

# **CHƯƠNG 1.**

## **KHÁI NIỆM CƠ BẢN VỀ**

## **CƠ SỞ DỮ LIỆU**

(Phần 2)

# KHÁI NIỆM CƠ BẢN VỀ CSDL

- ❖ Các khái niệm cơ bản: Cơ sở dữ liệu, Hệ quản trị CSDL, Hệ cơ sở dữ liệu
- ❖ Các hệ CSDL truyền thống
- ❖ Các thành phần của một hệ quản trị CSDL
- ❖ Sự cần thiết của việc thiết kế CSDL
- ❖ Các vai trò trong môi trường CSDL
- ❖ Mô hình trừu tượng 3 lớp
- ❖ Các ngôn ngữ cơ sở dữ liệu
- ❖ Phân loại các hệ CSDL

## SỰ CẦN THIẾT CỦA VIỆC THIẾT KẾ CSDL

- ❖ Đối với các hệ thống phục vụ người dùng, hoạt động thiết kế CSDL là cần thiết.
- ❖ Một CSDL được thiết kế không cẩn thận có thể tạo ra nhiều lỗi, dẫn đến những quyết định không đúng đắn, làm ảnh hưởng nghiêm trọng đến tổ chức.
- ❖ Mặt khác, một CSDL được thiết kế tốt, theo một cách hiệu quả, sẽ cung cấp những thông tin chính xác hỗ trợ cho quá trình tạo quyết định đúng đắn, dẫn tới thành công.

## CÁC VAI TRÒ TRONG MÔI TRƯỜNG CSDL

- ❖ ***Người quản trị dữ liệu (DA – Data Administrator):*** có trách nhiệm quản lý tài nguyên dữ liệu, bao gồm lập kế hoạch cho CSDL, phát triển và duy trì các chuẩn, chính sách và thủ tục, và thiết kế CSDL mức khái niệm/logic.
- ❖ ***Người quản trị CSDL (DBA – Database Administrator):*** có trách nhiệm với việc lưu trữ vật lý CSDL, bao gồm thiết kế và cài đặt CSDL vật lý, kiểm soát bảo mật và toàn vẹn dữ liệu, duy trì hệ điều hành, và đảm bảo thỏa mãn hiệu năng cho các ứng dụng người dùng.

*=> Vai trò của DBA liên quan đến nhiều đặc tính kỹ thuật hơn DA.*

## CÁC VAI TRÒ TRONG MÔI TRƯỜNG CSDL (cont.)

- ❖ **Người thiết kế CSDL:** trong các dự án thiết kế CSDL lớn, cần phân biệt 2 loại thiết kế:
  - **Người thiết kế CSDL logic** liên quan tới việc xác định CSDL (các thực thể và thuộc tính), các mối quan hệ giữa dữ liệu, và các ràng buộc đối với dữ liệu sẽ được lưu trữ trong CSDL.
  - **Người thiết kế CSDL vật lý** phụ thuộc nhiều vào hệ quản trị CSDL đích, và có thể có nhiều cách để cài đặt CSDL. Người thiết kế CSDL vật lý phải có hiểu biết đầy đủ về tính năng của hệ quản trị CSDL đích.
- ❖ **Người phát triển ứng dụng:** Có trách nhiệm xây dựng chương trình ứng dụng cung cấp các chức năng cần thiết cho người dùng cuối, sau khi CSDL đã được cài đặt.

## CÁC VAI TRÒ TRONG MÔI TRƯỜNG CSDL (cont.)

- ❖ ***Người dùng cuối:*** là “khách hàng” của CSDL, và có thể được phân thành 2 nhóm dựa theo cách mà họ sử dụng hệ thống:
  - Nhóm người dùng không biết đến khái niệm CSDL hoặc hệ quản trị CSDL: Truy nhập CSDL thông qua chương trình ứng dụng được viết riêng biệt, giúp cho các thao tác của người dùng đơn giản nhất có thể.
  - Nhóm người dùng nhận biết được cấu trúc CSDL và các phương tiện được cung cấp bởi hệ quản trị CSDL: Thường dùng các ngôn ngữ truy vấn bậc cao như SQL để thực hiện những thao tác được yêu cầu và thậm chí có thể viết những chương trình ứng dụng để phục vụ cho mục đích riêng.

