

**HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**

---

# **CƠ SỞ DỮ LIỆU PHÂN TÁN**

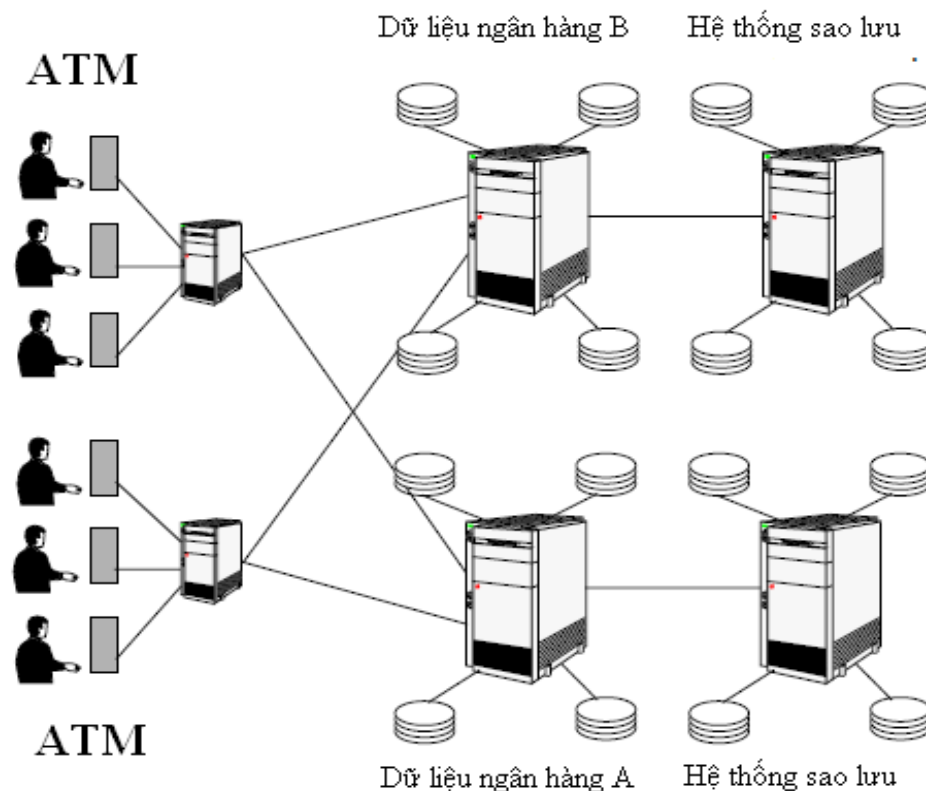
## **TỔNG QUAN CƠ SỞ DỮ LIỆU PHÂN TÁN**

# *Nội dung*

---

- ❑ Xử lý dữ liệu phân tán.
- ❑ Hệ cơ sở dữ liệu phân tán là gì.
- ❑ Khả năng của các hệ cơ sở dữ liệu phân tán.
- ❑ Các đặc điểm của cơ sở dữ liệu phân tán
- ❑ Các mô hình xử lý dữ liệu phân tán
- ❑ Kiến trúc hệ quản trị cơ sở dữ liệu phân tán.
- ❑ Một số vấn đề căn bản khi nghiên cứu cơ sở dữ liệu phân tán

Một **hệ phân tán** là một tập các thành phần tự trị hoạt động **cộng tác** với nhau thông qua **mạng truyền thông** và được thể hiện tới người dùng/ứng dụng như là một hệ thống **đơn lẻ**.



# Hệ thống phân tán

---

- ❑ Có hai khái niệm xử lý phân tán.
  - Mô hình tính toán Client/Server. Client truy nhập trực tiếp và xử lý dữ liệu trên Server.
  - Một giao dịch được chia ra xử lý trên nhiều hệ thống. Máy trung tâm sẽ giám sát và quản lý các tiến trình.
  - (per to per)

# *Hệ thống xử lý phân tán*

---

## ❑ *Các đối tượng phân tán bao gồm*

- phân tán chức năng: Các chức năng phân tán.
- Phân tán dữ liệu: Dữ liệu được phân tán trên nhiều vị trí khác nhau
- Điều khiển phân tán: Điều khiển các giao dịch khác nhau được phân tán trên nhiều vị trí

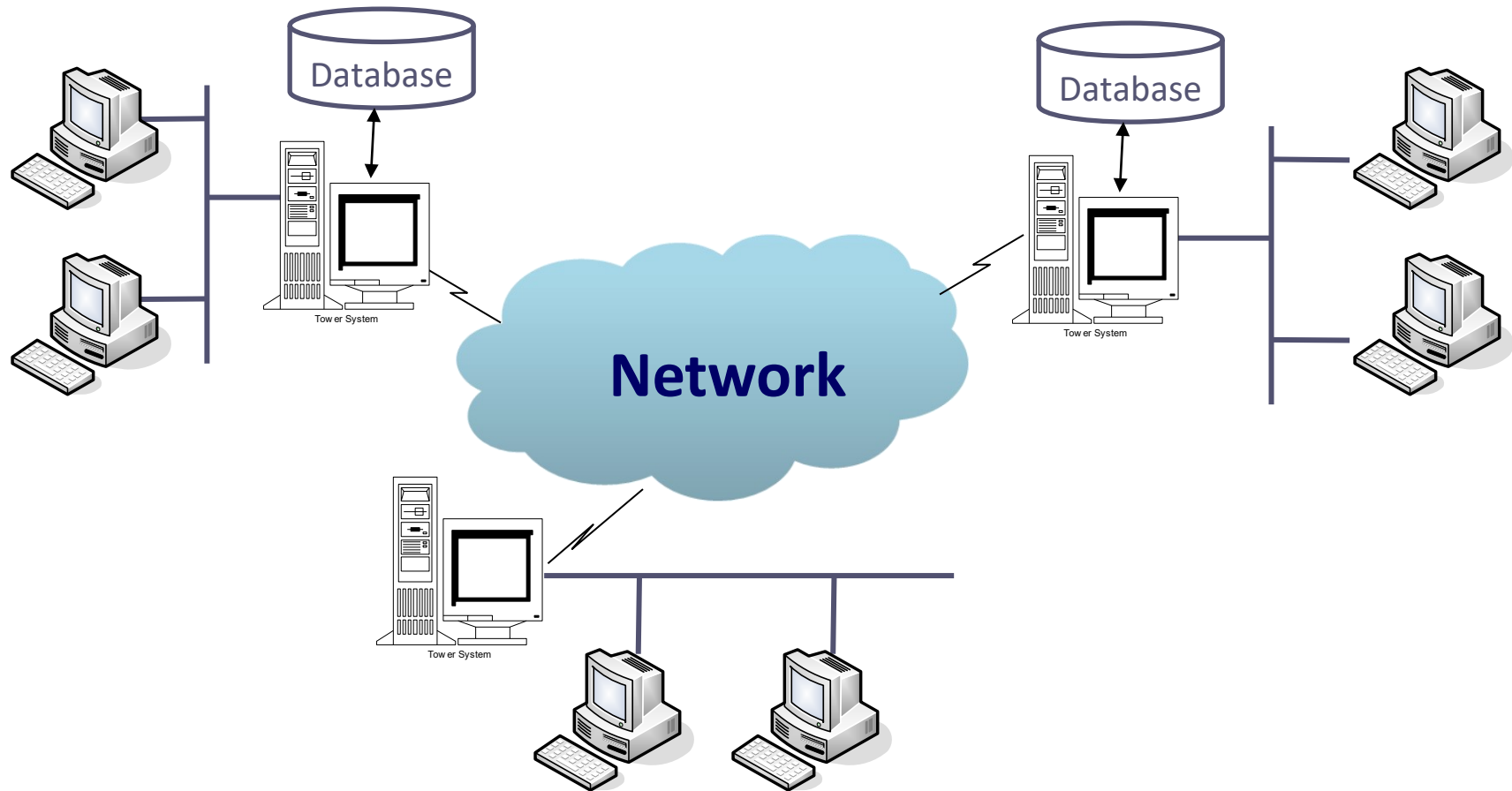
# ***Xử lý phân tán và các hệ thống xử lý phân tán***

---

*Hệ thống xử lý phân tán đáp ứng nhu cầu thu thập, lưu trữ, xử lý và trao đổi thông tin của các tổ chức kinh tế, xã hội có các hoạt động trên phạm vi rộng lớn. Yêu cầu có độ tin cậy cao, đáp ứng nhanh. Hệ cơ sở dữ liệu phân tán là nhân tố quan trọng, làm cho quá trình xử lý phân tán dễ dàng và có hiệu quả.*

# *Xử lý phân tán và các hệ thống xử lý phân tán*

---



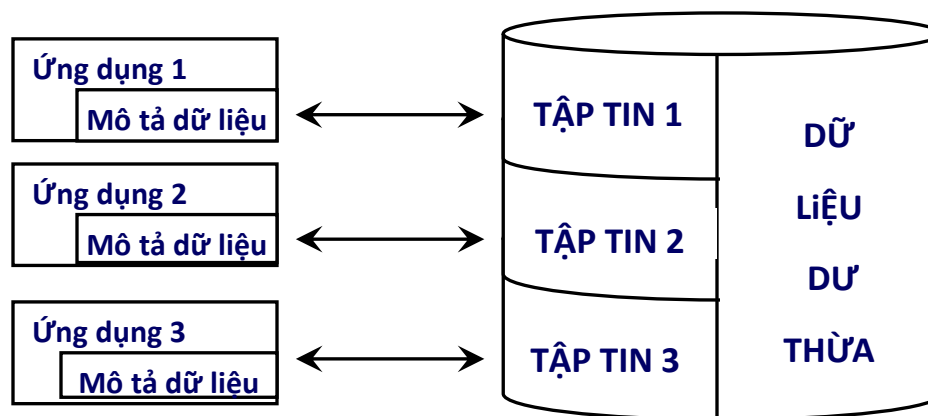
Hình 2.1 Hệ thống xử lý phân tán

# Các mô hình xử lý dữ liệu

---

## ❑ *Xử lý dữ liệu truyền thống*

- Mỗi ứng dụng định nghĩa một (nhiều) tệp dữ liệu.
- Ứng dụng khác nhau, tổ chức dữ liệu khác nhau
- Ứng dụng và tổ chức dữ liệu phụ thuộc lẫn nhau.
- Dư thừa dữ liệu



Hình 2.2: Xử lý dữ liệu truyền thống

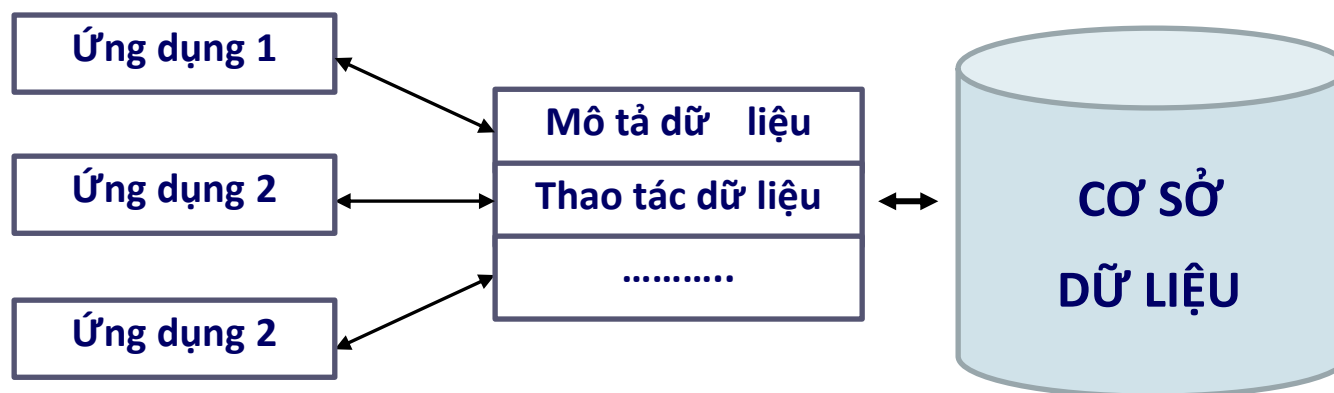


# Cơ sở dữ liệu phân tán

---

## ❑ *Xử lý dữ liệu tập trung:*

- Độc lập dữ liệu
- Tính bất biến ứng dụng đối với sự thay đổi cấu trúc lưu trữ và các chiến lược truy nhập dữ liệu.
- Không dư thừa dữ liệu



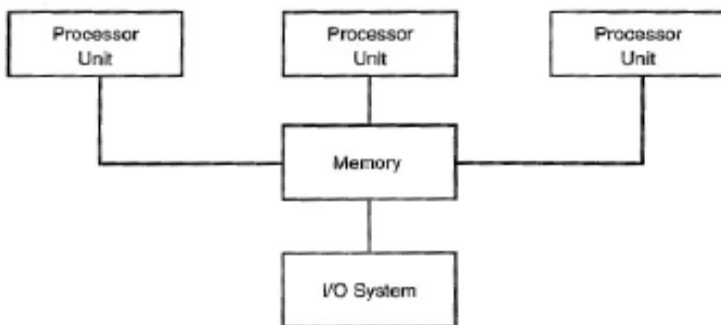
Hình 2.3 Xử lý cơ sở dữ liệu

## *Cơ sở dữ liệu phân tán*

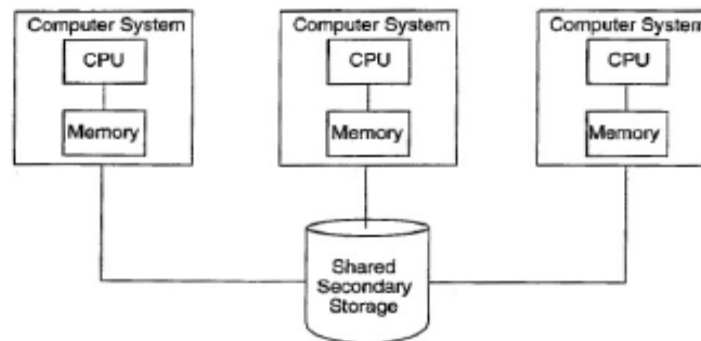
---

- ❑ Là một tập các cơ sở dữ liệu có quan hệ với nhau về mặt logic và được phân tán trên một mạng máy tính.
- ❑ Hệ QTCSDL phân tán là hệ thống phần mềm quản trị CSDLPT và làm cho sự phân tán trong suốt với NSD.
- ❑ Khái niệm CSDLPT nhấn mạnh hai khía cạnh:
  - *Tính phân tán*: dữ liệu phân tán trên nhiều vị trí. cơ sở dữ liệu địa phương (Local)
  - *Mối quan hệ logic*: Dữ liệu trên các vị trí phụ thuộc, ràng buộc lẫn nhau.

