

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

CƠ SỞ DỮ LIỆU PHÂN TÁN

TỔNG QUAN CƠ SỞ DỮ LIỆU PHÂN TÁN

Ts. Phan Thị Hà

Học Viên Công nghệ BCVT, Phan Thị
Hà, hathiphan@yahoo.com

✍ Nhóm 5: Nguyễn Đức Thanh 0949484756
nguyenducthanhtik@gmail.com

✍ Nhóm 8: Lê Xuân Hải 0964139918
xuanhaiptit@gmail.com

✍ Nhóm 2(nhóm thầy Nam) Võ Văn Đức
0985957029

✍ lovingyou9695@gmail.com

✍ Nhóm 4: Nguyễn Thị Diệu Linh
01669353568

Học Viện Công nghệ BCVT, Phan Thị
Hà, hathiphan@yahoo.com

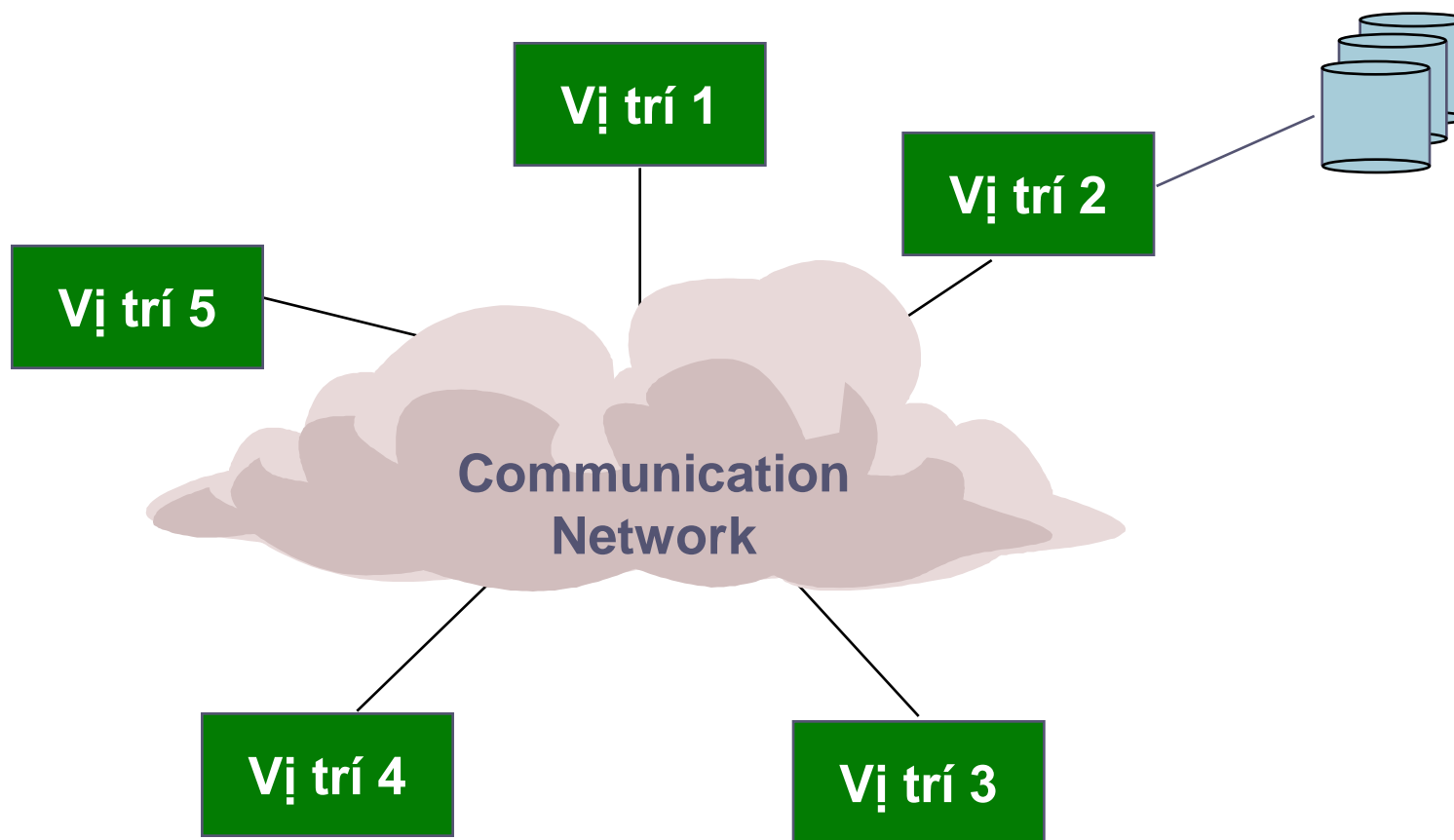
- ✓ **Nhóm 3**
- ✓ **Gmail: thanhtrungnguyen1208@gmail.com**
- ✓ **SDT: 01634473257**

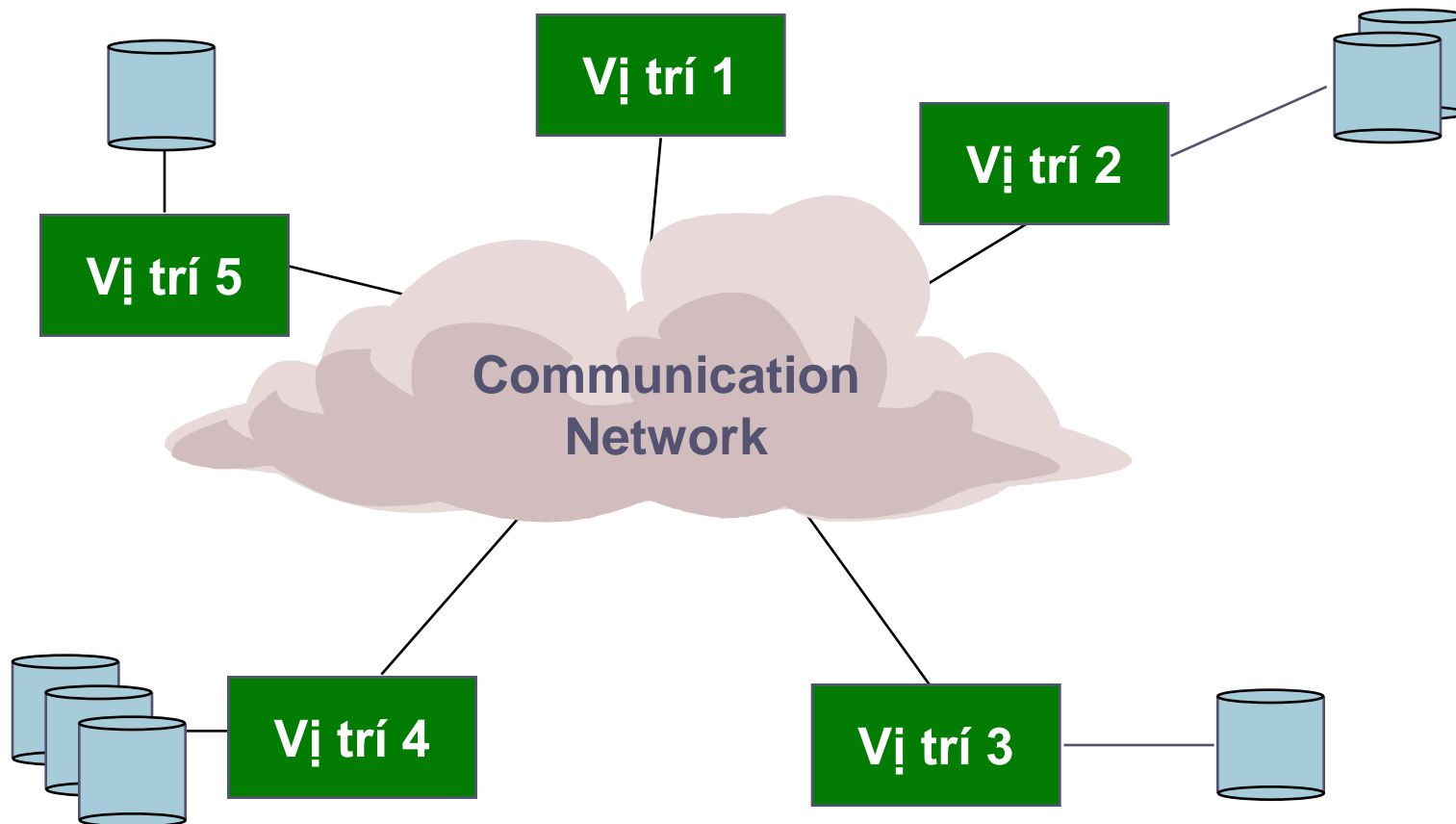
Nội dung

1. Giới thiệu

- ***Tại sao phải phân tán? - Phân tán cái gì?***
- Các tổ chức kinh tế có nhiều trụ sở phân tán ở nhiều vị trí địa lý khác nhau
- Làm thế nào để có thể quản lý các luồng dữ liệu và sử dụng chung dữ liệu, chương trình của các trụ sở trên những vị trí khác nhau đó
- => Nhu cầu dùng chung dữ liệu hợp nhất trên nhiều vị trí địa lý khác nhau đó chính là nhu cầu về hệ thống phân tán

