



Các khái niệm cơ bản về hệ cơ sở dữ liệu

Bởi:

Khoa CNTT ĐHSP KT Hưng Yên

Các cơ sở dữ liệu và các hệ cơ sở dữ liệu đã trở thành một thành phần chủ yếu trong cuộc sống hàng ngày của xã hội hiện đại. Trong vòng một ngày con người có thể có nhiều hoạt động cần có sự giao tiếp với cơ sở dữ liệu như: đến ngân hàng để rút tiền và gửi tiền, đăng ký chỗ trên máy bay hoặc khách sạn, truy cập vào thư viện đã tin học hoá để tìm sách báo, đặt mua tạp chí ở một nhà xuất bản... Tại các ngân hàng, các cửa hàng, người ta cũng cập nhật tự động việc quản lý tiền bạc, hàng hoá.

Tất cả các giao tiếp như trên được gọi là các ứng dụng của cơ sở dữ liệu truyền thống. Trong các cơ sở dữ liệu truyền thống, hầu hết các thông tin được lưu giữ và truy cập là văn bản hoặc số. Những năm gần đây, những tiến bộ về kỹ thuật đã đưa đến những ứng dụng mới của cơ sở dữ liệu. Các cơ sở dữ liệu đa phương tiện bây giờ có thể lưu trữ hình ảnh, phim và tiếng nói. Các hệ thống thông tin địa lý có thể lưu trữ và phân tích các bản đồ, các dữ liệu về thời tiết và các ảnh vệ tinh. Kho dữ liệu và các hệ thống phân tích trực tuyến được sử dụng trong nhiều công ty để lấy ra và phân tích những thông tin có lợi từ các cơ sở dữ liệu rất lớn nhằm đưa ra các quyết định. Các kỹ thuật cơ sở dữ liệu động và thời gian thực được sử dụng trong việc kiểm tra các tiến trình công nghiệp và sản xuất. Các kỹ thuật tìm kiếm cơ sở dữ liệu đang được áp dụng cho World Wide Web để cung cấp việc tìm kiếm các thông tin cần thiết cho người sử dụng bằng cách duyệt qua Internet.

Để hiểu được các cơ sở kỹ thuật của cơ sở dữ liệu chúng ta phải bắt đầu từ các cơ sở kỹ thuật của cơ sở dữ liệu truyền thống. Mục đích của giáo trình này là nghiên cứu các cơ sở kỹ thuật đó. Trong bài này chúng ta sẽ định nghĩa cơ sở dữ liệu, hệ quản trị cơ sở dữ liệu, mô hình cơ sở dữ liệu và các thuật ngữ cơ bản khác.

Cơ sở dữ liệu

Định nghĩa cơ sở dữ liệu

Cơ sở dữ liệu và kỹ thuật cơ sở dữ liệu đã có ảnh hưởng rất lớn đến việc sử dụng máy tính. Có thể nói rằng cơ sở dữ liệu đóng vai trò quan trọng trong mọi lĩnh vực có sử dụng

máy tính như giáo dục, thương mại, kỹ nghệ, khoa học, thư viện,.... Thuật ngữ cơ sở dữ liệu trở thành một thuật ngữ phổ dụng.

Một *cơ sở dữ liệu* là một tập hợp các dữ liệu có liên quan với nhau, được lưu trữ trên máy tính, có nhiều người sử dụng và được tổ chức theo một mô hình. *Dữ liệu* là những sự kiện có thể ghi lại được và có ý nghĩa.

Ví dụ, để quản lý việc học tập trong một môi trường đại học, các dữ liệu là các thông tin về sinh viên, về các môn học, điểm thi.... Chúng ta tổ chức các dữ liệu đó 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 các dữ liệu có liên quan đến nhau và mang nhiều ý nghĩa, đó là một cơ sở dữ liệu.

Các tính chất của một cơ sở dữ liệu

Một cơ sở dữ liệu có các tính chất sau:

1. Một cơ sở dữ liệu 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 cơ sở dữ liệu. Những thông tin được đưa vào trong cơ sở dữ liệu tạo thành một không gian cơ sở dữ liệu hoặc là một “thế giới nhỏ” (miniworld) .
2. Một cơ sở dữ liệu là một tập hợp dữ liệu liên kết với nhau một cách logic và mang một ý nghĩa cố hữu nào đó. Một cơ sở dữ liệu không phải là một tập hợp tùy tiện.
3. Một cơ sở dữ liệu 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 môi quan tâm của người sử dụng. Nói cách khác, một cơ sở dữ liệu 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 cơ sở dữ liệu có thể có cỡ tùy ý và có độ phức tạp thay đổi. Có những cơ sở dữ liệu chỉ gồm vài trăm bản ghi (như cơ sở dữ liệu phục vụ việc quản lý lương ở một cơ quan nhỏ), và có những cơ sở dữ liệu có dung lượng rất lớn (như các cơ sở dữ liệu phục vụ cho việc tính cước điện thoại, quản lý nhân sự trên một phạm vi lớn). Các cơ sở dữ liệu 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 cơ sở dữ liệu có thể được tạo ra và duy trì một cách thủ công và cũng có thể được tin học hoá. Một cơ sở dữ liệu tin học hoá được tạo ra và duy trì bằng bằng một nhóm chương trình ứng dụng hoặc bằng một *hệ quản trị cơ sở dữ liệu*.