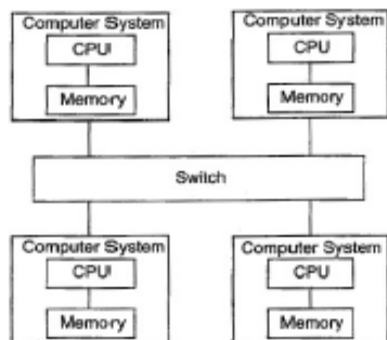
# Các hệ thống không phải là DDBS



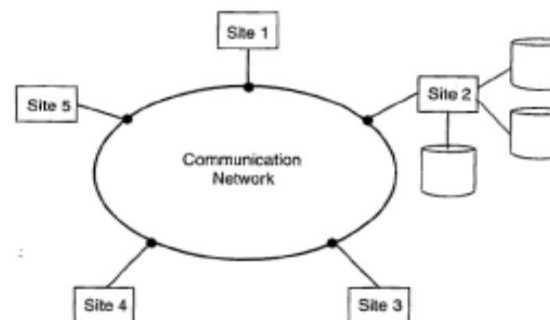
Shared Memory



Shared Disk

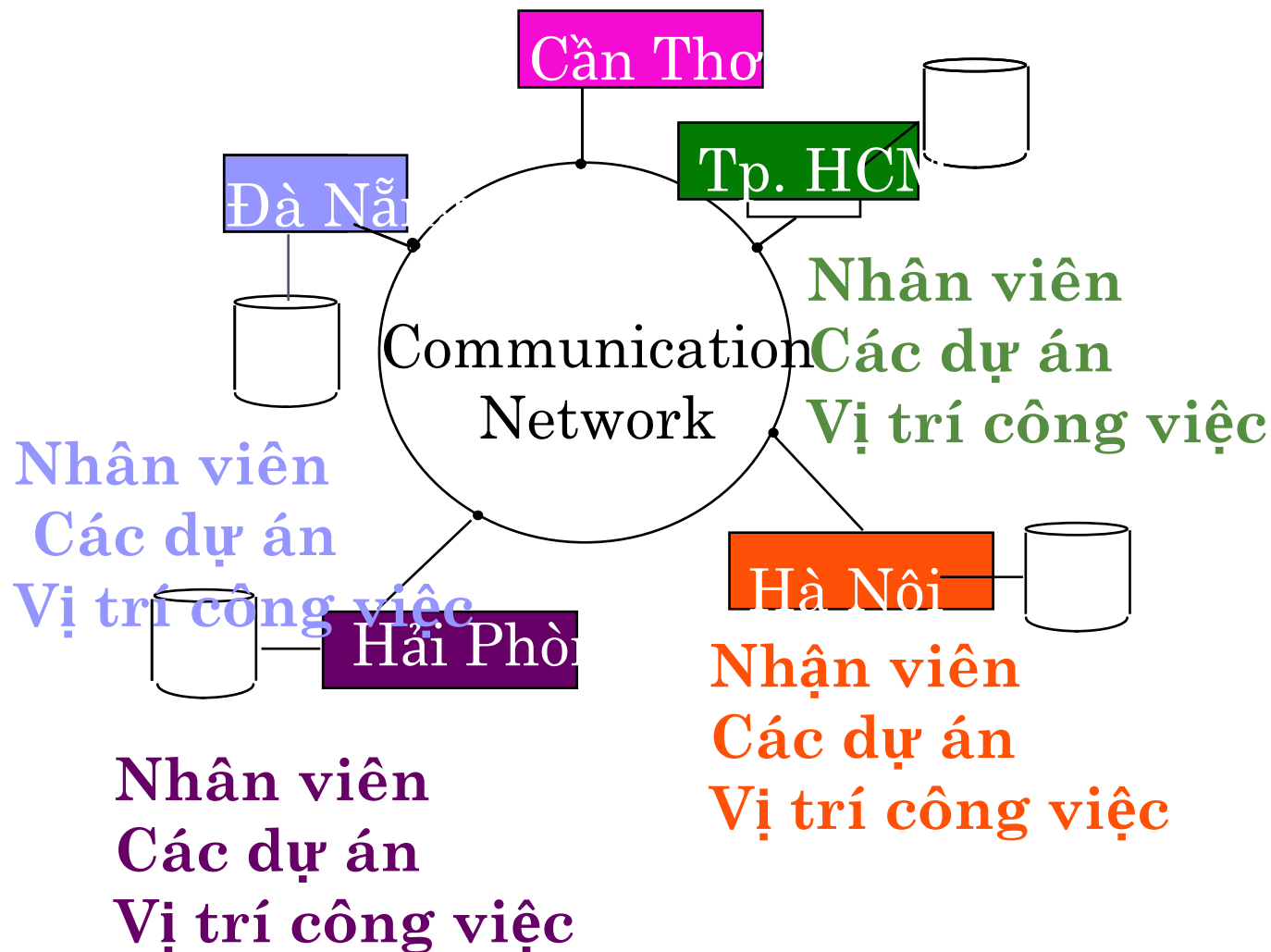


Shared Nothing



Central Databases

# Ví dụ



## *Cơ sở dữ liệu phân tán*

---

*Tóm lại, cơ sở dữ liệu phân tán là dữ liệu được phân tán trên mạng máy tính, đảm bảo cho việc tạo, truy nhập và duy trì dữ liệu phân tán, nhưng được thống nhất tổ chức như là một cơ sở dữ liệu tập trung duy nhất. Các trình ứng dụng truy nhập vào cơ sở dữ liệu phân tán như đang truy nhập vào cơ sở dữ liệu tập trung.*

# *Đặc điểm cơ sở dữ liệu phân tán*

## *(Có sự so sánh với dl tập trung)*

### ❑ *Điều khiển tập trung*

- CSDL tập trung: đảm bảo tính độc lập của dữ liệu, tránh dư thừa dữ liệu, **dl được quản lý tập trung, điều khiển bởi Người QTCSDL, ng QTCSDL đảm bảo độ an toàn dữ liệu**
- CSDL phân tán: không nhấn mạnh điều khiển tập trung. **Sự điều khiển được thực hiện bởi Người CSDL toàn cục (Global Database Administrator) và Người CSDL địa phương-cục bộ (Local Database Administrator)**
- **Người CSDL địa phương có quyền độc lập vị trí.**

# ***Đặc điểm cơ sở dữ liệu phân tán***

---

## ***Độc lập dữ liệu***

### ☐ CSDL Tập trung:

- Độc lập dữ liệu : DL trong suốt với ng lập trình=> Các ct không bị ảnh hưởng bởi về tổ chức lưu trữ vật lý của dữ liệu
- Tính chất độc lập được thể hiện thông qua mức khái niệm của kiến trúc nhiều mức.

### ☐ Trong CSDLPT

- Độc lập dữ liệu cũng là tính chất quan trọng;

## ***Độc lập dữ liệu (CSDLPT)***

-Bổ sung thêm sự trong suốt phân tán (Distribution Transparency):Chương trình ứng dụng được viết như trong CSDL tập trung=> tính đúng đắn của các trình ứng dụng không bị ảnh hưởng bởi sự di chuyển dữ liệu từ một vị trí này đến một vị trí khác. Tuy nhiên, các trình ứng dụng bị ảnh hưởng tốc độ khi phải trao đổi dữ liệu giữa các vị trí.

☛ Sự trong suốt phân tán được thể hiện bằng cách bổ sung vào kiến trúc nhiều mức của CSDL tập trung.



# *Đặc điểm CSDL phân tán*

---

## *Giảm dư thừa dữ liệu*

- ❑ Trong CSDL tập trung, sự dư thừa dữ liệu được giảm thiểu, tránh sự không nhất quán dữ liệu.
- ❑ Trong CSDL phân tán, sự dư thừa theo ngữ nghĩa sau:
  - Tăng tính cục bộ: ứng dụng cục bộ thực hiện nhanh hơn, nếu dữ liệu được nhân bản tại tất cả các vị trí.
  - Tăng tính sẵn sàng: Khi một vị trí có sự cố sẽ không làm ngưng sự thực hiện của các ứng dụng ở những vị trí khác nếu dữ liệu được nhân bản tại.

## ***Đặc điểm CSDL phân tán: Giảm dư thừa dữ liệu***

---

- ❑ Sự nhân bản dữ liệu dựa vào hai loại ứng dụng cơ bản: Ứng dụng chỉ đọc và ứng dụng cập nhật.
- ❑ Sự nhân bản giúp cho các ứng dụng chỉ đọc được thực hiện nhanh hơn, nhưng các ứng dụng cập nhật thực hiện chậm hơn vì phải cập nhật tại các vị trí
- ❑ Như vậy, sự nhân bản dữ liệu sẽ là một ưu điểm nếu hệ thống có rất nhiều ứng dụng chỉ đọc và có rất ít ứng dụng cập nhật. Trong trường hợp ngược lại thì sự nhân bản dữ liệu lại là một nhược điểm.

## ***Đặc điểm***

### ***Độ tin cậy giao dịch phân tán***

- ❑ Độ tin cậy giao dịch phân tán được cải thiện, vì nhân bản hạn chế được các lỗi trên vị trí riêng lẻ.
- ❑ Lỗi của cục bộ, hoặc lỗi truyền thông, không ảnh hưởng đến hoạt động bộ hệ thống.
- ❑ Nếu một số dữ liệu không thể truy nhập được, các giao dịch phân tán vẫn có thể truy nhập được tới phần khác trong cơ sở dữ liệu phân tán.

## ***Đặc điểm***

### ***Độ tin cậy giao dịch phân tán***

- ❑ Giao dịch là một đơn vị tính toán cơ bản, nhất quán và tin cậy, bao gồm một chuỗi các thao tác cơ sở dữ liệu được thực hiện từ trạng thái CSDL nhất quán này sang trạng thái nhất quán khác, ngay cả khi có một số giao dịch được thực hiện đồng thời và khi xảy ra lỗi.
- ❑ Khi hệ thống có lỗi, hệ QTCSDL đảm bảo việc thực thi đồng thời các giao dịch, không vi phạm tính nhất quán, với điều kiện là giao dịch được thực hiện chính xác, nghĩa là tuân theo các qui tắc toàn vẹn của CSDL.

## ***Đặc điểm CSDL phân tán***

### ***Cải tiến hiệu năng***

***Hiệu năng của CSDL PT được cải tiến dựa vào hai điểm:***

- ❑ ***Khả năng phân mảnh CSDL*** khái niệm và cục bộ hoá dữ liệu. Có hai ưu điểm:
  - ***Vì mỗi vị trí chỉ xử lý một phần cơ sở dữ liệu,*** sự tranh chấp CPU và các dịch vụ vào/ra không nghiêm trọng như trong môi trường tập trung.
  - ***Tính cục bộ làm giảm thời gian trễ truy nhập từ xa.***

***Lợi ích của việc phân mảnh và phân tán dữ liệu hợp lý sẽ làm giảm tranh chấp và giảm chi phí truyền thông.***

# ***Đặc điểm CSDL phân tán: Cải tiến hiệu năng***

---

***Hiệu năng của CSDL PT được cải tiến dựa vào hai điểm:***

- ❑ ***Tính song song của hệ thống phân tán*** có thể được khai thác để thực hiện
  - Song song liên truy vấn: là khả năng thực hiện nhiều truy vấn tại cùng thời điểm
  - Nội truy vấn song song là phương pháp tách một truy vấn đơn thành các truy vấn con và mỗi truy vấn con được thực hiện tại các trạm khác nhau, truy nhập các phần khác nhau của cơ sở dữ liệu phân tán.

## ***Đặc điểm CSDL phân tán: Cải tiến hiệu năng***

---

- ❑ Phân mảnh và nhân bản dữ liệu sao cho khai thác tốt nhất tính song song;

Trường hợp rất nhiều NSD cùng truy nhập vào một quan hệ chỉ được đọc, một số site yêu cầu xử lý song song, thì càng nhiều bản sao càng tăng tính sẵn sàng.

- ❑ Tăng bản sao làm giảm khối lượng thông tin trao đổi giữa các site. Tuy nhiên tính phức tạp tăng khi thực hiện cập nhật dữ liệu, phải đảm bảo tất cả các bản sao trên mạng phải thỏa các điều kiện toàn vẹn dữ liệu.

## ***Đặc điểm CSDL phân tán: Cải tiến hiệu năng***

---

*Tóm lại, việc nhân bản dữ liệu sẽ làm tăng hiệu quả các phép đọc, tăng tính sẵn sàng trong các giao tác đọc dữ liệu. Vấn đề điều khiển các giao tác tương tranh có sử dụng kỹ thuật bản sao trong các cơ sở dữ liệu phân tán rất phức tạp. Việc quản lý có thể đơn giản hóa bằng việc sinh ra các bản sao sơ cấp quan hệ.*



## ***Đặc điểm CSDL phân tán***

### ***Dễ dàng mở rộng***

---

Dễ dàng tăng kích thước dữ liệu, bằng cách tăng khả năng lưu trữ và xử lý của mạng. Phụ thuộc vào chi phí phân tán. Tuy nhiên, khả năng mở rộng hệ thống dễ dàng mang tính kinh tế, chi phí giảm.

## ***Đặc điểm CSDL phân tán:***

### ***Tính toàn vẹn, phục hồi và điều khiển tương tranh***

---

- ❑ Tính toàn vẹn, phục hồi và điều khiển tương tranh có quan hệ chặt chẽ với nhau thông qua các giao tác.
- ❑ Mỗi một giao tác là một nguyên tố, nghĩa là một dãy các thao tác hoặc được thực hiện hoàn toàn, hoặc không được thực hiện hay chỉ thực hiện một phần.
- ❑ Như vậy giao tác nguyên tố là phương tiện để đạt được tính toàn vẹn dữ liệu, vì các giao tác đảm bảo tất cả các thao tác biến đổi cơ sở dữ liệu từ trạng thái toàn vẹn này sang trạng thái toàn vẹn khác.

