

SAU BÀI HỌC NÀY EM SẼ:

- Hiểu được khái niệm và các thuộc tính cơ bản của cơ sở dữ liệu.



Theo em, việc lưu trữ dữ liệu phục vụ các bài toán quản lí có phải chỉ là việc chuyển các ghi chép trên giấy thành văn bản trên máy tính không?

1. YÊU CẦU TỔ CHỨC LƯU TRỮ DỮ LIỆU MỘT CÁCH KHOA HỌC

Hoạt động 1 Có cần lưu trữ bảng điểm lớp học không?

Giáo viên dạy mỗi môn học bắt buộc phải có một sổ điểm - bảng điểm môn học. Một bản sao của bảng điểm môn học được gửi cho giáo viên chủ nhiệm lớp. Hãy cùng thảo luận xem có cần lưu trữ bảng điểm lớp học không.



a) Hạn chế dư thừa trong lưu trữ dữ liệu

Ví dụ về bảng điểm môn học và bảng điểm của lớp ở Mục 1, Bài 10 được xem xét từ thực tế thầy cô giáo ghi chép dữ liệu trên giấy.

Khi ghi chép, lưu trữ trên giấy, giáo viên chủ nhiệm lớp thường lưu trữ cả các bảng điểm môn học và bảng điểm lớp học. Tuy nhiên, có thể thấy nhiều dữ liệu lặp lại trong các bảng điểm này và có thể dẫn đến sự không nhất quán dữ liệu, ví dụ:

Ở bảng điểm môn Toán, ĐĐG cuối kì của Nguyễn Kì Duyên là 9.

STT	Họ và tên	ĐĐG thường xuyên			ĐĐG giữa kì	ĐĐG cuối kì
...	...					
3	Nguyễn Kì Duyên	7		8	10	9
...	...					

Hình 11.1. Trích xuất bảng điểm môn Toán

Nhưng ở bảng điểm lớp học, ĐĐG cuối kì môn Toán của Nguyễn Kì Duyên có thể bị ghi nhầm là 8 (Hình 11.2).

STT	Họ và tên	Toán	Tin học	Vật lí	Hoá học	Sinh học	Ngữ văn	Lịch sử	Địa lí
...	...								
3	Nguyễn Kì Duyên	8	10	10	10	8	6	7	7
...	...								

Hình 11.2. Trích xuất bảng điểm lớp học có điểm trung bình

Thói quen cá nhân của người lưu trữ cũng có thể dẫn đến sự không nhất quán của dữ liệu khi lưu trữ thủ công. Ví dụ, điểm có thể ghi bằng chữ hoặc các kí hiệu riêng thay vì các cột điểm bằng số để nhận biết và tính toán.

Khi dữ liệu được lưu trữ trên máy tính, việc lập bảng điểm lớp học từ dữ liệu cơ sở (các bảng điểm môn học) có thể được thực hiện một cách dễ dàng. Vì thế, không cần lưu trữ bảng điểm lớp học do bảng này chỉ là một khung nhìn tổng hợp từ dữ liệu cơ sở bằng cách ghép các bảng điểm môn học.

Ví dụ trên cho thấy việc lưu trữ dữ liệu trên máy tính không đơn thuần là việc chuyển các văn bản trên giấy thành các tệp văn bản trên máy tính. Cần phải tổ chức việc lưu trữ sao cho có thể hạn chế trùng lặp làm dư thừa dữ liệu, khắc phục những lỗi không nhất quán về dữ liệu. Đây cũng là yêu cầu chung đối với việc lưu trữ dữ liệu của mọi bài toán quản lí.

b) Sự phụ thuộc phần mềm và dữ liệu

Việc lưu trữ dữ liệu không tách rời với việc khai thác thông tin nhờ các phần mềm ứng dụng vì khai thác thông tin chính là mục đích của việc lưu trữ dữ liệu. Để thấy rõ hơn sự cần thiết phải tổ chức lưu trữ dữ liệu một cách khoa học, ta sẽ xem xét một cách tiếp cận gắn kết chương trình - dữ liệu để khai thác thông tin.