2. Môi trường cơ sở dữ liệu tập trung

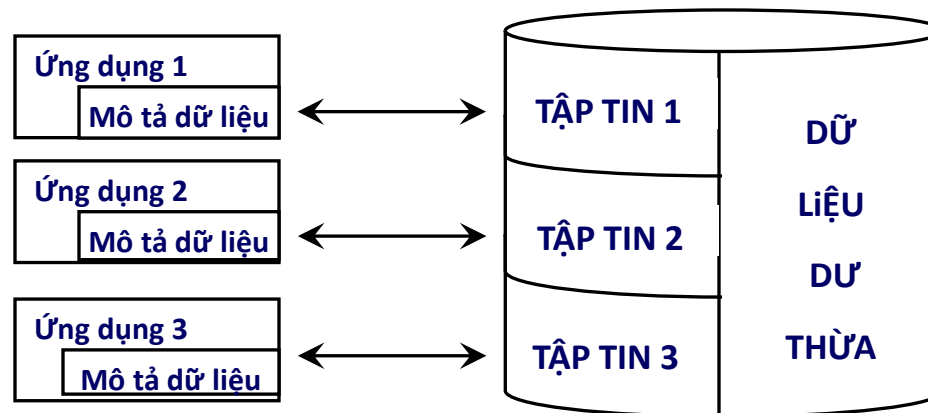




3. Hệ thống phân tán và Xử lý dữ liệu phân tán.

❑ *Xử lý dữ liệu truyền thống*

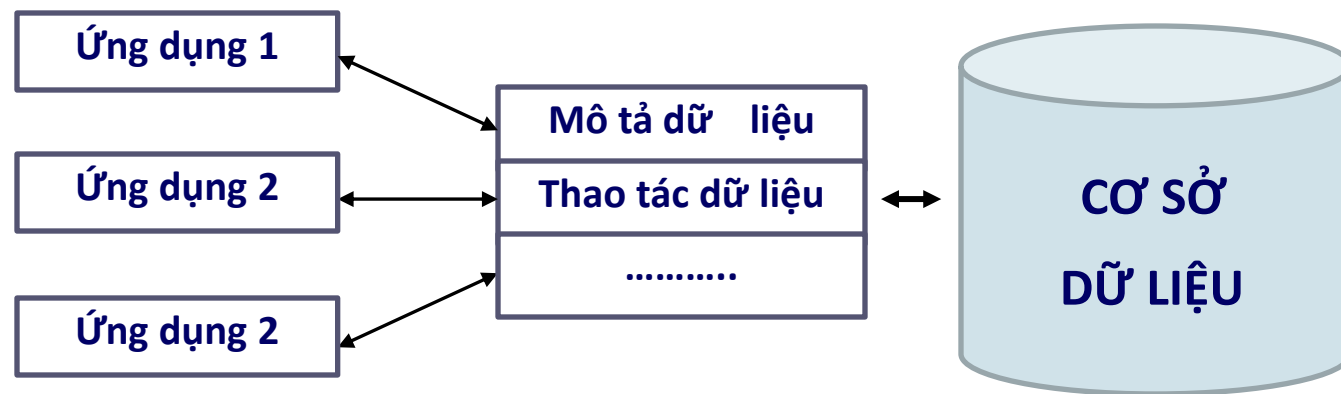
- Mỗi ứng dụng định nghĩa một (nhiều) tệp dữ liệu.
- Ứng dụng khác nhau, tổ chức dữ liệu khác nhau
- Ứng dụng và tổ chức dữ liệu phụ thuộc lẫn nhau.
- Dư thừa dữ liệu



Hình 2.2: Xử lý dữ liệu truyền thống

❑ *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



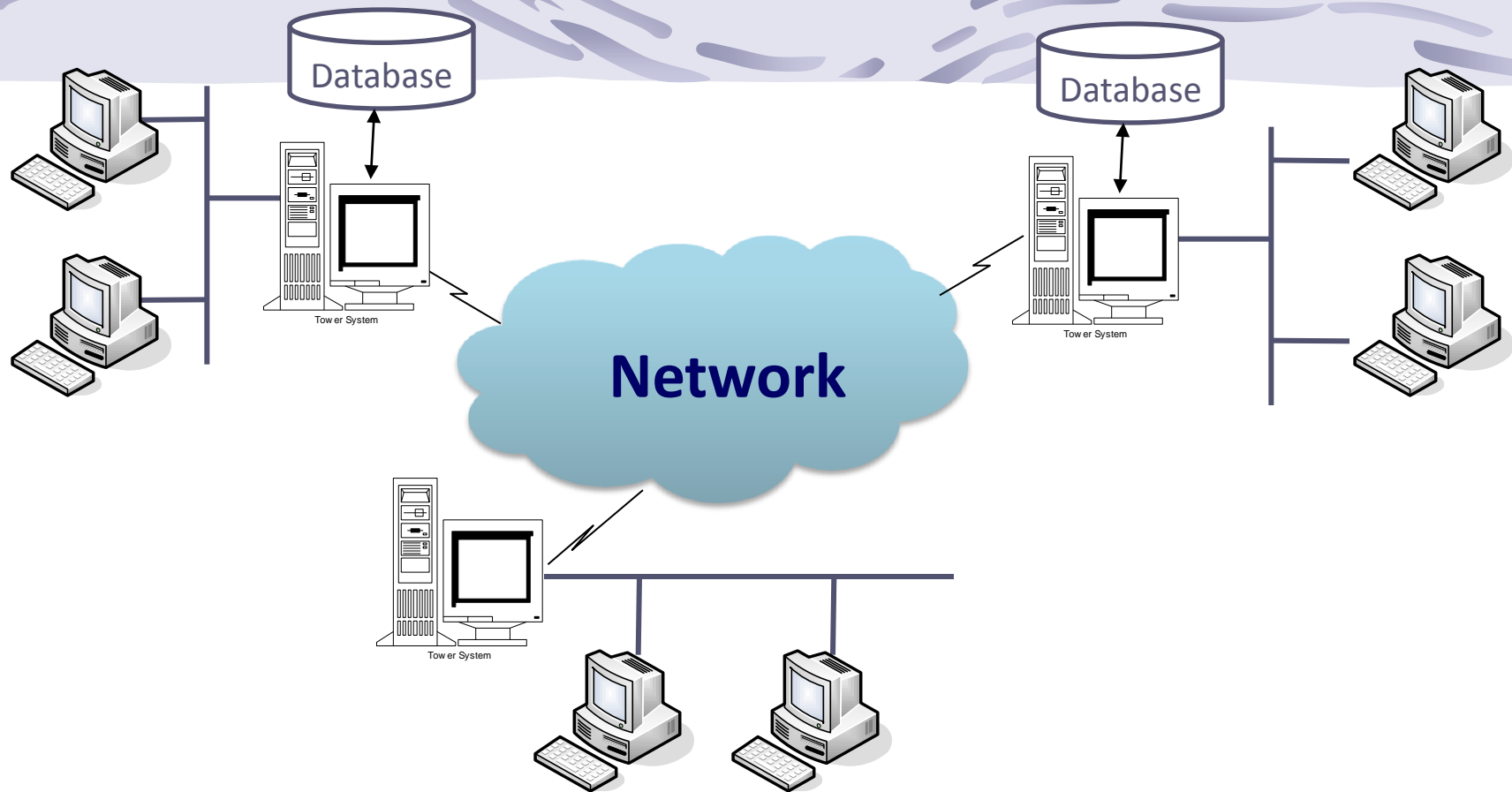
Hình 2.3. Xử lý cơ sở dữ liệu

❑ ***Khái niệm xử lý phân tán***

- Hệ xử lý phân tán là một tập các phần tử xử lý tự trị được kết nối với nhau bởi một mạng máy tính và thực hiện những công việc theo chức năng.

❑ **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.



Hình 2.1 Hệ thống xử lý phân tán

❑ *Hệ thống xử lý phân tán*

Mạng máy tính được cài đặt các cơ sở dữ liệu và các phần mềm hệ thống, tạo khả năng cho nhiều người sử dụng truy nhập chia sẻ nguồn thông tin chung.

=> 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.

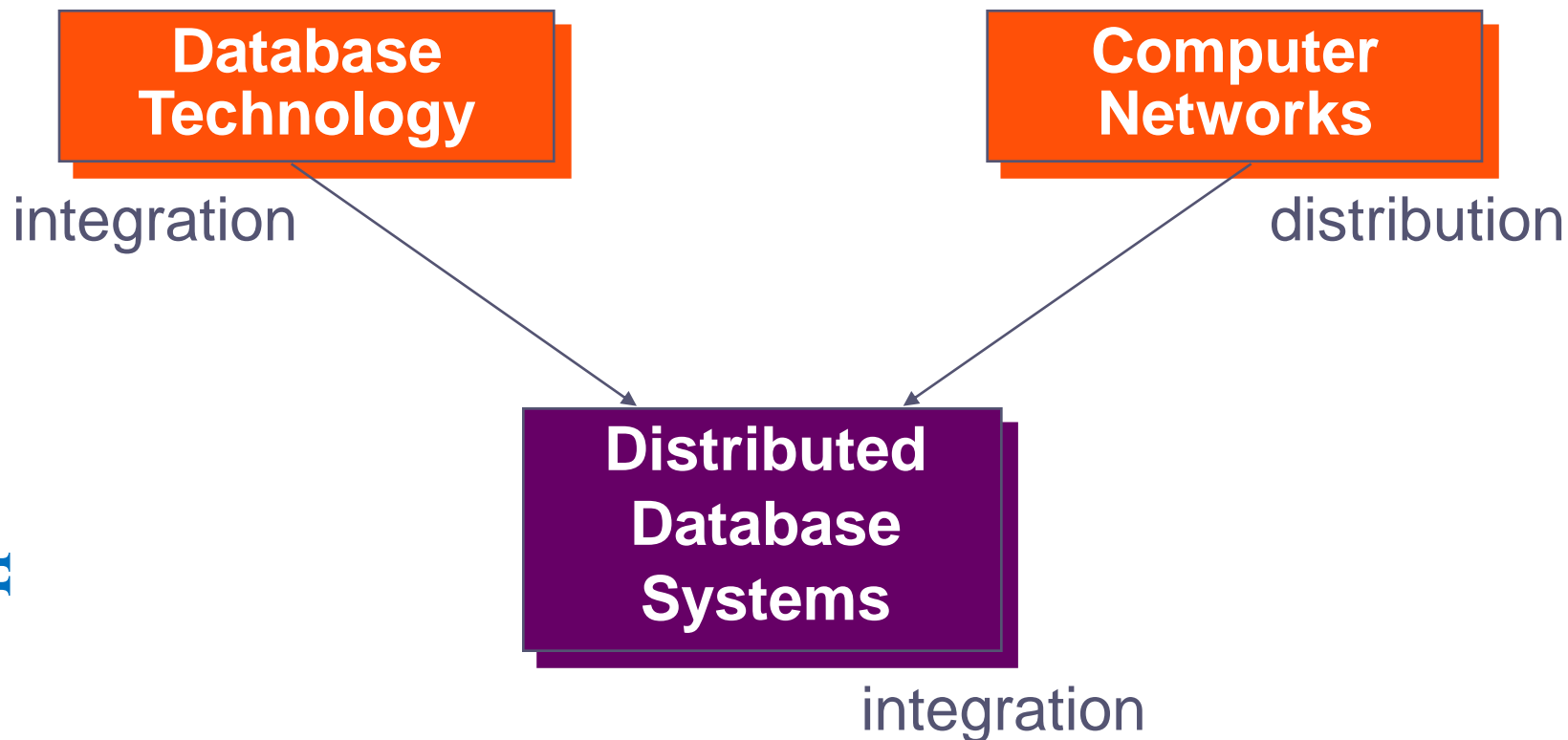
❑ *Các đối tượng phân tán bao gồm*


- Phân tán chức năng (Logic xử lý phân tán): Các chức năng xử lý đc phân tán.
- Phân tán dữ liệu: Dữ liệu được phân tán trên nhiều vị trí khác nhau, nhưng ràng buộc và liên quan đến nhau
- Điều khiển phân tán: Điều khiển các giao dịch khác nhau được phân tán trên nhiều vị

