

SAU BÀI HỌC NÀY EM SẼ:

- Hiểu được khái niệm hệ quản trị CSDL.
- Hiểu được khái niệm hệ CSDL.
- Phân biệt được CSDL tập trung và CSDL phân tán.



Một CSDL lưu dữ liệu trên hệ thống máy tính dưới dạng các tệp có cấu trúc được thiết kế để nhiều người dùng có thể cùng khai thác dữ liệu trong CSDL đó. Tuy nhiên không phải tất cả người dùng đều biết về cấu trúc các tệp lưu dữ liệu và tự viết chương trình khai thác dữ liệu. Theo em, có thể giải quyết vấn đề này như thế nào?

1. KHÁI NIỆM HỆ QUẢN TRỊ CƠ SỞ DỮ LIỆU

Hoạt động 1 Thảo luận về một phần mềm hỗ trợ thao tác dữ liệu

Để tạo ra, lưu trữ và sửa đổi một văn bản trên máy tính chúng ta cần một phần mềm soạn thảo văn bản. Để tạo ra và cập nhật một bảng tính điện tử chúng ta cần một phần mềm bảng tính.

Theo em, một phần mềm hỗ trợ làm việc với các CSDL cần thực hiện được những yêu cầu nào dưới đây?

- Cung cấp công cụ tạo lập CSDL.
- Cập nhật dữ liệu và tự động kiểm tra tính đúng đắn của dữ liệu.
- Hỗ trợ truy xuất dữ liệu.
- Cung cấp giao diện để ai cũng có thể xem nội dung của các bảng dữ liệu một cách dễ dàng.



Để hỗ trợ làm việc với các CSDL người ta đã xây dựng những bộ phần mềm chuyên dụng gọi là hệ quản trị cơ sở dữ liệu (Database management systems, sau đây sẽ gọi tắt là hệ QTCSDL) với các nhóm chức năng sau:

a) Nhóm chức năng định nghĩa dữ liệu

Với các chức năng này người dùng có thể:

- Khai báo CSDL với tên gọi xác định. Một hệ QTCSDL có thể quản trị nhiều CSDL.
- Tạo lập, sửa đổi kiến trúc bên trong mỗi CSDL.
- Nhiều hệ QTCSDL cho phép cài đặt các ràng buộc toàn vẹn dữ liệu để có thể kiểm soát tính đúng đắn của dữ liệu.

b) Nhóm chức năng cập nhật và truy xuất dữ liệu hay là nhóm *chức năng thao tác dữ liệu* với các chức năng sau:

- *Chức năng cập nhật dữ liệu.* CSDL sau khi được khởi tạo chưa có dữ liệu, cần phải nhập dữ liệu vào. Theo thời gian, do biến động của thế giới thực hoặc do sai sót khi nhập dữ liệu, dữ liệu trong CSDL không còn đúng nữa. Hệ QTCSDL cần cung cấp các chức năng thêm, xóa, sửa dữ liệu.

- *Chức năng truy xuất dữ liệu* theo những tiêu chí khác nhau.

c) Nhóm chức năng bảo mật, an toàn CSDL

- Không phải mọi hệ QTCSDL đều cung cấp công cụ để mọi người có thể dễ dàng đọc nội dung các bảng dữ liệu. Dữ liệu cần được bảo mật, chỉ cung cấp cho người có thẩm quyền. Do vậy, nhiều hệ QTCSDL cung cấp phương tiện kiểm soát quyền truy cập dữ liệu.

- Khi nhiều người được truy cập đồng thời vào CSDL sẽ nảy sinh ra vấn đề tranh chấp dữ liệu, chẳng hạn một người đang sửa trường dữ liệu của một bản ghi thì người kia ra lệnh xóa cả bản ghi. Trong những trường hợp như vậy, hệ QTCSDL cần cung cấp chức năng kiểm soát các giao dịch để đảm bảo tính nhất quán của dữ liệu.

- Hệ QTCSDL cung cấp các phương tiện thực hiện sao lưu dự phòng (backup) để đề phòng các sự cố gây mất dữ liệu và khôi phục dữ liệu khi cần thiết.

d) Nhóm chức năng giao diện lập trình ứng dụng cung cấp cho các nhà phát triển ứng dụng CSDL các phương thức và công cụ để họ có thể gửi được truy vấn đến CSDL từ ứng dụng mà họ phát triển, nhằm đáp ứng những nhu cầu công việc cụ thể.