Hệ quản trị cơ sở dữ liệu

Định nghĩa hệ quản trị cơ sở dữ liệu

Một *hệ quản trị cơ sở dữ liệu* là một tập hợp chương trình giúp cho người sử dụng tạo ra, duy trì và khai thác một cơ sở dữ liệu. Nó là một hệ thống phần mềm phổ dụng, làm cho quá trình định nghĩa, xây dựng và thao tác cơ sở dữ liệu trở nên dễ dàng cho các ứng dụng khác nhau.

Định nghĩa một cơ sở dữ liệu 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 cơ sở dữ liệu.

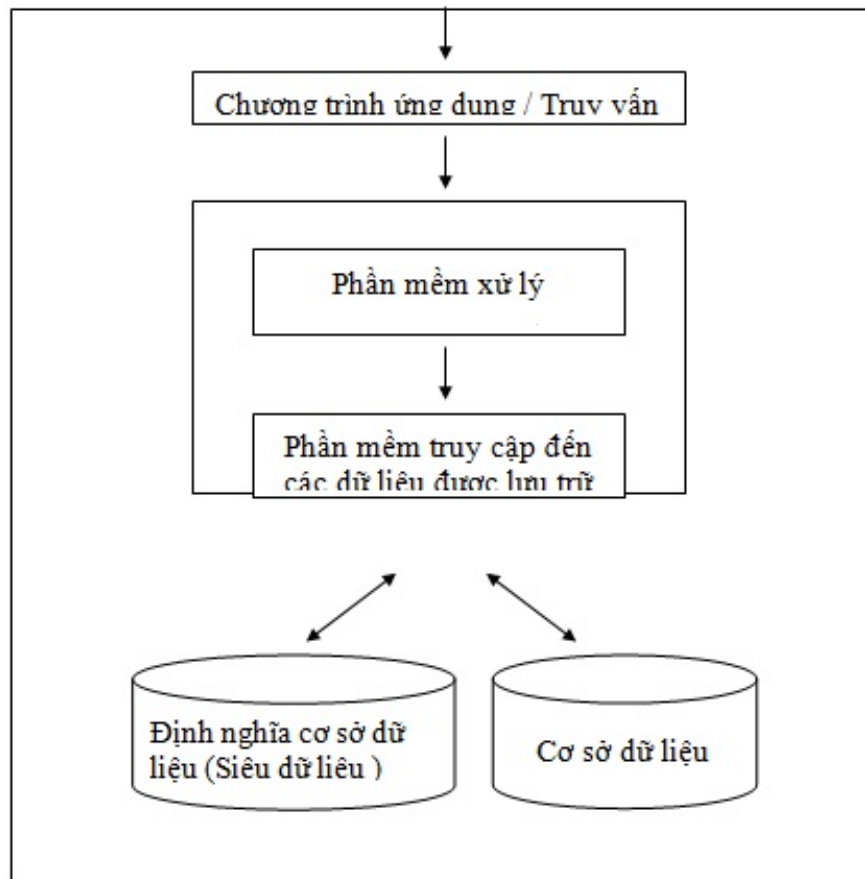
Xây dựng một cơ sở dữ liệu là quá trình lưu trữ các dữ liệu trên các phương tiện lưu trữ được hệ quản trị cơ sở dữ liệu kiểm soát.

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

Các hệ quản trị cơ sở dữ liệu dùng để thể hiện một cơ sở dữ liệu tin học hoá có thể là phổ dụng (là một phần mềm đóng gói) hoặc có thể là chuyên dụng (là một tập các phần mềm được tạo ra với một mục đích riêng).

Người ta gọi cơ sở dữ liệu và hệ quản trị cơ sở dữ liệu bằng một thuật ngữ chung là *hệ cơ sở dữ liệu*. Môi trường của một hệ cơ sở dữ liệu được mô tả bằng hình .

Người sử dụng / Người lập trình



Môi trường của một hệ cơ sở dữ liệu

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

Một hệ quản trị cơ sở dữ liệu hiện nay có các chức năng sau:

1. Lưu trữ các định nghĩa, các mối liên kết dữ liệu (gọi là siêu dữ liệu) vào một từ điển dữ liệu. Các chương trình truy cập đến cơ sở dữ liệu làm việc thông qua hệ quản trị cơ sở dữ liệu. Hệ quản trị cơ sở dữ liệu sử dụng dữ liệu trong từ điển dữ liệu để tìm kiếm các cấu trúc thành phần dữ liệu và các mối liên kết được yêu cầu. Mọi sự thay đổi trong các tệp cơ sở dữ liệu sẽ được tự động ghi lại vào từ điển dữ liệu. Như vậy, hệ quản trị cơ sở dữ liệu giải phóng người sử dụng khỏi việc lập trình cho các mối liên kết phức tạp trong mỗi chương trình, việc sửa đổi các chương trình truy cập đến tệp cơ sở dữ liệu đã bị sửa đổi. Nói cách khác, hệ quản trị cơ sở dữ liệu loại bỏ sự phụ thuộc giữa dữ liệu và cấu trúc ra khỏi hệ thống.
2. Tạo ra các cấu trúc phức tạp theo yêu cầu để lưu trữ dữ liệu. Nó giúp người sử dụng làm nhiệm vụ khó khăn là định nghĩa và lập trình cho các đặc trưng vật lý của dữ liệu.
3. Biến đổi các dữ liệu được nhập vào để phù hợp với các cấu trúc dữ liệu ở điểm 2. Như vậy, hệ quản trị cơ sở dữ liệu giúp người sử dụng phân biệt dạng logic và dạng vật lý của dữ liệu. Bằng việc duy trì sự độc lập dữ liệu, hệ quản trị cơ sở dữ liệu chuyển các