trí

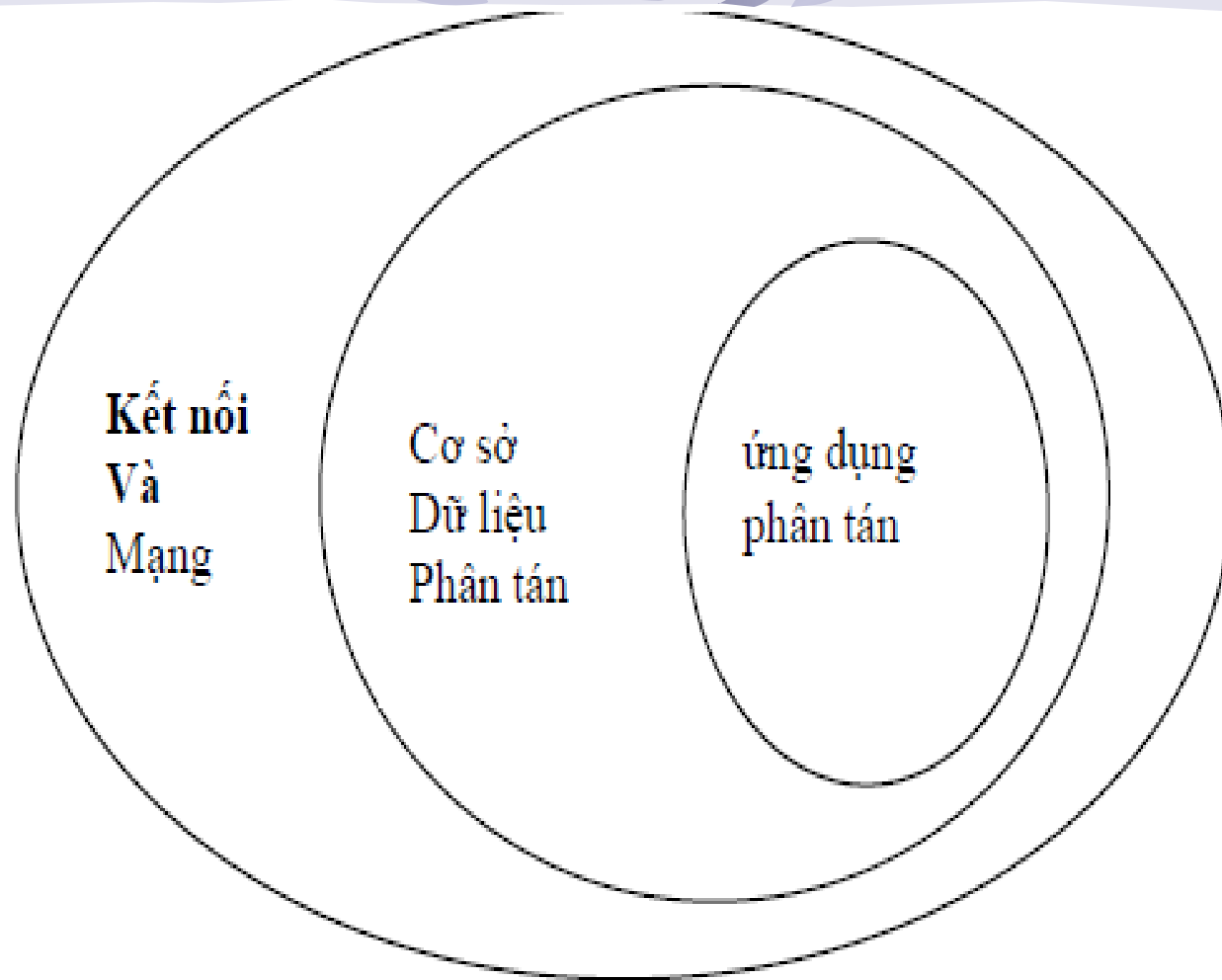
4. Cơ sở dữ liệu phân tán

- ❑ Là một tập các cơ sở dữ liệu có quan hệ với nhau về mặt logic và được phân tán trên nhiều vị trí của 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.





Chú ý: Mỗi vị trí có quyền tự quản cơ sở liệu cục bộ của mình và thực thi các ứng dụng cục bộ. Mỗi vị trí cũng phải tham gia vào việc thực thi ít nhất một ứng dụng toàn cục: yêu cầu truy xuất dữ liệu tại nhiều vị trí qua mạng.



Hình 1.2 Mối liên hệ giữa mạng máy tính, cơ sở dữ liệu phân tán và ứng dụng phân tán

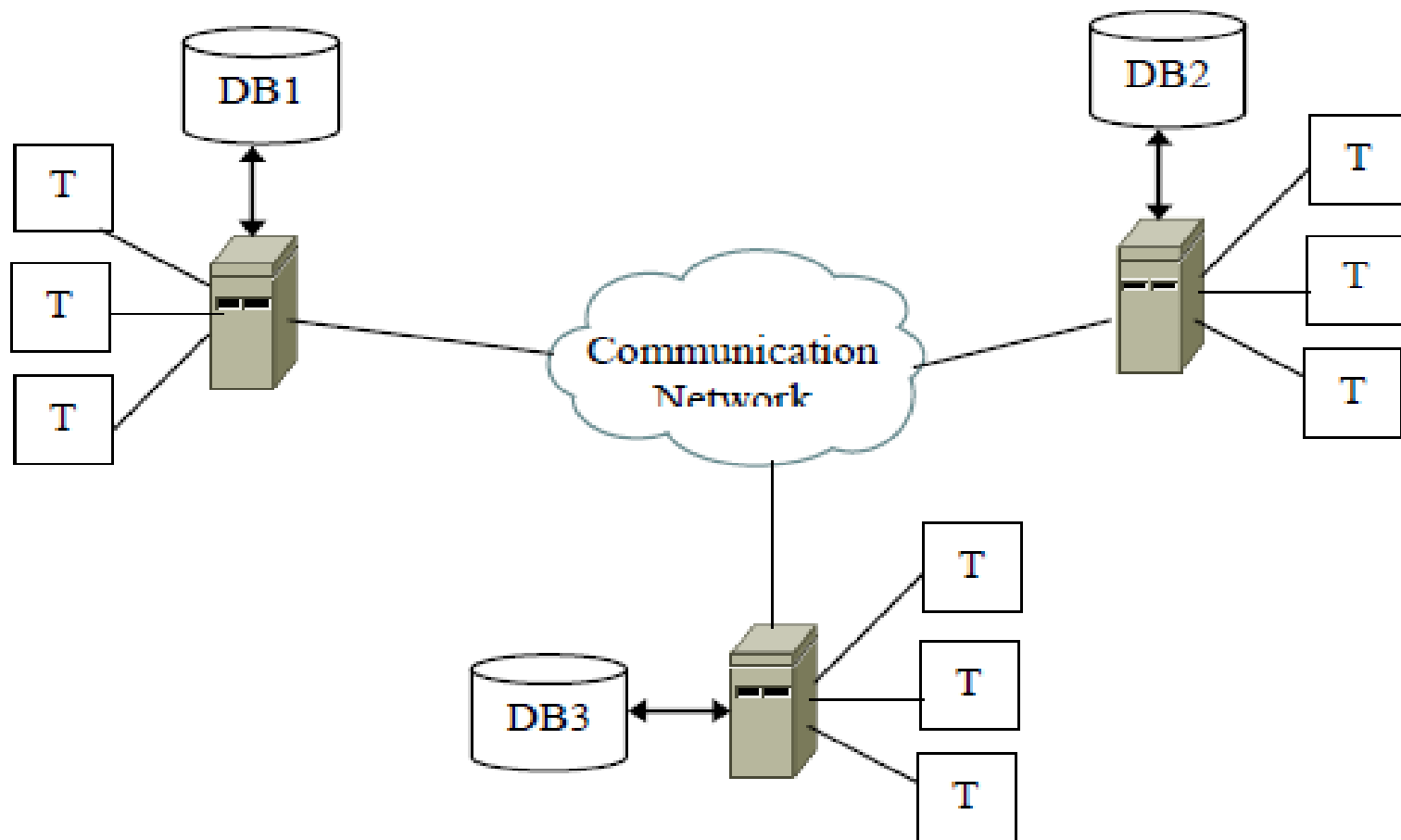
🌿 Vd CSDLPT: Xét một ngân hàng có ba chi nhánh nằm ở ba nơi khác nhau. Tại mỗi chi nhánh, một hệ thống máy tính điều khiển các trạm thu hay rút tiền và quản lý cơ sở dữ liệu về tài khoản. Mỗi hệ thống này với cơ sở dữ liệu tài khoản cục bộ tạo thành một site của cơ sở dữ liệu phân tán. Các hệ thống máy tính này được kết nối bởi một mạng truyền thông.

