

BÀI 7

CÁC LOẠI KIẾN TRÚC CỦA HỆ CƠ SỞ DỮ LIỆU

Học xong bài này, em sẽ:

- ✓ Phân biệt được CSDL tập trung và CSDL phân tán.
- ✓ Biết được một số kiến trúc thường gặp của hai loại hệ CSDL tập trung và hệ CSDL phân tán.



Theo em, CSDL của trường em được đặt trong một máy tính hay trong tất cả các máy tính có sử dụng CSDL đó? CSDL của một ngân hàng được đặt trong một máy tính hay nhiều máy tính?

1 Cơ sở dữ liệu tập trung và cơ sở dữ liệu phân tán

a) Cơ sở dữ liệu tập trung

Một CSDL tập trung được lưu trữ trên một máy tính (Hình 1). Việc quản lý, cập nhật được thực hiện tại chính vị trí này. Tùy trường hợp cụ thể, người dùng có thể truy cập và khai thác thông tin bằng chính máy tính chứa CSDL hay thông qua kết nối mạng (Internet, LAN, WAN,...).



Hình 1. Hệ cơ sở dữ liệu tập trung

Vì tất cả dữ liệu được lưu trữ tại một máy tính duy nhất nên việc truy cập và điều phối dữ liệu dễ dàng hơn, đây là một ưu điểm lớn. Bởi vậy, phần lớn các cơ quan, doanh nghiệp, tổ chức dùng hệ CSDL tập trung. Hệ thống quản lý học sinh của trường em là một hệ CSDL tập trung cỡ nhỏ. Hệ thống bán vé tàu hỏa của Tổng công ty Đường sắt Việt Nam cũng là một ví dụ về hệ CSDL tập trung.

Tuy nhiên, hệ CSDL tập trung cũng có những hạn chế. Trong quá trình khai thác, nếu CSDL tập trung gặp sự cố thì các chương trình ứng dụng CSDL không thể chạy được.

b) Cơ sở dữ liệu phân tán

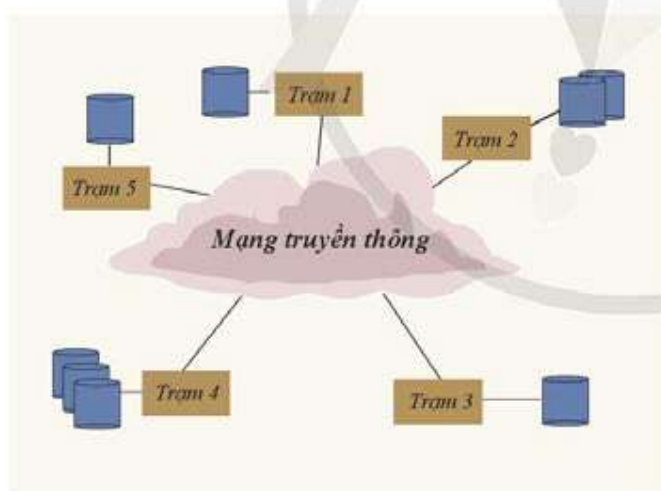


Theo em, các hệ thống thư điện tử trên Internet có thể sử dụng hệ CSDL tập trung không? Vì sao?

Một CSDL phân tán (Hình 2) là một tập hợp dữ liệu được lưu trữ phân tán trên các máy tính khác nhau của một mạng máy tính (mỗi máy tính như vậy được gọi là một *site* hay một *trạm* của mạng) cùng với những đặc điểm sau đây:

- Mỗi trạm có một CSDL được gọi là CSDL cục bộ của trạm này. Mỗi trạm thực hiện ít nhất một ứng dụng cục bộ, tức là chỉ sử dụng CSDL cục bộ để cho ra kết quả. Khả năng thực hiện ứng dụng cục bộ được gọi là xử lý độc lập.

- Mỗi trạm phải tham gia thực hiện ít nhất một ứng dụng toàn cục. Ứng dụng toàn cục là ứng dụng chạy tại một trạm và phải sử dụng CSDL của ít nhất hai trạm.



Hình 2. Hệ cơ sở dữ liệu phân tán



Cơ sở dữ liệu phân tán:

tập hợp dữ liệu được phân tán trên các máy tính khác nhau của một mạng máy tính. Mỗi nơi (site) của mạng máy tính có khả năng xử lý độc lập và thực hiện các ứng dụng cục bộ. Mỗi nơi cũng tham gia thực hiện ít nhất một ứng dụng toàn cục, yêu cầu truy xuất dữ liệu tại nhiều nơi bằng cách dùng hệ thống truyền thông con.

Ví dụ 1. Một ngân hàng có nhiều chi nhánh, ở mỗi thành phố có một chi nhánh, ở mỗi chi nhánh có dữ liệu quản lý các tài khoản của dân cư và đơn vị đăng kí kinh doanh tại thành phố đó. Thông qua mạng truyền thông, tập hợp dữ liệu của ngân hàng này tại các chi nhánh tạo thành một hệ CSDL phân tán. Người chủ của một tài khoản có thể thực hiện các giao dịch (chẳng hạn rút một khoản tiền trong tài khoản) ở

một chi nhánh nào đó (ví dụ ở Hà Nội), nhưng cũng có thể thực hiện giao dịch ở một chi nhánh đặt tại một thành phố khác (ví dụ ở Đà Nẵng).

