

Chương 1

Giới thiệu về cơ sở dữ liệu phân tán



Nội dung

- ❖ Định nghĩa CSDL phân tán.
- ❖ Các đặc điểm của CSDL phân tán so với CSDL tập trung.
- ❖ Các lý do sử dụng CSDL phân tán.
- ❖ Hệ quản trị CSDL phân tán.

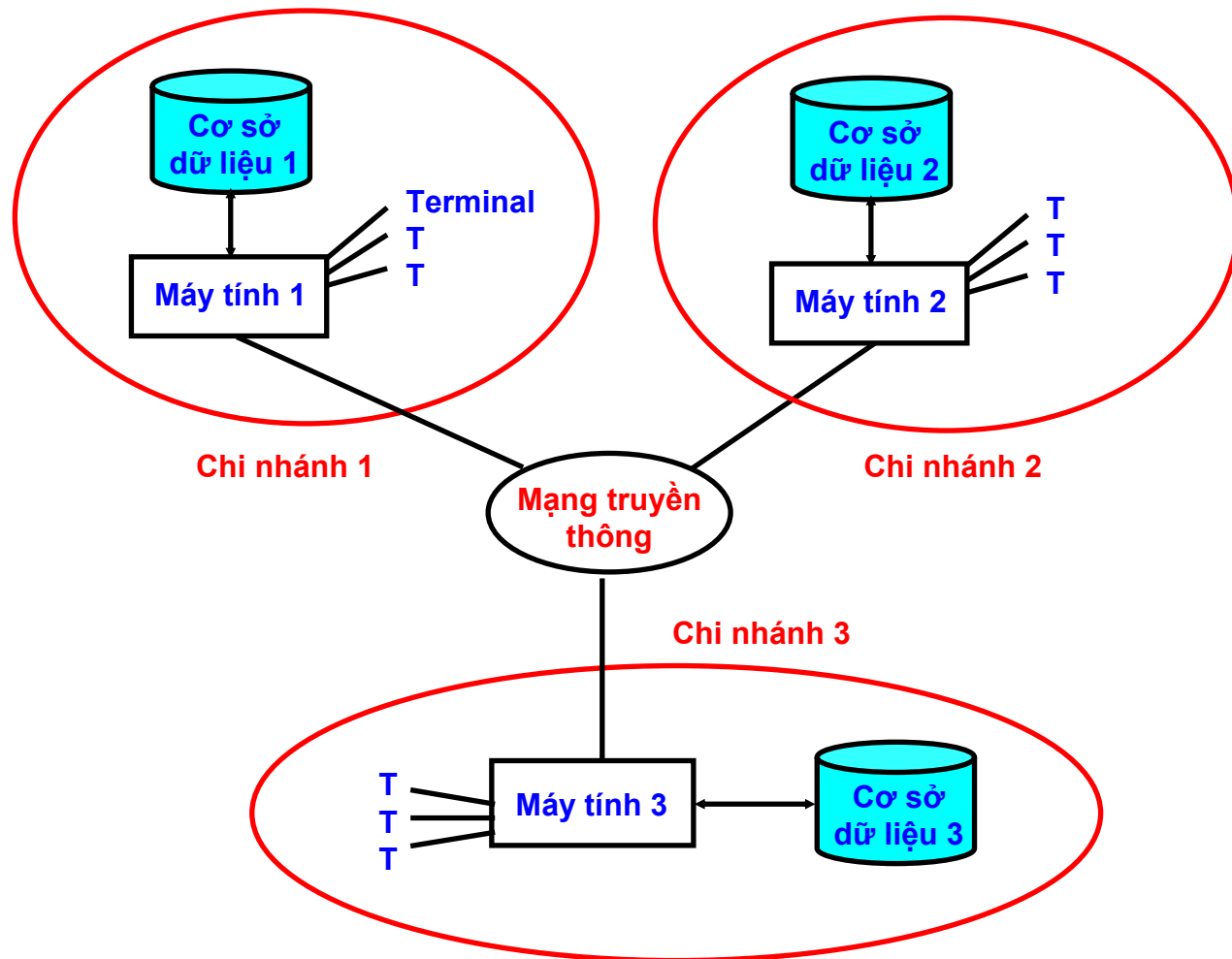
Định nghĩa cơ sở dữ liệu phân tán

❖ Định nghĩa 1

Cơ sở dữ liệu phân tán (*distributed database*) là sự tập hợp dữ liệu mà về mặt luận lý chúng **thuộc cùng một hệ thống** nhưng **được đặt ở nhiều nơi** (*site*) của một mạng máy tính.