🌿 Với những hoạt động thông thường, các yêu cầu từ các trạm chỉ cần truy xuất đến cơ sở dữ liệu tại chi nhánh của chúng. Vì thế ứng dụng này được gọi là ứng dụng cục bộ.

- ✎ Với ví dụ trên này sinh hai câu hỏi sau: mỗi nhánh chỉ lưu trữ cơ sở dữ liệu cục bộ có đủ chưa? Cơ sở dữ liệu phân tán có phải là một tập các cơ sở dữ liệu cục bộ?
- ✎ Để trả lời các câu hỏi này chúng ta tìm hiểu xem việc xử lý trên cơ sở dữ liệu cục bộ khác gì trên cơ sở dữ liệu phân tán. Về mặt kỹ thuật, chúng ta thấy cần có các ứng dụng mà truy xuất dữ liệu đang đặt ở nhiều nhánh. Các ứng dụng này được gọi là ứng dụng toàn cục hay ứng dụng phân tán.

Một ứng dụng toàn cục thông thường trong ví dụ trên là việc chuyển tiền từ một tài khoản này đến tài khoản khác. Ứng dụng này yêu cầu cập nhật cơ sở dữ liệu ở cả hai nhánh.

Hơn nữa ứng dụng toàn cục giúp cho người sử dụng không phân biệt được dữ liệu đó cục bộ hay từ xa. Đó là tính trong suốt dữ liệu trong cơ sở dữ liệu phân tán. Và đương nhiên khi ứng dụng toàn cục truy cập dữ liệu cục bộ sẽ nhanh hơn ứng dụng từ xa điều này nói lên sự nhân bản dữ liệu ở các nơi cũng làm tăng tốc độ xử lý chương trình.



Hình 1.1 Cơ sở dữ liệu phân tán của ngân hàng có ba chi nhánh

5. Đặc điểm cơ sở dữ liệu phân tán :Điều khiển tập trung

□ Điều khiển tập trung

- tập trung: đảm bảo tính độc lập của dữ liệu. tránh tổn thất thông tin, khả năng đk tập trung trên các tài nguyên thông tin. Sự điều khiển được thực hiện bởi Người QTCSDL
- phân tán:-Cấu trúc điều khiển phân cấp: quản trị cơ sở dữ liệu toàn cục và quản trị cơ sở dữ liệu cục bộ phân tán.Sự điều khiển được thực hiện bởi Người QTCSDL toàn cục (Global Database Administrator) và Người QT 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

□ *Độc lập dữ liệu*

- Tập trung: Thể hiện thông qua kiến trúc nhiều mức, các mức này có những mô tả các nhau về dl và những ax biến đổi giữa các mức. Tức là tổ chức dl trong suốt với các lập trình viên. Các chương trình được viết có cái nhìn “quan niệm” về dữ liệu

Lợi ích: Các Ctud không bị thay đổi khi thay đổi tổ chức lưu trữ vật lý của dl.

- phân tán: Là sự trong suốt phân tán (Distribution Transparency), nghĩa là tính đúng đắn của c các chương trình được viết như cơ sở dữ liệu không hề được phân tán.=> Tuy nhiên, tốc độ thực hiện của các trình ứng dụng bị ảnh hưởng khi phải trao đổi dữ liệu giữa các vị trí.

Đặc điểm CSDL phân tán: Giảm dư thừa dữ liệu

□ Giảm dư thừa dữ liệu

-tập trung, Giảm thiểu sự dư thừa dữ liệu do:

Tính nhất quán dữ liệu cao

Tiết kiệm dung lượng nhớ..

- phân tán, sự dư thừa theo ngữ nghĩa sau: Giảm thiểu dư thừa dl đảm bảo tính nhất quán

- Tăng tính cục bộ: Dữ liệu có thể được nhân bản tại tất cả các vị trí=>ứng dụng cục bộ thực hiện nhanh hơn.

- Tăng tính sẵn sàng: Dữ liệu được nhân bản tại vt nào đó. Khi một vị trí có sự cố thì sẽ không làm ngưng sự thực hiện của các ứng dụng ở những vị trí khác .=>Từ đó vấn đề quản

lý nhất quán dữ liệu sẽ phức tạp hơn

Chú ý: Khi nào sử dụng “dự thừa dữ liệu” nhân bản có nhiều ưu điểm hơn???

- 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

Đặc điểm CSDL phân tán:

□ . Các cấu trúc vật lý phức tạp và truy xuất hiệu quả


TT: Các cấu trúc vật lý phức tạp giúp cho việc truy xuất dữ liệu được hiệu quả.

PT: Các cấu trúc vật lý phức tạp giúp liên lạc dữ liệu trong cơ sở dữ liệu phân tán .

Đặc điểm : Tính toàn vẹn, phục hồi, đồng thời

TT: Dựa vào giao tác.

PT: Dựa vào giao tác phân tán.



=> Từ các đặc trưng trên, chúng ta thấy việc chọn lựa cơ sở dữ liệu phân tán sẽ thích hợp hơn đối với các ứng dụng phát triển trong một hệ thống mạng diện rộng do giảm được chi phí truyền thông để truy xuất dữ liệu.

Tại sao cần có cơ sở dữ liệu phân tán ?

- *Lý do tổ chức kinh tế?*
- *Lý do kết nối các cơ sở dữ liệu hiện có?*
- *Lý do tăng trưởng tổ chức*
- *Lý do tải truyền thông*
- *Đánh giá về hiệu suất*
- *Độ tin cậy và tính hiệu quả*

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.

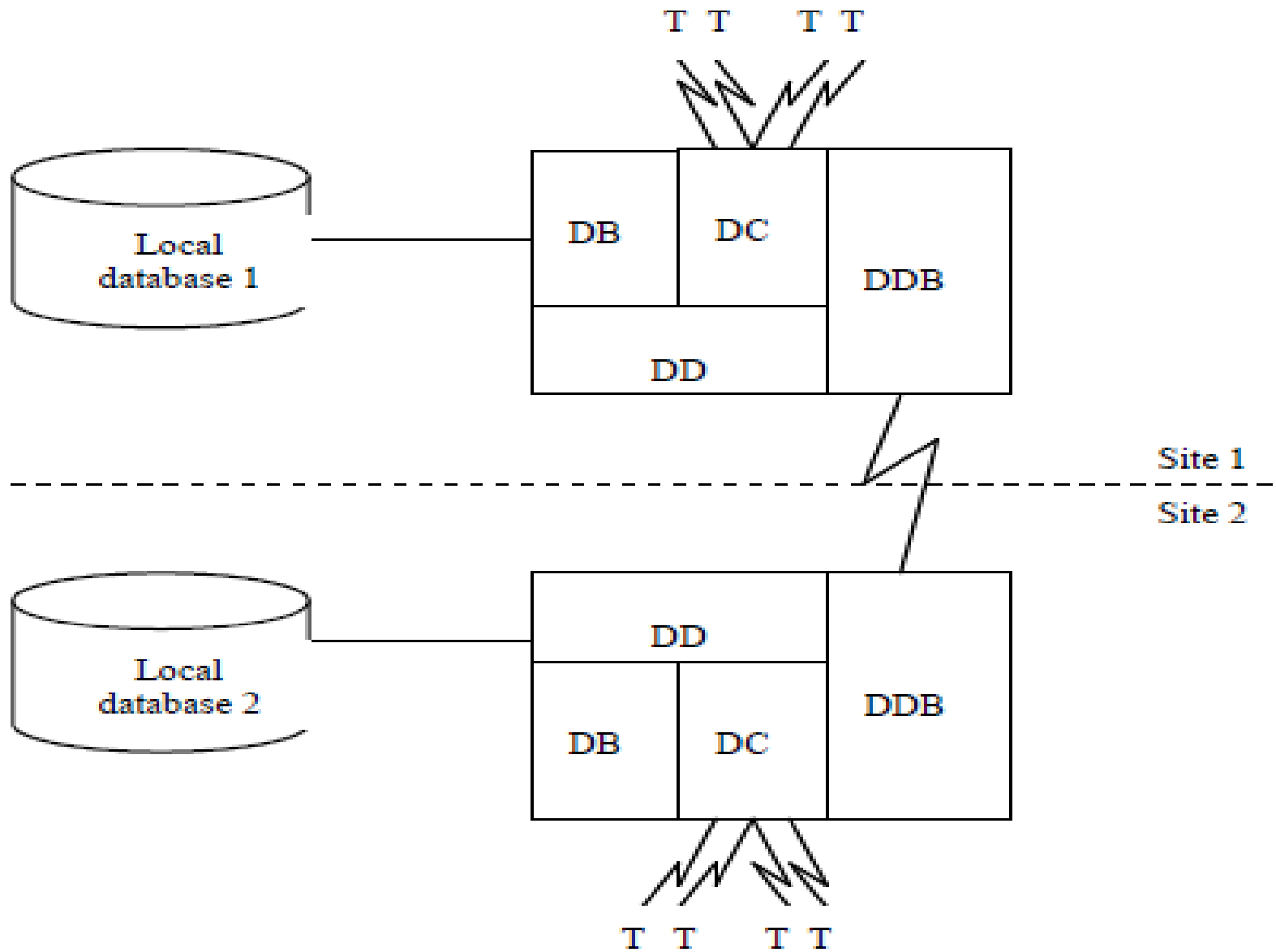
Chi phí xử lý tăng: Sự trao đổi các thông báo và xử lý phối hợp giữa các vị trí sẽ tăng chi phí xử lý hơn trong các hệ thống tập trung.