Ví dụ 2. Hệ thống tìm kiếm Google có hệ CSDL phân tán. Mỗi yêu cầu có thể được thực hiện bởi hàng trăm máy tính thu thập dữ liệu web và trả về các kết quả có liên quan.

Đối với người dùng, Google dường như là một hệ thống, nhưng nó thực sự là nhiều máy tính làm việc cùng nhau và truy xuất dữ liệu tại nhiều trạm để hoàn thành một nhiệm vụ duy nhất (trả lại kết quả cho truy vấn tìm kiếm).

So với hệ CSDL tập trung, hệ CSDL phân tán có một số ưu điểm chính:

- Sự phân tán dữ liệu về mặt vật lý phù hợp với các tổ chức, doanh nghiệp lớn hoạt động trải rộng về mặt địa lý, phù hợp với các dịch vụ phủ rộng trên toàn cầu, ví dụ như: các hệ thống dịch vụ dựa trên web, hệ thống thương mại điện tử,...

- Tính sẵn sàng và tính tin cậy của dữ liệu cao hơn. Tính sẵn sàng phục vụ cao là do những dữ liệu được đơn vị nào sử dụng nhiều nhất sẽ được lưu trữ và quản lý tại đơn vị đó, thêm nữa khi có sự cố không truy cập được dữ liệu tại một trạm thì vẫn có thể khai thác bản sao của dữ liệu đặt tại một trạm khác. Cũng như vậy, về tính tin cậy, khi một trạm gặp sự cố, có thể khôi phục được dữ liệu tại đây do có bản sao của nó được lưu trữ và vận hành tại một hay vài trạm khác nữa.

- Mở rộng các tổ chức một cách linh hoạt. Có thể thêm trạm mới vào mạng máy tính mà không ảnh hưởng đến hoạt động của các trạm sẵn có.

Tuy nhiên, hệ CSDL phân tán có một số hạn chế so với hệ CSDL tập trung:

- Chi phí cao hơn do hệ thống phức tạp hơn, hệ thống phải làm ẩn đi sự phân tán dữ liệu đối với người dùng.

- Khó khăn hơn trong đảm bảo tính nhất quán dữ liệu và tính an ninh, đồng thời rất khó cung cấp một cái nhìn thống nhất cho người dùng vì dữ liệu đặt tại nhiều địa điểm khác nhau.

2 Các loại kiến trúc của các hệ cơ sở dữ liệu

Mỗi hệ CSDL bao gồm 3 lớp:

- Lớp CSDL.
- Lớp hệ quản trị CSDL.
- Lớp các ứng dụng CSDL.

Nói về kiến trúc của một hệ CSDL là muốn nhìn hệ thống đó dưới cách phân chia nó thành các thành phần chức năng để có thể hiểu và chỉnh sửa, thay thế mỗi thành phần đó một cách khá độc lập. Dưới đây giới thiệu sơ lược một số kiến trúc phổ biến của hai loại hệ CSDL tập trung và hệ CSDL phân tán.

a) Kiến trúc phổ biến của hệ CSDL tập trung

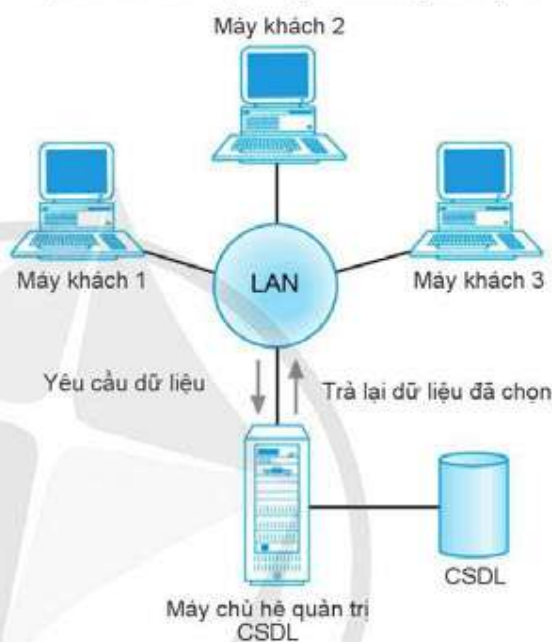
Nhìn chung các hệ CSDL tập trung theo kiến trúc khách – chủ (Client – Server), các thành phần của hệ quản trị CSDL gồm thành phần yêu cầu tài nguyên (dữ liệu) và

thành phần cung cấp tài nguyên (dữ liệu) không nhất thiết phải cài đặt trên cùng một máy tính.

Thành phần cung cấp tài nguyên thường đặt tại một máy chủ (server). Thành phần yêu cầu tài nguyên có thể cài đặt tại nhiều máy khác trên mạng, ta gọi là máy khách (client).