## ***Đặc điểm CSDL phân tán:***

### ***Tính toàn vẹn, phục hồi và điều khiển tương tranh***

---

- ❑ Tác động sự cố kỹ thuật và tương tranh có ảnh hưởng đến tính nguyên tố của giao tác.
- ❑ Sự cố kỹ thuật làm cho hệ thống dừng giao tác đang thực hiện, vì vậy vi phạm phải mang tính nguyên tố.
- ❑ Sự tương tranh của các giao tác cho phép một giao tác nào đó quan sát một trạng thái chuyển tiếp không toàn vẹn, tạo ra bởi một giao tác khác trong quá trình thực hiện nó.

## ***Đặc điểm CSDL phân tán:***

### ***Tính toàn vẹn, phục hồi và điều khiển tương tranh***

---

- ❑ Sự phục hồi liên quan đến đảm bảo tính nguyên tố của giao tác khi có sự cố về kỹ thuật, khi có một số vị trí (site) tham gia thực hiện giao tác có thể bị hỏng.
- ❑ Điều khiển tương tranh nhằm đảm bảo tính nguyên tố của giao tác khi xuất hiện sự tương tranh giao tác. Vấn đề điều khiển tương tranh trong môi trường phân tán phức tạp hơn nhiều trong môi trường tập trung.

## Ưu điểm csdlpt

- **Chia sẻ dữ liệu và điều khiển phân tán:** Người sử dụng tại một vị trí này có thể truy xuất dữ liệu (được phép) ở vị trí khác. Hơn nữa việc quản trị cơ sở dữ liệu có thể được phân tán và thực hiện tự quản tại mỗi vị trí.
- **Độ tin cậy và tính sẵn sàng:** Nếu một vị trí bị hỏng thì các vị trí còn lại trong hệ thống cơ sở dữ liệu phân tán vẫn tiếp tục hoạt động. Nếu dữ liệu được nhân bản ở một số vị trí thì một giao dịch cần truy xuất một mục dữ liệu có thể tìm thấy ở bất kỳ vị trí nào trong số vị trí đó. Như thế sự cố tại một vị trí không ảnh hưởng đến hệ thống.
- **Tăng tốc độ xử lý truy vấn:** Nếu một truy vấn cần dữ liệu ở một số vị trí thì có thể chia câu truy vấn đó thành các câu truy vấn con rồi thực thi nó song song tại các vị

## Nhược điểm

**Chi phí phát triển phần mềm:** Việc phát triển một hệ thống cơ sở dữ liệu phân tán khá phức tạp vì thế cần chi phí lớn.

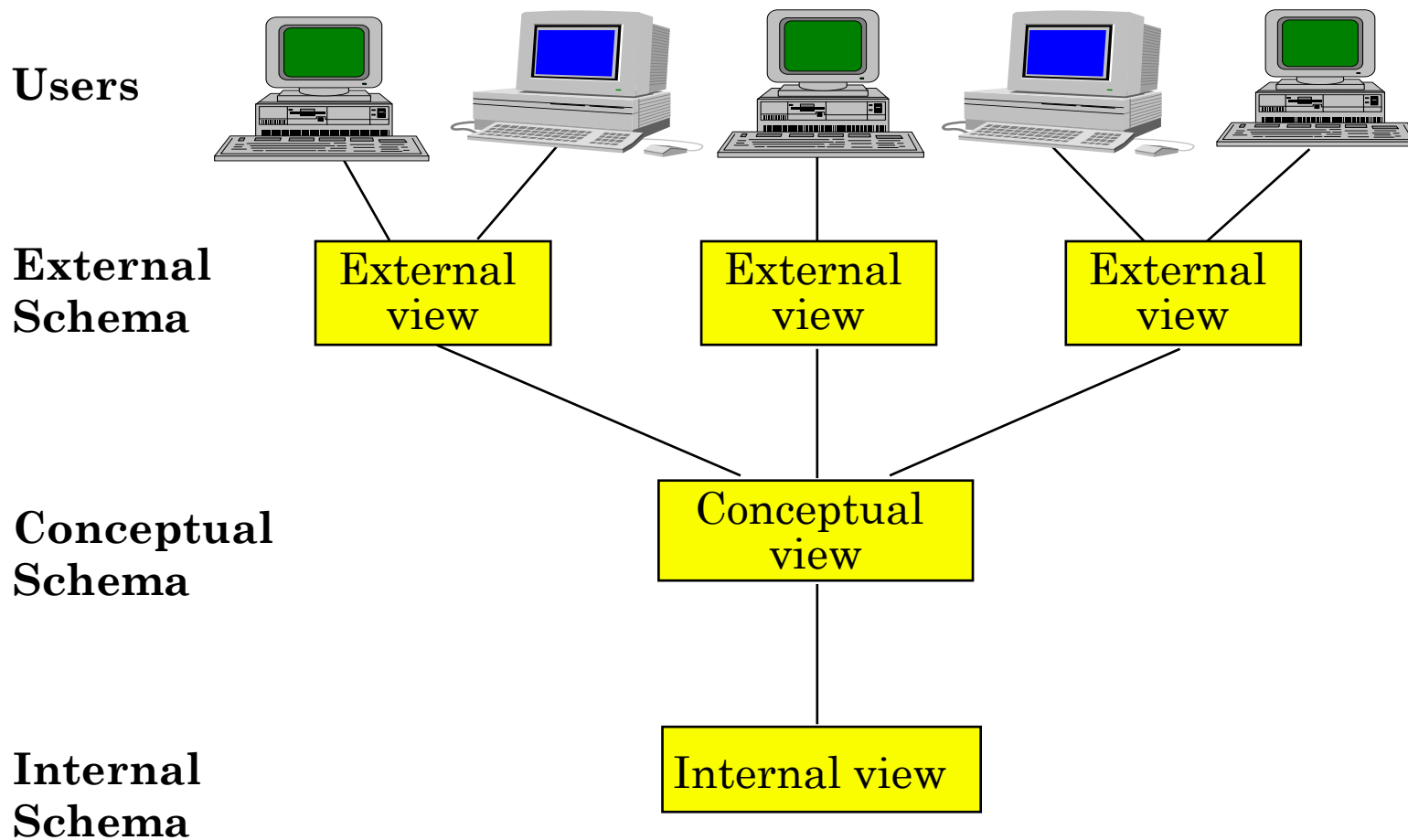
**Khó phát hiện lỗi:** Việc phát hiện lỗi và đảm bảo tính đúng đắn của các thuật toán song song sẽ rất khó khăn.

**Chi phí xử lý tăng:** Sự trao đổi các thông báo và xử lý phối hợp giữa các vị trí sẽ tăng chi phí xử lý hơn trong các hệ thống tập trung.

## *Mô hình tham chiếu cơ sở dữ liệu phân tán*

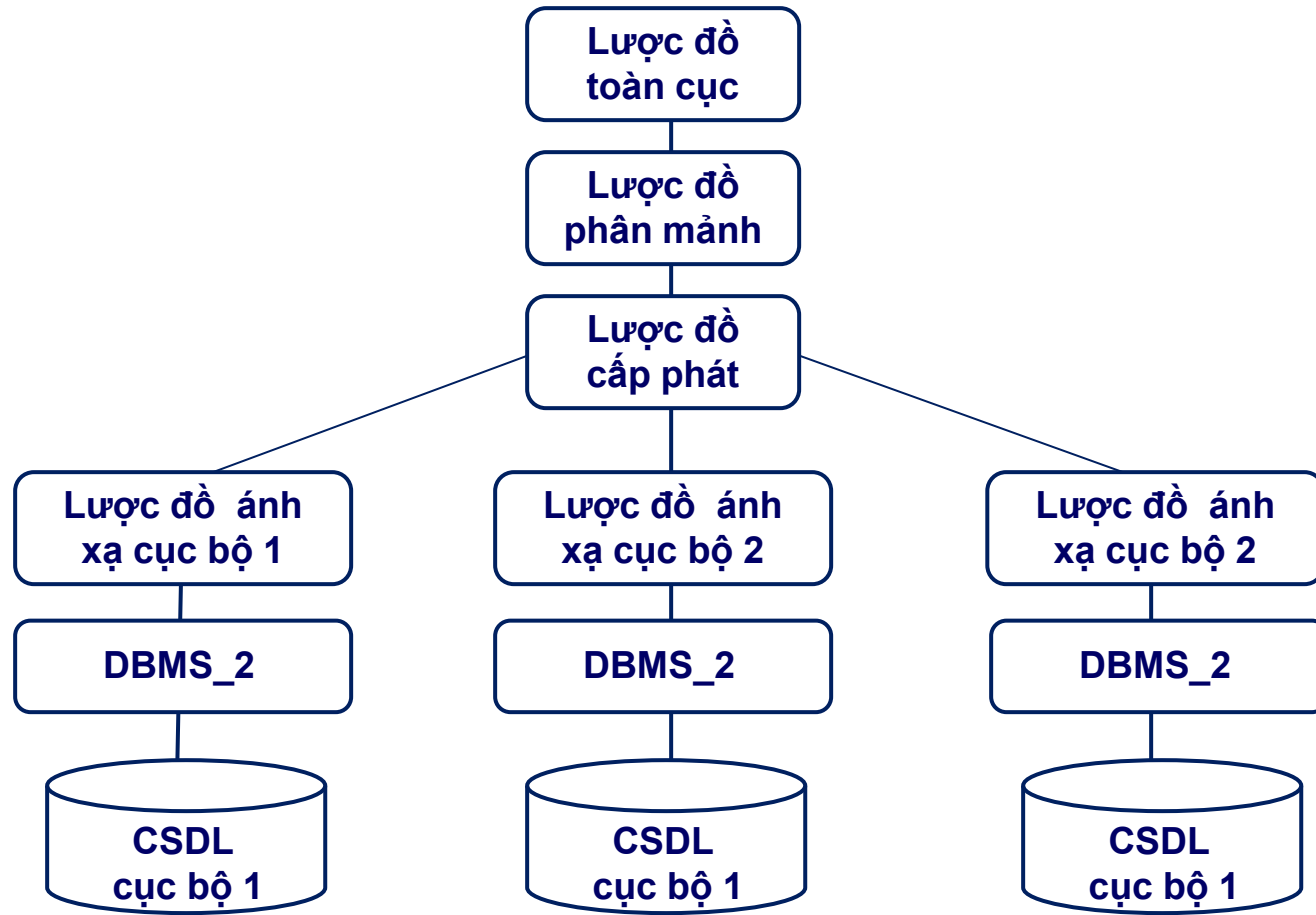
---

- ❑ Mô hình kiến trúc cơ sở dữ liệu phân tán dựa trên sự mở rộng của mô hình kiến trúc ba mức của cơ sở dữ liệu tập trung ANSI/SPARC.
- ❑ Mô hình kiến trúc cơ sở dữ liệu phân tán bao gồm lược đồ tổng thể, lược đồ phân mảnh và lược đồ cấp phát.





# Mô hình tham chiếu cơ sở dữ liệu phân tán



Hình 2.4 Mô hình tham chiếu của cơ sở dữ liệu phân tán

## *Lược đồ toàn cục*

---

- ❑ Lược đồ toàn cục định nghĩa tất cả dữ liệu được chứa trong CSDL phân tán như trong CSDL tập trung.
- ❑ Lược đồ toàn cục được định nghĩa chính xác như định nghĩa lược đồ cơ sở dữ liệu tập trung.
- ❑ Tuy nhiên, mô hình dữ liệu lược đồ toàn cục cần phải tương thích với việc định nghĩa các ánh xạ tới các mức của cơ sở dữ liệu phân tán.
- ❑ Sơ đồ toàn cục bao gồm thông tin về các thực thể, tập các phụ thuộc hàm và mối quan hệ giữa các thực thể.

## *Lược đồ phân mảnh*

---

- ❑ Từ quan hệ toàn cục có thể chia thành nhiều quan hệ con được gọi là các mảnh, tách biệt với nhau nhau
- ❑ Ánh xạ giữa các quan hệ toàn cục và phân mảnh được định nghĩa bởi lược đồ phân mảnh. Ánh xạ này là mỗi quan hệ một-nhiều.
- ❑ Các mảnh được chỉ ra bằng tên của quan hệ toàn cục với một chỉ số (chỉ số phân mảnh)
- ❑ Các mảnh có thể được cài đặt tại một hay nhiều vị trí khác nhau trên mạng.

# *Lược đồ phân mảnh*

---

- ❑ Các kiểu phân mảnh dữ liệu bao gồm:
  - Phân mảnh ngang
  - Phân mảnh dọc và một kiểu phân mảnh là
  - Sự kết hợp của phân mảnh ngang và phân mảnh dọc, phức tạp hơn
- ❑ Trong các kiểu phân mảnh, một mảnh được định nghĩa bằng một biểu thức đại số quan hệ, các toán hạng là các quan hệ toàn cục.