yêu cầu logic thành các lệnh định vị một cách vật lý và lấy ra các dữ liệu yêu cầu. Điều đó cũng có nghĩa là hệ quản trị cơ sở dữ liệu tạo khuôn dạng cho các dữ liệu được lấy ra để làm cho nó phù hợp với mong muốn logic của người sử dụng.

4. Tạo ra một hệ thống bảo mật và áp đặt tính bảo mật và riêng tư trong cơ sở dữ liệu.

5. Tạo ra các cấu trúc phức tạp cho phép nhiều người sử dụng truy cập đến dữ liệu

1. Cung cấp các thủ tục sao lưu và phục hồi dữ liệu để đảm bảo sự an toàn và toàn vẹn dữ liệu.

7. Xúc tiến và áp đặt các quy tắc an toàn để loại bỏ vấn đề toàn vẹn dữ liệu. Điều đó cho phép ta làm tối thiểu sự dư thừa dữ liệu và làm tối đa tính nhất quán dữ liệu.

8. Cung cấp việc truy cập dữ liệu thông qua một ngôn ngữ truy vấn. Một ngôn ngữ truy vấn là một ngôn ngữ phi thủ tục cho phép người sử dụng chỉ ra cái gì cần phải làm mà không cần phải chỉ ra nó được làm như thế nào. Các hệ quản trị cơ sở dữ liệu cũng cung cấp việc truy cập dữ liệu cho những người lập trình thông qua các ngôn ngữ thủ tục.

Các đặc trưng của giải pháp cơ sở dữ liệu

Trước khi khái niệm cơ sở dữ liệu ra đời, hệ thống tệp (file) là một phương pháp được áp dụng trong việc quản lý. Một tệp có thể được xem là một cặp hồ sơ lưu trữ các thông tin liên quan đến từng công việc riêng biệt. Ví dụ, trong một cơ quan, bộ phận tài vụ sẽ có một cặp hồ sơ liên quan đến lương của các nhân viên, bộ phận tổ chức có cặp hồ sơ liên quan đến vấn đề nhân sự... Việc xử lý để lấy ra các thông tin như là các thống kê về lương, về quá trình công tác... lúc đầu được thực hiện một cách thủ công. Dần dần, khối lượng thông tin ngày càng lớn, việc xử lý thông tin ngày càng phức tạp, người ta sử dụng máy tính vào việc quản lý. Các cặp hồ sơ được chuyển thành các tệp trên máy tính và việc xử lý thông tin được thực hiện bằng cách lập trình (trong một ngôn ngữ lập trình thế hệ 3).

Việc quản lý theo giải pháp hệ thống tệp có rất nhiều nhược điểm. Thứ nhất, đó là sự dư thừa thông tin: cùng một thông tin được lưu trữ nhiều lần (chẳng hạn, danh sách nhân viên có mặt trong tệp lương và cũng có mặt cả trong tệp nhân sự). Điều đó gây ra việc lãng phí bộ nhớ và dễ gây sai sót trong khi cập nhật dữ liệu, dễ sinh ra các dữ liệu không đúng đắn. Thứ hai, đó là việc phụ thuộc giữa chương trình ứng dụng và dữ liệu. Mỗi khi có sự thay đổi cấu trúc tệp và các dữ liệu trong tệp, chương trình ứng dụng khai thác thông tin trên tệp đó cũng thay đổi theo. Điều đó gây ra khó khăn lớn cho việc bảo trì.

Giải pháp cơ sở dữ liệu ra đời đã giải quyết được những nhược điểm đó. Cụ thể, giải pháp cơ sở dữ liệu có những đặc trưng sau:

1. Bản chất tự mô tả của hệ cơ sở dữ liệu.

Một đặc trưng cơ bản của giải pháp cơ sở dữ liệu là hệ thống cơ sở dữ liệu không chỉ gồm có bản thân cơ sở dữ liệu mà còn có cả định nghĩa hoặc mô tả đầy đủ về cấu trúc cơ sở dữ liệu và các ràng buộc. Định nghĩa này được lưu trữ trong từ điển hệ thống, nó chứa các thông tin như là cấu trúc của mỗi tệp, kiểu và dạng lưu trữ của từng mục dữ liệu. Các thông tin được lưu giữ trong từ điển gọi là siêu dữ liệu (meta-data) và chúng mô tả cấu trúc của dữ liệu nguyên thủy (hình 1-1). Phần mềm hệ quản trị cơ sở dữ liệu và những người sử dụng cơ sở dữ liệu sử dụng từ điển để lấy thông tin về cấu trúc của cơ sở dữ liệu.

