

Quản trị dữ liệu

Chương 1: Giới thiệu

Mục tiêu của chương

- ◆ Sau chương này, sinh viên có thể hiểu được:
 - Khái niệm về HQTCSDL và các khái niệm cơ bản liên quan đến HQTCSDL
 - Cấu trúc tổng thể của một HQTCSDL.
 - Các kiến trúc của một HQTCSDL

Tóm tắt nội dung

- ◆ Giới thiệu HQT CSDL
- ◆ Cách nhìn dữ liệu
- ◆ Mô hình dữ liệu
- ◆ Ngôn ngữ Cơ sở dữ liệu
- ◆ Quản trị giao dịch
- ◆ Quản trị lưu trữ
- ◆ Nhà quản trị CSDL
- ◆ Người sử dụng CSDL
- ◆ Cấu trúc hệ thống tổng thể
- ◆ Kiến trúc Hệ CSDL.

Giới thiệu Hệ CSDL (1)

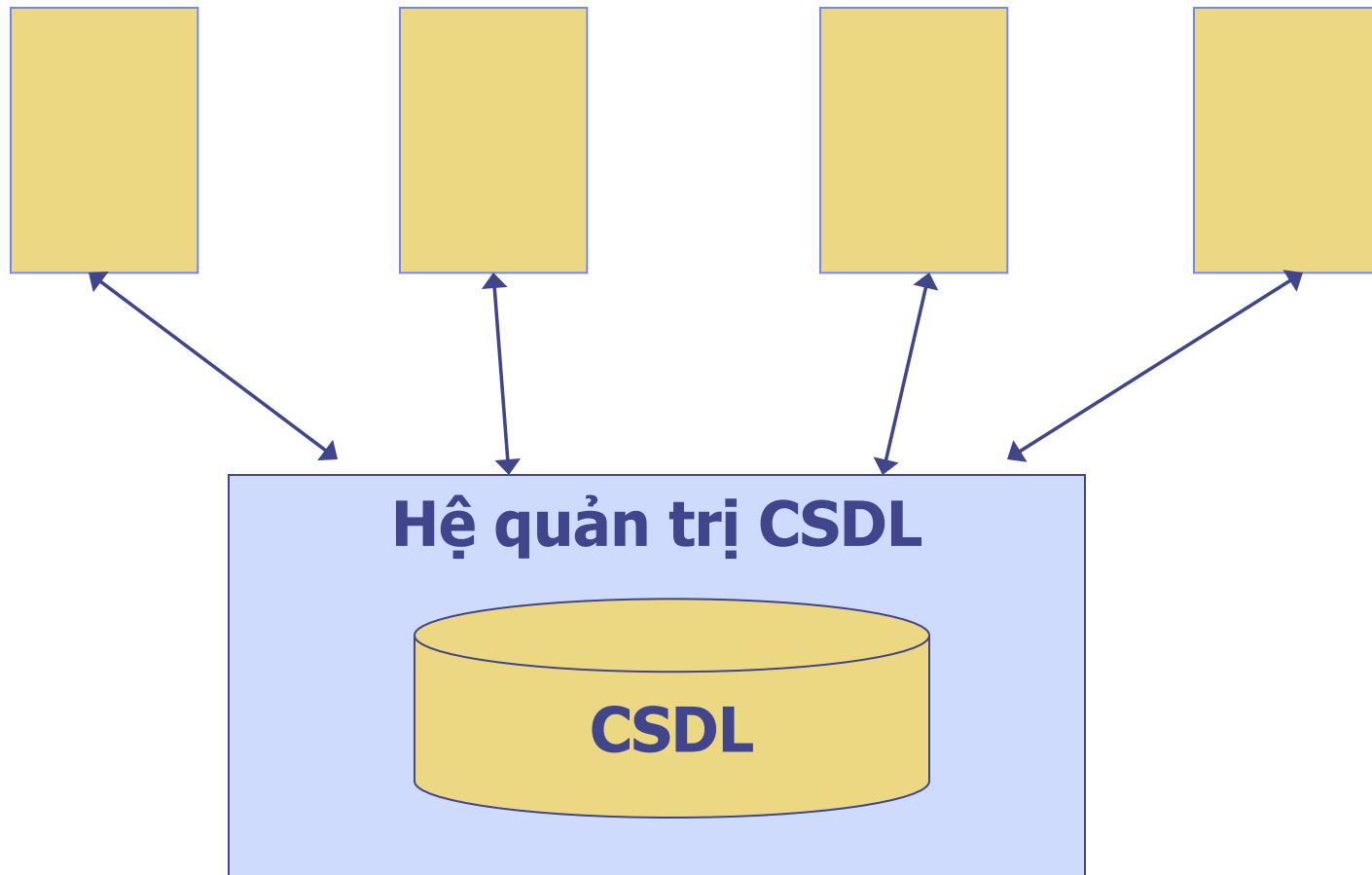
- ◆ Việc lưu trữ thông tin có tổ chức trong hệ thống file thông thường có một số hạn chế:
 - Dư thừa dữ liệu và tính bất nhất
 - Khó khăn trong việc truy xuất
 - Sự cô lập dữ liệu
 - Tính nguyên tử
 - Truy xuất cạnh tranh
 - An toàn
- ◆ Các bất lợi trên đã gợi mở cho sự phát triển các DBMS

Giới thiệu Hệ CSDL (2)

- ◆ CSDL: Là tập hợp các dữ liệu có quan hệ một cách luận lý với nhau (và sự mô tả đầy đủ về những dữ liệu này), được thiết kế nhằm đảm bảo yêu cầu là hệ thống có tổ chức, giảm thiểu dữ liệu dư thừa.
- ◆ DBMS: là hệ thống phần mềm cho phép người dùng định nghĩa, tạo ra và quản lý CSDL. Ngoài ra nó còn cho phép người dùng truy cập đến CSDL đó.