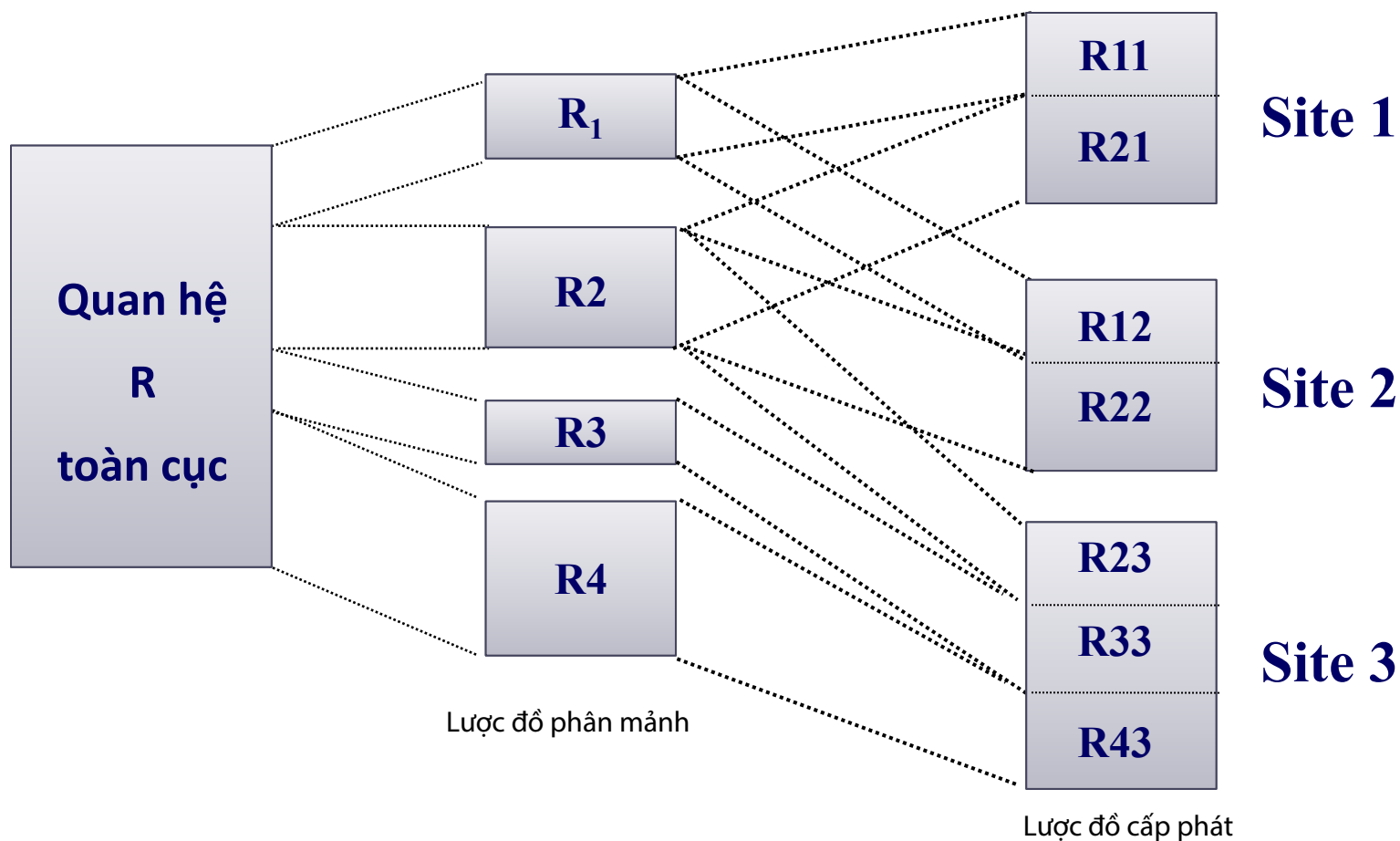
## *Lược đồ cấp phát*

---

- ❑ Các mảnh được chứa ở một hay nhiều site trên mạng.
- ❑ Lược đồ cấp phát chứa thông tin các mảnh được chứa trên những site nào.
- ❑ Các mảnh của một quan hệ lưu trên site  $j$  tạo thành một mô hình vật lý của quan hệ toàn cục lên site  $j$ .
- ❑ Ký hiệu  $R_{ij}$  là mảnh thứ  $i$  của quan hệ  $R$  trên site  $j$ .

# *Lược đồ cấp phát*

---



Hình 2.5: Các mảnh và mô hình vật lý cho một quan hệ

## *Lược đồ cấp phát*

---

- ❑ Quan hệ toàn cục  $R$  được phân thành 4 mảnh quan hệ con  $R_1, R_2, R_3$  và  $R_4$ , và được cấp phát trên 3 site của mạng máy tính, tạo nên ba mô hình vật lý:
  - Trên site 1: Một bản sao của  $R_1$  và một bản sao của  $R_2$ . Ký hiệu  $R_{11}$  và  $R_{21}$
  - Trên site 2: Một bản sao của  $R_1$  và một bản sao của  $R_2$ . Ký hiệu  $R_{12}$  và  $R_{22}$
  - Trên site 3: Một bản sao của  $R_2, R_3$  và  $R_4$ . Ký hiệu  $R_{23}$ ,  $R_{33}$  và  $R_{43}$

## *Phân mảnh và cấp phát dữ liệu*

---

- ❑ Trong suốt trong phân mảnh là mức trong suốt cao nhất, NSD làm việc trên mỗi quan hệ tổng thể.
- ❑ Trong suốt trong cấp phát là mức trong suốt thấp hơn và yêu cầu người sử dụng chỉ được làm việc trên các mảnh địa phương, thay cho làm việc trên quan hệ toàn cục mà không biết mảnh đó ở đâu.
- ❑ Cần phân biệt sự khác nhau giữa khái niệm phân mảnh và cấp phát. Khái niệm phân mảnh dữ liệu khác với khái niệm cấp phát tối ưu dữ liệu.



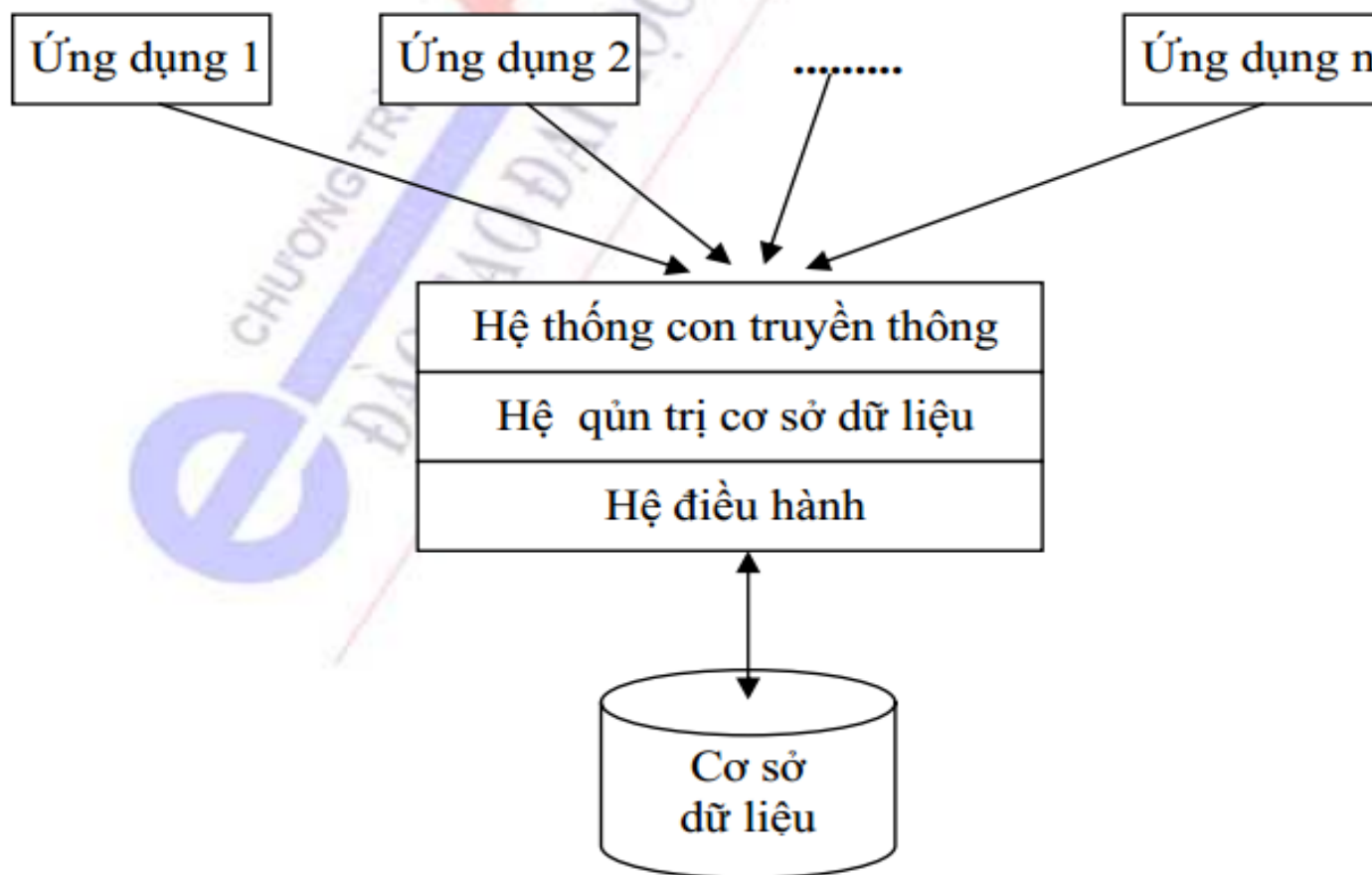
## ***Lược đồ ánh xạ cục bộ***

- **Ba mức đầu của các site** là độc lập, ko phụ thuộc vào mô hình dữ liệu của HQTCSDL cục bộ
- Ở mức thấp hơn: phụ thuộc vào kiểu của hệ QTCSDL cục bộ. Cần phải có ánh xạ của mô hình vật lý thành các đối tượng thao tác bởi hệ QTCSDL cục bộ. Trong hệ thống không thuần nhất có các kiểu khác nhau của AX cục bộ tại các site khác nhau. Yếu tố quan trọng để thiết kế kiến trúc này là:
  - PM và CP dữ liệu
  - QL dư thừa dl
  - Sự độc lập của các DBMS cục bộ

## ***Hệ QTCSDL ở các site cục bộ***

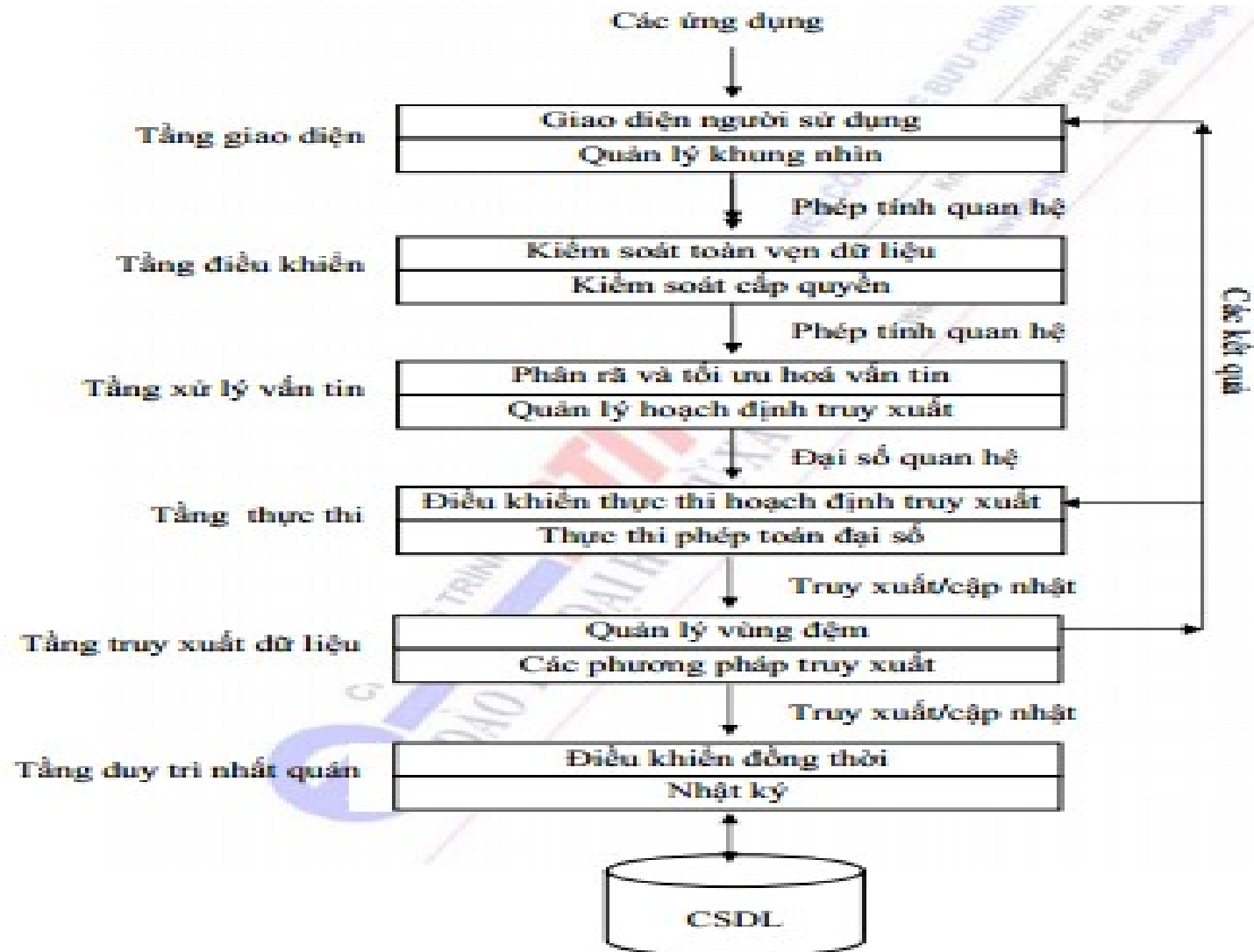
- Hệ CSDLPT đồng nhất: Lược đồ của 1 site đc đn như trong hệ CSDL tập trung
- Hệ CSDLPT **Không** đồng nhất: Lược đồ ánh xạ cục bộ dùng để phối hợp các kiểu khác nhau của hệ QTCSDL

# Hệ QTCSDL QH



Hình 1.7: Kiến trúc tổng quát của mô hình hệ quản trị cơ sở dữ liệu quan hệ

# Chức năng của hệ QTCSDL quan hệ



Hình 1.8: Các tầng chức năng của một hệ quản trị cơ sở dữ liệu quan hệ

# Hệ quản trị cơ sở dữ liệu phân tán

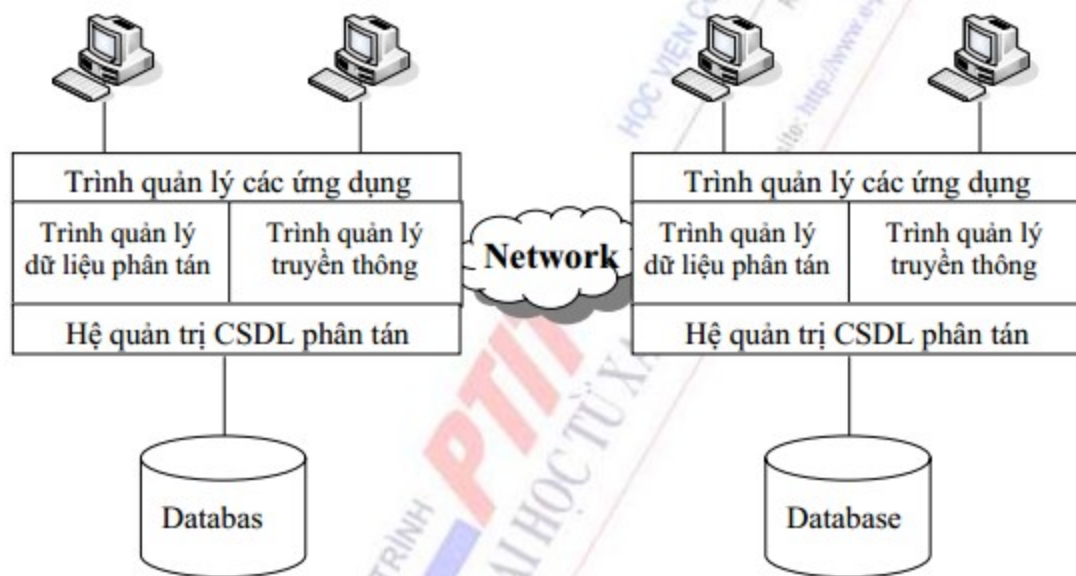
- Hệ quản trị cơ sở dữ liệu phân tán hỗ trợ việc tạo lập, quản lý, điều khiển và duy trì cơ sở dữ liệu phân tán, làm cho sự phân tán trong suốt với người sử dụng
- Chúng chứa các thành phần bổ sung mở rộng các khả năng của các hệ quản trị cơ sở dữ liệu tập trung như hỗ trợ sự truyền thông và sự cộng tác giữa các hệ quản trị cơ sở dữ liệu trên các địa điểm khác nhau qua mạng máy tính.