## ƯU ĐIỂM CỦA CÁC HỆ QUẢN TRỊ CSDL

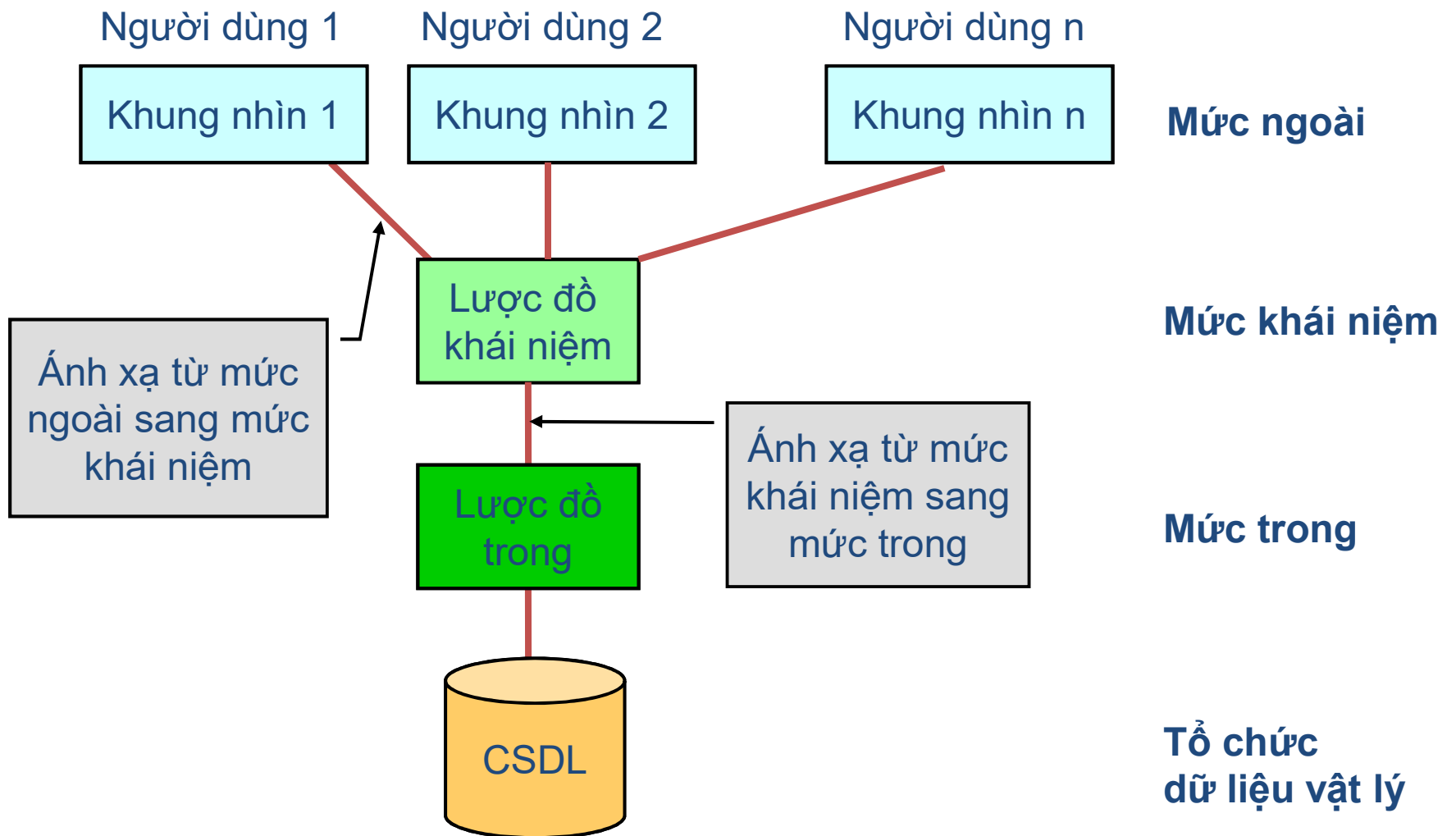
Kiểm soát dư thừa dữ liệu	Kinh tế khi tăng số lượng dữ liệu
Đảm bảo tính nhất quán dữ liệu	Cân bằng những yêu cầu bị xung đột
Thêm thông tin từ cùng dữ liệu	Cải tiến việc truy nhập dữ liệu
Hỗ trợ tính sẵn dùng của khối lượng lớn dữ liệu	Tăng hiệu suất của hệ thống
Chia sẻ dữ liệu	Cải tiến việc bảo trì
Cải tiến sự toàn vẹn dữ liệu	Tăng xử lý đồng thời
Cải tiến tính bảo mật dữ liệu	Cải tiến sao lưu và khôi phục dữ liệu
Ép buộc chuẩn hóa dữ liệu	Cải tiến đáp ứng truy vấn

## NHƯỢC ĐIỂM CỦA CÁC HỆ QUẢN TRỊ CSDL

Phức tạp
Kích thước lớn
Chi phí mua và bảo trì
Thêm giá thành cho các phần cứng hỗ trợ
Chi phí chuyển đổi hệ thống
Hạn chế hiệu năng (trong một số trường hợp cụ thể)
Ảnh hưởng lớn khi có lỗi



# MÔ HÌNH TRỪU TƯỢNG 3 LỚP



# MỨC NGOÀI

- ❖ **Mức ngoài** là khung nhìn của người sử dụng CSDL, mô tả phần CSDL tương ứng với người dùng đó.
- ❖ Mỗi người dùng có một khung nhìn khác nhau về thế giới thực, được biểu diễn theo cách thân thiện với họ trong CSDL.  
**=> Mức ngoài bao gồm một số khung nhìn khác nhau về CSDL.**
- ❖ Mức ngoài chỉ bao gồm các thực thể, thuộc tính, quan hệ trong thế giới thực mà người dùng quan tâm. (Những thực thể, thuộc tính, quan hệ khác có thể tồn tại, nhưng người dùng sẽ không cần biết đến sự tồn tại của chúng.)

## MỨC NGOÀI (cont.)

- ❖ Thường xảy ra trường hợp các khung nhìn ngoài khác nhau sẽ có những biểu diễn khác nhau đối với cùng dữ liệu.

**Ví dụ:** Một khung nhìn có thể thể hiện ngày tháng dưới dạng (ngày, tháng, năm), trong khi một khung nhìn khác lại thể hiện dưới dạng (tháng, ngày, năm).

- ❖ Một số khung nhìn có thể bao gồm các dữ liệu dẫn xuất. Các dữ liệu này sẽ không được lưu thực trong CSDL mà chúng sẽ được tạo ra khi cần.

**Ví dụ:** Một khung nhìn cần biết tuổi của một người. Tuy nhiên, dữ liệu về tuổi không cần thiết phải được lưu trong CSDL vì nó được cập nhật hàng ngày: Tuổi sẽ được tính dựa theo dữ liệu ngày sinh của người đó và ngày hiện tại trong hệ thống.

## MỨC KHÁI NIỆM

- ❖ **Mức khái niệm** là khung nhìn của người thiết kế CSDL, mô tả dữ liệu nào được lưu trong CSDL và mối quan hệ giữa chúng.
- ❖ Người quản trị CSDL nhìn thấy toàn bộ cấu trúc logic của CSDL. Cấu trúc này thể hiện khung nhìn hoàn chỉnh về những yêu cầu dữ liệu của tổ chức mà không liên quan tới bất kỳ phương thức lưu trữ nào.

