phân chia thời gian, người nào có yêu cầu trước thì có quyền truy xuất dữ liệu trước.
- Hệ quản trị CSDL cũng phải có cơ chế sao lưu (Backup) và phục hồi (Restore) dữ liệu khi có sự cố xảy ra.
- Hệ quản trị CSDL phải cung cấp một giao diện (Interface) tốt, dễ sử dụng, dễ hiểu cho những người sử dụng không chuyên.
- Hệ quản trị CSDL phải bảo đảm tính độc lập giữa dữ liệu và chương trình: Khi có sự thay đổi dữ liệu (như sửa đổi cấu trúc lưu trữ các bảng dữ liệu...) thì các chương trình ứng dụng đang chạy trên CSDL đó vẫn không cần phải được viết lại, hay cũng không làm ảnh hưởng đến những người sử dụng khác.

4. Hệ cơ sở dữ liệu

Người ta gọi CSDL và hệ quản trị CSDL bằng một thuật ngữ chung là hệ CSDL. Môi trường của một hệ CSDL được mô tả bằng hình vẽ dưới đây:



Hình: Hệ cơ sở dữ liệu

Hệ CSDL là một hệ thống gồm người dùng, các phần mềm ứng dụng, hệ quản trị CSDL và CSDL. Trong đó, người dùng chỉ can thiệp trực tiếp với các chương trình ứng dụng, các chương trình ứng dụng sẽ thông qua hệ quản trị CSDL để truy cập CSDL và thực hiện các công việc bên dưới, đáp ứng yêu cầu của người dùng.

II. MÔ HÌNH CƠ SỞ DỮ LIỆU

1. Khái niệm mô hình cơ sở dữ liệu

Các loại cấu trúc CSDL và mối liên hệ giữa chúng đóng vai trò rất lớn trong việc xác định tính hiệu quả của hệ quản trị CSDL. Vì vậy, thiết kế CSDL trở thành hoạt động chính trong môi trường CSDL.

Việc thiết kế CSDL được thực hiện đơn giản hơn nhiều khi ta sử dụng các mô hình. Các mô hình là sự trừu tượng đơn giản của các sự kiện trong thế giới thực. Các trừu tượng như vậy cho phép ta khảo sát các đặc điểm của các thực thể và các mối liên hệ được tạo ra giữa các thực thể đó. Việc thiết kế các mô hình tốt sẽ đưa ra các CSDL tốt và trên cơ sở đó sẽ có các ứng dụng tốt.

Một mô hình CSDL là một tập hợp các khái niệm, các luật hay quy tắc dùng để biểu diễn, mô tả một cách trừu tượng cấu trúc của CSDL. Cấu trúc của một CSDL là các kiểu dữ liệu,

các mối liên kết và các ràng buộc phải tuân theo trên các dữ liệu. Nhiều mô hình còn có thêm một tập hợp các phép toán cơ bản để đặc tả các thao tác trên CSDL.

Thông qua mô hình dữ liệu, người thiết kế sẽ mô tả toàn cảnh CSDL được thiết kế bao gồm:

- Các đối tượng, thực thể được quản lý
- Các mối quan hệ giữa các đối tượng
- Các ràng buộc dữ liệu thể hiện các quy tắc quản lý ảnh hưởng đến các đối tượng quản lý.

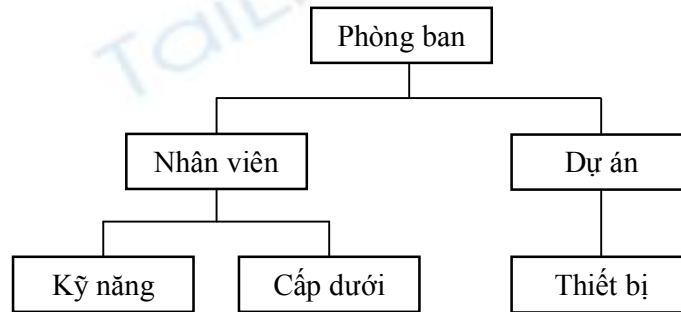
2. Các mô hình cơ sở dữ liệu

Có nhiều mô hình được đề nghị, ở đây chúng ta sẽ khảo sát các mô hình đặc trưng nhất:

a. Mô hình phân cấp

Mô hình phân cấp hay mô hình CSDL dạng cây được tổ chức theo cấu trúc từ trên xuống dưới giống như cây lộn ngược. Mỗi nút tương ứng với một kiểu dữ liệu, có thể có một hoặc nhiều trường, mô tả thực thể và một nhánh cây tạo nên một liên kết giữa kiểu dữ liệu này với kiểu dữ liệu khác. Mỗi nút đều có một nút cha và nhiều nút con, trừ nút gốc không có nút cha.

Ví dụ mô hình phân cấp quản lý nhân sự của một Công ty:



Hình: Mô hình phân cấp

Tuy nhiên, mô hình chỉ thể hiện được quan hệ 1-n (một-nhiều), tức là chỉ mô tả được trường hợp nút cha có nhiều nút con như một phòng thì có thể có nhiều nhân viên hay một phòng có thể có nhiều dự án, còn trường hợp ngược lại thì không. Chẳng hạn nếu một dự án mà thuộc về nhiều phòng thì dự án phải được lưu ở nhiều nơi khác nhau. Điều này gây dư thừa dữ liệu và lãng phí không gian lưu trữ.