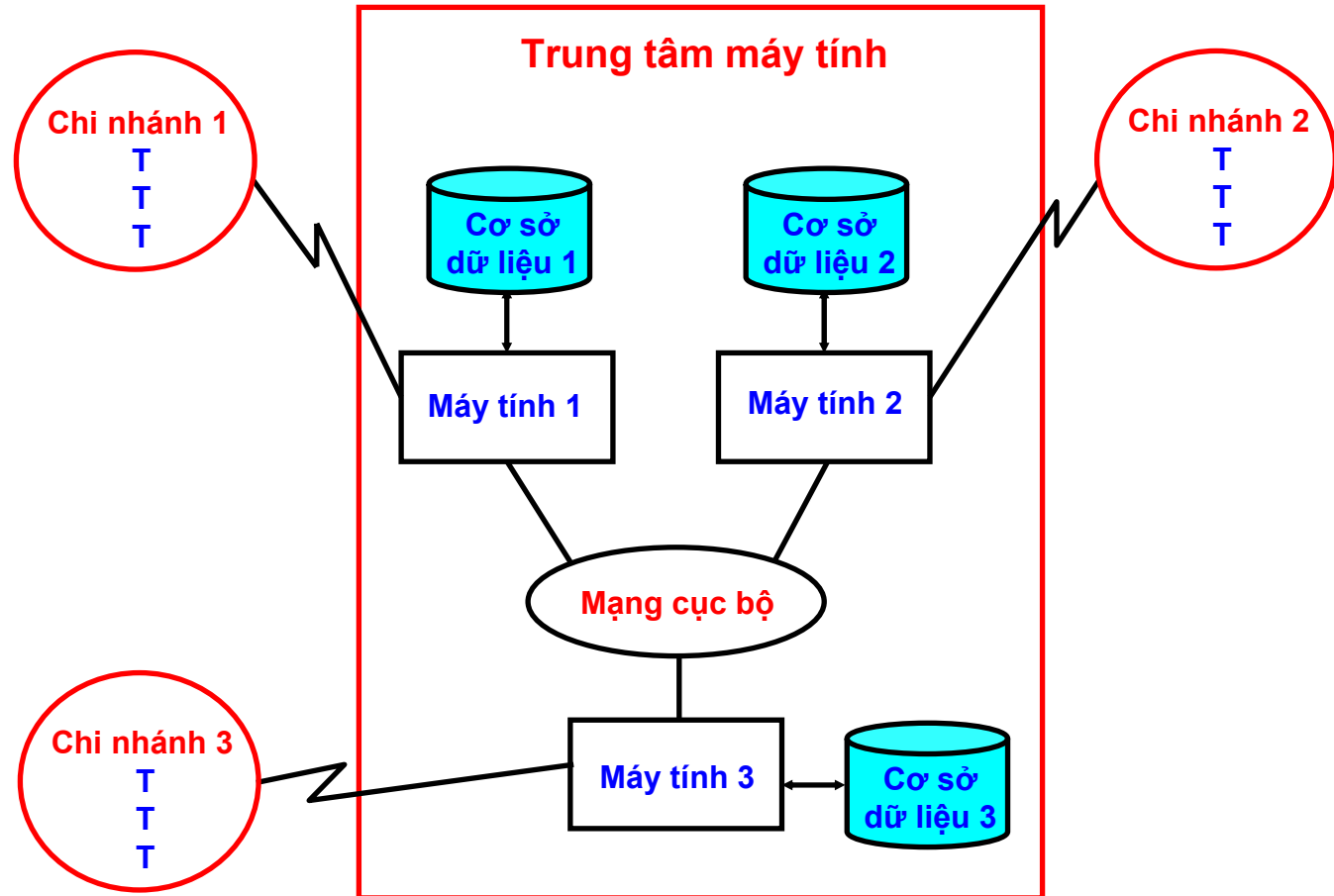
- ▶ **Sự phân tán dữ liệu** (*data distribution*): dữ liệu phải được phân tán ở nhiều nơi.
- ▶ **Sự tương quan luận lý** (*logical correlation*): dữ liệu của các nơi được sử dụng chung để cùng giải quyết một vấn đề.