## MỨC KHÁI NIỆM (cont.)

- ❖ Mức khái niệm hỗ trợ từng khung nhìn ngoài, trong đó bất kỳ dữ liệu nào chuyển tới người dùng đều phải được lưu lại hoặc được sinh ra từ mức khái niệm.
- ❖ Mức khái niệm không liên quan tới bất kỳ thông tin nào về việc lưu trữ dữ liệu.

**Ví dụ:** Một thực thể được định nghĩa ở dạng số nguyên tại mức khái niệm, nhưng số byte chứa số nguyên đó sẽ không được chỉ ra ở đây.

## MỨC TRONG

- ❖ **Mức trong** thể hiện biểu diễn về mặt vật lý của CSDL trong máy tính, mô tả cách thức lưu trữ dữ liệu trong CSDL.
- ❖ Mô tả cài đặt vật lý cần thiết để đạt được tối ưu về thời gian thực thi và việc sử dụng không gian lưu trữ.
- ❖ Bao gồm các cấu trúc dữ liệu và tổ chức tệp lưu trữ dữ liệu trong các thiết bị nhớ.
- ❖ Có giao diện với các phương thức truy nhập của hệ điều hành (các kỹ thuật quản lý tệp để lưu trữ và lấy các bản ghi dữ liệu) để đưa dữ liệu vào các thiết bị nhớ, xây dựng chỉ mục, lấy dữ liệu, ...

# MỨC VẬT LÝ

- ❖ **Mức vật lý** nằm dưới mức trong, được quản lý bởi hệ điều hành dưới chỉ dẫn của hệ quản trị CSDL.
- ❖ Chức năng của hệ quản trị CSDL và hệ điều hành tại mức vật lý là không có ranh giới rõ ràng và thay đổi từ hệ thống này sang hệ thống khác.
- ❖ Một số hệ quản trị CSDL tận dụng ưu điểm của nhiều phương thức truy nhập của hệ điều hành, trong khi một số hệ quản trị CSDL khác lại chỉ sử dụng những phương thức cơ bản và tự tạo ra tổ chức tệp của riêng chúng.
- ❖ Mức vật lý dưới hệ quản trị CSDL gồm các mục chỉ được biết đến bởi hệ điều hành. **Ví dụ:** làm thế nào để tạo một chuỗi các thực thi? Liệu các trường của các bản ghi trong CSDL có được lưu trữ bởi các byte liền nhau trong ổ đĩa hay không?

# LƯỢC ĐỒ, ÁNH XẠ VÀ THỂ HIỆN

❖ **Lược đồ CSDL** là thiết kế tổng thể của CSDL.

Có 3 loại lược đồ khác nhau được định nghĩa theo các mức trừu tượng của mô hình 3 lớp:

- Trên cùng có rất nhiều **lược đồ ngoài** (hay gọi là **lược đồ con**) tương ứng với các khung nhìn dữ liệu khác nhau.
- Tại mức khái niệm có một **lược đồ khái niệm**, mô tả tất cả các thực thể, thuộc tính và các quan hệ cùng với những ràng buộc toàn vẹn.
- Mức trừu tượng thấp nhất có một **lược đồ trong**, mô tả toàn bộ mô hình trong, gồm định nghĩa các bản ghi được lưu trữ, phương thức biểu diễn, ...



## LƯỢC ĐỒ, ÁNH XẠ VÀ THỂ HIỆN (cont.)

*=> Hệ quản trị CSDL có trách nhiệm ánh xạ các loại lược đồ với nhau.*

Hệ quản trị CSDL phải kiểm tra được tính nhất quán của các lược đồ; nghĩa là phải kiểm tra xem mỗi lược đồ ngoài có được suy ra từ lược đồ khái niệm hay không, và sử dụng các thông tin trong lược đồ khái niệm để ánh xạ giữa các lược đồ ngoài và lược đồ trong.

## LƯỢC ĐỒ, ÁNH XẠ VÀ THỂ HIỆN (cont.)

- ❖ Lược đồ khái niệm liên kết với lược đồ trong thông qua **ánh xạ mức khái niệm/mức trong**.

=> Ánh xạ này giúp cho hệ quản trị CSDL tìm ra được bản ghi thực tế hoặc kết nối các bản ghi trong bộ lưu trữ vật lý để tạo bản ghi logic trong lược đồ khái niệm, cùng với một số ràng buộc được gắn với các thao tác cho bản ghi logic đó.

- ❖ Lược đồ ngoài liên kết với lược đồ khái niệm thông qua **ánh xạ mức ngoài/mức khái niệm**.

=> Ánh xạ này giúp cho hệ quản trị CSDL ánh xạ các tên trong khung nhìn của người dùng thành các phần có liên quan trong lược đồ khái niệm.