2. Sự độc lập giữa chương trình và dữ liệu.

Trong hệ thống tệp, cấu trúc của các tệp cơ sở dữ liệu được nhúng vào trong các chương trình truy cập, vì vậy bất kỳ một thay đổi nào về cấu trúc của một tệp cũng đòi hỏi phải thay đổi tất cả các chương trình truy cập đến tệp đó. Ngược lại, các chương trình truy cập của hệ quản trị cơ sở dữ liệu không đòi hỏi việc thay đổi như thế. Cấu trúc của các tệp dữ liệu được lưu trữ trong từ điển tách rời với các chương trình truy cập. Tính chất này gọi là sự độc lập dữ liệu – chương trình.

3. Hỗ trợ các khung nhìn dữ liệu nhiều thành phần.

Một cơ sở dữ liệu có nhiều người sử dụng, mỗi một người có thể đòi hỏi một phối cảnh hoặc một khung nhìn (view) khác nhau. Một khung nhìn có thể là một tập con của cơ sở dữ liệu hoặc nó có thể chứa các dữ liệu ảo, đó là các dữ liệu được trích ra từ các tệp cơ sở dữ liệu khác nhau nhưng không được lưu trữ một cách rõ ràng. Một hệ quản trị cơ sở dữ liệu nhiều người sử dụng phải cung cấp nhiều công cụ để định nghĩa các khung nhìn nhiều thành phần.

4. Chia sẻ dữ liệu và nhiều người sử dụng.

Một hệ quản trị cơ sở dữ liệu nhiều người sử dụng phải cho phép nhiều người sử dụng truy cập đồng thời đến cơ sở dữ liệu. Hệ quản trị cơ sở dữ liệu phải có phần mềm kiểm tra cạnh tranh để đảm bảo rằng các người sử dụng cập nhật đến cùng một cơ sở dữ liệu phải được thực hiện theo cách được kiểm tra để cho kết quả của các cập nhật là đúng đắn.

Ví dụ về một cơ sở dữ liệu

Chúng ta hãy xem xét một cơ sở dữ liệu mà nhiều người đã quen biết: cơ sở dữ liệu TRƯỜNG. Cơ sở dữ liệu này lưu giữ các thông tin liên quan đến sinh viên, các môn học, điểm... trong một môi trường đại học. Cơ sở dữ liệu được tổ chức thành 5 bảng, mỗi bảng lưu trữ các bản ghi dữ liệu cùng một kiểu. Bảng SINHVIÊN lưu giữ dữ liệu về các sinh viên, bảng MÔN HỌC lưu giữ các dữ liệu về các môn học, bảng HỌC PHẦN

Các khái niệm cơ bản về hệ cơ sở dữ liệu

lưu giữ các dữ liệu về các học phần của các môn học, bảng ĐIỂM lưu giữ điểm của từng học phần của các sinh viên và bảng BIẾTTRƯỚC lưu giữ thông tin về các môn học cần biết trước của các môn học. Cấu trúc của cơ sở dữ liệu và một vài mẫu dữ liệu ví dụ được trình bày ở hình.

SINHVIÊN	Mã số SV	Họ tên SV	Lớp	Chuyên ngành
	17	Nguyễn Nam	K45T	Tinh học
	8	Lê Bắc	K45C	Công nghệ TT

MÔN HỌC	Mã số MH	Tên MH	Số đvht	Khoa
	101	Tinh học cơ sở	8	Công nghệ
	102	Cấu trúc DL và GT	5	Công nghệ
	103	Toán rời rạc	5	Công nghệ
	104	Cơ sở dữ liệu	3	Công nghệ

HỌC PHẦN	Mã số HP	Mã số MH	Học kỳ	Năm	Tên giảng viên
	1011	101	1	2001	Vân
	1012	101	2	2002	Vân
	1031	103	1	2001	Hoàng
	1032	103	2	2002	Hoàng
	1020	102	3	2002	Lân
	1040	104	4	2002	Huy

Cơ sở dữ liệu TRƯỜNG

ĐIỂM	Mã số SV	Mã số HP	Điểm
	17	1031	8
	17	1020	6
	8	1031	9
	8	1011	10
	8	1020	7
	8	1040	9

BIẾT TRƯỚC	Mã số MH	Mã số MH biết trước
	104	102
	104	103
	102	101

Cơ sở dữ liệu TRƯỜNG

Để định nghĩa cơ sở dữ liệu này, chúng ta phải chỉ ra cấu trúc của các bản ghi của mỗi tệp (bảng) bằng cách đặc tả các kiểu khác nhau của các phần tử dữ liệu sẽ được lưu trữ trong mỗi bản ghi. Theo hình 1-2, mỗi bản ghi SINHVIÊN bao gồm các dữ liệu để biểu diễn Mã số sinh viên, Họ tên sinh viên, Lớp, Chuyên ngành. Mỗi bản ghi MÔN HỌC bao gồm các dữ liệu để biểu diễn Tên môn học, Mã số môn học, Số đơn vị học trình, Khoa,... Chúng ta phải chỉ ra một kiểu dữ liệu cho mỗi phần tử dữ liệu bên trong các bản ghi. Ví dụ, ta có thể đặc tả Họ tên sinh viên là một dãy ký tự có độ dài nhỏ hơn hoặc bằng 30, Mã số sinh viên là một số nguyên,....