Việc lưu trữ dữ liệu điểm các môn học trên máy tính đòi hỏi cần có những phần mềm hỗ trợ cập nhật dữ liệu điểm và khai thác thông tin từ những dữ liệu ấy. Có thể chỉ ra một số thành phần cần có của phần mềm (thường được gọi là các mô đun phần mềm) đó là: cập nhật điểm môn học (Điểm môn học), quản lí danh sách lớp học (Danh sách lớp) và lập bảng điểm lớp học từ dữ liệu điểm môn học (Lập bảng điểm). Nếu viết các mô đun phần mềm này bằng một ngôn ngữ lập trình, ví dụ Python, giải pháp lưu trữ đơn giản nhất thường được nghĩ đến sử dụng trực tiếp hệ thống tệp, ví dụ tệp văn bản (text).

Danh sách lớp học sẽ gồm các dòng, mỗi dòng có nội dung lần lượt là: tên lớp, số thứ tự, họ và tên của một học sinh, ngăn cách nhau bởi dấu phẩy, ví dụ:

11A, 1, Dương Hồng Anh

11A, 2, Lương Việt Anh

11A, 3, Nguyễn Kì Duyên

...

Bảng điểm môn học sẽ gồm các dòng, mỗi dòng có nội dung lần lượt là: tên lớp, số thứ tự, họ và tên của một học sinh, 4 ĐĐG thường xuyên, 1 ĐĐG giữa kì và 1 ĐĐG cuối kì, ngăn cách nhau bởi dấu phẩy (.). Ví dụ:

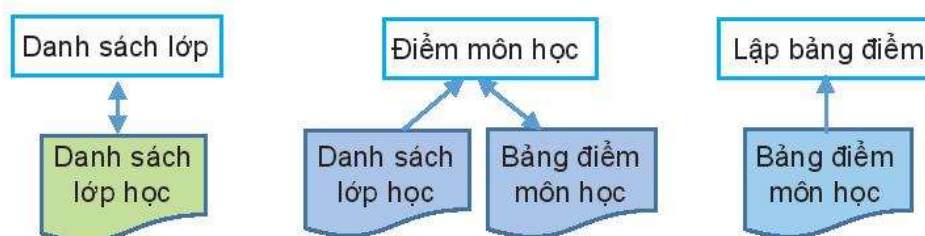
11A, 1, Dương Hồng Anh, 8, 7, , 7, 9, 8

11A, 2, Lương Việt Anh, , 5, 6, 5, 5, 5

11A, 3, Nguyễn Kì Duyên, 7, , 8, 0, 10, 9

...

Hình 11.3 là sơ đồ cho thấy quan hệ giữa các mô đun phần mềm và các tệp dữ liệu.



Hình 11.3. Các mô đun phần mềm và quan hệ với các tệp dữ liệu

Khi viết mỗi mô đun phần mềm, người lập chương trình phải biết cấu trúc của các tệp dữ liệu để lần lượt đọc từng dòng, rồi tách các thành phần dữ liệu tương ứng theo dấu phẩy,...

Mô đun *Danh sách lớp* chịu trách nhiệm đọc và ghi dữ liệu cập nhật vào Danh sách lớp học. Mô đun *Điểm môn học* phải đọc dữ liệu từ Danh sách lớp học, tạo lập tệp Bảng điểm môn học nếu chưa có, cập nhật tệp Bảng điểm môn học nếu xuất hiện những dòng mới trong Danh sách lớp học; đọc và ghi dữ liệu cập nhật vào Bảng điểm môn học. Mô đun *Lập bảng điểm* phải đọc dữ liệu từ Bảng điểm môn học, xử lý dữ liệu để tạo và kết xuất ra bảng điểm lớp học.

Khi thay đổi cấu trúc các dòng ghi dữ liệu thì các mô đun phần mềm liên quan bắt buộc phải chỉnh sửa theo. Ví dụ quy định mới về điểm có phần lẻ thập phân, dùng dấu phẩy làm dấu ngăn cách phần nguyên và phần thập phân của điểm dẫn tới việc phải dùng dấu khác dấu phẩy để ngăn cách các thành phần của dòng dữ liệu. Khi đó, người viết chương trình buộc phải sửa mô đun *Điểm môn học* và *Lập bảng điểm* ở những câu lệnh phân tích, tách dòng dữ liệu.