## LƯỢC ĐỒ, ÁNH XẠ VÀ THỂ HIỆN (cont.)

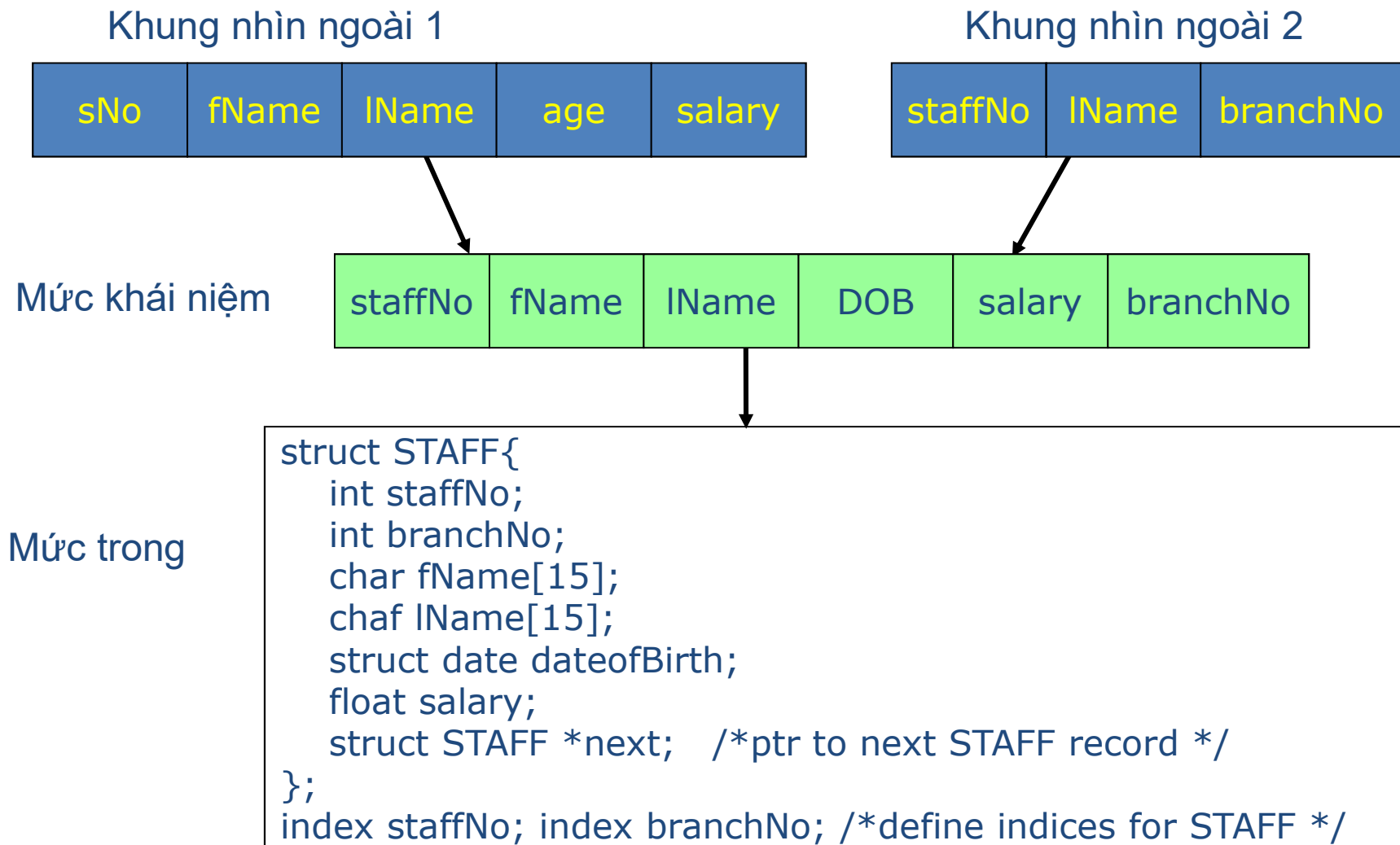
- ❖ Các cơ sở dữ liệu thay đổi theo thời gian khi thông tin được thêm vào hay bị xóa đi.

*=> Tập hợp các thông tin được lưu trữ trong CSDL tại một thời điểm đặc biệt được gọi là một **thể hiện** của cơ sở dữ liệu đó.*