Để xây dựng cơ sở dữ liệu TRƯỜNG, chúng ta lưu giữ các dữ liệu để biểu diễn mỗi sinh viên, mỗi môn học,... vào các tệp thích hợp. Để ý rằng các bản ghi trong các tệp khác nhau có thể có mối quan hệ với nhau. Ví dụ, bản ghi đối với Nguyễn Nam trong tệp SINHVIÊN có liên quan đến hai bản ghi trong tệp ĐIỂM. Các bản ghi này chỉ ra điểm của Nguyễn Nam trong hai học phần. Tương tự như vậy, các bản ghi trong tệp có mối quan hệ với các bản ghi trong tệp MÔN HỌC... Thông thường một cơ sở dữ liệu chứa nhiều kiểu bản ghi và chứa nhiều mối quan hệ giữa các tệp.

Thao tác cơ sở dữ liệu bao gồm việc truy vấn và cập nhật cơ sở dữ liệu. Truy vấn cơ sở dữ liệu là đưa ra các yêu cầu đối với cơ sở dữ liệu để lấy ra các thông tin cần thiết. Ví dụ, chúng ta có thể có các truy vấn như: “Liệt kê các môn học và điểm thi của sinh viên Nguyễn Nam”, “Đưa ra danh sách các sinh viên thi trượt môn cơ sở dữ liệu”. Cập nhật cơ sở dữ liệu bao gồm việc thêm vào cơ sở dữ liệu bản ghi, xoá bỏ các bản ghi hoặc sửa đổi các giá trị trong các bản ghi. Các truy vấn và các cập nhật phải được đặc tả trong ngôn ngữ hệ cơ sở dữ liệu một cách chính xác trước khi chúng được xử lý.

Mô hình cơ sở dữ liệu

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

Việc thiết kế cơ sở dữ liệu 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 cơ sở dữ liệu tốt và trên cơ sở đó sẽ có các ứng dụng tốt. Ngược lại, mô hình không tốt sẽ đưa đến thiết kế cơ sở dữ liệu tồi và dẫn đến các ứng dụng không đúng.

Một *mô hình cơ sở dữ liệu* là một tập hợp các khái niệm dùng để biểu diễn các cấu trúc của cơ sở dữ liệu. Cấu trúc của một cơ sở dữ liệu 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 cơ sở dữ liệu.

Các loại mô hình cơ sở dữ liệu

Có rất nhiều mô hình dữ liệu đã được đề nghị. Chúng ta có thể phân loại các mô hình dữ liệu dựa trên các khái niệm mà chúng sử dụng để mô tả các cấu trúc cơ sở dữ liệu.

Các mô hình dữ liệu bậc cao hoặc mô hình dữ liệu mức quan niệm cung cấp các khái niệm gắn liền với cách cảm nhận dữ liệu của nhiều người sử dụng. Các mô hình này tập trung vào bản chất logic của biểu diễn dữ liệu, nó quan tâm đến cái (đối tượng) được biểu diễn trong cơ sở dữ liệu chứ không phải cách biểu diễn dữ liệu.

Các mô hình dữ liệu bậc thấp hoặc các mô hình dữ liệu vật lý cung cấp các khái niệm mô tả chi tiết về việc các dữ liệu được lưu trữ trong máy tính như thế nào. Các khái niệm do mô hình dữ liệu vật lý cung cấp nói chung có ý nghĩa đối với các chuyên gia máy tính chứ không có ý nghĩa mấy đối với người sử dụng thông thường. Ở giữa hai loại mô hình này là một lớp các mô hình dữ liệu thể hiện, chúng cung cấp những khái niệm mà người sử dụng có thể hiểu được và không xa với cách tổ chức dữ liệu bên trong máy tính. Người ta còn gọi loại mô hình dữ liệu này là loại mô hình dữ liệu mức logic. Các mô hình dữ liệu thể hiện che giấu một số chi tiết về việc lưu trữ dữ liệu nhưng có thể được cài đặt trực tiếp trên hệ thống máy tính.

Trong những bài sau chúng ta sẽ nghiên cứu một mô hình dữ liệu mức quan niệm, mô hình thực thể - liên kết, gọi tắt là mô hình ER (Entity – Relationship Model). Mô hình này sử dụng các khái niệm thực thể, thuộc tính, mối liên kết, để diễn đạt các đối tượng của thế giới thực. Một thực thể diễn đạt một đối tượng hoặc một khái niệm của thế giới thực. Ví dụ, một thực thể là một nhân viên hoặc một dự án được mô tả trong cơ sở dữ

liệu. Một thuộc tính diễn đạt một đặc trưng nào đó của thực thể. Chẳng hạn, họ tên, lương... là các thuộc tính của thực thể nhân viên. Một mối liên kết giữa hai hay nhiều thực thể diễn đạt một mối quan hệ qua lại giữa các thực thể. Ví dụ, giữa thực thể nhân viên và thực thể dự án có mối liên kết một nhân viên làm việc trên một dự án. Mô hình dữ liệu hướng đối tượng cũng là một mô hình dữ liệu bậc cao. Nó sử dụng các khái niệm như lớp, phương thức, thông điệp...