Hệ quản trị cơ sở dữ liệu phân tán

Hệ quản trị cơ sở dữ liệu phân tán hỗ trợ việc tạo và duy trì cơ sở dữ liệu phân tán. Các hệ quản trị cơ sở dữ liệu phân tán hiện nay được phát triển bởi các nhà sản xuất các hệ quản trị cơ sở dữ liệu tập trung. 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 cho việc xây dựng một cơ sở dữ liệu phân tán là :

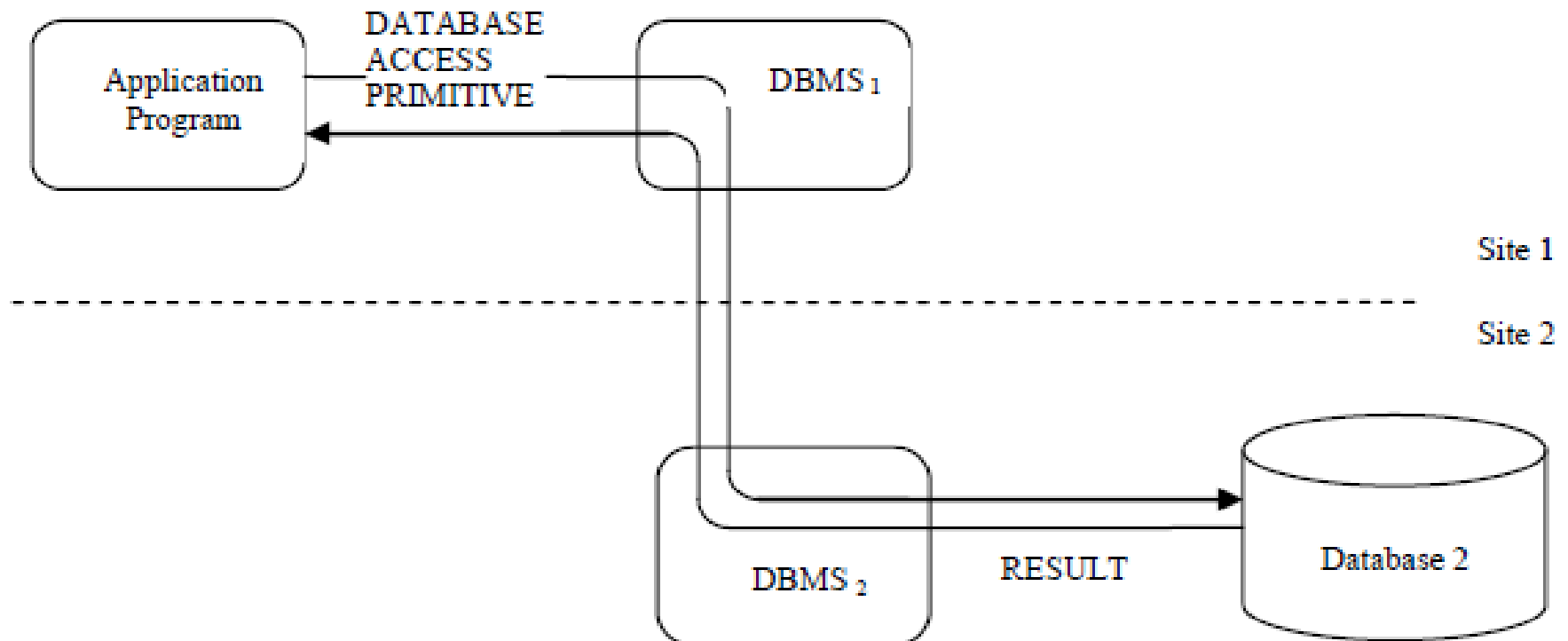
- 1. Thành phần quản trị cơ sở dữ liệu (DB: Database Management)
- 2. Thành phần truyền dữ liệu (DC: Data Communication)
- 3. Tự điển dữ liệu (DD : Data Dictionnary) mở rộng để biểu diễn thông tin về sự phân tán dữ liệu trên mạng.
- 4. Thành phần cơ sở dữ liệu phân tán (DDB: Distributed Database)



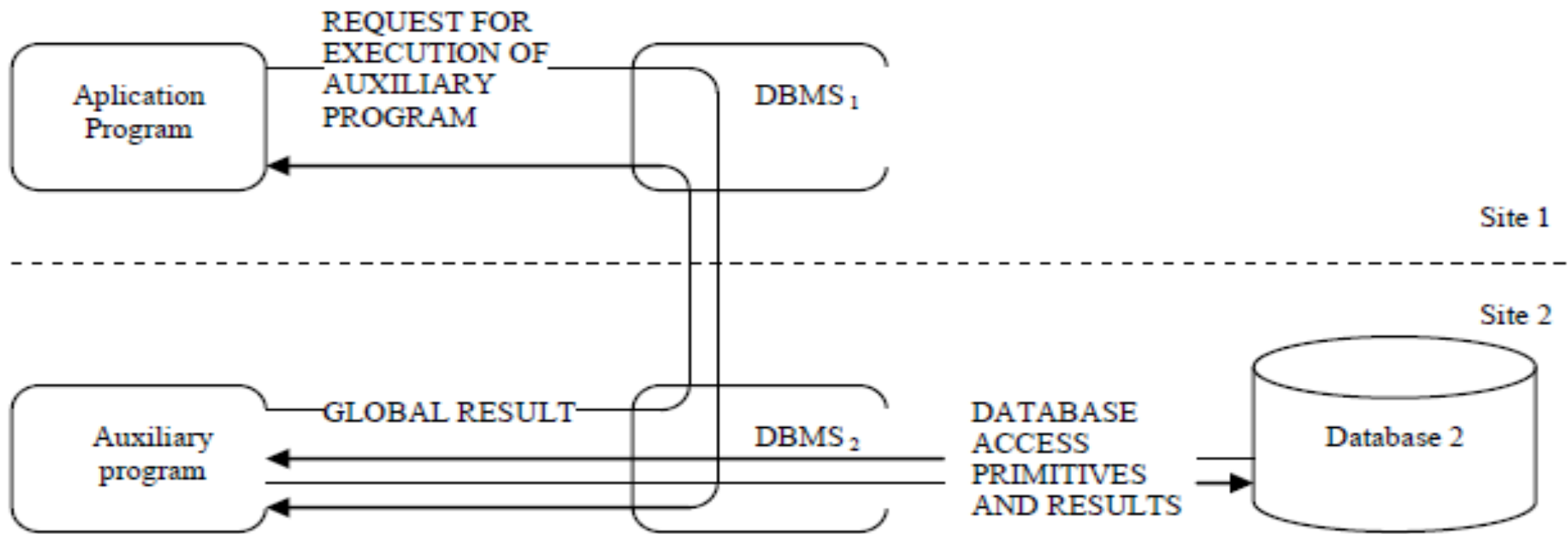
Các dịch vụ được hỗ trợ cho hệ quản trị CSDLPT

- Dịch vụ truy xuất cơ sở dữ liệu từ xa: tính chất này là một tính chất quan trọng nhất và được cung cấp bởi tất cả các hệ thống có thành phần cơ sở dữ liệu phân tán.
- Mức độ trong suốt của sự phân tán: tính chất này được hỗ trợ bởi các hệ thống khác nhau vì đó là sự cân bằng các yếu tố để đạt được sự kết hợp tốt nhất giữa sự trong suốt phân tán và hiệu suất.
- Hỗ trợ việc quản trị và điều khiển cơ sở dữ liệu: tính chất này bao gồm các công cụ để giám sát cơ sở dữ liệu, lấy thông tin về việc sử dụng cơ sở dữ liệu, cung cấp một cái nhìn toàn cục về các file dữ liệu lưu trữ trên các vị trí khác nhau.
- Hỗ trợ cho việc điều khiển đồng thời và phục hồi các giao tác phân tán