Như vậy, tình trạng phụ thuộc giữa chương trình và dữ liệu dẫn tới việc nếu thay đổi cách lưu trữ dữ liệu phải sửa đổi phần mềm làm cho việc thiết kế, bảo trì, phát triển phần mềm mất nhiều thời gian và công sức. Bên cạnh đó, bài toán quản lý, nhất là khi liên quan tới những lĩnh vực lớn hoặc có nhiều biến động, lại thường xuất hiện các yêu cầu khai thác thông tin đa dạng khác nhau từ dữ liệu lưu trữ. Do vậy việc tổ chức dữ liệu độc lập để phần mềm không cần “nhìn thấy” chi tiết về cách lưu trữ mà vẫn sử dụng được dữ liệu là một trong các ý tưởng quan trọng để hình thành nên khoa học về cơ sở dữ liệu (CSDL) được nêu trong mục tiếp theo.

Dữ liệu cần được tổ chức lưu trữ một cách độc lập với việc xây dựng phát triển phần mềm, đảm bảo dễ dàng chia sẻ, dễ dàng bảo trì phát triển, đồng thời đảm bảo hạn chế tối đa việc dữ liệu lặp lại, gây dư thừa dữ liệu và hỗ trợ đảm bảo tính nhất quán dữ liệu.



1. Hãy giải thích yêu cầu về tính nhất quán dữ liệu trong lưu trữ dữ liệu.
2. Tại sao cần tổ chức lưu trữ dữ liệu độc lập với phần mềm?

2. CƠ SỞ DỮ LIỆU VÀ MỘT SỐ THUỘC TÍNH CƠ BẢN

Hoạt động 2 Sự cần thiết phải lưu trữ dữ liệu theo một cấu trúc xác định

Hãy so sánh cách thức ghi chép và lưu trữ kết quả điểm môn học nêu trong Mục 1 với cách ghi chép và lưu trữ dưới dạng bảng. Theo em, cách nào là phù hợp hơn? Thông qua ví dụ bảng điểm môn học hãy chỉ ra một vài lí do cần lưu trữ dữ liệu theo một cấu trúc xác định.



a) Khái niệm CSDL

Không thể tiếp tục với cách tiếp cận có sự phụ thuộc lẫn nhau giữa dữ liệu và phần mềm như mô tả ở trên. Cần phải chọn cách tiếp cận khác, theo đó việc lưu trữ dữ liệu phải là một vấn đề độc lập cần được xem xét một cách khái quát. Nhu cầu tổ chức dữ liệu sao cho việc khai thác dữ liệu thuận tiện, giảm được công sức và thời gian làm phần mềm là lí do chính cần tổ chức lưu trữ dữ liệu dưới dạng các CSDL.

CSDL là một tập hợp dữ các liệu có liên quan với nhau, được lưu trữ một cách có tổ chức trên hệ thống máy tính.

Bảng điểm các môn học khi được lưu trữ trong máy tính có thể được xem là ví dụ về CSDL. Thông tin về tài khoản ngân hàng bao gồm tên chủ tài khoản, số căn cước công dân, số dư có trong tài khoản,... có thể được tổ chức thành CSDL để quản lí và phục vụ khách hàng một cách nhanh chóng và tiện lợi,...

b) Một số thuộc tính cơ bản của CSDL

Hoạt động 2 cho ta hình dung về *tính cấu trúc* của CSDL. Để hình dung nhất về tính cấu trúc của CSDL khi dữ liệu được lưu trữ dưới dạng bảng gồm các hàng và các cột.

- **Tính không dư thừa:** Trong phần trên ta đã thấy cần hạn chế việc lưu trữ các dữ liệu trùng lặp cũng như những dữ liệu dễ dàng có được thông qua việc khai thác thông tin từ dữ liệu đã có. Thuộc tính này được gọi là *tính không dư thừa* của CSDL. Dưới đây ta sẽ xem xét thêm một số thuộc tính quan trọng khác của CSDL.