**Ví dụ:** Giá trị của một bản ghi tại một thời điểm được gọi là một thể hiện của CSDL.

# VÍ DỤ CHUYỂN ĐỔI LỢC ĐỒ

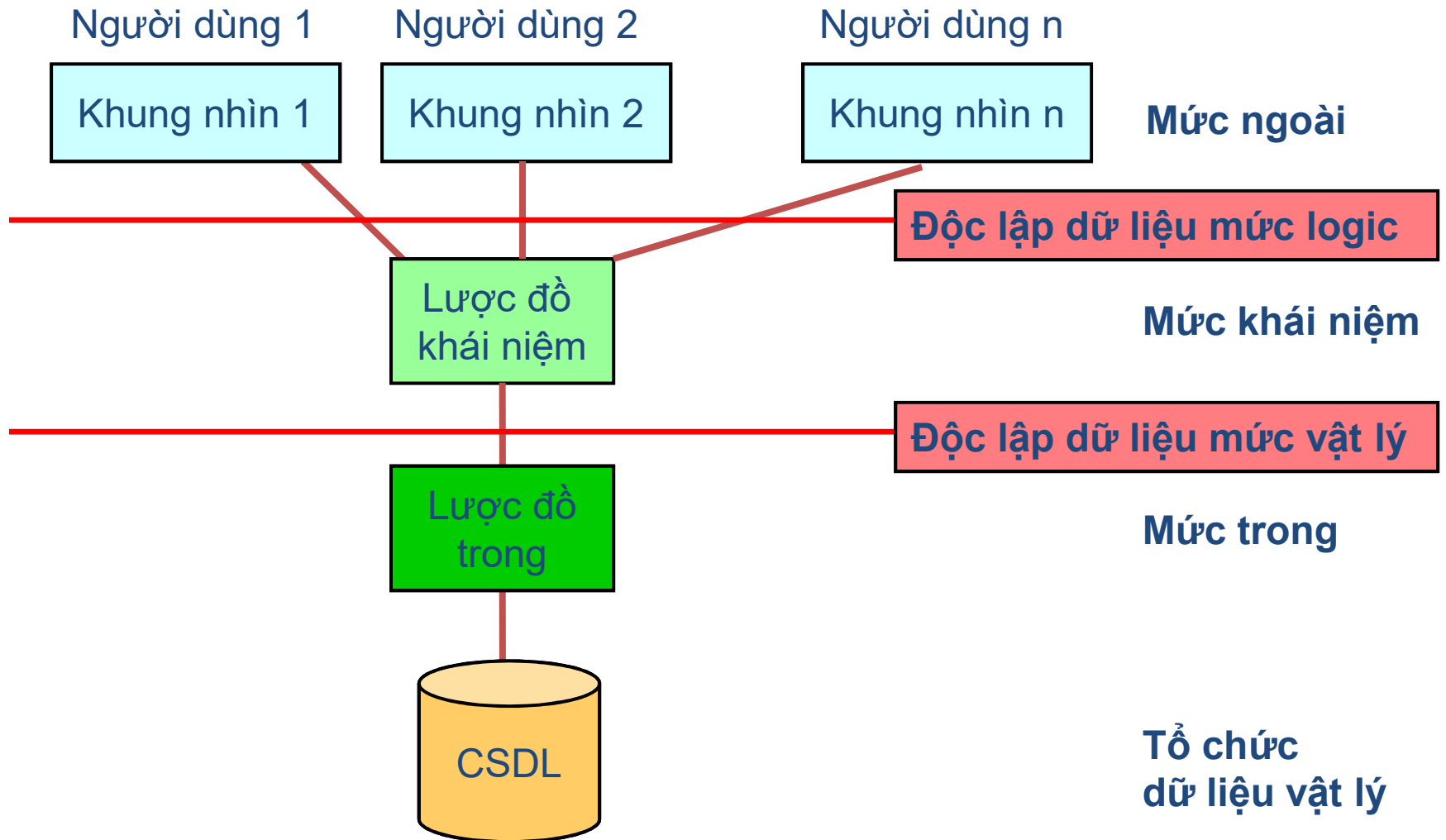
## GIỮA 3 MỨC TRỪU TƯỢNG



# TÍNH ĐỘC LẬP DỮ LIỆU

- ❖ Một trong những mục tiêu chính của kiến trúc 3 lớp là cung cấp **tính độc lập dữ liệu**, nghĩa là các mức cao hơn không bị ảnh hưởng bởi bất kỳ sự thay đổi nào từ các mức thấp hơn.
- ❖ Có 2 loại độc lập dữ liệu:
  - **Độc lập dữ liệu mức logic (mức khái niệm):** các lược đồ ngoài không bị ảnh hưởng bởi sự thay đổi của lược đồ khái niệm.
  - **Độc lập dữ liệu mức vật lý:** lược đồ khái niệm không bị ảnh hưởng bởi sự thay đổi của lược đồ trong.

# TÍNH ĐỘC LẬP DỮ LIỆU (cont.)



# CÁC NGÔN NGỮ CSDL

- ❖ Một ngôn ngữ con dữ liệu bao gồm 2 phần: Một **ngôn ngữ định nghĩa dữ liệu** (DDL - Data Definition Language) và một **ngôn ngữ thao tác dữ liệu** (DML - Data Manipulation Language).
- ❖ DDL được dùng để xác định lược đồ CSDL, và DML dùng để đọc và cập nhật CSDL.
- ❖ Các ngôn ngữ này được gọi là *ngôn ngữ con dữ liệu* vì chúng không bao gồm các cấu trúc lập trình cần thiết cho việc tính toán như là các cấu trúc điều khiển hoặc câu lệnh lặp (được cung cấp bởi các ngôn ngữ lập trình bậc cao).

## CÁC NGÔN NGỮ CSDL (cont.)

- ❖ Hầu hết các hệ quản trị CSDL đều có một môi trường cho phép nhúng các ngôn ngữ con dữ liệu vào trong một ngôn ngữ lập trình bậc cao như COBOL, Pascal, C, C++, Java hay Visual Basic (được gọi là các *ngôn ngữ chủ*).
- ❖ Nhiều ngôn ngữ con dữ liệu cũng cung cấp một phiên bản tương tác hoặc không nhúng vào ngôn ngữ mà có thể được đưa vào trực tiếp từ một thiết bị đầu cuối.



## NGÔN NGỮ ĐỊNH NGHĨA DỮ LIỆU (DDL)

- ❖ DDL là ngôn ngữ cho phép người quản trị CSDL hoặc người dùng mô tả và đặt tên các thực thể, thuộc tính và các quan hệ cần thiết cho ứng dụng, cùng với những ràng buộc về bảo mật và toàn vẹn liên quan.
- ❖ Kết quả của việc thực thi/biên dịch câu lệnh DDL là một tập các bảng được lưu trong các tệp đặc biệt, được gọi là **danh mục hệ thống (system catalog)** (hay **từ điển dữ liệu/ thư mục dữ liệu**).

## NGÔN NGỮ THAO TÁC DỮ LIỆU (DML)

- ❖ DML là ngôn ngữ cung cấp một tập các thao tác hỗ trợ cho các phép toán thao tác dữ liệu cơ bản trên dữ liệu được lưu trong CSDL.
- ❖ Các thao tác của DML bao gồm:
  - Chèn dữ liệu mới vào CSDL
  - Sửa đổi dữ liệu đã được lưu trữ trong CSDL
  - Lấy dữ liệu từ CSDL
  - Xóa dữ liệu trong CSDL
- ❖ Phần thao tác lấy dữ liệu ra được gọi là **ngôn ngữ truy vấn**.

## NGÔN NGỮ THAO TÁC DỮ LIỆU (cont.)

- ❖ Các DML được phân biệt bởi cấu trúc lấy dữ liệu bên trong nó, gồm hai loại chính: **có thủ tục** và **không thủ tục**.
  - **DMLs có thủ tục** là các ngôn ngữ trong đó người dùng có thông báo với hệ thống những dữ liệu nào cần và cách thức chính xác để lấy nó ra.
  - **DMLs không có thủ tục** là các ngôn ngữ trong đó người dùng chỉ thông báo cho hệ thống dữ liệu nào cần và hệ thống sẽ tự xác định cách thức lấy dữ liệu đó ra.
- ❖ DMLs có thủ tục thường được nhúng vào các ngôn ngữ lập trình bậc cao.
- ❖ DMLs có thủ tục có xu hướng tập trung vào từng bản ghi đơn còn DMLs không thủ tục có xu hướng thực hiện trên một tập các bản ghi.