Điều quan trọng là các chức năng nói trên của hệ QTCSDL được xây dựng một cách tổng quát, theo mô hình CSDL không phụ thuộc các CSDL và ứng dụng cụ thể.

Những hệ QTCSDL được dùng phổ biến hiện nay là Oracle, MySQL, SQL Server, DB2, PostGreSQL, SQLite,... Cũng có thể nhắc đến Microsoft Access, tuy nhiên do có nhiều hạn chế về hiệu suất, hệ QTCSDL này không được sử dụng cho các ứng dụng phức tạp, đặc biệt với các ứng dụng trực tuyến đang trở nên phổ biến.

- Hệ QTCSDL là phần mềm cung cấp phương thức để lưu trữ, cập nhật và truy xuất dữ liệu của CSDL, bảo mật và an toàn dữ liệu.
- Hệ QTCSDL cũng cung cấp giao diện lập trình ứng dụng cho các nhà phát triển ứng dụng và người dùng.



1. Nêu những khó khăn trong việc khai thác CSDL nếu không có hệ QTCSDL.
2. Tóm tắt các nhóm chức năng của hệ QTCSDL.

2. HỆ CƠ SỞ DỮ LIỆU

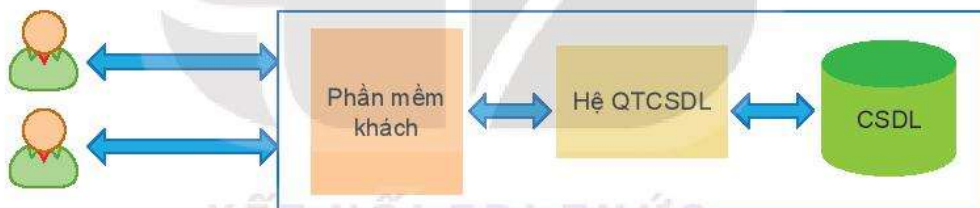
Hoạt động 2 Thảo luận về tra cứu trực tuyến điểm thi

Khi lên mạng để tra cứu điểm thi vào lớp 10, thông thường trên màn hình chỉ yêu cầu nhập vài dữ liệu tối thiểu. Ví dụ sau khi nhập số báo danh gần như ngay lập tức em nhận được đầy đủ thông tin họ tên, trường lớp, điểm thi cùng kết quả xét tuyển. Vậy, ngoài CSDL điểm thi cần có những gì để có thể cung cấp cho em thông tin như vậy?

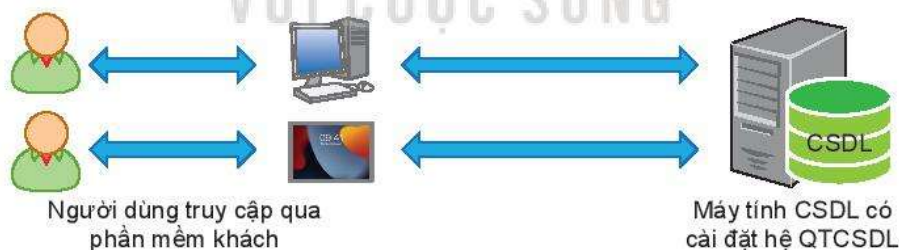


Câu trả lời dễ suy đoán là đã có một phần mềm tổ chức giao tiếp với người dùng (tra cứu điểm thi) làm cầu nối giữa người dùng với CSDL điểm thi.

Vì mục đích hỗ trợ nhiều người dùng, các hệ QTCSDL nhiều người dùng thường được xây dựng theo mô hình hai thành phần, có thể được cài đặt trên cùng một máy tính hay được cài đặt trên các máy tính khác nhau. Một thành phần có nhiệm vụ chính là thực hiện các nhiệm vụ tính toán, xử lý dữ liệu (được gọi là phần “chủ”), thành phần còn lại (được gọi là phần mềm “khách”) có nhiệm vụ tổ chức giao diện tương tác với người dùng, kết nối với phần “chủ”, gửi các yêu cầu tính toán xử lý dữ liệu đến phần “chủ” và nhận lại, hiển thị kết quả tính toán xử lý dữ liệu. Hai thành phần này luôn được cung cấp trong một gói cài đặt hệ QTCSDL và thành phần chủ thường được gọi luôn là hệ QTCSDL.