Kiến trúc 1 tầng (1-Tier Architecture) là kiến trúc đơn giản nhất, toàn bộ CSDL được lưu trữ tại một máy tính và cũng chỉ được khai thác tại máy tính này. Máy tính như vậy vừa là máy chủ CSDL vừa là máy khách duy nhất khai thác CSDL. Tuy nhiên, kiến trúc đơn giản này không phù hợp cho các ứng dụng phức tạp.

Kiến trúc 2 tầng (2-Tier Architecture) là kiến trúc có CSDL được lưu trữ ở một máy chủ trên mạng (được xem là tầng 2), thành phần trình bày dữ liệu cho người khai thác được cài đặt trên máy khách kết nối được với mạng (được xem là tầng 1). Máy khách có thể là PC, máy tính bảng hay điện thoại di động,... Tuy nhiên, hiệu suất hoạt động của hệ thống này sẽ kém trong trường hợp có nhiều máy khách cùng khai thác CSDL.



Hình 3. Mô hình Khách – Chủ 2 tầng của CSDL tập trung

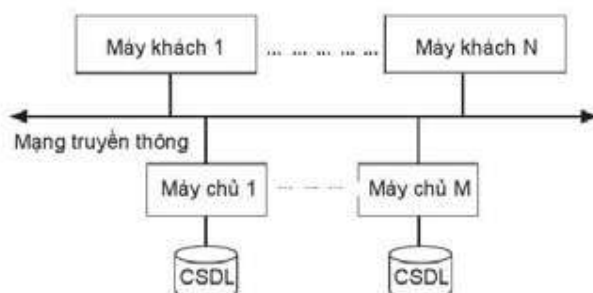
Kiến trúc 3 tầng (3-Tier Architecture) là kiến trúc mở rộng của kiến trúc 2 tầng. Tầng 1 vẫn là thành phần trình bày dữ liệu. Tầng 3 là máy chủ chứa CSDL. Tầng 2 nằm giữa gọi là tầng ứng dụng, hoạt động như một phương tiện để trao đổi dữ liệu đã được xử lý một phần giữa máy chủ và máy khách. Tầng trung gian này chứa các chương trình ứng dụng thường xử lý các vấn đề nghiệp vụ trước khi chuyển dữ liệu qua lại giữa tầng 1 và tầng 3. Loại kiến trúc này thường được sử dụng trong trường hợp các ứng dụng web lớn.

b) Các kiến trúc phổ biến của hệ CSDL phân tán

Hệ CSDL phân tán có một số mô hình kiến trúc phổ biến: mô hình ngang hàng (peer to peer), mô hình khách – chủ cho hệ CSDL phân tán.

Kiến trúc ngang hàng cho hệ CSDL phân tán có mỗi máy tính hoạt động như một máy khách và máy chủ để truyền tải các dịch vụ CSDL. Các máy tính ngang hàng với nhau trong khả năng chia sẻ nguồn tài nguyên dữ liệu của nó với các máy khác và cùng ngang hàng trong khả năng điều phối các hoạt động.

Kiến trúc khách – chủ cho hệ CSDL cũng là kiến trúc khách – chủ như đã biết, nhưng khác với ở hệ CSDL tập trung, hệ CSDL phân tán có nhiều máy chủ CSDL (Hình 4).



Hình 4. Mô hình khách – chủ 2 tầng của CSDL phân tán với nhiều máy chủ và nhiều máy khách



Hãy nêu đặc điểm quan trọng nhất để phân biệt một hệ CSDL tập trung với một hệ CSDL phân tán.



Dựa vào quy mô và đặc điểm tổ chức của mình mà các doanh nghiệp lựa chọn xây dựng cho mình loại hệ CSDL (tập trung hay phân tán) và mô hình kiến trúc phù hợp. Em hãy giải thích và lấy vài ví dụ để minh họa.



Trong các câu sau đây, những câu nào đúng?

- CSDL luôn chỉ được lưu trữ và khai thác tại một máy tính.
- Trong hệ CSDL tập trung, việc quản lý và cập nhật dữ liệu dễ dàng hơn so với hệ CSDL phân tán.
- Trong tất cả các hệ CSDL, hễ có sự cố không truy cập được một máy chủ CSDL thì toàn bộ hệ thống CSDL đó ngừng hoạt động.
- Một hệ CSDL phân tán đắt hơn so với một hệ CSDL tập trung vì nó phức tạp hơn nhiều.

Tóm tắt bài học

- ✓ Điểm khác biệt quan trọng giữa CSDL tập trung và CSDL phân tán là: CSDL tập trung có toàn bộ dữ liệu được lưu trữ trên một máy tính, trong khi đó CSDL phân tán có dữ liệu phân tán trên các máy tính khác nhau của một mạng máy tính và mỗi máy tính khai thác CSDL đều tham gia ít nhất một ứng dụng toàn cục.
- ✓ Kiến trúc khách – chủ là kiến trúc phổ biến của các hệ CSDL tập trung, tùy theo ứng dụng mà có kiến trúc theo mô hình 1 tầng, 2 tầng hay nhiều tầng hơn.
- ✓ Có vài loại mô hình kiến trúc phổ biến của các hệ CSDL phân tán: khách – chủ (cho CSDL phân tán), ngang hàng,...