Giới thiệu Hệ CSDL (3)

Các ứng dụng



Giới thiệu Hệ CSDL (4)

◆ DBMS:

- Hệ thống phần mềm thoả mãn hai yêu cầu cơ bản:
 - ◆ Có khả năng quản lý dữ liệu một cách lâu dài.
 - ◆ Có khả năng truy cập một lượng dữ liệu lớn một cách nhanh chóng và hiệu quả.
- Một số đòi hỏi khác:
 - ◆ Hỗ trợ ít nhất là một mô hình dữ liệu
 - ◆ Hỗ trợ các ngôn ngữ lập trình cấp cao
 - ◆ Quản lý giao dịch - điều khiển cạnh tranh
 - ◆ Quản lý truy cập: kiểm tra tính hợp lệ của truy cập và kiểm tra quyền hạn truy cập
 - ◆ Có hệ thống phục hồi lỗi tự động

Giới thiệu Hệ CSDL (5)

- ◆ Các tiện ích mà DBMS cung cấp:
 - Cho phép định nghĩa CSDL thông qua ngôn ngữ DDL.
 - Cho phép chèn, cập nhật, xoá và truy vấn dữ liệu thông qua ngôn ngữ DML
 - Điều khiển các truy cập đến CSDL, thông qua:
 - ◆ Hệ thống an ninh (dùng mật khẩu...)
 - ◆ Hệ thống bảo đảm toàn vẹn DL nhằm duy trì tính nhất quán của dữ liệu được lưu trữ
 - ◆ Hệ thống phục hồi
 - Xây dựng cơ chế view, cho phép người dùng có được cái nhìn riêng về CSDL mà họ yêu cầu.
 - ...

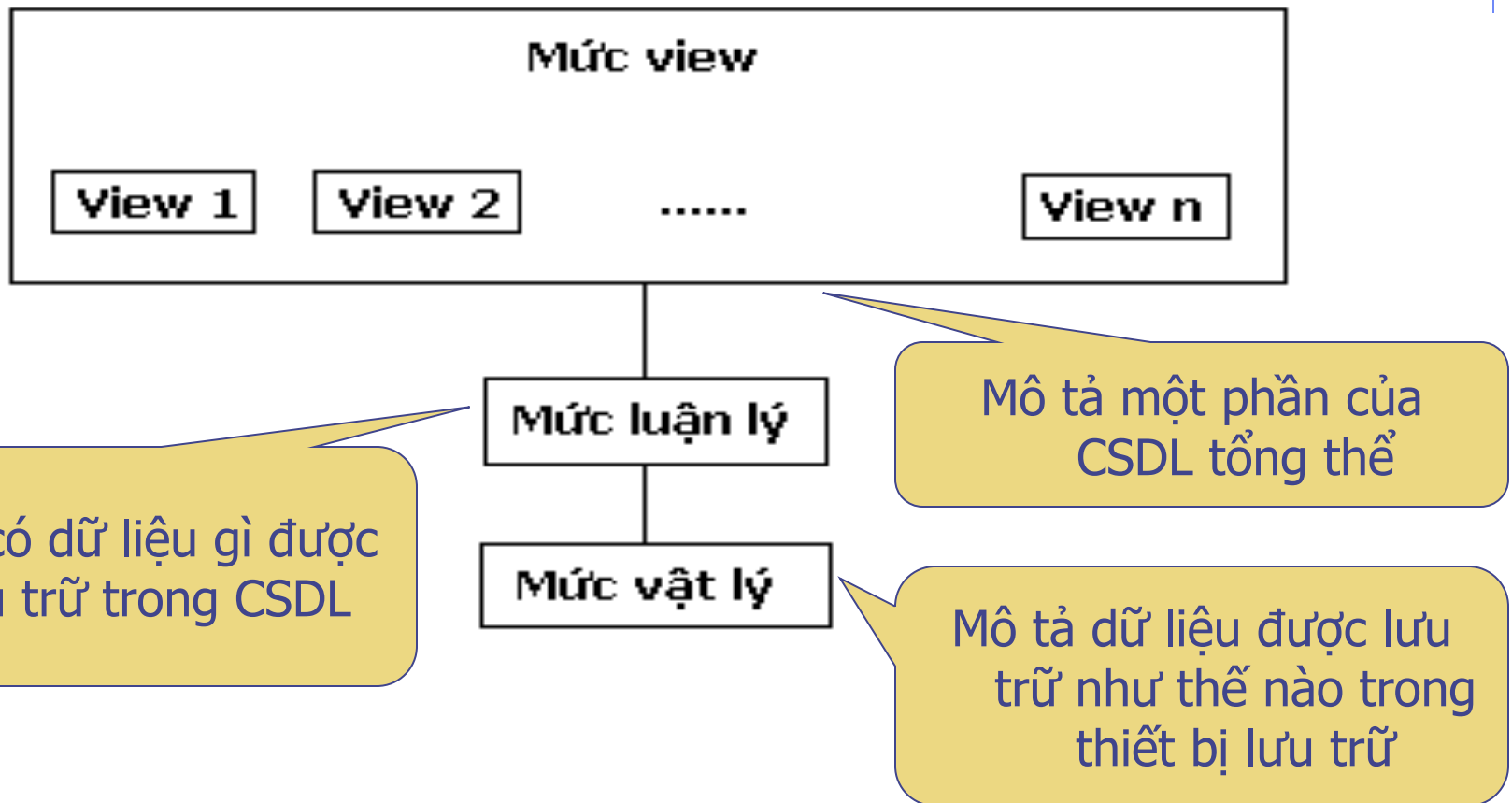
Giới thiệu Hệ CSDL (6)

◆ Các bất lợi của DBMS:

- Sự phức tạp
- Kích thước lớn
- Giá cả: của phần mềm DBMS, của phần cứng được yêu cầu tương ứng, cho qua trình nâng cấp hệ thống
- Hiệu năng của hệ thống giảm
- Khả năng hỏng hóc cao hơn

Cách nhìn dữ liệu (1)

- ◆ Người ta **che dấu sự phức tạp** của CSDL bằng cách thiết kế nó thành các lớp trừu tượng



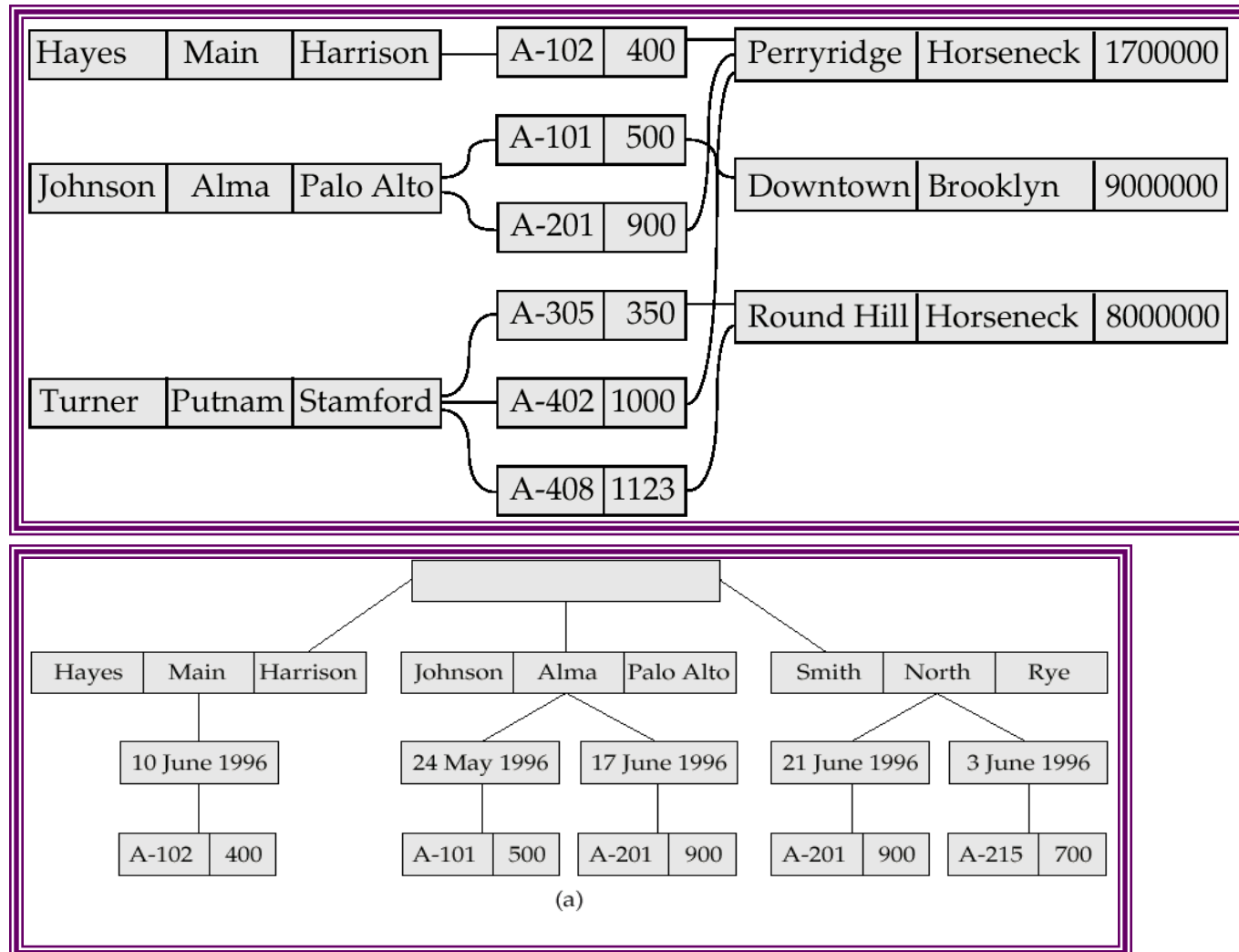
Cách nhìn dữ liệu (2)

- ◆ **Sơ đồ** (schema): là thiết kế tổng thể của một CSDL. Các DBMS có một vài sơ đồ, được phân tương ứng với các mức trừu tượng: sơ đồ vật lý, sơ đồ luận lý và sơ đồ con.
- ◆ **Thể hiện** (instance): Tập các thông tin được lưu trữ trong cơ sở dữ liệu tại một thời điểm.
- ◆ Khả năng sửa đổi một định nghĩa ở một mức không ảnh hưởng đến mức cao hơn được gọi là ***sự độc lập dữ liệu***.
 - ***Độc lập dữ liệu vật lý***: khả năng sửa đổi sơ đồ vật lý mà không viết lại chương trình.
 - ***Độc lập dữ liệu luận lý***: khả năng thay đổi sơ đồ luận lý mà không phải viết lại chương trình.