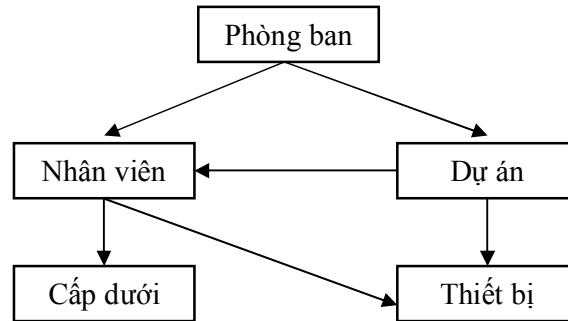
Điểm nổi bật trong các thủ tục truy xuất đến một đối tượng trong mô hình phân cấp là đường dẫn đi từ gốc đến phần tử cần xét trong cây phân cấp.

b. Mô hình mạng

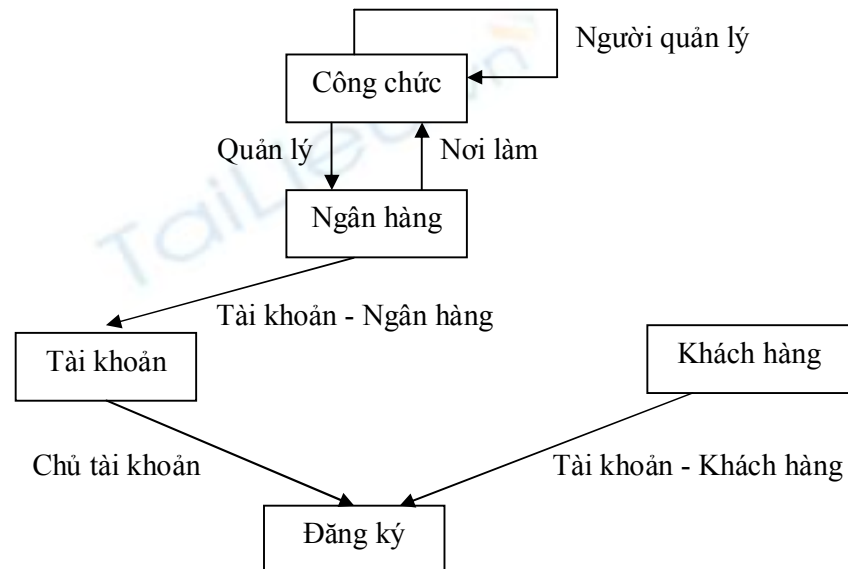
Mô hình mạng được biểu diễn như một đồ thị có hướng. Dữ liệu được biểu diễn bởi một tập các bản ghi, các mối quan hệ được biểu diễn bởi các mối nối (links) có thể được xem như những con trỏ. Xuất phát từ một đối tượng (biểu diễn bằng một bản ghi), có thể nối với một nút bất kỳ để biểu diễn một liên kết 1-n thông qua con trỏ liên kết, và có thể có nhiều mối quan hệ đến những đối tượng khác

Sự khác nhau chính giữa hệ thống mạng và hệ thống phân cấp là mô hình mạng không ràng buộc về số và hướng của các liên kết thiết lập giữa các nút.

Ví dụ 1: Mô hình quản lý nhân sự của một Công ty



Ví dụ 2:



c. Mô hình quan hệ

Cơ sở dữ liệu quan hệ được xây dựng theo lý thuyết do E.F.Codd giới thiệu năm 1970. Thuật ngữ “quan hệ” là do bảng dữ liệu 2 chiều được Codd gọi là bảng quan hệ. Mô hình quan hệ từ năm 1980 đã trở thành mô hình được dùng rộng rãi để phát triển hệ quản trị CSDL.

Đặc điểm của mô hình dữ liệu quan hệ:

- Dữ liệu được thể hiện trong các bảng. Các bảng gọi là các “quan hệ”. Mỗi bảng gồm các dòng và các cột, mỗi cột có một tên duy nhất. Các dòng gọi là các “bộ”, cột là “thuộc tính”.
- Mỗi dòng cho thông tin về một đối tượng cụ thể, bao gồm một bộ các giá trị tương ứng với các cột, mỗi giá trị thể hiện thông tin về một thuộc tính của đối tượng đó, tên thuộc tính này chính là tên cột mà giá trị đó được hiển thị.
- Mối liên kết giữa các đối tượng được thể hiện bằng mối liên kết giữa các bảng nhờ vào sự xuất hiện trùng lặp của một số thuộc tính ở hơn một bảng.

Ví dụ: Ta có các bảng dữ liệu trong CSDL sau:

KHACH_HANG	MaKH	HoTen	Tuoi	ĐịaChí
	Khách hàng 1	Nguyễn Văn Quang	16	68, Nguyễn Du
	Khách hàng 2	Hồ Hải Hà	18	23/10, Nguyễn Trãi
	Khách hàng 3	Dương Quốc Toàn	21	32/7, Nguyễn Trãi