Hình 12.1. Phần mềm khách và hệ QTCSDL cài đặt trên cùng một máy tính



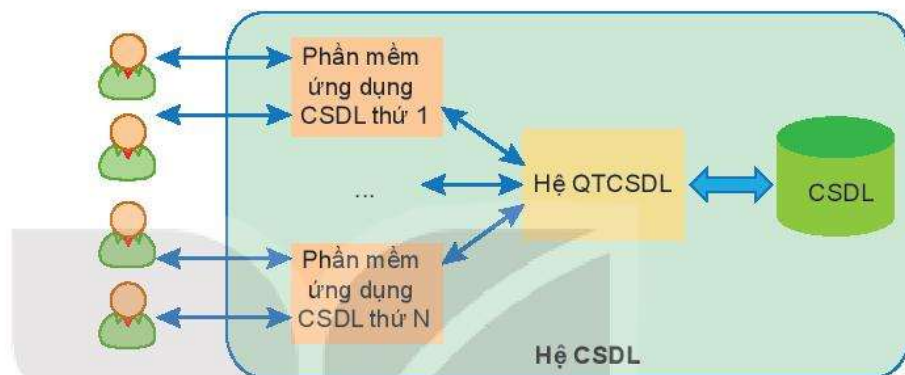
Hình 12.2. Hệ QTCSDL và phần mềm khách cài đặt trên các máy tính khác nhau

Các hệ QTCSDL cũng cung cấp công cụ (giao diện lập trình ứng dụng) để người dùng có thể tự viết phần mềm khách chuyên biệt theo nhu cầu. Phần mềm khách chuyên biệt như thế được gọi là phần mềm ứng dụng CSDL.

Nói một cách khái quát thì phần mềm ứng dụng CSDL là phần mềm được xây dựng tương tác với hệ QTCSDL nhằm mục đích hỗ trợ người dùng khai thác thông tin từ CSDL một cách thuận tiện theo các yêu cầu xác định. Có thể có nhiều phần mềm ứng dụng CSDL được xây dựng với các mục tiêu yêu cầu khác nhau cùng truy xuất, khai thác thông tin từ một CSDL.

Khi lên mạng tra cứu điểm thi vào lớp 10, em đã tương tác với một phần mềm ứng dụng CSDL tra cứu điểm thi được thiết kế với giao diện đơn giản, dễ dàng nhất cho người dùng. Sau khi tiếp nhận thông tin (ví dụ số báo danh), nó sẽ kết nối với hệ QTCSDL quản lý điểm thi, yêu cầu lấy ra những thông tin cần thiết (họ tên, trường lớp, điểm tất cả các môn thi cùng kết quả xét tuyển) và hiển thị kết quả nhận được cho em. Ngoài phần mềm ứng dụng CSDL tra cứu điểm thi, có thể còn có nhiều phần mềm ứng dụng CSDL nữa được xây dựng để khai thác thông tin từ CSDL điểm thi, như phần mềm xét tuyển, phần mềm thống kê đánh giá kết quả toàn bộ kì thi,...

Các ứng dụng mua bán trực tuyến, đặt xe công nghệ, thanh toán điện tử,... mà các em thường gặp trong cuộc sống đều là các phần mềm ứng dụng CSDL của một hệ thống CSDL cụ thể.



Hình 12.3. Hệ cơ sở dữ liệu

- Phần mềm ứng dụng CSDL là phần mềm được xây dựng tương tác với hệ QTCSDL nhằm mục đích hỗ trợ người dùng khai thác thông tin từ CSDL một cách thuận tiện theo các yêu cầu xác định
- Một hệ thống gồm ba thành phần: CSDL, hệ QTCSDL và các phần mềm ứng dụng CSDL được gọi là một hệ CSDL.



Hệ QTCSDL và hệ CSDL khác nhau như thế nào?

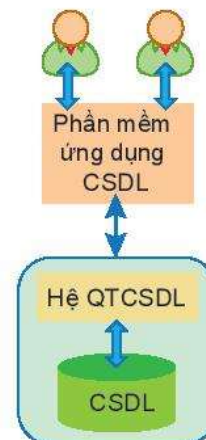
3. HỆ CƠ SỞ DỮ LIỆU TẬP TRUNG VÀ PHÂN TÂN



a) Hệ CSDL tập trung

Hệ CSDL mà CSDL được lưu trữ tập trung trên một máy tính được gọi là *hệ cơ sở dữ liệu tập trung*.