# Các thành phần cơ bản cần thiết của hệ QTCSDLPT tại mỗi trạm (site) cần cài đặt:

1. Thành phần quản trị cơ sở dữ liệu (DB: Database Management)
2. Thành phần truyền dữ liệu (DC: Data Communication)
3. Tự điển dữ liệu (DD : Data Dictionary) mở rộng để biểu diễn thông tin về sự phân tán dữ liệu trên mạng.
4. Thành phần cơ sở dữ liệu phân tán (DDB:Distributed Database)

# Các dịch vụ của hệ thống

- Các ứng dụng truy cập từ xa
- Cung cấp các mức phân tán
- Hỗ trợ quản trị và điều khiển CSDL
- Khả năng mở rộng các hệ thống khác nhau
- Cung cấp khả năng thực hiện đồng thời và phục hồi các giao tác phân tán

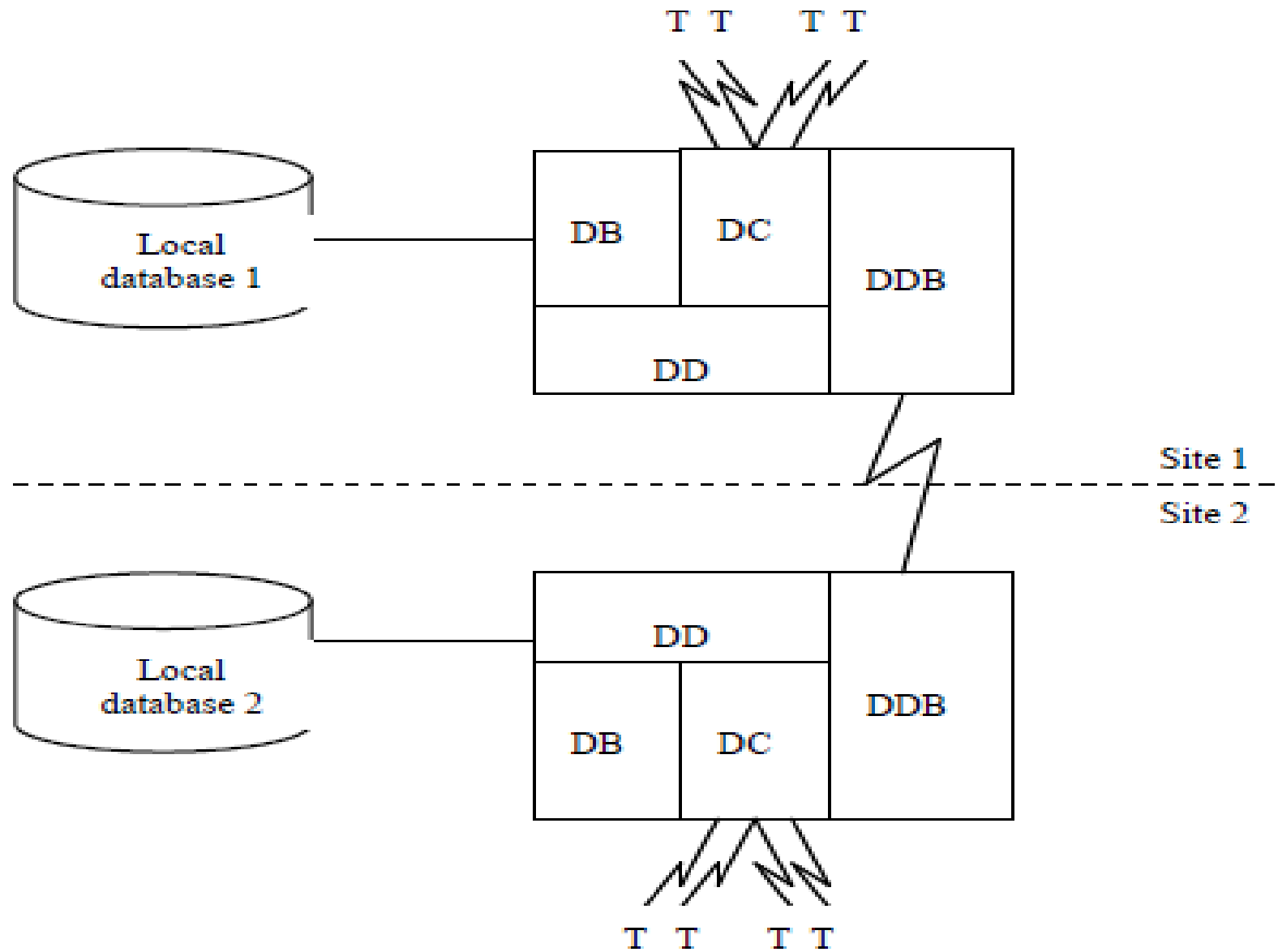


Hình 1.9 Hệ quản trị CSDL phân tán

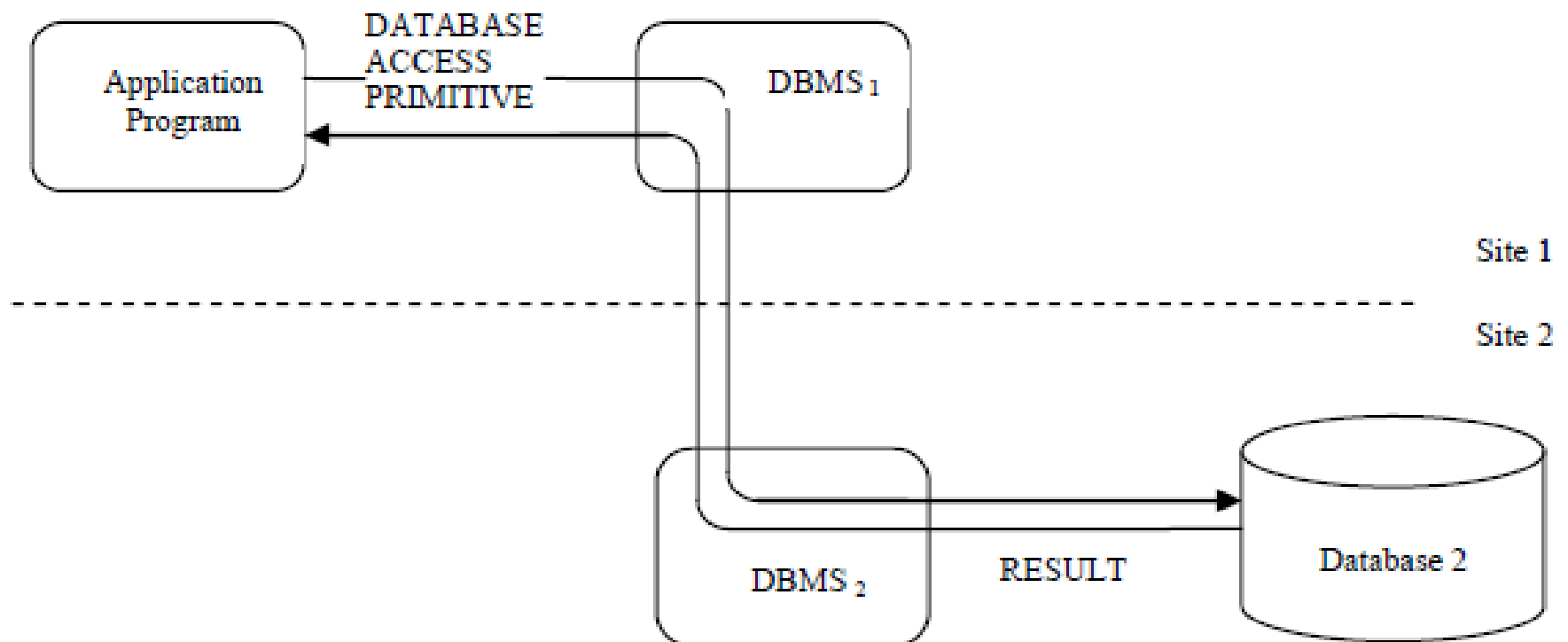
## *Truy nhập dữ liệu từ xa trực tiếp*

- ❑ Các trình ứng dụng trên Client yêu cầu truy nhập trực tiếp vào các hệ cơ sở dữ liệu từ xa. Yêu cầu này được máy chủ CSDL thực hiện và gửi kết quả về Client.
- ❑ Tính trong suốt phân tán cao. Có thể thực hiện được bằng việc cung cấp các file chung và có thể truy nhập tự động theo địa chỉ các truy nhập trước đó.
- ❑ Tuy nhiên việc điều khiển tương tranh và khôi phục các tiến trình phân tán khi số người sử dụng tăng lên là khó khăn và phức tạp.





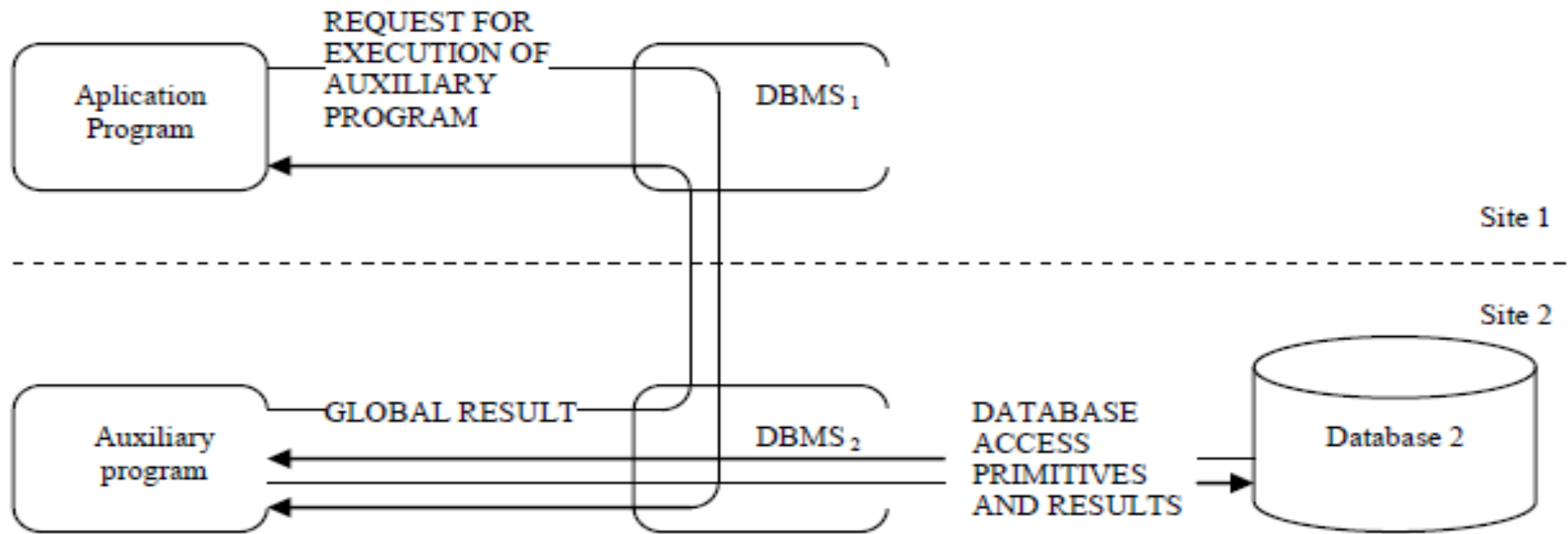
# Các kiểu truy xuất đến cơ sở dữ liệu phân tán



(a) Truy xuất từ xa qua các lệnh có sẵn của hệ quản trị cơ sở dữ liệu

# *Truy nhập dữ liệu từ xa qua trung gian Middleware*

- ❑ Middleware là bộ phần mềm trung gian, thực hiện việc điều khiển các tiến trình truyền thông và điều khiển cấp phát tài nguyên cho các tiến trình theo yêu cầu của các trình ứng dụng Client.