## NGÔN NGỮ THỂ HỆ THỨ TƯ

- ❖ *Ngôn ngữ thể hệ thứ tư* là một ngôn ngữ lập trình rất nhanh. Những yêu cầu được thực hiện với hàng trăm dòng lệnh trong ngôn ngữ thể hệ thứ ba sẽ được thể hiện chỉ trong một vài dòng lệnh của ngôn ngữ thể hệ thứ tư.
- ❖ Ngôn ngữ thể hệ thứ ba là loại có thủ tục, còn ngôn ngữ thể hệ thứ tư là loại không có thủ tục.
- ❖ Ngôn ngữ thể hệ thứ tư gồm các ngôn ngữ làm việc trên bảng tính và trên cơ sở dữ liệu.

*Ví dụ:* Ngôn ngữ SQL.

# PHÂN LOẠI CÁC HỆ CSDL

## *Có 2 loại kiến trúc CSDL:*

- ❖ Hệ cơ sở dữ liệu tập trung:
  - Hệ CSDL cá nhân
  - Hệ CSDL trung tâm
  - Hệ CSDL khách/chủ (client/server)
- ❖ Hệ cơ sở dữ liệu phân tán
  - Hệ CSDL phân tán thuần nhất
  - Hệ CSDL phân tán không thuần nhất

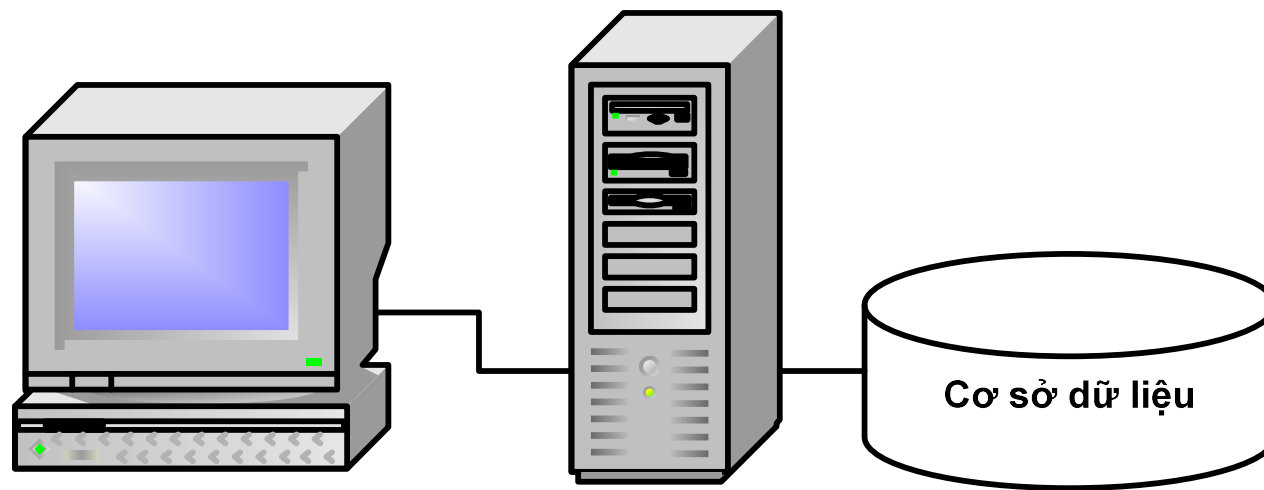
## HỆ CSDL TẬP TRUNG

- ❖ Là các hệ CSDL mà trong đó phần CSDL được lưu trữ tại một vị trí nhất định.
- ❖ Người dùng tại các trạm từ xa nói chung có thể truy nhập CSDL thông qua các công cụ truyền thông dữ liệu.
- ❖ Cung cấp một sự kiểm soát lớn đối với việc truy nhập và cập nhật dữ liệu.
- ❖ Dễ bị lỗi do phụ thuộc vào tính sẵn sàng của các tài nguyên.

## HỆ CSDL CÁ NHÂN

- ❖ Là hệ CSDL nhỏ, trong đó người quản trị CSDL chính là người viết chương trình ứng dụng, đồng thời cũng là người dùng cuối.
- ❖ Ứng dụng: trong các tổ chức nhỏ và vừa, ví dụ: quản lý nhân sự ở một đơn vị hành chính.
- ❖ Việc phát triển và sử dụng các hệ CSDL cá nhân là khá đơn giản và dễ dàng.
- ❖ Có nguy cơ phải chịu rủi ro, vì CSDL chỉ được lưu trữ tại một trạm đơn lẻ.
- ❖ Dữ liệu khó được chia sẻ cho nhiều ứng dụng khác nhau.