Hệ CSDL tập trung bao gồm cả những CSDL một người dùng trên một máy (như các CSDL của Microsoft Access); người dùng vừa là người thiết kế, tạo lập và bảo trì CSDL, vừa là người viết phần mềm ứng dụng CSDL, vừa là người dùng đầu cuối hệ thống, khai thác thông tin theo những mục tiêu đã được đặt ra.



Hình 12.4. Hệ CSDL tập trung

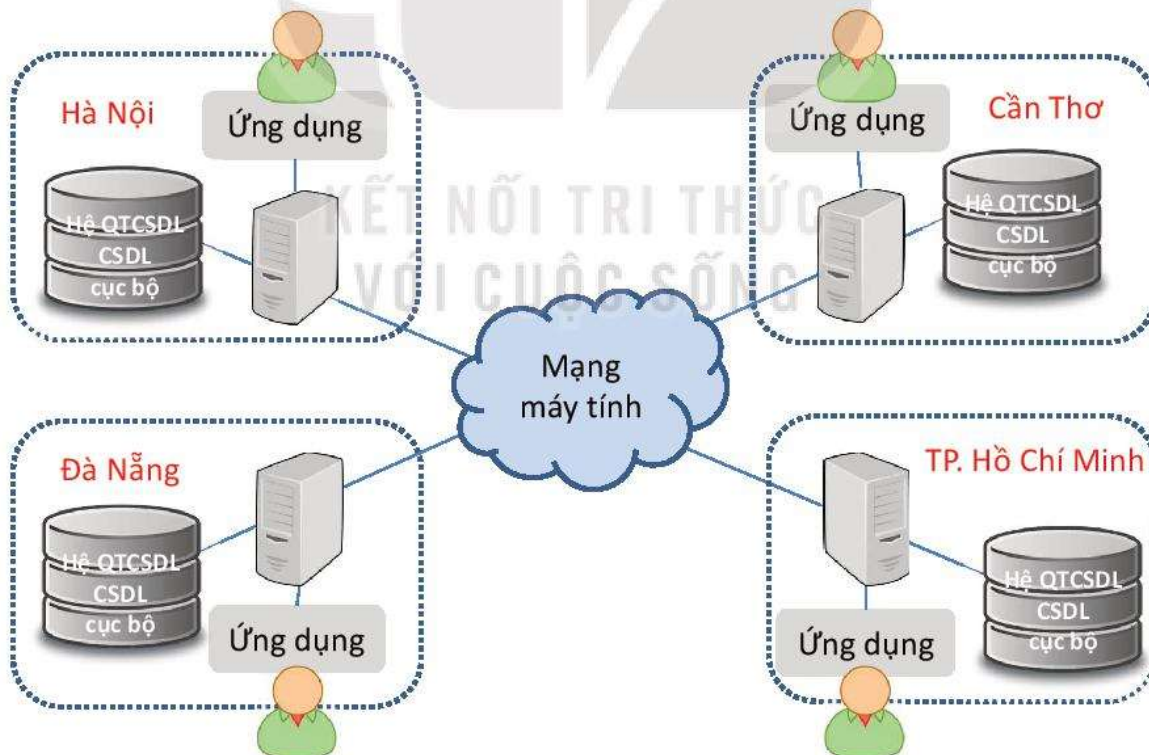
b) Hệ CSDL phân tán

Khi một tổ chức có nhiều đơn vị phân tán về mặt địa lý, họ có thể chọn giải pháp tổ chức hệ CSDL phân tán để giải quyết các bài toán quản lý liên quan. Khác với hệ QTCSDL tập trung, quản lý dữ liệu tập trung trong một CSDL đặt trên một máy tính, hệ CSDL phân tán cho phép người dùng truy cập dữ liệu được lưu trữ ở nhiều máy tính khác nhau của mạng máy tính. Như vậy, để thiết lập một hệ CSDL phân tán, trước tiên cần xây dựng một CSDL phân tán tương ứng.

Ví dụ: Một doanh nghiệp có trụ sở chính tại Hà Nội và các chi nhánh tại Đà Nẵng, TP. Hồ Chí Minh, Cần Thơ. Mỗi máy tính tại trụ sở Hà Nội, Đà Nẵng, TP. Hồ Chí Minh, Cần Thơ hình thành một trạm; các trạm này được kết nối với nhau bởi một mạng máy tính.