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



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

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

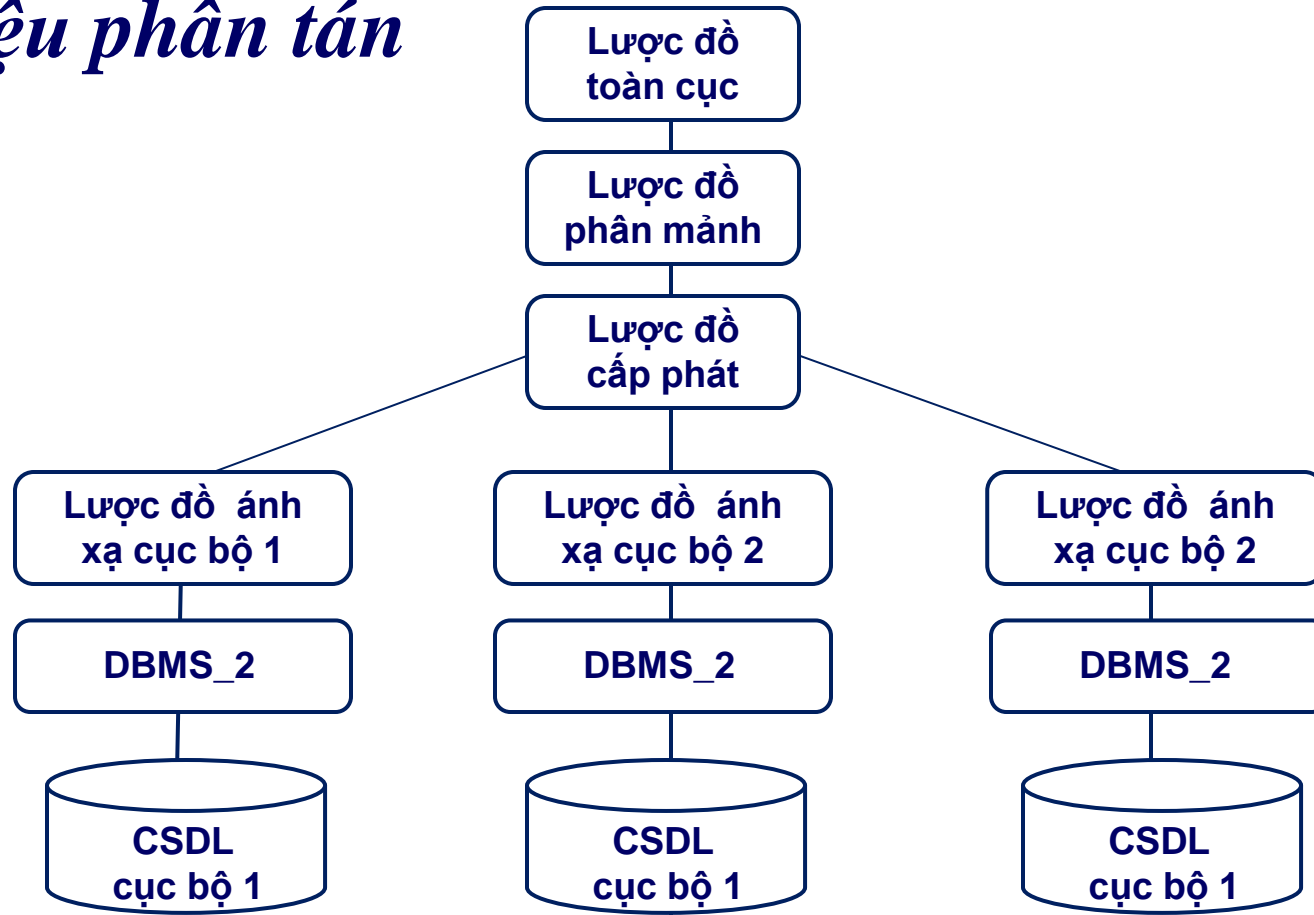
- Các CS 1 cung cấp sự trong suốt phân tán nhiều hơn trong khi cách tiếp cận thứ hai có thể linh động hơn nếu nhiều truy xuất cơ sở dữ liệu được yêu cầu vì ứng dụng hỗ trợ có thể thực hiện tất cả các truy xuất yêu cầu và chỉ gửi kết quả về.
- HQTCSDLT trong hệ thống là chúng cùng loại hay khác loại. Các hệ quản trị cơ sở dữ liệu phân tán khác loại phải thêm vấn đề thông dịch giữa các mô hình dữ liệu khác nhau, các cấu trúc dữ liệu khác nhau. Đây là một vấn đề rất khó giải quyết, nên nó được khắc phục bằng cách hỗ trợ sự truyền thông giữa các thành phần truyền thông dữ liệu (data communication component DC) khác nhau.
- Bài toán này cũng đã được công ty Microsoft giải quyết bằng các thành phần truy xuất dữ liệu (Microsoft Data Access Components (MDAC)). Cho nên một hệ thống bao gồm các hệ quản trị cơ sở dữ liệu cục bộ khác nhau sẽ thích hợp hơn cho việc phát triển hệ thống thông tin một cách linh động và tự trị

7. Mô hình kiến trúc của cơ sở dữ liệu phân tán

(Mô hình tham chiếu)

- ❑ 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.

a. Mô hình kiến trúc tham chiếu cho hệ cơ sở dữ liệu phân tán



Hình 2.4 Mô hình kt tham chiếu cho 1 cơ sở dữ liệu

Lược đồ toàn cục (Mô hình tham chiếu 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 bao gồm thông tin về các thực thể, tập các phụ thuộc hàm và mối quan hệ giữa các thực thể .

- 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 ảnh xạ tới các mức của

Lược đồ phân mảnh (Mô hình tham chiếu phân mảnh)

- ❑ Các mảnh: Quan hệ toàn cục có thể chia thành nhiều quan hệ con được gọi là các mảnh, tách biệt với nhau nhau .(Các phân mảnh là các thành phần của các quan hệ toàn cục mà được lưu trữ vật lý tại một hay một số địa điểm).
- ❑ Lược đồ phân mảnh:Là ánh xạ giữa quan hệ toàn cục và các 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

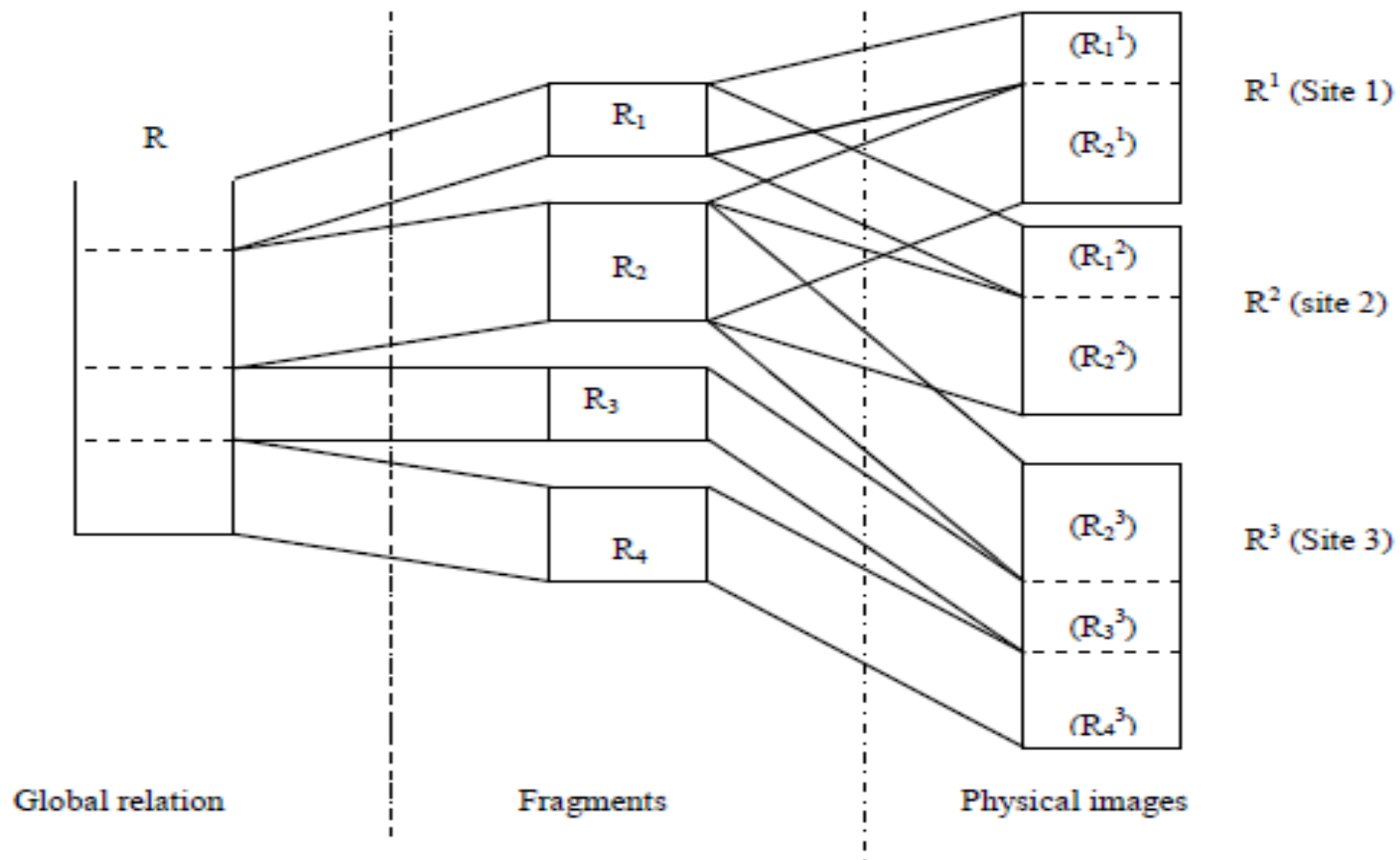
- ❑ Các kiểu phân mảnh dữ liệu bao gồm:
 - Phân mảnh ngang
 - Phân mảnh dọc
 - Sự kết hợp của phân mảnh ngang và phân mảnh dọc, phức tạp hơn
- ❑ Một mảnh được định nghĩa bằng một biểu thức đại số quan hệ, các toán hạng là các quan hệ toàn cục.

Lược đồ cấp phát=>

- ❑ Lược đồ cấp phát: chứa thông tin các mảnh được chứa trên những site nào (Các mảnh được chứa ở một hay nhiều site trên mạng).
- ❑ Các mảnh là các phần logic của một quan hệ tổng thể được định vị vật lý trên một hay nhiều trạm.
- ❑ Tất cả các mảnh được liên kết với cùng một quan hệ tổng thể R và được định vị tại cùng một trạm j cấu thành ảnh vật lý quan hệ tổng thể R tại trạm j. Do đó ta có thể ánh xạ một-một giữa một ảnh vật lý và một cặp (quan hệ tổng thể, site). Các ảnh vật lý có thể chỉ ra bằng tên của một quan hệ tổng thể và một chỉ mục trạm.

Lược đồ cấp phát

- ❑ Ký hiệu R_{ij} là mảnh thứ i của quan hệ R trên site j .
- ❑ Quan hệ toàn cục R được phân thành 3 mảnh quan hệ con R_1, R_2, R_3, R_4 , và được cấp phát trên 3 site của mạng máy tính, tạo nên ba mô hình vật lý R^1, R^2, R^3 :
 - Trên site 1: Một bản sao của R_1 và một bản sao của R_2 . Ký hiệu R_1^1 và R_2^1
 - Trên site 2: Một bản sao của R_1 và một bản sao của R_2 . Ký hiệu R_1^2 và R_2^2
 - Trên site 3: Một bản sao của R_2, R_3 và R_4 . Ký hiệu R_2^3, R_3^3, R_4^3



Hình 1.6 Các phân mảnh và các ảnh vật lý đối với một quan hệ toàn cục

Lược đồ ánh xạ cục bộ

✎ . Ở mức thấp hơn, cần ánh xạ các ảnh vật lý đến các đối tượng được thao tác bởi các hệ quản trị cơ sở dữ liệu cục bộ. Ánh xạ này được gọi là lược đồ ánh xạ cục bộ và nó phụ thuộc vào kiểu của hệ quản trị cơ sở dữ liệu cục bộ; vì thế trong hệ thống không đồng nhất, có nhiều kiểu ánh xạ cục bộ tại các vị trí khác nhau.