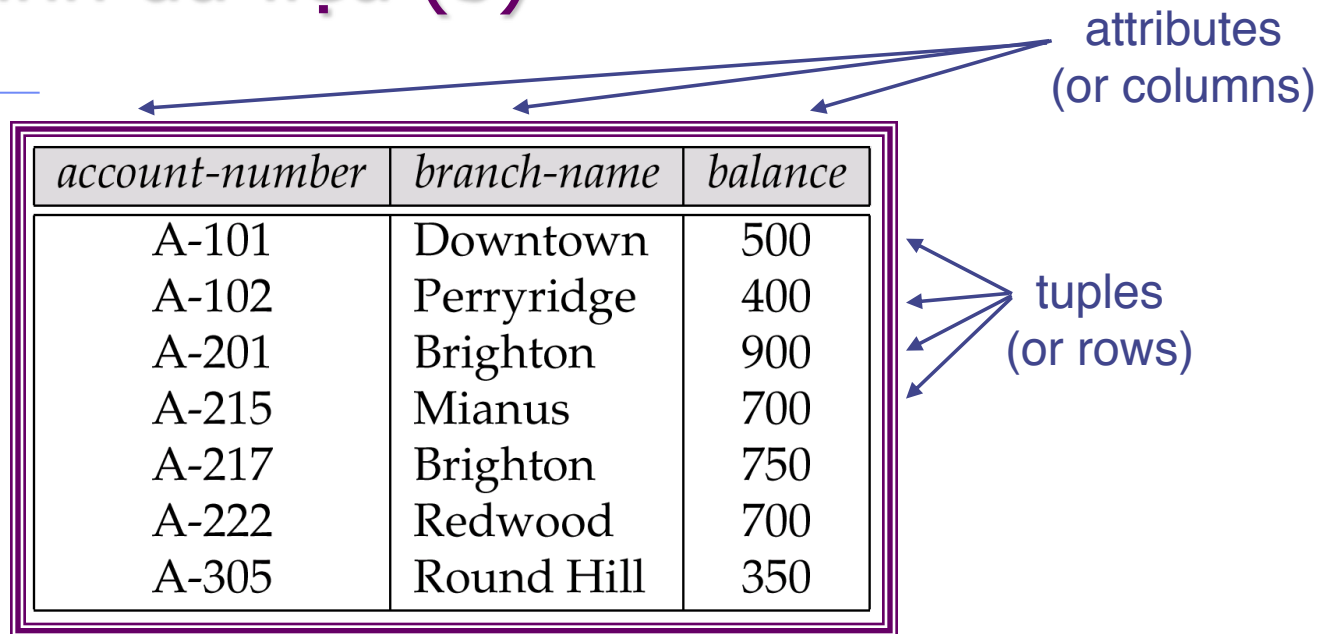
Mô hình dữ liệu (1)

- ◆ Mô hình dữ liệu là một bộ các công cụ quan niệm để mô tả dữ liệu, quan hệ dữ liệu và các ràng buộc.
 - Mô hình luận lý dựa trên đối tượng: dữ liệu được mô tả dưới dạng các đối tượng. Thường được dùng mô tả dữ liệu ở mức luận lý và mức view: mô hình thực thể quan hệ, mô hình hướng đối tượng.
 - Mô hình luận lý dựa trên mẫu tin: dữ liệu được mô tả dưới dạng các mẫu tin. Dùng mô tả dữ liệu ở mức luận lý hoặc mức view: mô hình quan hệ, mô hình mạng, mô hình phân cấp.
 - Mô hình dữ liệu vật lý: dùng mô tả dữ liệu ở mức vật lý: mô hình hợp nhất, mô hình khung – bộ nhớ.

Mô hình dữ liệu (2)



Mô hình dữ liệu (3)



<i>account-number</i>	<i>branch-name</i>	<i>balance</i>
A-101	Downtown	500
A-102	Perryridge	400
A-201	Brighton	900
A-215	Mianus	700
A-217	Brighton	750
A-222	Redwood	700
A-305	Round Hill	350

<i>customer-name</i>	<i>account-number</i>
Hayes	A-102
Johnson	A-101
Johnson	A-201
Jones	A-217
Lindsay	A-222
Smith	A-215
Turner	A-305

Ngôn ngữ CSDL

◆ Ngôn ngữ định nghĩa dữ liệu (DDL):

- Cho phép xác định sơ đồ cơ sở dữ liệu.
- Kết quả biên dịch các câu lệnh DDL là tập hợp các bảng được lưu trong một file đặc biệt được gọi là **từ điển dữ liệu** (data dictionary) hay **danh bạ dữ liệu** (data directory).
- Từ điển dữ liệu sẽ được tham khảo đến mỗi khi dữ liệu hiện hành được đọc hoặc thay đổi

◆ Ngôn ngữ thao tác dữ liệu (DML):

- Cho phép truy xuất hoặc thao tác dữ liệu.
- Phần ngôn ngữ DML liên quan đến việc tìm lại thông tin được gọi là ngôn ngữ vấn tin (query language).
- Gồm 2 loại: DML có thủ tục và không thủ tục.

Quản trị giao dịch (1)

- ◆ Giao dịch: là tập các thao tác trên CSDL tạo thành một đơn vị công việc logic.
- ◆ Ví dụ: các thao tác
 $R(A); A := A - x; W(A); R(B); B := B + x; W(B)$
→ Tạo thành một GD chuyển khoản $x(\text{đ})$ từ A sang B
- ◆ Khi một GD xảy ra, để đảm bảo không ảnh hưởng đến tính chính xác của dữ liệu trong CSDL, nó đòi hỏi GD phải thỏa 4 tính chất: nguyên tử (**A**tomicity), nhất quán (**C**onsistency), cô lập (**I**solate), bền vững (**D**urability).



Quản trị giao dịch (2)

◆ Ý nghĩa của các tính chất:

- Tính nguyên tử: đòi hỏi toàn thể hoặc không.
- Tính bền vững: đòi hỏi về sự tồn tại bền vững của DL.
- Tính cô lập: đòi hỏi các GD trong hệ thống không cạnh tranh dữ liệu lẫn nhau.
- Tính nhất quán: đòi hỏi tính chính xác.

◆ Việc đảm bảo các tính chất của GD:

- Tính nguyên tử + bền vững: bộ điều khiển phục hồi.
- Tính cô lập: bộ điều khiển cạnh tranh.
- Tính nhất quán: người sử dụng.