Các mô hình dữ liệu thể hiện là các mô hình được sử dụng thường xuyên nhất trong các hệ cơ sở dữ liệu thương mại. Ba mô hình nổi tiếng thuộc loại này là mô hình quan hệ, mô hình mạng và mô hình phân cấp. Các mô hình mạng và phân cấp ra đời trước và được sử dụng rộng rãi trong quá khứ (trước 1970). Vào đầu những năm 70, mô hình quan hệ ra đời. Do tính ưu việt của nó, mô hình quan hệ dần dần thay thế các mô hình mạng và phân cấp. Chúng ta sẽ nghiên cứu về mô hình quan hệ trong những bài sau. Các mô hình dữ liệu vật lý mô tả cách lưu trữ dữ liệu trong máy tính giới thiệu các thông tin như khuôn dạng bản ghi, sắp xếp bản ghi, đường truy cập.

Lược đồ và trạng thái cơ sở dữ liệu

Trong một mô hình dữ liệu cần phải phân biệt rõ giữa mô tả của cơ sở dữ liệu và bản thân cơ sở dữ liệu. Mô tả của một cơ sở dữ liệu được gọi là lược đồ cơ sở dữ liệu, nó được xác định rõ trong quá trình thiết kế cơ sở dữ liệu và không bị thay đổi thường xuyên. Đa số các mô hình dữ liệu có các quy ước hiển thị các lược đồ. Hiển thị của một lược đồ được gọi là biểu đồ của lược đồ đó. Một biểu đồ lược đồ chỉ thể hiện một vài khía cạnh của lược đồ như là các kiểu bản ghi, các mục dữ liệu và một số kiểu ràng buộc. Các khía cạnh khác không được thể hiện trong biểu đồ lược đồ.

Các dữ liệu trong một cơ sở dữ liệu có thể thay đổi một cách thường xuyên. Các dữ liệu trong một cơ sở dữ liệu tại một thời điểm cụ thể được gọi là một trạng thái cơ sở dữ liệu hoặc là ảnh (snapshot) của cơ sở dữ liệu. Nhiều trạng thái quan hệ có thể được xây dựng để làm tương ứng với một lược đồ cơ sở dữ liệu cụ thể. Mỗi khi chúng ta chèn vào hoặc loại bỏ một bản ghi, sửa đổi giá trị của một mục dữ liệu trong một bản ghi, chúng ta đã làm thay đổi trạng thái của cơ sở dữ liệu sang một trạng thái khác.

Việc phân biệt giữa lược đồ cơ sở dữ liệu và trạng thái cơ sở dữ liệu là rất quan trọng. Khi chúng ta định nghĩa một cơ sở dữ liệu mới, ta chỉ đặc tả lược đồ cơ sở dữ liệu cho hệ quản trị cơ sở dữ liệu. Tại thời điểm này, trạng thái của cơ sở dữ liệu là một trạng thái rỗng, không có dữ liệu. Chúng ta nhận được trạng thái ban đầu của cơ sở dữ liệu khi ta nhập dữ liệu lần đầu tiên. Từ đó trở đi, mỗi khi một phép toán cập nhật được thực hiện đối với cơ sở dữ liệu, chúng ta nhận được một trạng thái cơ sở dữ liệu khác. Tại mọi thời điểm, cơ sở dữ liệu có một trạng thái hiện tại. Hệ quản trị cơ sở dữ liệu có trách nhiệm đảm bảo rằng mỗi trạng thái cơ sở dữ liệu là một trạng thái vững chắc, nghĩa là một trạng thái thoả mãn cấu trúc và các ràng buộc được đặc tả trong lược đồ. Vì vậy, việc đặc tả một lược đồ đúng đắn cho hệ quản trị cơ sở dữ liệu là một việc làm cực kỳ

quan trọng và lược đồ phải được thiết kế một cách cẩn thận. Hệ quản trị cơ sở dữ liệu lưu trữ các mô tả của các cấu trúc lược đồ và các ràng buộc – còn gọi là siêu dữ liệu – vào trong từ điển (catalog) của hệ quản trị sao cho phần mềm hệ quản trị cơ sở dữ liệu có thể tham khảo đến lược đồ khi nó cần. Đôi khi người ta còn gọi lược đồ là mục tiêu (intension) và trạng thái cơ sở dữ liệu là mở rộng (extension) của lược đồ.

Con người trong hệ cơ sở dữ liệu

Với một cơ sở dữ liệu lớn, rất nhiều người tham gia vào việc thiết kế, sử dụng và duy trì cơ sở dữ liệu. Những người liên quan đến hệ cơ sở dữ liệu được chia thành hai nhóm chính. Nhóm thứ nhất gồm những người mà công việc của họ liên quan hàng ngày đến cơ sở dữ liệu, đó là những người quản trị cơ sở dữ liệu, thiết kế cơ sở dữ liệu, sử dụng cơ sở dữ liệu, phân tích hệ thống và lập trình ứng dụng. Nhóm thứ hai gồm những người làm việc để duy trì môi trường hệ cơ sở dữ liệu nhưng không quan tâm đến bản thân cơ sở dữ liệu, đó là những người thiết kế và cài đặt hệ quản trị cơ sở dữ liệu, phát triển công cụ, thao tác viên và bảo trì.