Cơ sở dữ liệu phân tán là tập hợp dữ liệu được phân tán trên các trạm khác nhau của một mạng máy tính. Dữ liệu được lưu trữ tại mỗi trạm hình thành **một CSDL cục bộ** của trạm này. Mỗi trạm **phải thực hiện các ứng dụng cục bộ**. Ứng dụng cục bộ là ứng dụng chạy tại một trạm và chỉ sử dụng dữ liệu cục bộ của trạm này để cho ra kết quả cuối cùng. Mỗi trạm phải tham gia thực hiện ứng dụng toàn cục, còn được gọi là **ứng dụng phân tán**. Ứng dụng toàn cục là ứng dụng chạy tại một trạm và sử dụng dữ liệu của ít nhất hai trạm để cho ra kết quả cuối cùng.

Các chương trình ứng dụng chạy tại mỗi trạm, các hệ quản trị CSDL tại mỗi trạm và các CSDL của mỗi trạm hình thành **hệ CSDL phân tán**.



Hình 12.5. Hệ cơ sở dữ liệu phân tán

- Hệ CSDL mà CSDL được lưu trữ tập trung trên một máy tính được gọi là **hệ cơ sở dữ liệu tập trung**.
- Hệ CSDL phân tán cho phép người dùng truy cập dữ liệu được lưu trữ ở nhiều máy tính khác nhau trên mạng máy tính.



So với hệ CSDL tập trung, thiết kế và triển khai hệ CSDL phân tán phức tạp, khó khăn hơn trong đảm bảo tính nhất quán và bảo mật dữ liệu, chi phí duy trì cao hơn. Tuy nhiên, hệ CSDL phân tán có ưu điểm:

- Dễ dàng mở rộng, luôn có thể bổ sung thêm trạm dữ liệu vào hệ thống khi cần mà không làm ảnh hưởng đến hoạt động của các trạm dữ liệu đang hoạt động.
- Tính sẵn sàng và độ tin cậy được nâng cao. Hệ thống hoạt động ổn định, hạn chế tối đa việc mất mát dữ liệu dù có thể có trạm dữ liệu gặp sự cố vì dữ liệu có thể được sao lưu nhiều bản đặt ở các trạm dữ liệu khác.

Những hệ CSDL lớn dùng trong những tổ chức lớn, có nhiều người truy xuất trên phạm vi địa lý rộng lớn thường được tổ chức ở dạng hệ cơ sở dữ liệu phân tán để tối ưu hoá được tốc độ, giảm tải đường truyền khi truy xuất, cập nhật dữ liệu. Ví dụ điển hình là các CSDL của Facebook, Google, Amazon,... được cài đặt trên nhiều máy chủ ở nhiều quốc gia và hỗ trợ khai thác thông tin theo mô hình khách chủ trên nền web.

Lưu ý: Cần phân biệt Hệ CSDL phân tán với hệ CSDL tập trung nhưng xử lý dữ liệu phân tán. Một ví dụ cụ thể của mô hình CSDL tập trung nhưng xử lý phân tán là mô hình máy chủ tập. Toàn bộ CSDL và phần mềm ứng dụng CSDL được đặt trên máy chủ tập. Các máy tính tham gia vào hệ thống được gọi là máy trạm. Mỗi khi có yêu cầu thao tác từ máy trạm, toàn bộ CSDL và phần mềm được chuyển qua đường truyền của mạng về bộ nhớ RAM của máy trạm. Công việc xử lý dữ liệu được thực hiện trên máy trạm, khi kết thúc toàn bộ dữ liệu lại được lưu về máy chủ tập.



Hệ CSDL tập trung và hệ CSDL phân tán khác nhau như thế nào?



LUYỆN TẬP

1. Hãy lập danh sách các chức năng của hệ QTCSDL trong từng nhóm chức năng của hệ QTCSDL.
2. Hãy phân tích điểm mạnh và điểm yếu của CSDL phân tán so với CSDL tập trung.



VẬN DỤNG

1. Cho ví dụ về một hệ CSDL trên thực tế, chỉ rõ những thành phần của nó.
2. Hãy tìm hiểu qua Internet tên một số hệ quản trị CSDL quan hệ thông dụng.