Quản trị giao dịch (3)

- ◆ Nếu không có **sự cố** thì tất cả giao dịch đều thành công và tính nguyên tử được bảo đảm.
- ◆ Nếu có sự cố, CSDL phải được hoàn lại trạng thái trước khi giao dịch xảy ra.
- ◆ Khi một số giao dịch **cạnh tranh** cập nhật CSDL:
 - Tính nhất quán có thể bị vi phạm
 - Bộ điều khiển cạnh tranh có nhiệm vụ điều khiển trao đổi giữa các giao dịch, bảo đảm tính thống nhất của CSDL

Các vai trò trong môi trường CSDL

◆ Nhà quản trị:

- Định nghĩa sơ đồ
- Định nghĩa cấu trúc lưu trữ và phương pháp truy xuất
- Sửa đổi sơ đồ và sửa đổi tổ chức vật lý
- Cấp quyền truy xuất CSDL
- Đặc tả ràng buộc toàn vẹn

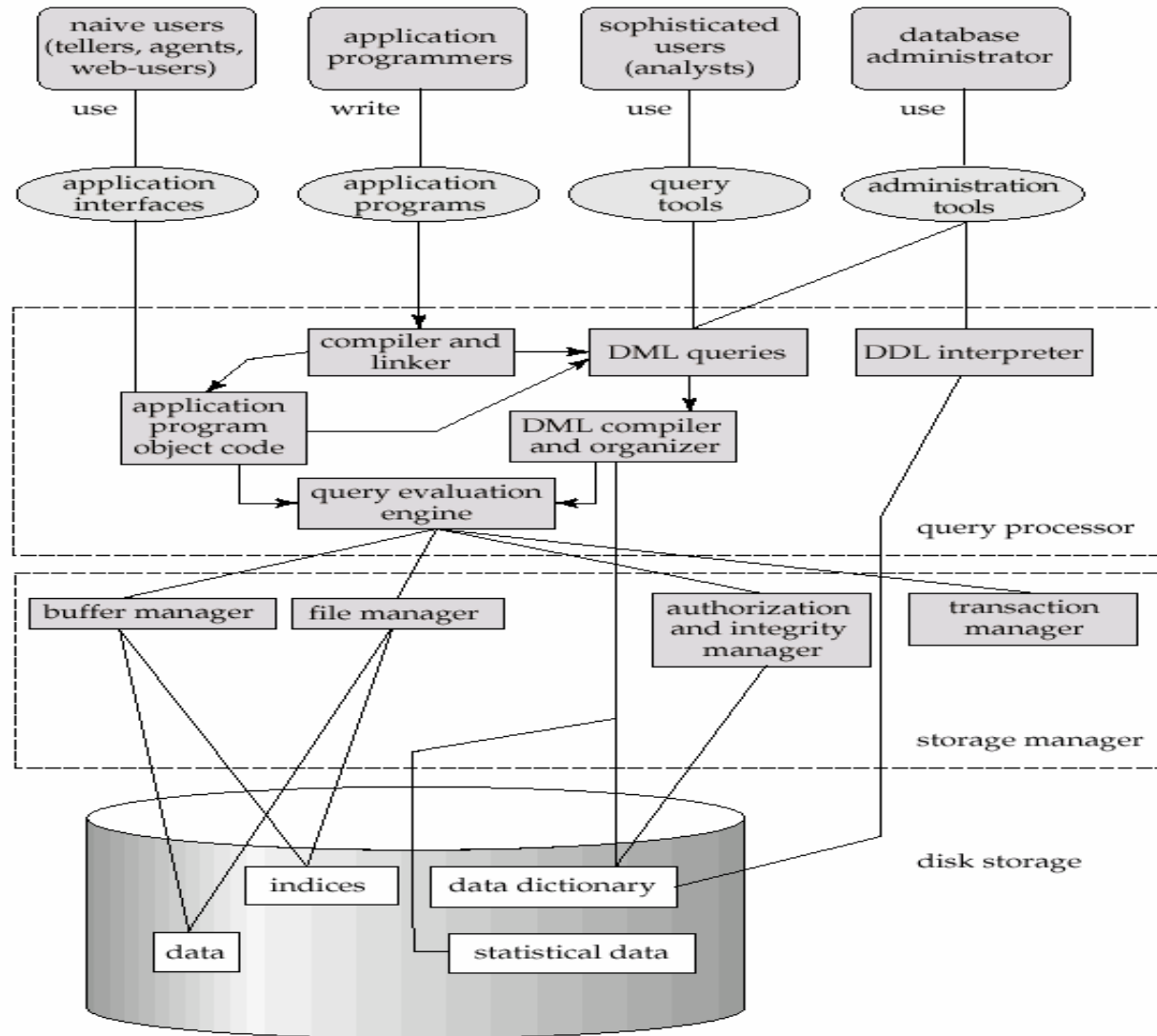
◆ Người lập trình ứng dụng

◆ Người sử dụng thành thạo (Sophisticated user)

◆ Người sử dụng chuyên biệt (Specialized user)

◆ Người sử dụng ngây thơ (Naive user)