Người quản trị hệ cơ sở dữ liệu (Database Administrator – DBA)

Trong một tổ chức có nhiều người cùng sử dụng các tài nguyên, cần phải có một người giám sát và quản lý. Trong môi trường hệ cơ sở dữ liệu, các tài nguyên là cơ sở dữ liệu, hệ quản trị cơ sở dữ liệu và các phần mềm liên quan. Người quản trị hệ cơ sở dữ liệu là người chịu trách nhiệm quản lý các tài nguyên đó. Người này chịu trách nhiệm về việc cho phép truy cập cơ sở dữ liệu, tổ chức và hướng dẫn việc sử dụng cơ sở dữ liệu, cấp các phần mềm và phần cứng theo yêu cầu.

Người thiết kế cơ sở dữ liệu (Database Designer)

Người này chịu trách nhiệm xác định các dữ liệu sẽ được lưu giữ trong cơ sở, chọn các cấu trúc thích hợp để biểu diễn và lưu giữ các dữ liệu đó. Những nhiệm vụ này được thực hiện trước khi cơ sở dữ liệu được cài đặt và phổ biến. Người thiết kế có trách nhiệm giao thiệp với những người sử dụng tương lai để hiểu được các đòi hỏi của họ và đưa ra một thiết kế thoả mãn các yêu cầu đó. Anh ta cũng có nhiệm vụ giao thiệp với các nhóm người sử dụng và có khả năng hỗ trợ các yêu cầu của các nhóm.

Những người sử dụng (End User)

Những người sử dụng là những người mà công việc của họ đòi hỏi truy cập đến cơ sở dữ liệu để truy vấn, cập nhật và sinh ra các thông tin. Có thể chia những người sử dụng thành hai nhóm chính: những người sử dụng thụ động (tức là những người sử dụng không có nhiều kiến thức về hệ cơ sở dữ liệu) và những người sử dụng chủ động (là những người có hiểu biết tốt về hệ cơ sở dữ liệu).

Các khái niệm cơ bản về hệ cơ sở dữ liệu

Chức năng công việc của những người sử dụng thụ động (chiếm phần lớn những người sử dụng) gắn liền với việc truy vấn và cập nhật thường xuyên cơ sở dữ liệu bằng cách sử dụng các câu hỏi và các cập nhật chuẩn (gọi là các giao tác định sẵn) đã được lập trình và kiểm tra cẩn thận. Những người này chỉ cần học một ít về các phương tiện do hệ quản trị cơ sở dữ liệu cung cấp và hiểu các kiểu giao tác chuẩn đã được thiết kế và cài đặt là đủ.

Những người sử dụng chủ động có hiểu biết tốt về hệ cơ sở dữ liệu, họ có thể tự cài đặt các ứng dụng riêng của mình để làm thoả mãn các yêu cầu phức tạp của họ.

Người phân tích hệ thống và lập trình ứng dụng

Người phân tích hệ thống xác định các yêu cầu của những người sử dụng (chủ yếu là những người sử dụng thụ động) để đặc tả các chương trình phù hợp với yêu cầu của họ.

Người viết chương trình ứng dụng thể hiện các đặc tả của những người phân tích thành chương trình, sau đó kiểm thử, sửa lỗi làm tài liệu và bảo trì các giao tác định sẵn.

Người thiết kế và cài đặt hệ quản trị dữ liệu

Đó là những người thiết kế, cài đặt các mô đun, giao diện của hệ quản trị cơ sở dữ liệu thành các phần mềm đóng gói. Một hệ quản trị cơ sở dữ liệu là một hệ thống phần mềm phức tạp bao gồm nhiều thành phần (mô đun). Đó là các mô đun cài đặt từ điển dữ liệu, ngôn ngữ truy vấn, bộ xử lý giao diện, truy cập dữ liệu, kiểm tra cạnh tranh, phục hồi và an toàn. Hệ quản trị cơ sở dữ liệu phải giao tiếp với các hệ thống phần mềm khác như hệ điều hành và các chương trình dịch cho nhiều ngôn ngữ khác nhau.

Những người phát triển công cụ

Là những người thiết kế và cài đặt các công cụ (tool), đó là các phần mềm đóng gói làm dễ việc thiết kế và sử dụng cơ sở dữ liệu.

Các thao tác viên và những người bảo trì

Là những người chịu trách nhiệm về việc chạy và bảo trì phần cứng và phần mềm của hệ thống.

Ngôn ngữ cơ sở dữ liệu và giao diện

Các ngôn ngữ hệ quản trị cơ sở dữ liệu

Một khi việc thiết kế cơ sở dữ liệu đã hoàn thành, cần phải chọn một hệ quản trị cơ sở dữ liệu để cài đặt cơ sở dữ liệu. Trong các hệ quản trị cơ sở dữ liệu hiện nay thường có

Các khái niệm cơ bản về hệ cơ sở dữ liệu

các ngôn ngữ: ngôn ngữ định nghĩa dữ liệu (data definition language – DDL) và ngôn ngữ thao tác dữ liệu (data manipulation language – DML).

Ngôn ngữ định nghĩa dữ liệu được sử dụng để định nghĩa các lược đồ. Hệ quản trị cơ sở dữ liệu có một chương trình dịch ngôn ngữ DDL, nhiệm vụ của nó là xử lý các câu lệnh DDL để xác định mô tả của cấu trúc lược đồ và lưu trữ mô tả lược đồ vào từ điển của hệ quản trị cơ sở dữ liệu.