Lược đồ ánh xạ cục bộ=>

- Thực hiện ánh xạ các ảnh vật lý đến các đối tượng được thao tác bởi hệ quản trị cơ sở dữ liệu địa phương (Tất cả các mảnh của một quan hệ tổng thể trên cùng một trạm tạo ra ảnh vật lý)
- Ánh xạ này được gọi là ánh xạ cục bộ, nó phụ thuộc vào kiểu của DBMS.
- (Trong ht đồng nhất lđ độc lập của mỗi site đc đn sử dụng cùng 1 mô hình như DBMS cục bộ, Trong hệ thống không đồng nhất tại mỗi site thì ánh xạ cục bộ khác nhau, khi đó lược đồ axcb còn dùng để phối hợp các kiểu khác nhau của các DBMS không đồng nhất)

Ba đối tượng quan trọng nhất của mô hình kiến trúc tham chiếu CSDLPT

- Sự tách biệt giữa sự phân mảnh dữ liệu sự định vị dữ liệu
- Điều khiển dư thừa dữ liệu
- Tính độc lập ở các hệ quản trị cơ sở dữ liệu cục bộ.

Sự tách biệt quan niệm phân mảnh dữ liệu và quan niệm định vị dữ liệu (Trong suốt phân tán và trong suốt định vị)

- Giúp tách biệt hai mức khác nhau của sự trong suốt đó là sự trong suốt phân tán và sự trong suốt định vị.
- Sự trong suốt phân tán là mức độ cao nhất của sự trong suốt và bao gồm các yếu tố mà người sử dụng và các lập trình viên làm việc trên các quan hệ toàn cục.
- Sự trong suốt định vị là mức thấp hơn và yêu cầu người sử dụng và các lập trình viên làm việc trên các phân mảnh thay vì trên các quan hệ toàn cục. Tuy nhiên họ không cần biết các phân mảnh này lưu trữ ở đâu.
- Sự tách biệt hai quan niệm phân mảnh và định vị rất phù hợp trong thiết kế cơ sở dữ liệu phân tán vì sự xác định các thành phần thích hợp của dữ liệu được nhận biết từ bài toán định vị tối

Điều khiển tường minh sự dư thừa dữ liệu:

- ☛ Sự dư thừa dữ liệu tại mức phân mảnh. Ví dụ trong hình 1.6 hai ảnh vật lý R^2_2 và R^3_2 trùng lặp nghĩa là chúng chứa chung dữ liệu (nhân bản).
- ☛ Định nghĩa các phân mảnh một cách tách biệt khi xây dựng các khối vật lý cho phép tham khảo tường minh đến từng phần trùng lặp này tức phân mảnh nhân bản R_2 . Điều khiển sự dư thừa dữ liệu rất hữu dụng trong một số khía cạnh quản trị cơ sở dữ liệu phân tán

Tính độc lập tại các hệ quản trị cơ sở dữ liệu cục

- Cho phép nghiên cứu một số vấn đề quản trị cơ sở dữ liệu mà không quan tâm đến mô hình dữ liệu cụ thể tại các hệ quản trị cơ sở dữ liệu cục bộ (Sự trong suốt ánh xạ cục bộ)
- Một kiểu trong suốt khác liên quan chặt chẽ tới sự trong suốt định vị là sự trong suốt nhân bản. Sự trong suốt nhân bản có nghĩa là người sử dụng không nhận thấy được sự nhân bản của các phân mảnh.

Kiến trúc của hệ quản trị CSDL phân tán

☛ Các kiến trúc hệ thống là:

Hệ khách/chủ (client/server)

Các hệ cơ sở dữ liệu phân tán ngang hàng

Các phức hệ cơ sở dữ liệu

Hệ khách/chủ (client/server)

- Phân biệt chức năng cần được cung cấp và chia những chức năng này thành hai lớp: chức năng chủ, chức năng khách.
- Cung cấp một kiến trúc hai mức, tạo dễ dàng cho việc quản lý mức độ phức tạp của các hệ quản trị cơ sở dữ liệu hiện đại và độ phức tạp của việc phân tán dữ liệu.

=> 1. Phải chú ý là máy chủ thực hiện phần lớn các công việc quản lý dữ liệu. Điều này có nghĩa là mọi việc xử lý và tối ưu hóa vấn tin, quản lý giao tác và quản lý thiết bị lưu trữ đều được thực hiện tại máy chủ.

2. Khách, ngoài giao diện và ứng dụng, sẽ có một module quản trị cơ sở dữ liệu, khách chịu trách nhiệm quản lý dữ liệu được gửi đến và đôi khi cả việc quản lý các khoá chốt giao tác.



➤ Kiến trúc khách/chủ được biểu diễn trong hình sau.

Client

Hệ	Giao tiếp người dùng	Chương trình ứng dụng	...
Điều	Client DBMS		
Hành	Phần mềm truyền thông		

Truy vấn

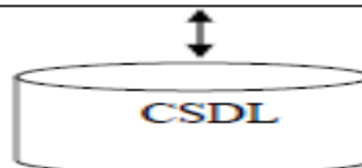
SQL

Kết quả

quan hệ

Server

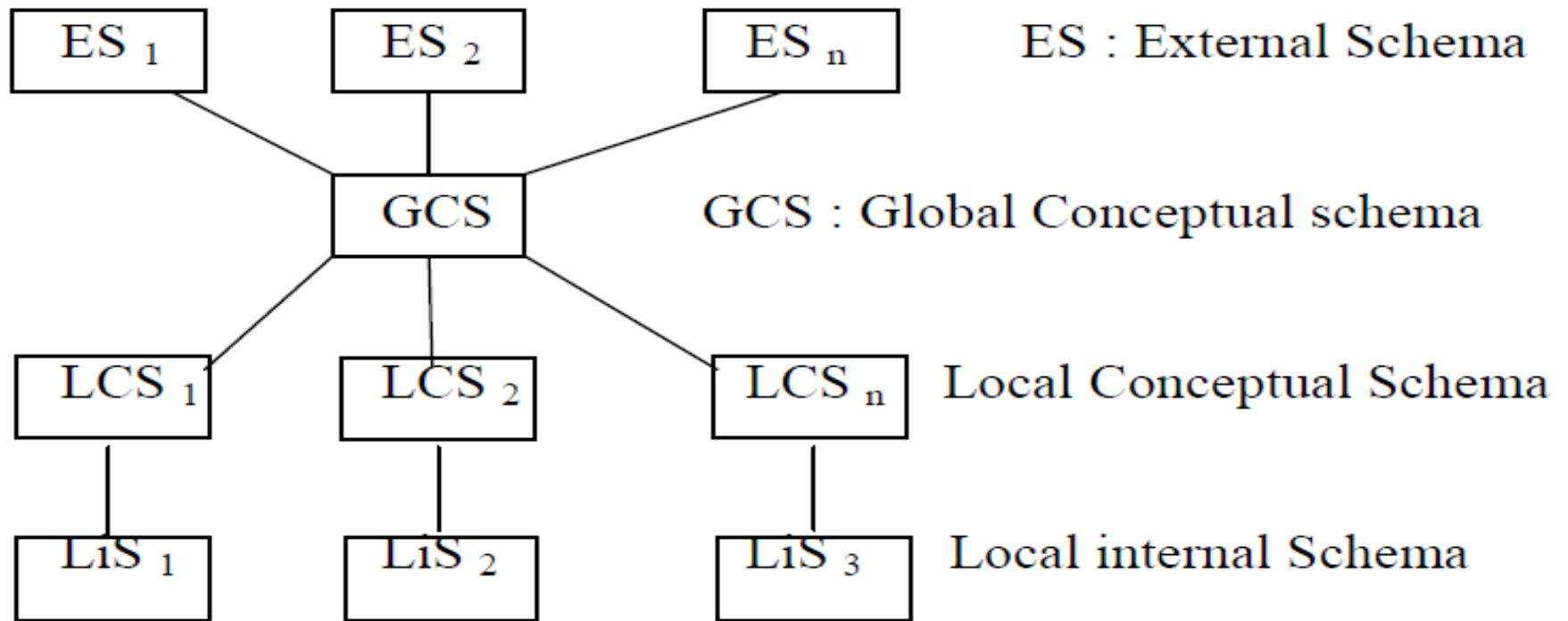
Hệ	Phần mềm truyền thông
	Bộ điều khiển ngữ nghĩa dữ liệu
	Bộ tối ưu hóa câu hỏi
	Bộ quản lý giao tác
Điều	Bộ phục hồi
	Bộ xử lý hỗ trợ run time
Hành	



Hệ phân tán ngang hàng

- ✦ LIS (Local Internal Schema): định nghĩa nội tại riêng tại mỗi vị trí được gọi là lược đồ cục bộ
- ✦ GCS (Global Conceptual Schema): Hình ảnh của mô hình dữ liệu toàn cục được mô tả bằng lược đồ quan niệm toàn cục
- ✦ LCS (Local Conceptual Schema): Để xử lý hiện tượng nhân bản và phân mảnh, mô tả việc tổ chức logic của dữ liệu tại mỗi vị trí,. Công việc này được mô tả ở (tầng thứ ba) lược đồ quan niệm cục bộ
- ✦ ES (External Schema): Các ứng dụng và truy xuất cơ sở dữ liệu được hỗ trợ qua lược đồ ngoài

Mô hình kiến trúc tham chiếu (tam khảo) CSDLPT trong hệ ngang hàng



Hình 1.8 Kiến trúc tham khảo cơ sở dữ liệu phân tán

Hệ phân tán ngang hàng

- ✎ Các thành phần cụ thể của một hệ quản trị cơ sở dữ liệu phân tán gồm hai thành phần chính (minh họa ở hình 1.9):
 - Bộ phận giao tiếp người sử dụng (user processor)
 - Bộ phận xử lý dữ liệu (data processor)

Các thành phần của bộ phận giao tiếp người sử dụng gồm:

- ✎ - Bộ phận giao tiếp (user interface handler): chịu trách nhiệm dịch các câu lệnh người sử dụng và định dạng dữ liệu kết quả để chuyển cho người sử dụng.
- ✎ - Bộ phận kiểm soát dữ liệu ngữ nghĩa (semantic data controller): sử dụng các ràng buộc toàn vẹn và thông tin quyền hạn, được định nghĩa như thành phần của lược đồ quan niệm toàn cục để kiểm tra xem các câu truy vấn có thể xử lý được hay không.
- ✎ - Bộ phận phân rã và tối ưu hoá vấn tin toàn cục (global query optimizer and decomposer): xác định như một chiến lược hoạt động nhằm giảm thiểu chi phí, phiên dịch các câu vấn tin toàn cục thành các câu vấn tin cục bộ bằng cách sử dụng các lược đồ quan niệm toàn cục, lược đồ quan niệm cục bộ và thư mục toàn cục. Bộ phận tối ưu vấn tin toàn cục còn chịu trách nhiệm tạo ra một chiến lược thực thi tốt nhất cho phép nối phân tán.