Cấu trúc hệ thống tổng thể



Kiến trúc DBS

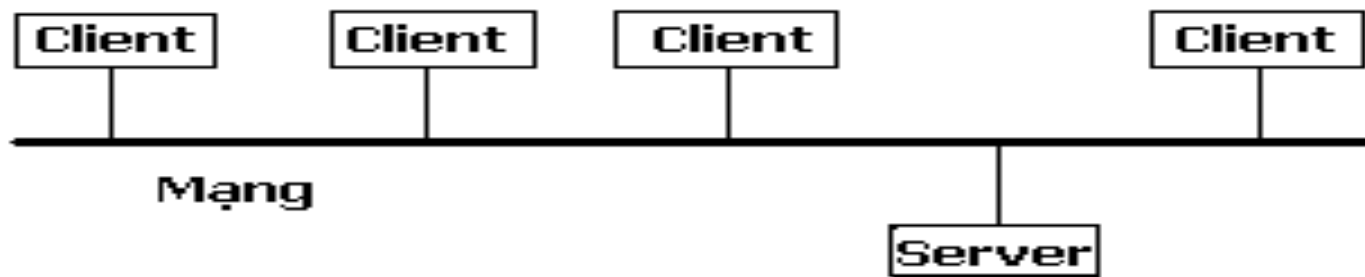
- ◆ Kiến trúc DBS bị ảnh hưởng nhiều bởi hệ thống máy nền:
 - Mạng máy tính: Sự phân chia công việc trên server và client dẫn đến việc hình thành DBS dạng Client-Server
 - Hệ thống song song: Giúp tăng tốc độ thực hiện các công việc nhờ vào cơ chế thực thi tác vụ song song. Sự cần thiết xử lý vấn đề song song dẫn đến sự phát triển DBS song song
 - Hệ phân tán: Lưu nhiều bản sao của CSDL trên nhiều site khác nhau. Dữ liệu thường trú tại nơi chúng được sinh ra, nhưng vẫn có thể truy xuất được chúng từ những site khác.

Kiến trúc DBMS - Hệ thống tập trung

- ◆ Là những hệ CSDL chạy trên máy đơn, tất cả các chức năng được tập trung tại một máy.
- ◆ Bao gồm hệ thống đơn người dùng và đa người dùng (nối kết thông qua các terminal).
- ◆ Trong các hệ đơn người dùng:
 - Không hỗ trợ các truy vấn song song.
 - Chức năng phục hồi hoặc là thiếu hoặc chỉ là sự sao chép dự phòng đơn giản.

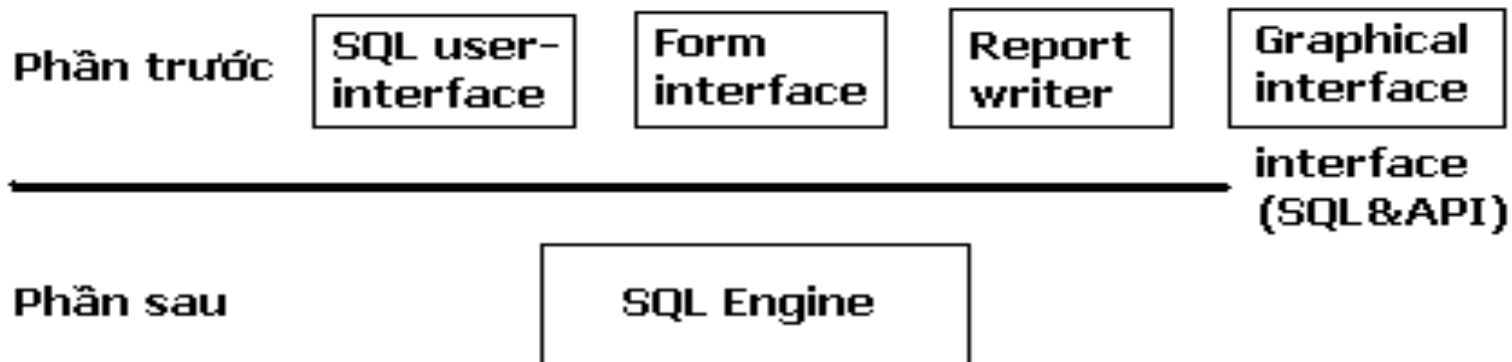
Kiến trúc DBMS - Hệ thống Client-Server (1)

- ◆ Các máy PC trở nên rẻ hơn, nhanh hơn, mạnh hơn.
- ◆ Chức năng của DBS được chia làm hai phần:
 - **Phần trước (front-end):** Bao gồm các công cụ như tạo form, soạn report, tạo giao diện đồ hoạ người dùng (**Client**)
 - **Phần sau (back-end):** quản trị truy xuất CSDL, định giá câu vấn tin và tối ưu hoá, điều khiển truy cập cạnh tranh và phục hồi sau lỗi (**Server**)



Kiến trúc DBMS - Hệ thống Client-Server (2)

- ◆ Giao diện của phần trước và phần sau thông qua SQL hoặc một giao diện lập trình ứng dụng.



- ◆ Một số giao diện thông dụng: ODBC, DAO, RDO, ADO, JDBC,...

Kiến trúc DBMS - Hệ thống Client-Server (2)

- ◆ Server giao dịch (server văn tin): cung cấp giao diện để các client có thể gửi đến các yêu cầu thực hiện một GD.
- ◆ Server dữ liệu: cho phép các Client đưa ra các yêu cầu đọc hoặc cập nhật DL theo đơn vị file hoặc trang.
 - Gửi trang đổi lại với gửi hạng mục
 - Khoá (lock)
 - Trữ dữ liệu (data caching)
 - Trữ chốt (lock caching)

Kiến trúc DBMS - Hệ thống song song (1)