## HỆ CSDL CÁ NHÂN (Cont.)



*Hệ cơ sở dữ liệu cá nhân*

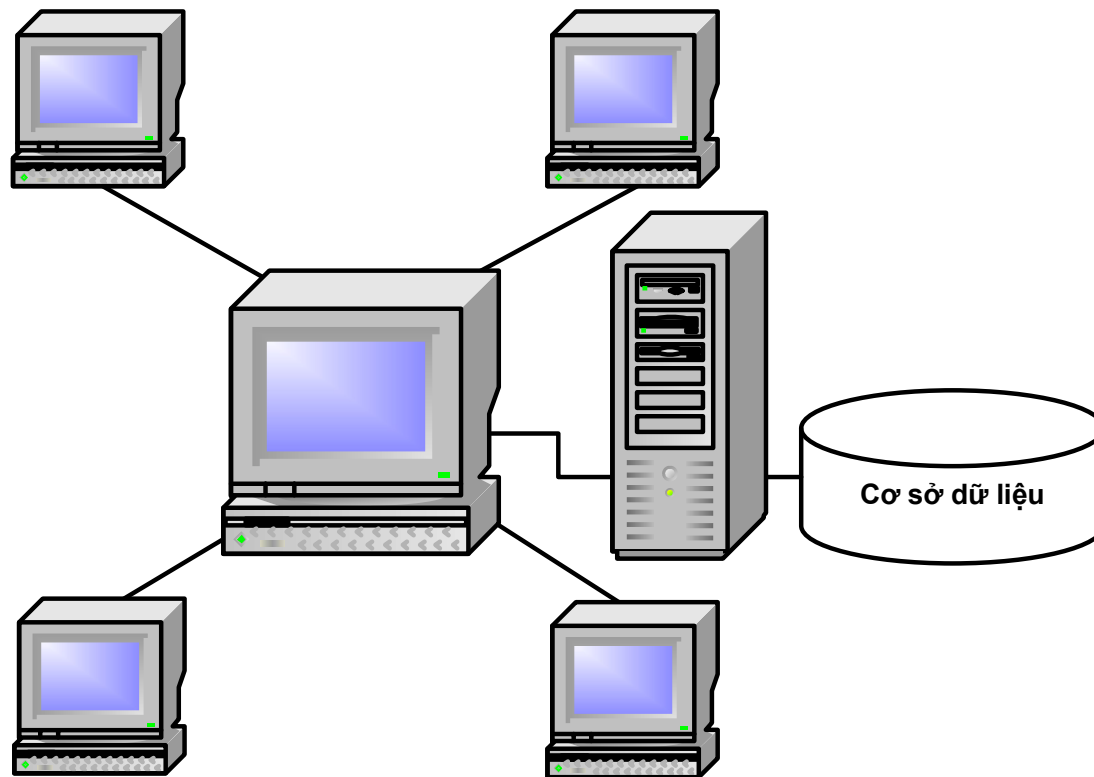


# HỆ CSDL TRUNG TÂM

- ❖ Trong các tổ chức lớn, dữ liệu mà hầu hết các ứng dụng có thể truy nhập được lưu trữ trên một máy tính trung tâm.
- ❖ Người dùng từ xa có thể truy nhập CSDL này thông qua các thiết bị đầu cuối và các kết nối truyền thông dữ liệu.
- ❖ CSDL trung tâm thường lưu trữ các CSDL tích hợp rất lớn và được nhiều người dùng truy nhập.
- ❖ Việc sử dụng thường có cường độ lớn với hàng trăm đến hàng nghìn giao dịch trong một giây.

**Ví dụ:** hệ thống đặt vé máy bay, hoặc các hệ thống ngân hàng.

## HỆ CSDL TRUNG TÂM (cont.)

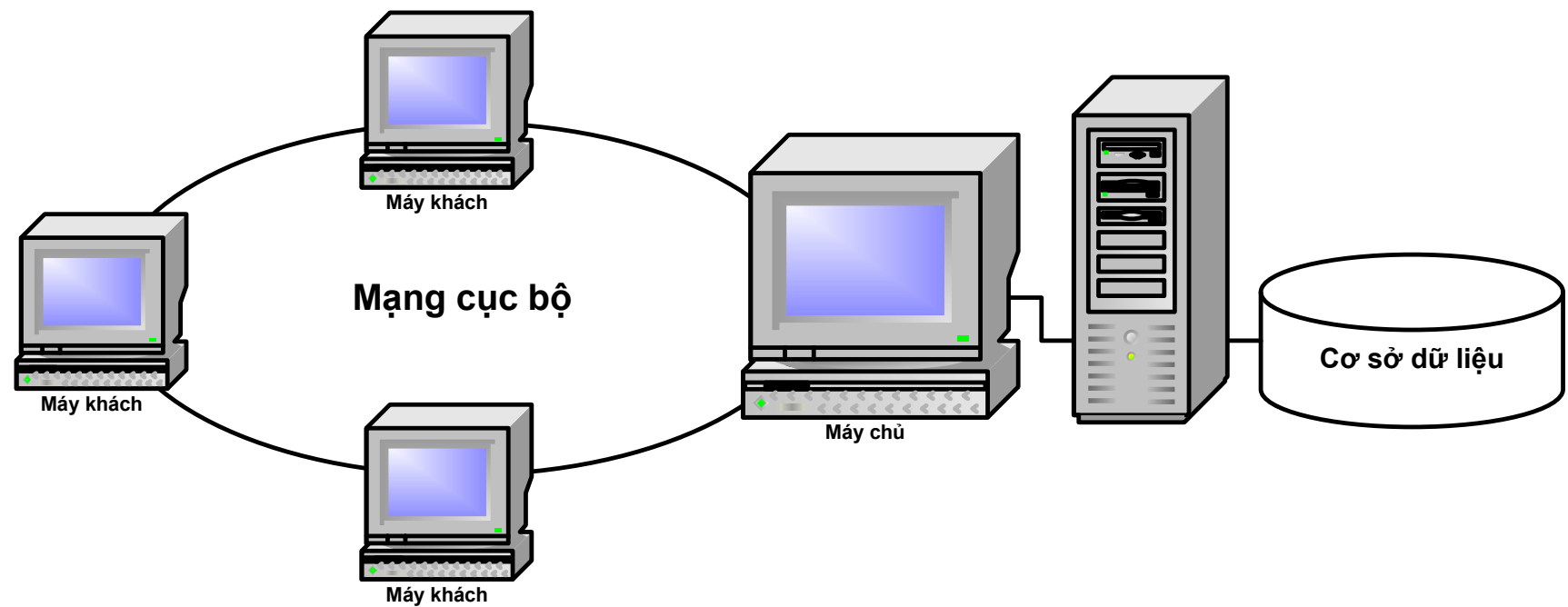


*Hệ cơ sở dữ liệu trung tâm*

## HỆ CSDL KHÁCH/CHỦ

- ❖ Mục đích chính của kiến trúc khách/chủ là cho phép các ứng dụng máy khách truy nhập dữ liệu được quản lý bởi máy chủ. Giao diện người dùng và logic của chương trình ứng dụng được xử lý bên máy khách, trong khi xử lý CSDL được thực hiện bên máy chủ.
- ❖ Máy chủ không cần có cấu hình quá mạnh như trong các hệ CSDL trung tâm.