Ngôn ngữ thao tác cơ sở dữ liệu được sử dụng để thao tác cơ sở dữ liệu. Các thao tác chính gồm có lấy ra, chèn vào, loại bỏ và sửa đổi các dữ liệu. Có hai kiểu ngôn ngữ thao tác dữ liệu chính: ngôn ngữ thao tác dữ liệu mức cao hoặc ngôn ngữ phi thủ tục hoặc ngôn ngữ thao tác dữ liệu mức thấp.

Ngôn ngữ thao tác dữ liệu mức cao có thể được sử dụng để diễn đạt các phép toán cơ sở dữ liệu một cách ngắn gọn. Phần lớn các hệ quản trị cơ sở dữ liệu cho phép nhập các lệnh của ngôn ngữ thao tác dữ liệu mức cao theo cách lặp (nghĩa là sau khi nhập một lệnh, hệ thống sẽ thực hiện lệnh đó rồi mới nhập lệnh tiếp theo) hoặc được nhúng vào một ngôn ngữ lập trình vạn năng. Trong trường hợp nhúng vào ngôn ngữ khác, các lệnh của ngôn ngữ thao tác dữ liệu phải được xác định bên trong chương trình sao cho một chương trình tiền dịch có thể nhận ra chúng và được hệ quản trị cơ sở dữ liệu xử lý.

Ngôn ngữ thao tác cơ sở dữ liệu mức thấp hoặc ngôn ngữ thủ tục phải được nhúng vào trong một ngôn ngữ lập trình vạn năng. Ngôn ngữ thao tác cơ sở dữ liệu kiểu này thường rút ra các bản ghi hoặc các đối tượng riêng rẽ và xử lý chúng một cách riêng rẽ. Vì vậy, chúng cần phải sử dụng các cấu trúc ngôn ngữ lập trình như vòng lặp, điều kiện,... để rút ra từng bản ghi một từ một tập các bản ghi. Ngôn ngữ thao tác dữ liệu mức thấp được gọi là ngôn ngữ “một lần một bản ghi”. Các ngôn ngữ thao tác dữ liệu mức cao có thể dùng một lệnh để rút ra một lúc nhiều bản ghi nên chúng được gọi là ngôn ngữ “một lần một tập hợp”.

Các loại giao diện hệ quản trị cơ sở dữ liệu

Các hệ quản trị cơ sở dữ liệu cung cấp rất nhiều loại giao diện người dùng thân thiện. Các loại giao diện chính gồm có:

Giao diện dựa trên bảng chọn: Các giao diện này cung cấp cho người sử dụng danh sách các lựa chọn, gọi là bảng chọn (menu) và hướng dẫn người sử dụng diễn đạt một yêu cầu từ đầu đến cuối. Các bảng chọn làm cho người sử dụng không cần nhớ các lệnh và cú pháp của ngôn ngữ truy vấn. Các bảng chọn thả xuống đã trở thành kỹ thuật phổ biến trong các giao diện dựa trên cửa sổ. Chúng thường được sử dụng trong các giao diện quét, cho phép người sử dụng nhìn thấy nội dung của một cơ sở dữ liệu theo cách không có cấu trúc.

Giao diện dựa trên mẫu biểu: Các giao diện này hiển thị một mẫu biểu cho người sử dụng. Những người sử dụng có thể điền vào tất cả các ô của mẫu biểu để nhập các dữ liệu mới hoặc họ chỉ điền vào một số ô còn hệ quản trị cơ sở dữ liệu sẽ đưa ra các dữ liệu phù hợp cho các ô khác. Các mẫu biểu thường được thiết kế và được lập trình cho các người dùng đơn giản. Một số hệ thống có các tiện ích giúp người sử dụng từng bước xây dựng một mẫu biểu trên màn hình.

Giao diện đồ họa: Một giao diện đồ họa (GUI) thường hiển thị một lược đồ cho người sử dụng dưới dạng biểu đồ. Người dùng có thể thực hiện một truy vấn bằng cách thao tác trên biểu đồ. Trong nhiều trường hợp, GUI sử dụng cả các bảng chọn và các mẫu biểu. Đa số các GUI sử dụng các công cụ trợ như chuột, phím để kích các phần của sơ đồ.

Giao diện cho người quản trị hệ thống: Đa số các hệ quản trị cơ sở dữ liệu có các lệnh ưu tiên, chỉ có những người quản trị hệ thống mới sử dụng các lệnh đó. Đó là các lệnh tạo ra các tài khoản (account), đặt các tham số cho hệ thống, cấp các tài khoản, thay đổi lược đồ hoặc tổ chức lại các cấu trúc lưu trữ của cơ sở dữ liệu.

Câu hỏi ôn tập

1. Định nghĩa các thuật ngữ: cơ sở dữ liệu, hệ quản trị cơ sở dữ liệu, hệ cơ sở dữ liệu, từ điển cơ sở dữ liệu, mô hình cơ sở dữ liệu.
2. Nêu các tính chất của một cơ sở dữ liệu
3. Nêu các chức năng của một hệ quản trị cơ sở dữ liệu
4. Giải thích các đặc trưng của giải pháp cơ sở dữ liệu
5. Định nghĩa mô hình cơ sở dữ liệu và phân loại
6. Liệt kê các người có liên quan đến hệ cơ sở dữ liệu.