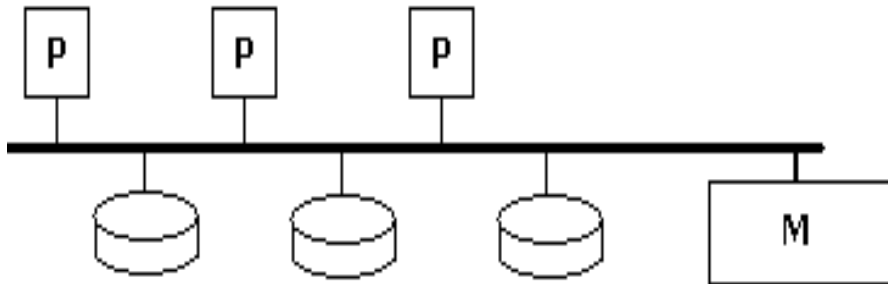
- ◆ Cải tiến tốc độ xử lý và tốc độ I/O bằng cách sử dụng nhiều CPU và đĩa song song
- ◆ Có hai biện pháp chính để đánh giá hiệu năng của một DBS:
 - Năng lực truyền qua (throughput): Số công việc được hoàn tất trong một thời gian đã cho.
 - Thời gian đáp ứng (response time): Lượng thời gian cần thiết để hoàn thành một công việc từ khi nó được đệ trình.
- ◆ Một hệ thống xử lý các giao dịch lớn có thể cải tiến thời gian đáp ứng và năng lực truyền qua bởi việc thực hiện song song các công việc con (subtask) của mỗi giao dịch

Kiến trúc DBMS - Hệ thống song song (2)

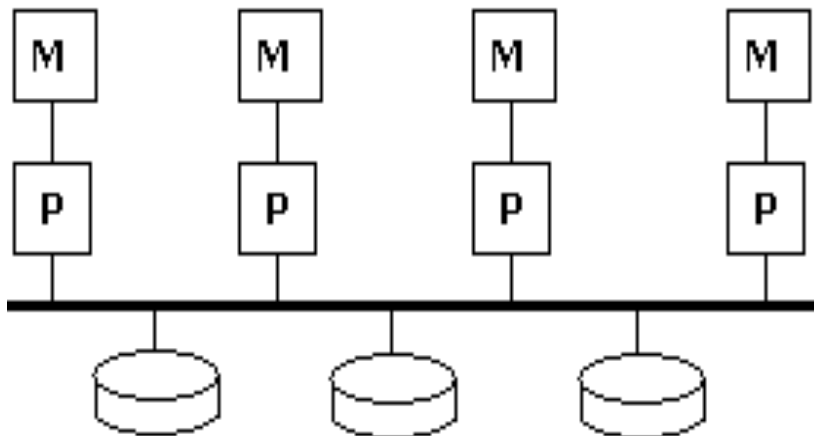
- ◆ Tăng tốc độ: thực hiện một công việc đã cho trong một khoảng thời gian ngắn hơn bằng cách tăng bậc song song. Công thức: T_s/T_L .
- ◆ Tăng quy mô: liên quan đến việc xử lý các công việc có quy mô lớn trong một khoảng thời gian ổn định. Công thức: T_s/T_L .
- ◆ Tăng quy mô là một độ đo quan trọng trong việc đo lường hiệu quả của các hệ CSDL song song.
- ◆ Có một số nhân tố ảnh hưởng đến tính hiệu quả của các hệ CSDL song song:
 - Chi phí khởi động
 - Sự giao thoa
 - Sự lệch.

Kiến trúc DBMS - Hệ thống song song (3)

◆ Bộ nhớ chia sẻ:



◆ Đĩa chia sẻ:



Ưu điểm:

- ◆ Tốc độ liên lạc giữa các P là hiệu quả

Hạn chế:

- ◆ Không thể hỗ trợ >64 P. Hoặc là bus dễ bị nghẽn
- ◆ Mỗi P thường có cache lớn để tránh tham khảo bộ nhớ chính khi có thể

Ưu điểm:

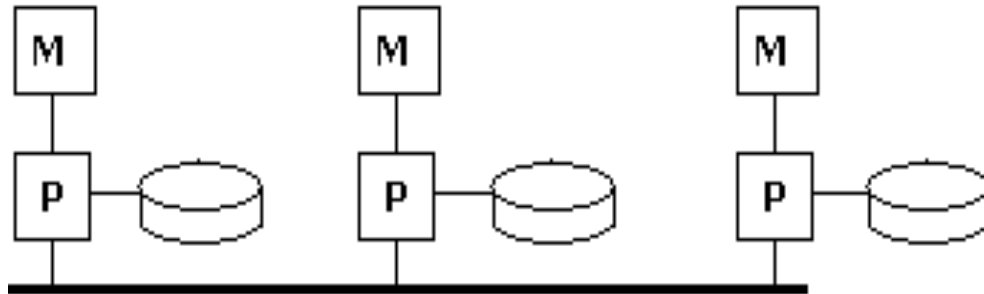
- ◆ Bus bộ nhớ không bị nghẽn
- ◆ Lượng thứ lỗi: khi một P bị hỏng, các P khác vẫn hoạt động bình thường

Hạn chế:

- ◆ Việc hợp nhất các hệ thống con các đĩa dễ bị nghẽn.
- ◆ Tốc độ liên lạc giữa các P chậm.

Kiến trúc DBMS - Hệ thống song song (4)

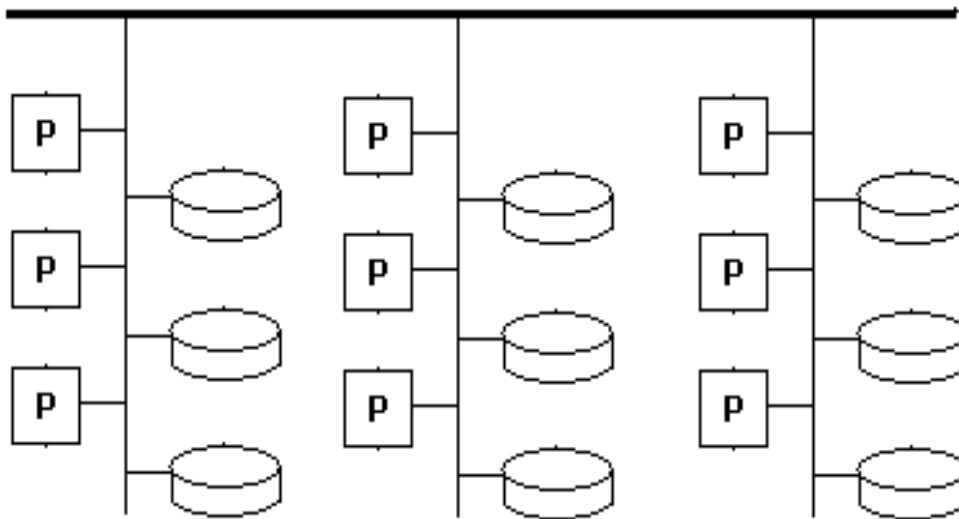
◆ Không chia sẻ:



Hạn chế

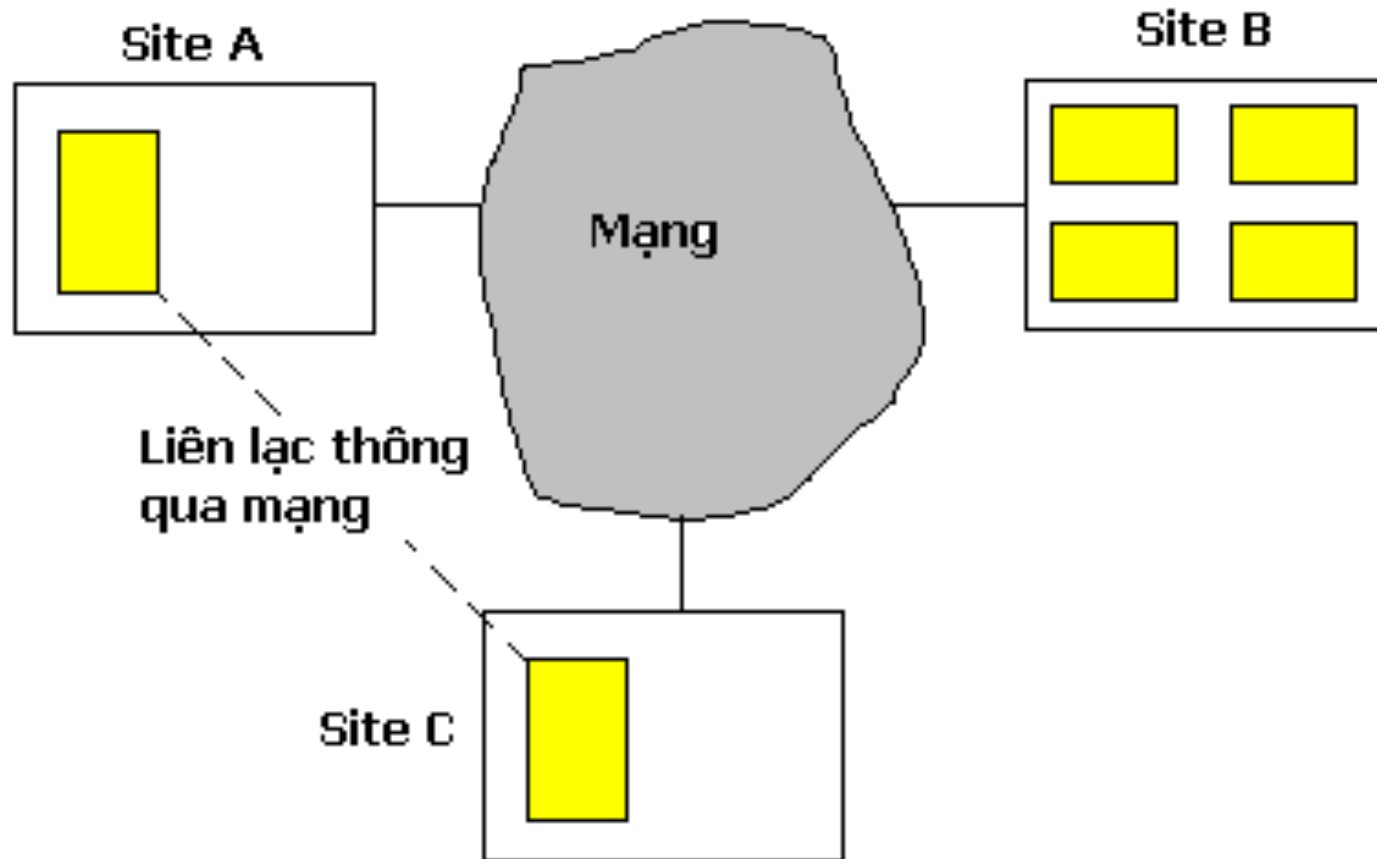
- ◆ Vấn đề phát sinh khi các CPU truy xuất các đĩa không cục bộ
- ◆ Việc truyền các quan hệ kết quả qua mạng

◆ Phân cấp:



Tổ hợp các đặc trưng của các kiến trúc chia sẻ bộ nhớ, chia sẻ đĩa và không chia sẻ gì

Kiến trúc DBMS - Hệ thống phân tán (1)



Kiến trúc DBMS - Hệ thống phân tán (2)

- ◆ Các máy tính trong hệ thống phân tán rất đa dạng về kích cỡ cũng như chức năng. Từ workstation đến mainframe
- ◆ Các máy tính liên lạc với nhau thông qua nhiều phương tiện liên lạc khác nhau: mạng tốc độ cao, đường điện thoại
- ◆ Sự sai khác của hệ thống phân tán và hệ thống song song không chia sẻ gì là:
 - CSDL phân tán được tách biệt về mặt địa lý, được quản trị tách biệt và có sự hợp nhất chậm
 - Giao dịch trong hệ phân tán được phân biệt thành giao dịch cục bộ và giao dịch toàn thể



Questions?