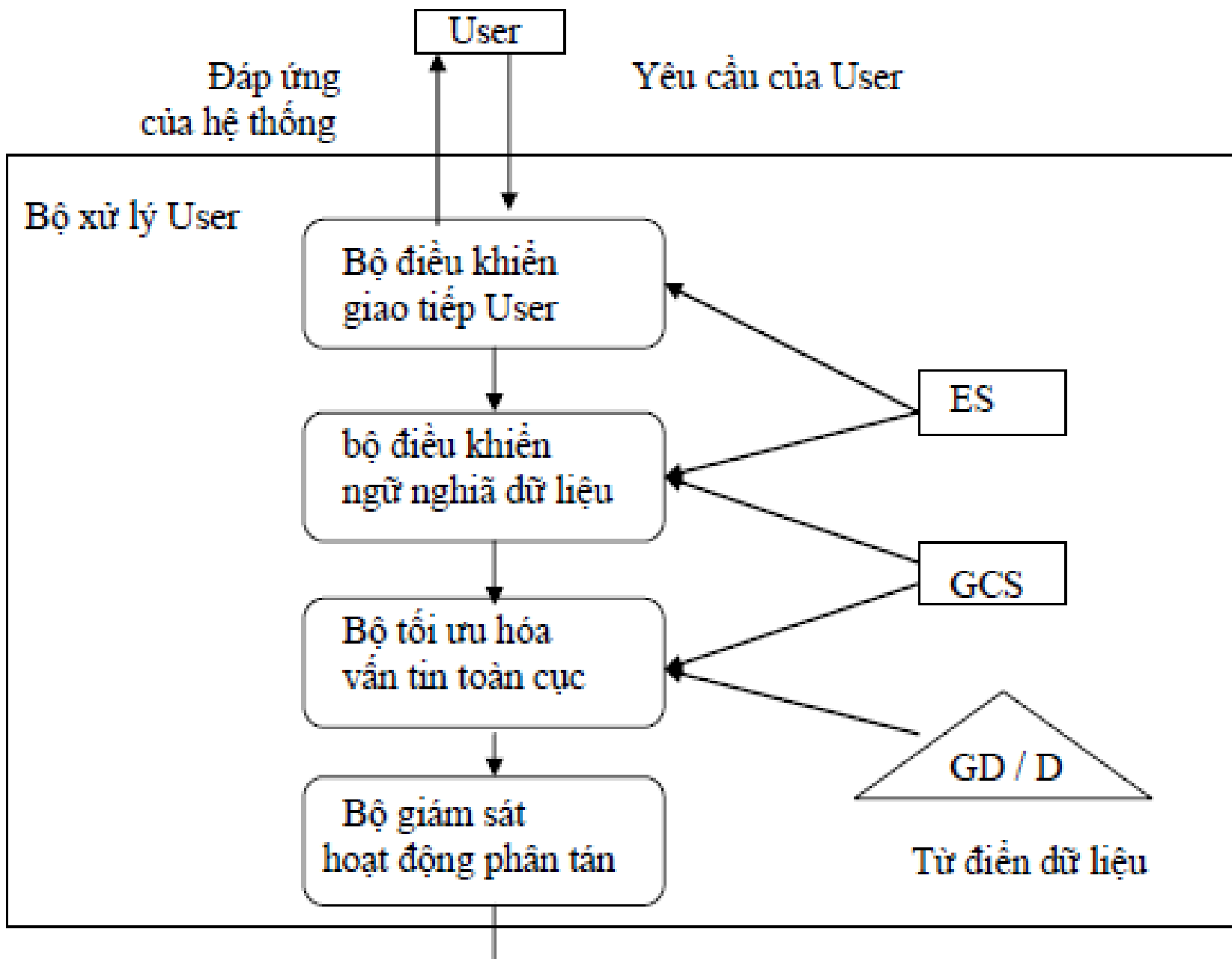
Bộ phận giám sát hoạt động phân tán (distributed execution monitor): điều phối việc thực hiện phân tán các yêu cầu người sử dụng và cũng được gọi là bộ quản lý giao tác phân tán (distributed transaction manager).

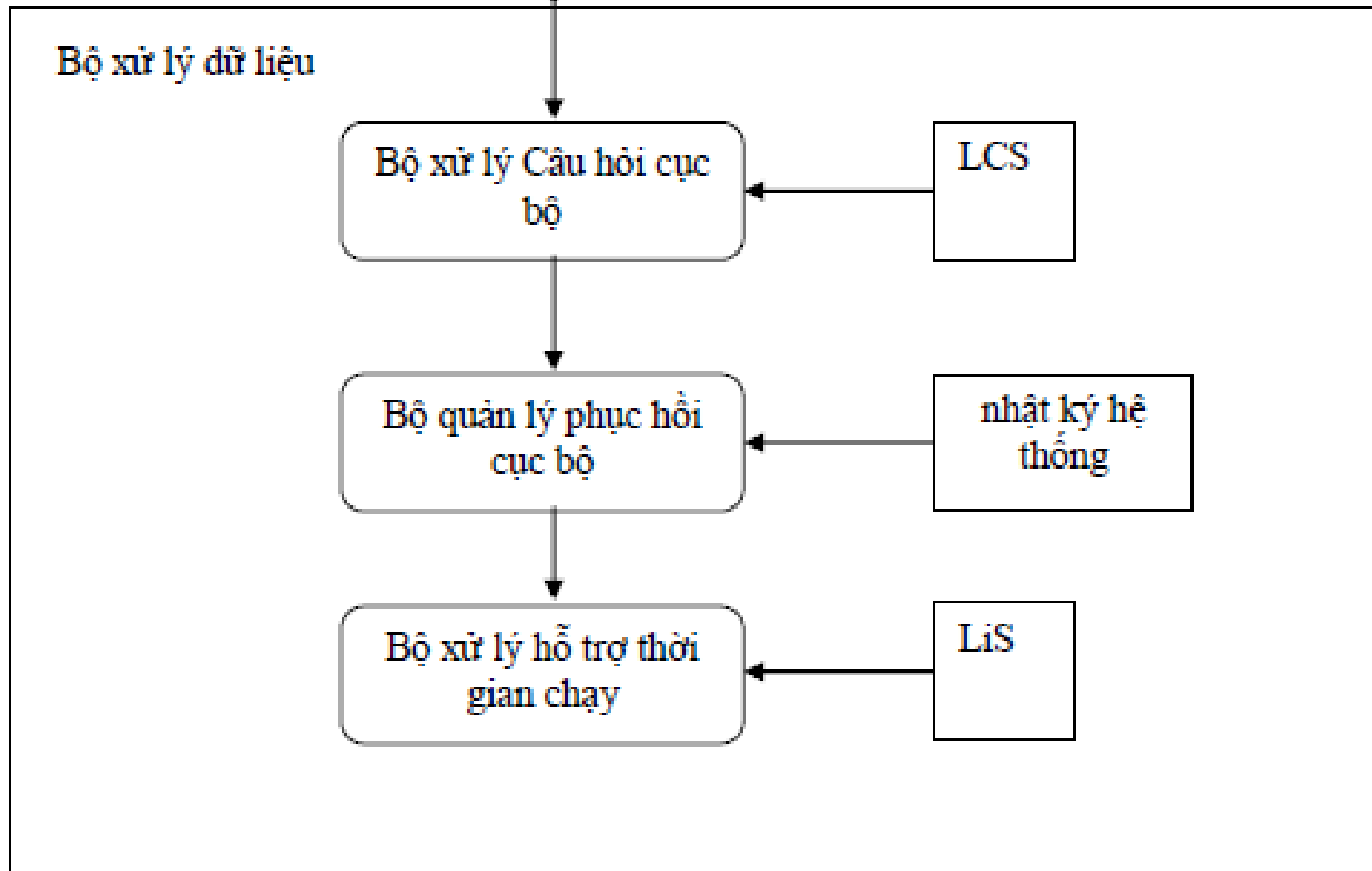
Thành phần chủ yếu thứ hai của hệ quản trị cơ sở dữ liệu phân tán là bộ xử lý dữ liệu (data processor), bao gồm các thành phần:

- Bộ phận tối ưu hoá vấn tin cục bộ (local query optimizer): thường hoạt động như bộ chọn đường truy xuất, chịu trách nhiệm chọn ra một đường truy xuất thích hợp nhất để truy xuất các mục dữ liệu.

- Bộ phận khôi phục cục bộ (local recovery manager) bảo đảm cho các cơ sở dữ liệu cục bộ vẫn duy trì được tính nhất quán ngay cả khi có sự cố xảy ra.

- Bộ phận hỗ trợ lúc thực thi (run-time support processor): truy xuất cơ sở dữ liệu tùy thuộc vào các lệnh trong lịch biểu do bộ phận tối ưu vấn tin tạo ra. Nó chính là bộ giao tiếp với hệ điều hành và chứa bộ quản lý vùng đệm cơ sở dữ liệu, chịu trách nhiệm quản lý vùng đệm và quản lý việc truy xuất dữ liệu.





Hình 1.9 Kiến trúc của hệ quản trị cơ sở dữ liệu phân tán.