(b) Truy xuất từ xa qua chương trình hỗ trợ

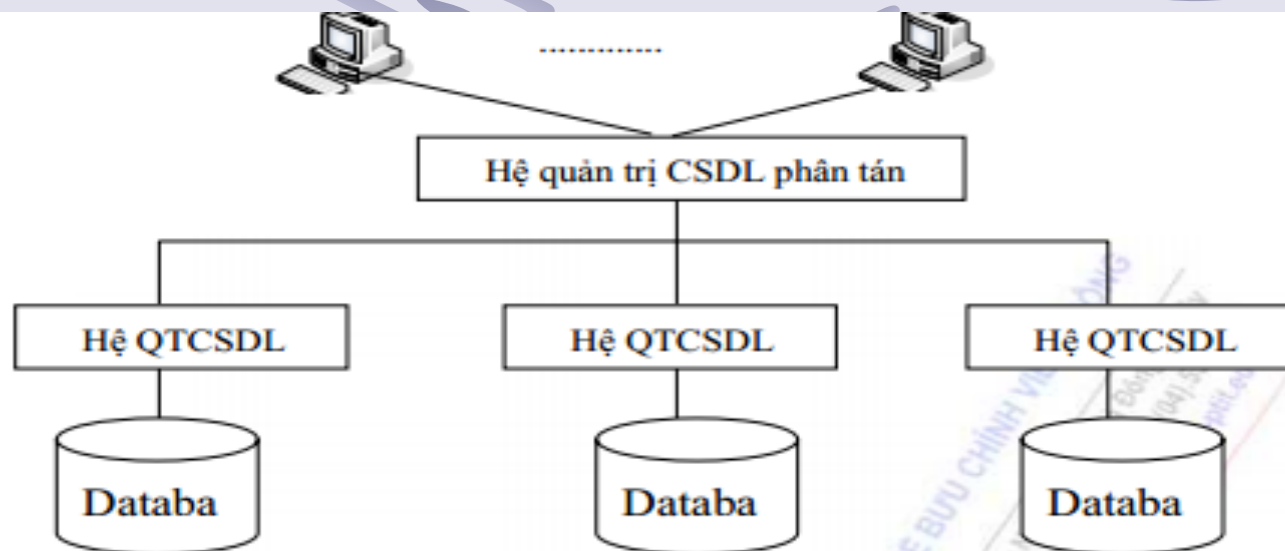
Hình 1.4 Các kiểu truy xuất đến cơ sở dữ liệu phân tán

# Hệ QTCSDL thuần nhấn

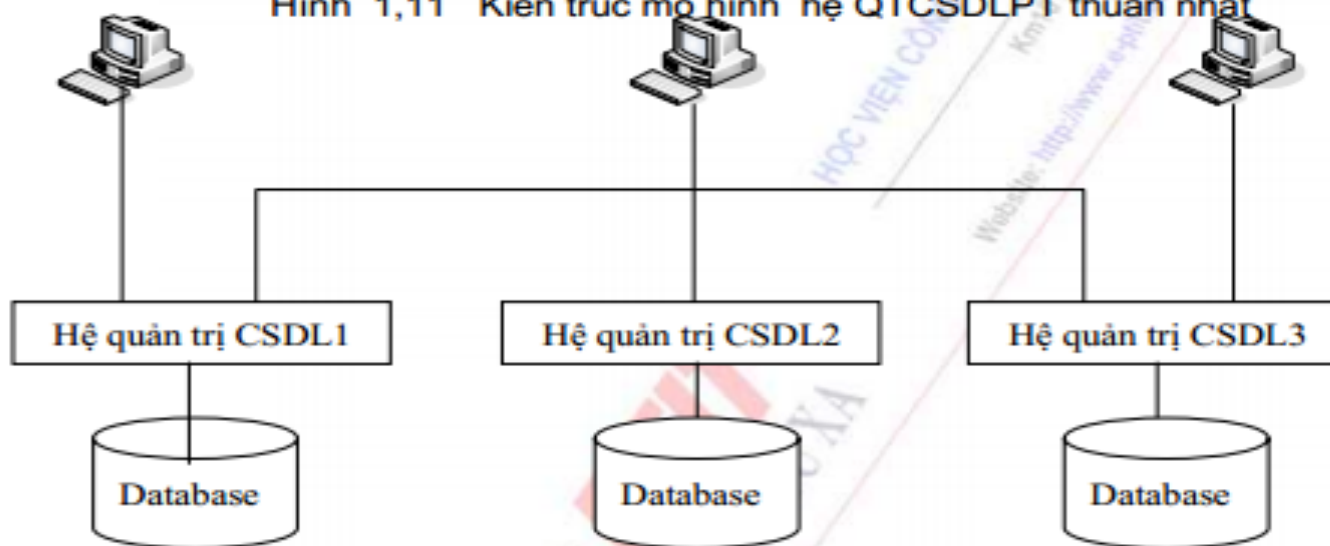
- Mọi site đều cài đặt cùng một hệ QTCSDL

# Hệ QTCSDL Không thuần nhận

- Mọi site đều Không cài đặt cùng một hệ QTCSDL
- Hệ QTCSDLPT ko thuần nhất phải tích hợp thêm việc chuyển đổi của các mô hình dữ liệu khác nhau giữa các hệ QTCSDL để thống nhất quản lý

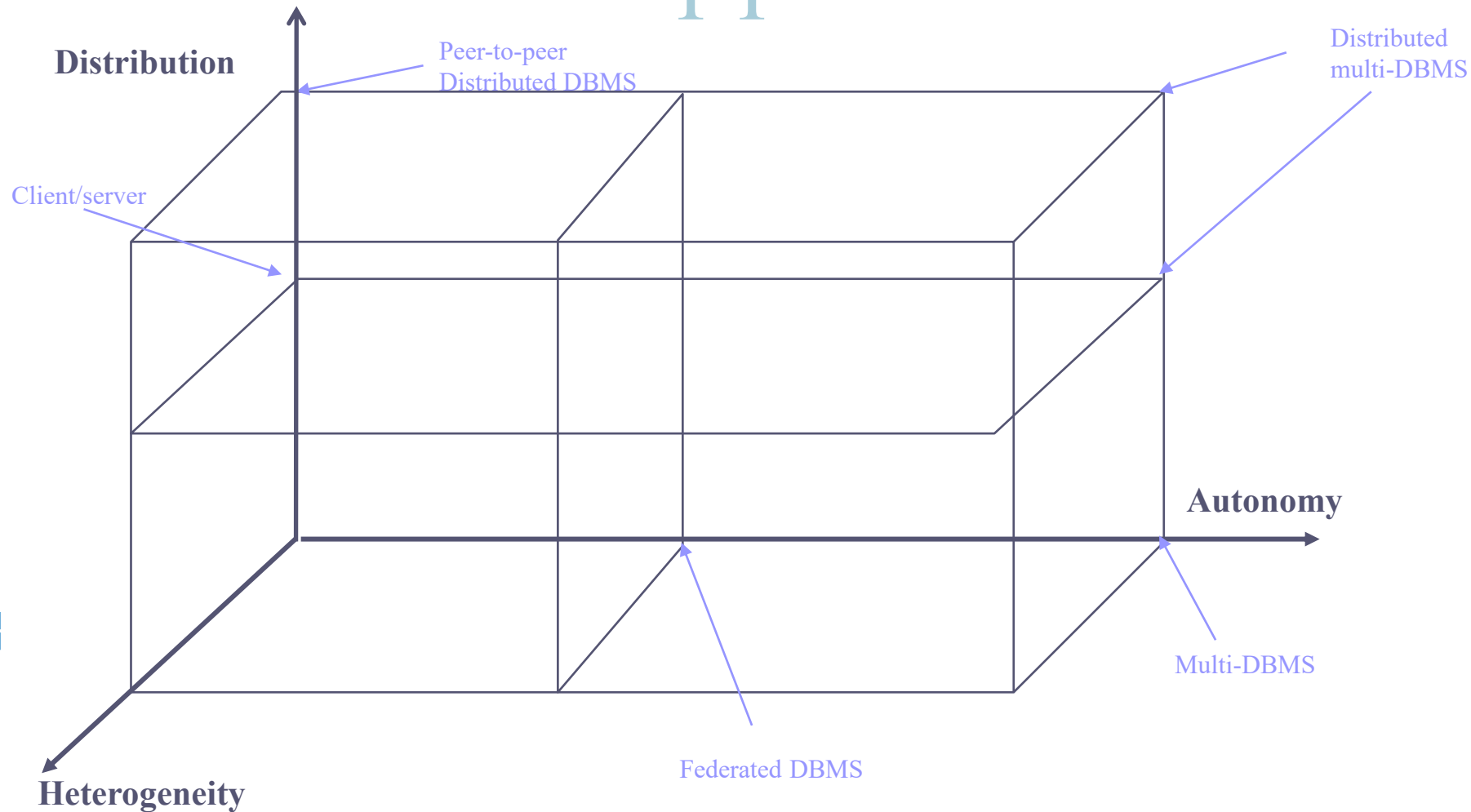


Hình 1.11 Kiến trúc mô hình hệ QTCSDLPT thuần nhất



Hình 1.12 Kiến trúc mô hình hệ QTCSDLPT không thuần nhất

PT





- Tính phân tán
  - Các thành phần của hệ thống nằm trên cùng một máy hay không
- Tính không đồng nhất
  - Các mức khác nhau (phần cứng, truyền thông, Hệ điều hành)
  - DBMS: Mô hình dữ liệu, ngôn ngữ truy vấn, thuật toán quản lý giao dịch
- Tính tự chủ, tự trị
  - Tự chủ về thiết kế: Khả năng của DBMS trong quyết định các vấn đề liên quan đến thiết kế của bản thân nó.
  - Tự chủ về truyền thông: Khả năng của DBMS trong quyết định có/không và cách thức liên lạc với các DBMSs khác.
  - Tự chủ về thực thi: Khả năng của DBMS trong thực thi các hoạt động theo các riêng của mỗi DBMS

## Các kiến trúc logic của hệ QTCSDLPT

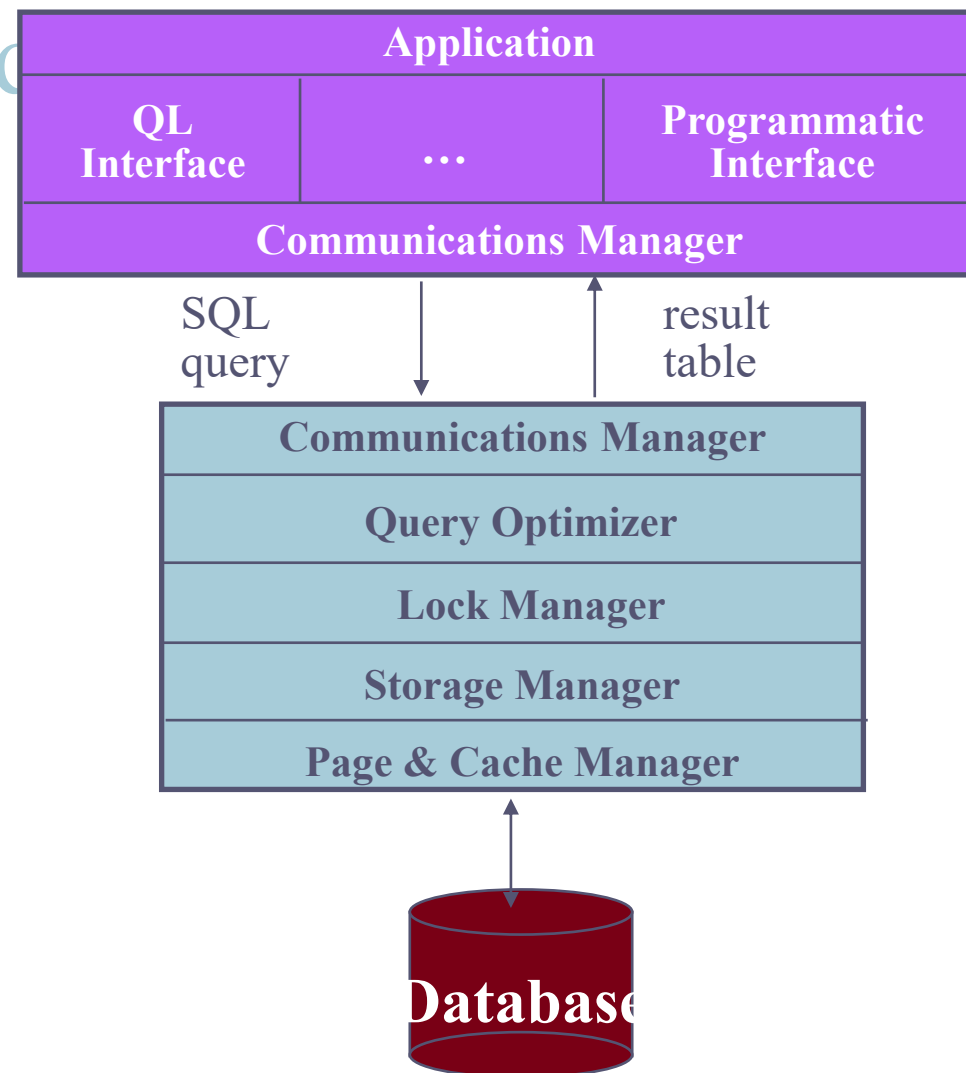
Có 3 kiểu kiến trúc cho HQTCSDLPT:

- Client Server
- Per to per
- Đa hệ CSDL

# Kiến trúc khách/chủ cho CSDL PT (dựa trên c

Ý tưởng chung: Chia các chức năng thành 2 lớp

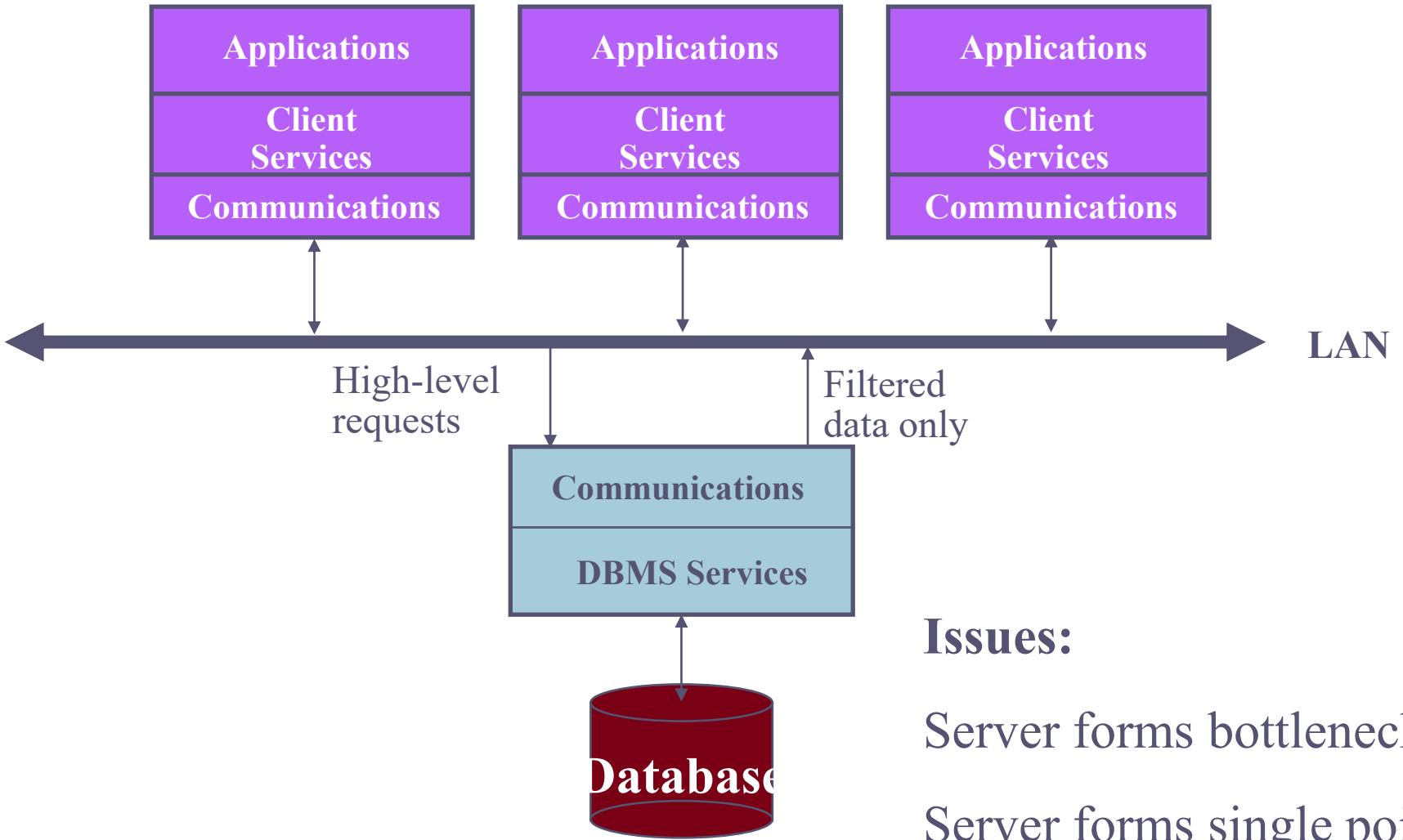
- Các chức năng phía chủ
  - Quản lý dữ liệu, tối ưu xử lý truy vấn, quản lý giao dịch v.v
- Các chức năng phía khách
  - Có thể bao gồm một phần chức năng quản lý dữ liệu của nó, không chỉ UI
- Cung cấp kiến trúc 2 mức
- Phân chia hiệu quả hơn công việc
- Các loại kiến trúc C/S:
  - Multiple client/single server
  - Multiple client/multiple server



# Multiclients/Single Server

www.ptit.edu.v

n



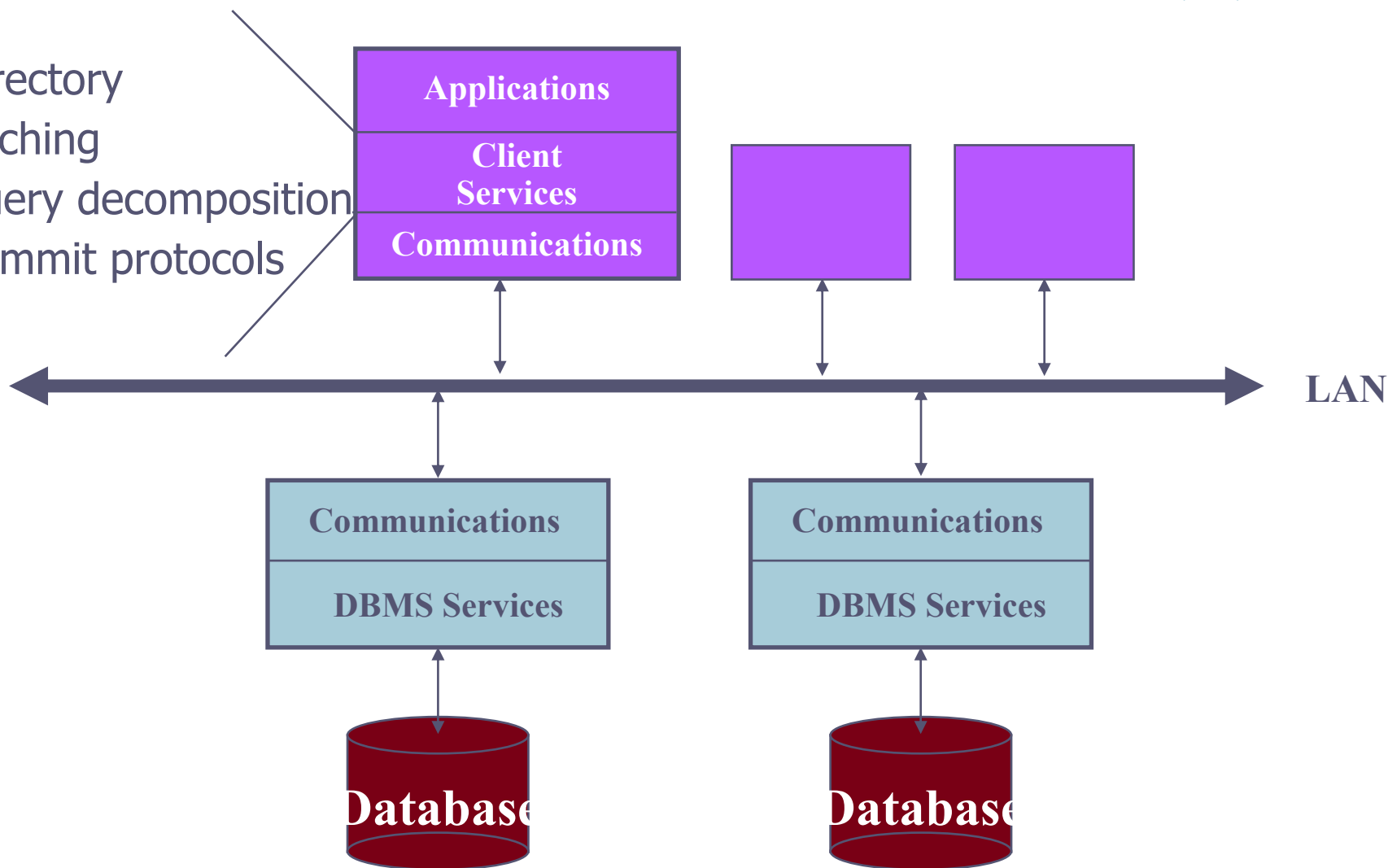
## Issues:

Server forms bottleneck

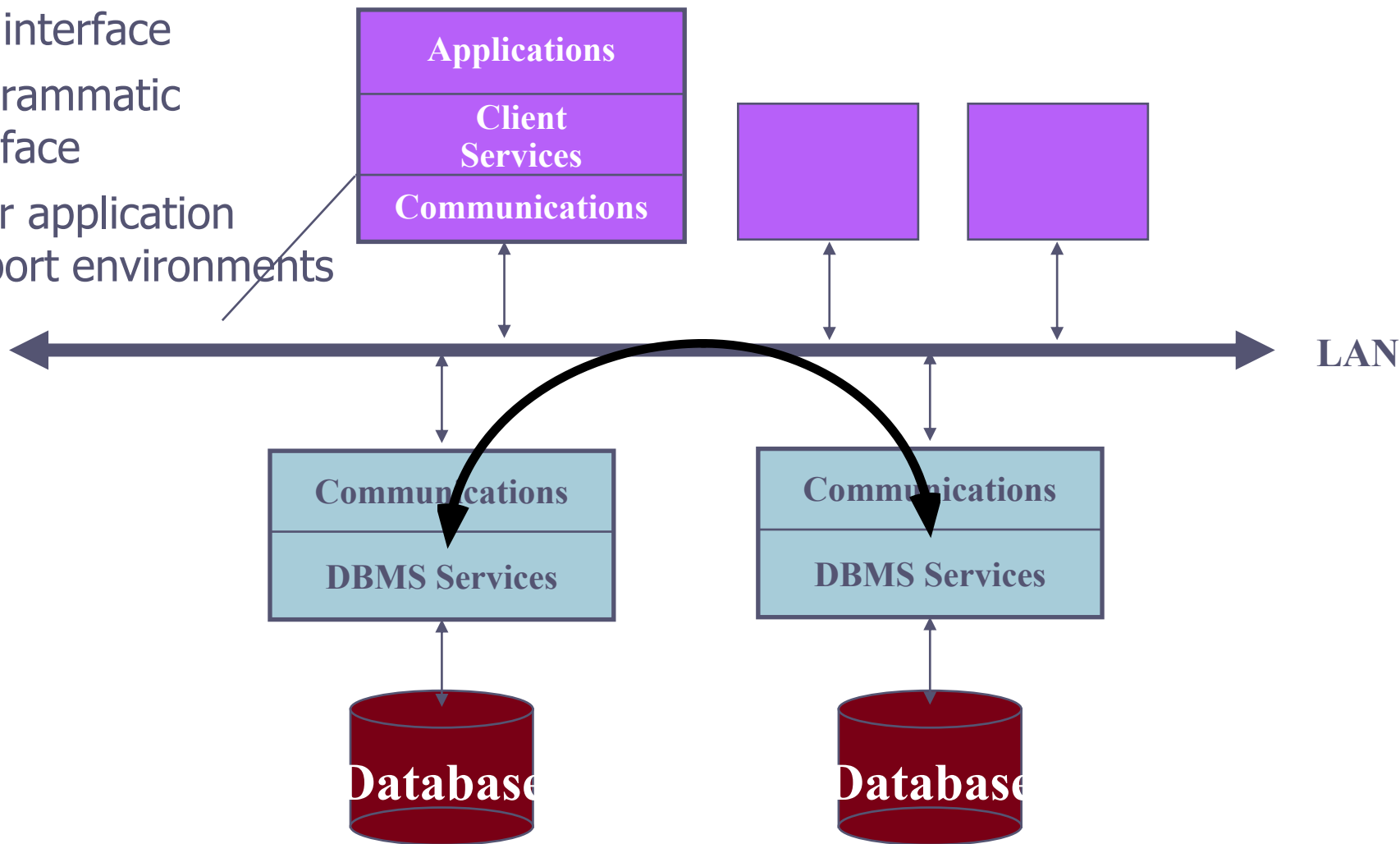
Server forms single point of failure

# Multi Clients/ Multi Servers (1)

- directory
- caching
- query decomposition
- commit protocols

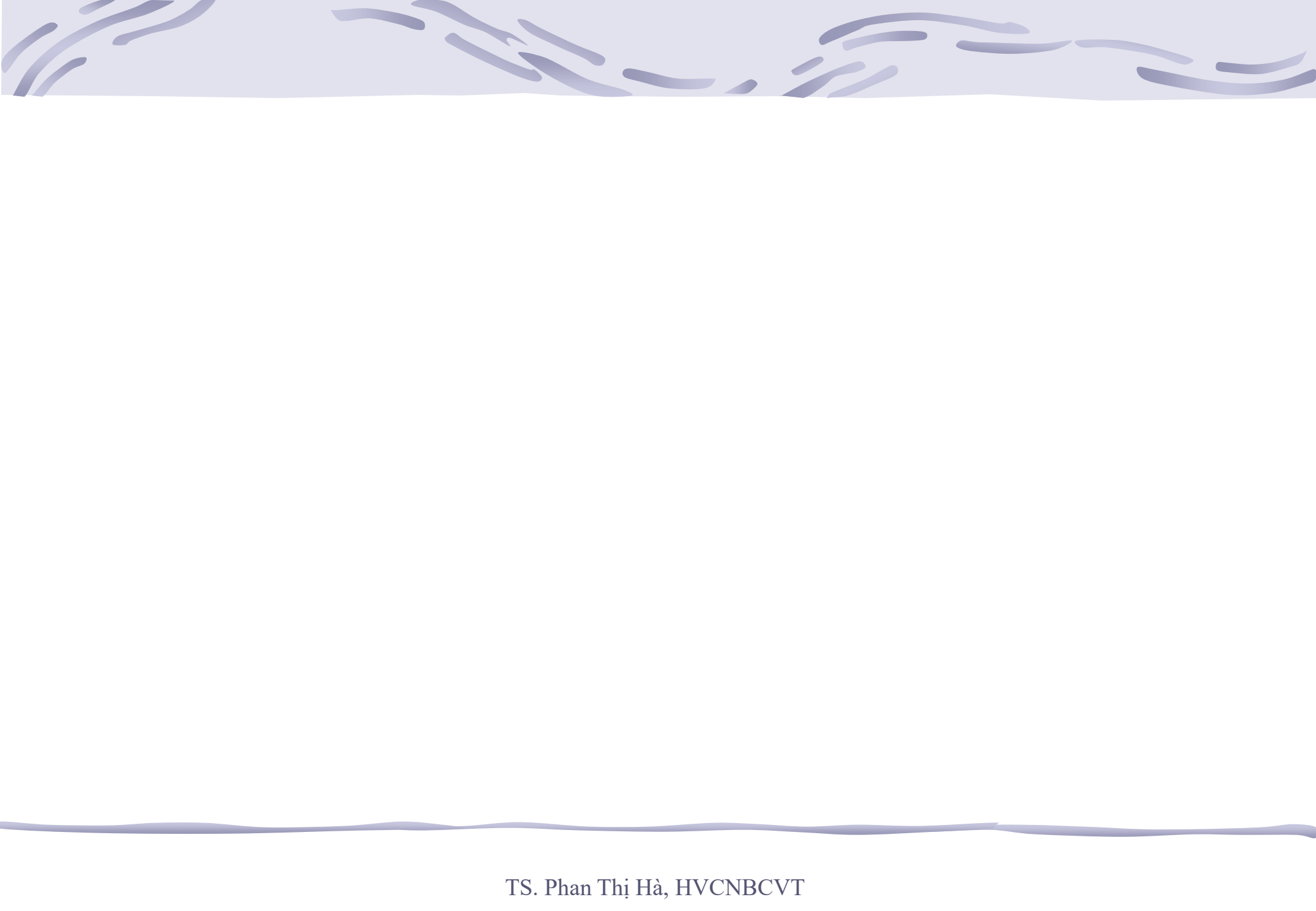


- SQL interface
- programmatic interface
- other application support environments