- **Tính độc lập dữ liệu:**

Ở Mục 1, các em có thể thấy sự cần thiết phải đảm bảo tính độc lập dữ liệu với phần mềm. Trong phạm vi kiến thức phổ thông, tính độc lập dữ liệu có thể được hiểu là khả năng các mô đun phần mềm ứng dụng không cần phải cập nhật khi thay đổi cách thức tổ chức hoặc lưu trữ dữ liệu. Trong thực tế, các CSDL được thiết kế nhằm phục vụ nhiều người dùng với nhiều mục đích quản lí khác nhau, không phụ thuộc vào bài toán quản lí cụ thể hay phương tiện lưu trữ, xử lí dữ liệu, không đòi hỏi họ phải biết được các chi tiết kĩ thuật liên quan đến CSDL. Do vậy, việc đảm bảo độc lập dữ liệu là một trong các thuộc tính quan trọng nhất của CSDL.

- **Tính toàn vẹn:** Các giá trị dữ liệu lưu trữ phải thoả mãn những ràng buộc cụ thể tùy thuộc vào thực tế mà nó phản ánh, ví dụ điểm đánh giá học tập phải là số nguyên (hay số thập phân) không âm và nhỏ hơn hoặc bằng 10, tùy theo quy định về đánh giá điểm học tập của tổ chức.

- **Tính nhất quán:** Trong Mục 1 ta đã biết sơ suất khi cập nhật dữ liệu trong trường hợp CSDL tổ chức không tốt có thể làm ảnh hưởng tới *tính nhất quán* của dữ liệu. Khái niệm về tính nhất quán còn đòi hỏi dữ liệu trong các CSDL được đảm bảo đúng đắn sau các thao tác cập nhật dữ liệu, kể cả khi xảy ra sự cố ngay trong quá trình cập nhật. Ví dụ phải có cơ chế để đảm bảo không xảy ra hiện tượng số tiền đã bị trừ bớt trong tài khoản chuyển đi nhưng lại chưa xuất hiện trong tài khoản nhận về; hoặc hai đại lí vé máy bay bán cùng một vé ngồi cho hai khách hàng khác nhau,... Để xảy ra những trường hợp như vậy đều là sự vi phạm tới tính nhất quán của dữ liệu.

- **Tính bảo mật và an toàn:** Điều này có nghĩa là dữ liệu phải được bảo vệ an toàn, ngăn chặn được những truy xuất trái phép, chống được việc sao chép dữ liệu không hợp lệ. CSDL cần được tổ chức sao cho không phải ai cũng có quyền truy cập hay cập nhật dữ liệu. Ví dụ với CSDL Bảng điểm môn học không phải ai cũng có thể vào sửa chữa điểm; với CSDL tài khoản ngân hàng người không có thẩm quyền không được truy xuất để lấy thông tin cá nhân hay sửa đổi số dư tài khoản,... Bên cạnh đó, dữ liệu phải được bảo vệ an toàn, không dễ bị sai lạc, mất mát, có thể khôi phục dù có xảy ra các sự cố liên quan tới phần cứng hay phần mềm của máy tính.

- CSDL là một tập hợp dữ liệu có liên quan với nhau, được lưu trữ một cách có tổ chức trong hệ thống máy tính.
- Một số thuộc tính cơ bản của CSDL bao gồm: tính cấu trúc; tính không dư thừa; tính độc lập; tính toàn vẹn; tính nhất quán; tính bảo mật và an toàn.



Hãy nêu ví dụ minh họa cho một vài thuộc tính cơ bản của CSDL.



LUYỆN TẬP

1. Khi lưu trữ trên máy tính, theo em, có cần lưu trữ cột điểm trung bình trong bảng điểm môn học không?
2. Hãy lấy một ví dụ minh họa cho sự cần thiết của việc lưu trữ dữ liệu độc lập với phần mềm khai thác dữ liệu.



VẬN DỤNG

Thư viện là nơi em có thể đến để đọc hay mượn sách. Hãy đề xuất các dữ liệu cần quản lí của một thư viện.