## HỆ CSDL KHÁCH/CHỦ (cont.)



*Hệ cơ sở dữ liệu khách/chủ*

## HỆ CSDL PHÂN TÁN

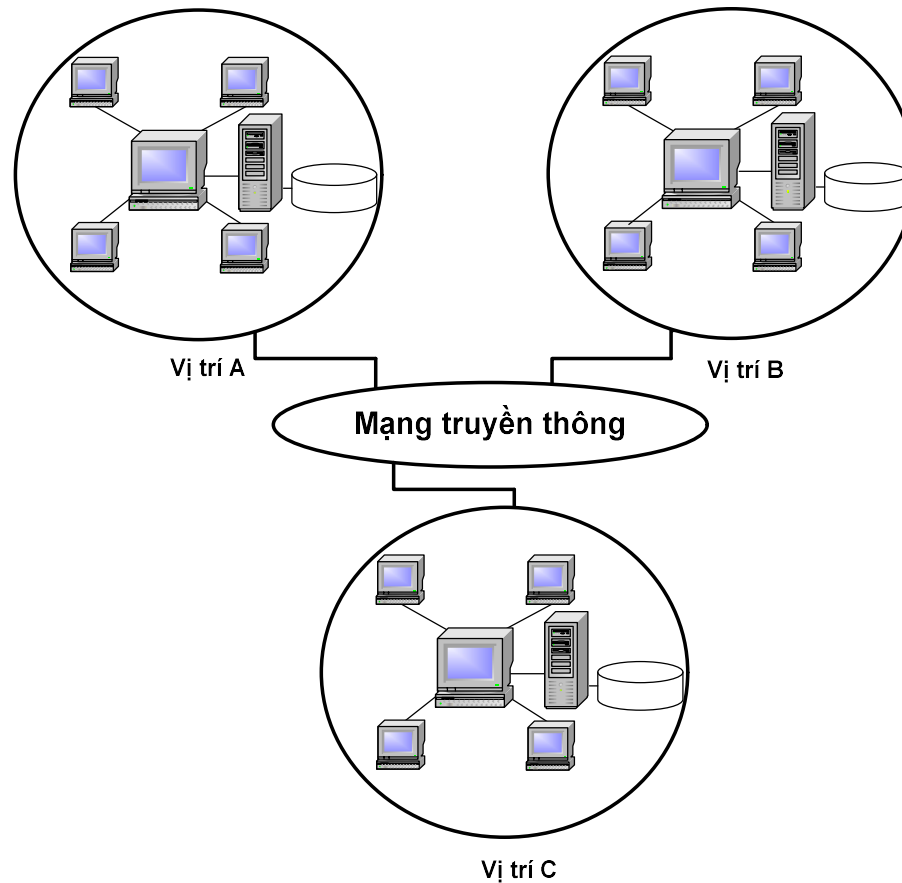
- ❖ Hiện nay, nhiều tổ chức phân bố trên nhiều vị trí địa lý khác nhau (các thành phố hay các quốc gia khác nhau).

=> Việc xây dựng các hệ CSDL tập trung là không thực tế và không kinh tế.

- ❖ CSDL phân tán là một CSDL logic đơn lẻ được trải ra về mặt vật lý trên nhiều máy tính ở nhiều vị trí địa lý khác nhau.

**Ví dụ:** CSDL của một ngân hàng sẽ được phân bố theo các chi nhánh tại từng địa phương.

# HỆ CSDL PHÂN TÁN (cont.)



*Hệ cơ sở dữ liệu phân tán*

## HỆ CSDL PHÂN TÁN THUẦN NHẤT

- ❖ Các hệ điều hành máy tính tại mỗi vị trí địa lý là như nhau hoặc có khả năng tương thích cao.
- ❖ Các mô hình dữ liệu được sử dụng tại mỗi vị trí địa lý là như nhau. Mô hình quan hệ được sử dụng chung nhất đối với các hệ CSDL phân tán ngày nay.
- ❖ Các hệ quản trị CSDL được sử dụng tại mỗi vị trí địa lý là như nhau hoặc có khả năng tương thích cao.
- ❖ Dữ liệu tại các vị trí khác nhau có định nghĩa và khuôn dạng chung.

## HỆ CSDL PHÂN TÁN KHÔNG THUẦN NHẤT

- ❖ Các máy tính khác nhau và các hệ điều hành khác nhau có thể được sử dụng tại các vị trí địa lý khác nhau.
- ❖ Các mô hình dữ liệu khác nhau và các hệ quản trị CSDL khác nhau cũng có thể được sử dụng.
- ❖ Dữ liệu trên các vị trí thường không tương thích. Có khác biệt về cú pháp và ngữ nghĩa.

*=> Khi có nhu cầu chia sẻ dữ liệu, giải pháp là phát triển một hệ CSDL mới hoàn toàn, hợp nhất tất cả các hệ CSDL đang tồn tại. Tuy nhiên, giải pháp này rất khó khăn về mặt kỹ thuật và kinh tế.*