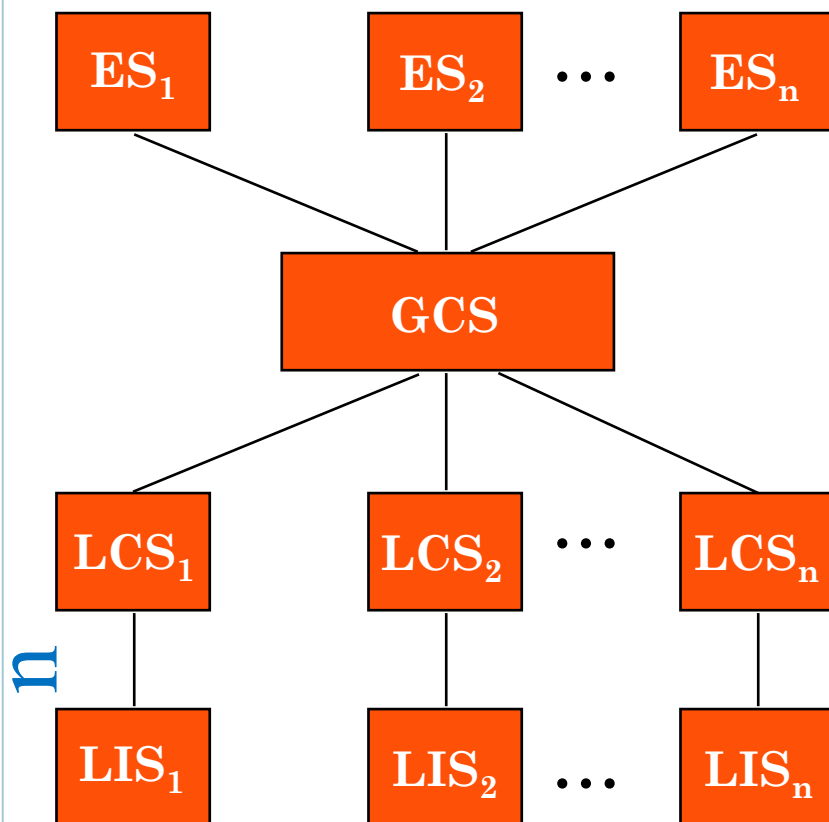


# Ưu điểm của kiến trúc Khách/Chủ

- Phân chia hiệu quả công việc
- Dễ dàng mở rộng tài nguyên
- Khả năng sử dụng các công cụ quen thuộc trên các máy khách
- Truy cập đến dữ liệu từ xa thông qua chuẩn
- Các chức năng đầy đủ của DBMS được cung cấp tới các trạm khách
- Tỷ lệ giá thành/hiệu năng tốt hơn







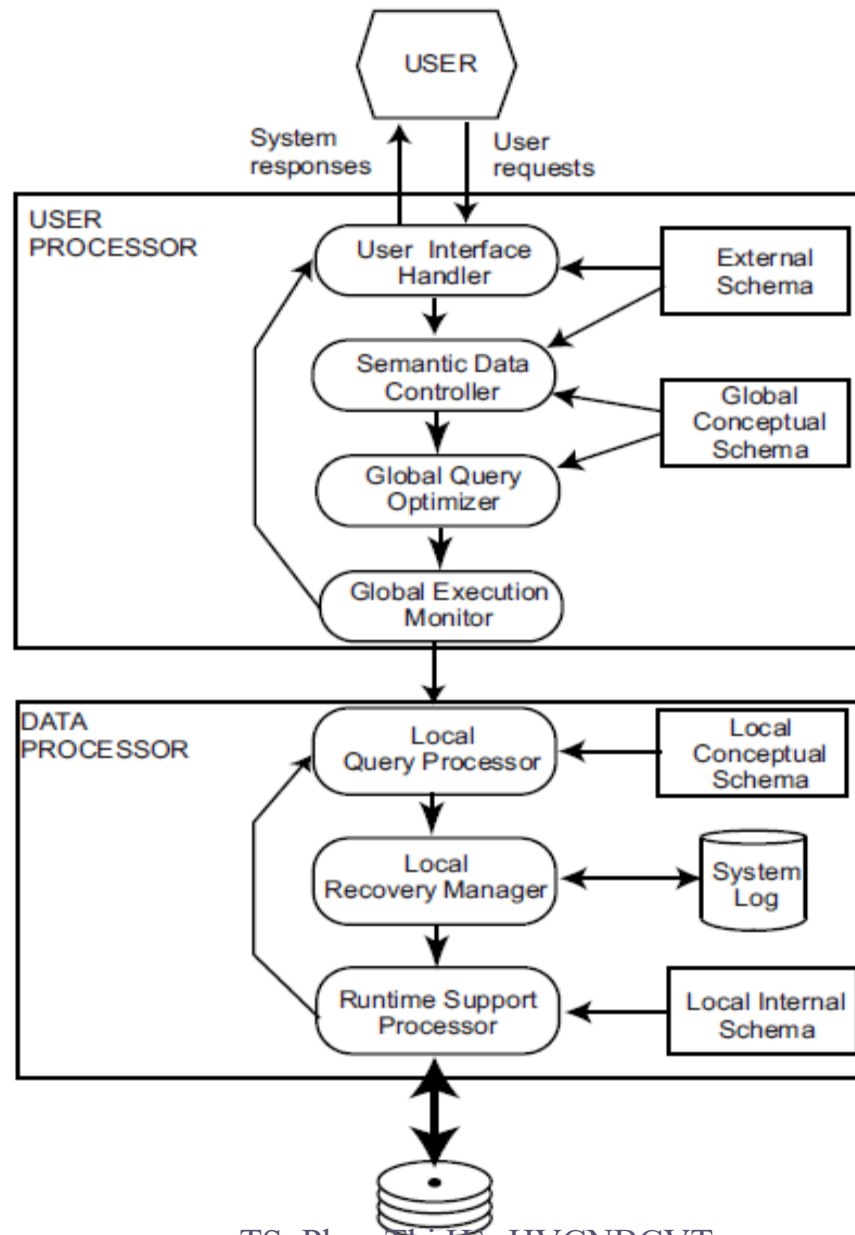
ES: External Schema

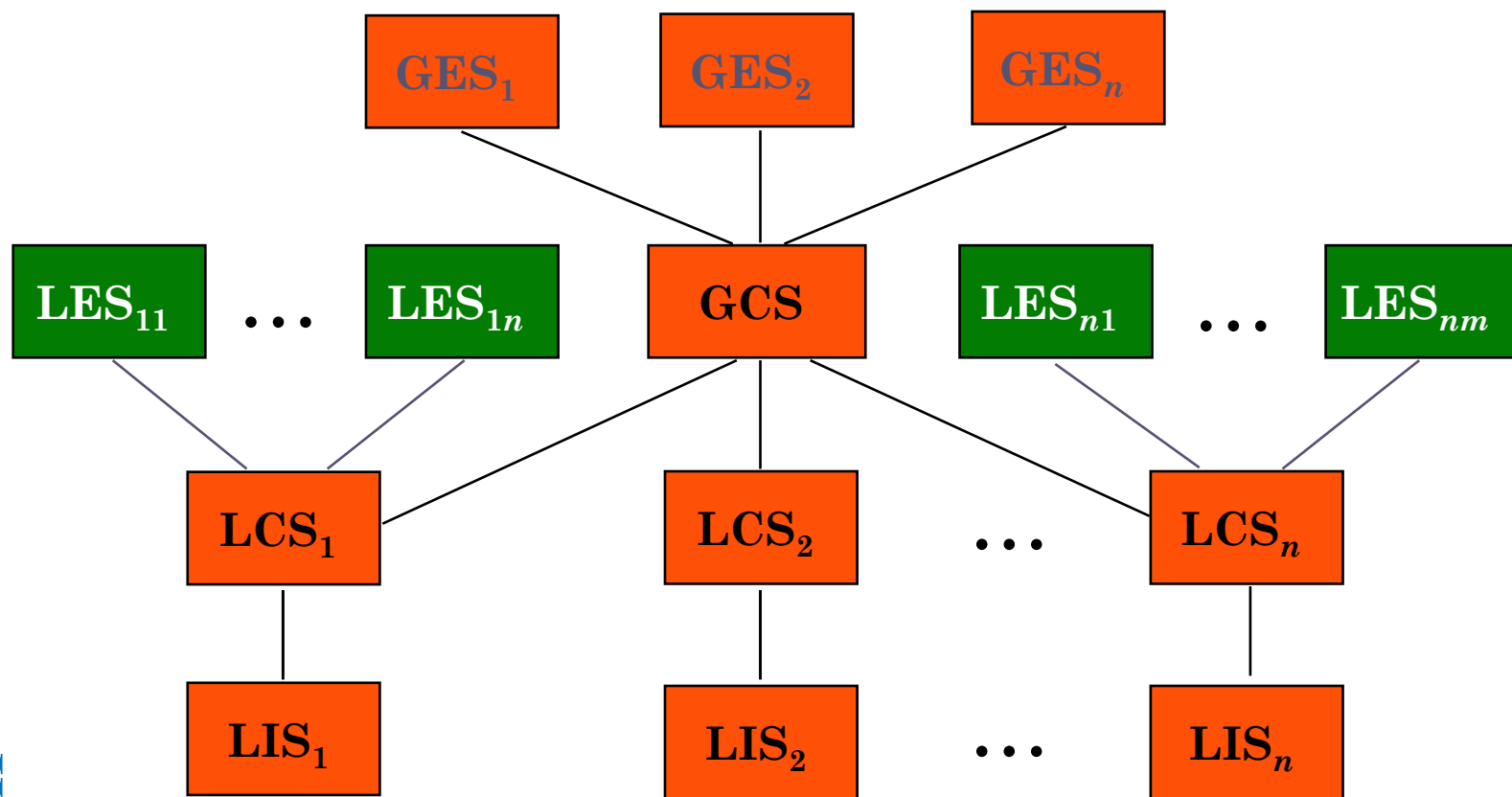
GCS: Global Conceptual Schema

LCS: Local Conceptual Schema

LIS: Local Internal Schema

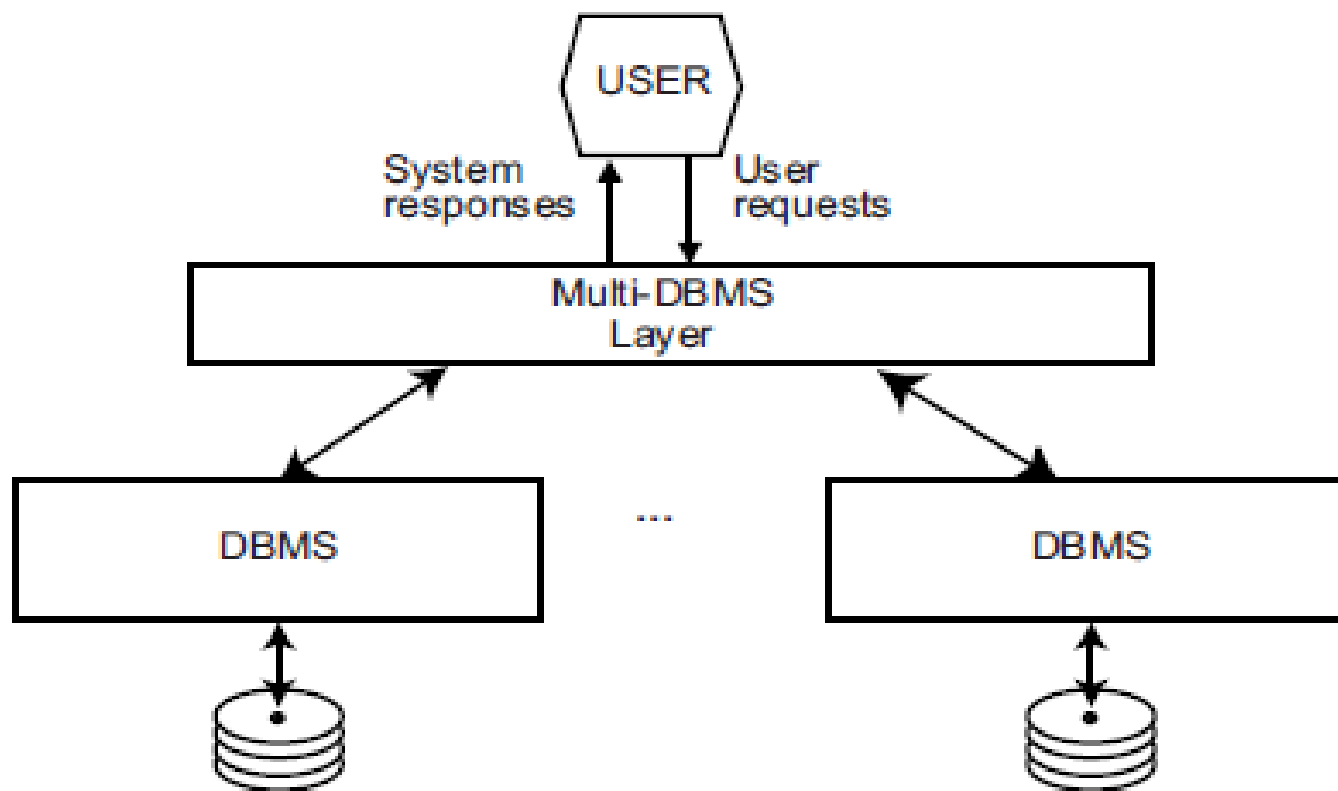
# Kiến trúc ngang hàng theo thành phần





- GES: Global External Schema
- LES: Local External Schema

- LCS: Local Conceptual Schema
- LIS: Local Internal Schema



- Khái niệm chung về CSDL Phân tán
- Các triển vọng của CSDL PT
- Các vấn đề nghiên cứu trong CSDL PT
- Kiến trúc hệ thống
- Kiến trúc
  - Ba chiều cho kiến trúc Hệ CSDLPT: Tính phân tán, Tính tự trị, Tính đồng nhất
- Các kiến trúc khác nhau
  - Các hệ thống ngang hàng
  - Các hệ thống đa Hệ Quản trị CSDL
  - Các hệ thống khách/chủ