Định nghĩa cơ sở dữ liệu phân tán



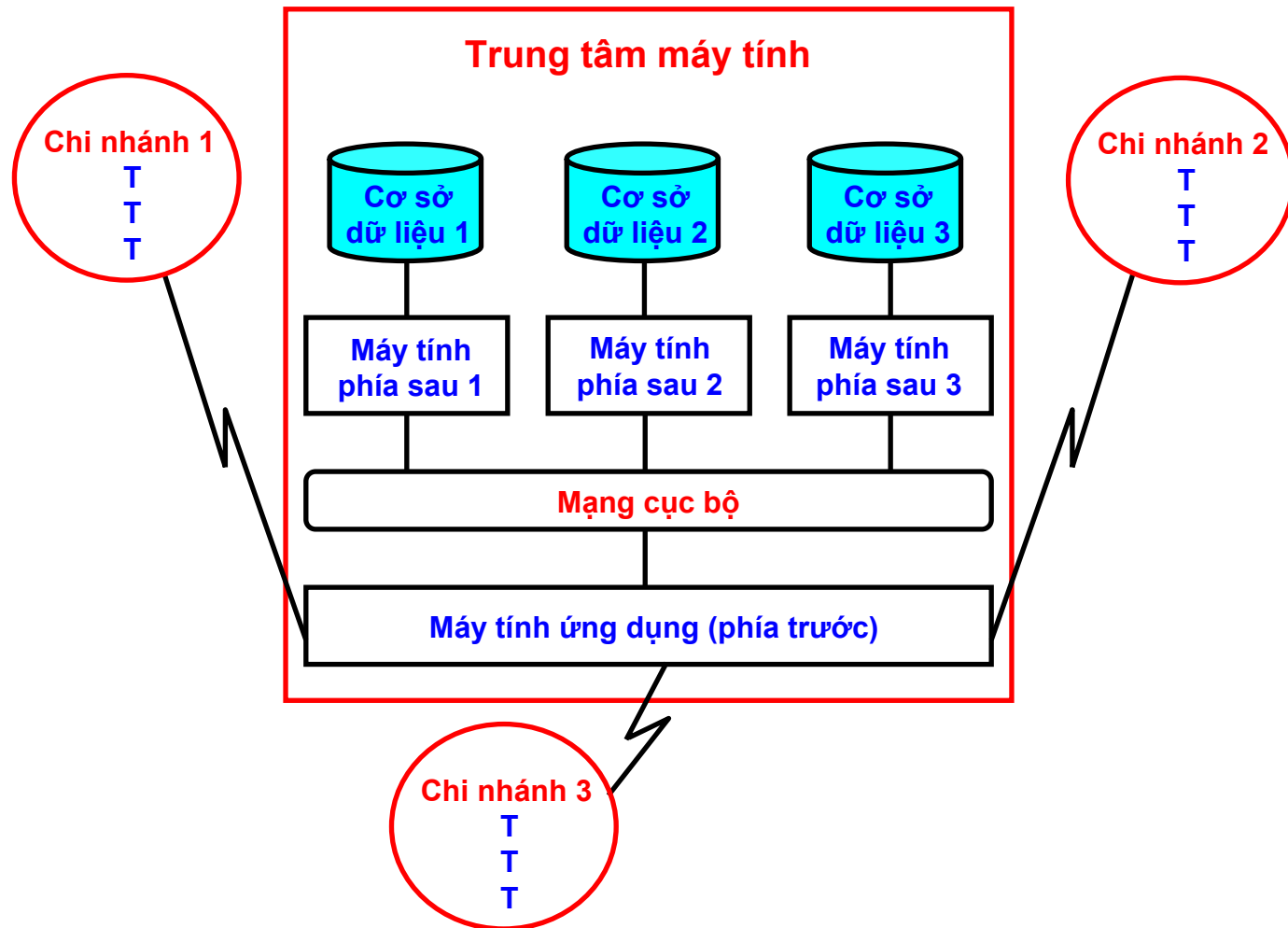
Hình 1.1. Cơ sở dữ liệu phân tán trên một mạng phân tán địa lý.

Định nghĩa cơ sở dữ liệu phân tán



Hình 1.2. Cơ sở dữ liệu phân tán trên một mạng cục bộ.

Định nghĩa cơ sở dữ liệu phân tán



Hình 1.3. Hệ thống đa xử lý (*multiprocessor system*).

Định nghĩa cơ sở dữ liệu phân tán

❖ Định nghĩa 2

Cơ sở dữ liệu phân tán là sự tập hợp dữ liệu **được phân tán** trên các máy tính khác nhau của một mạng máy tính. Mỗi nơi của mạng máy tính có khả năng xử lý tự trị và có thể **thực hiện** các ứng dụng cục bộ. Mỗi nơi cũng **tham gia thực hiện** ít nhất một ứng dụng toàn cục, mà nơi này yêu cầu truy xuất dữ liệu ở nhiều nơi bằng cách dùng hệ thống truyền thông con.

Định nghĩa cơ sở dữ liệu phân tán

❖ Định nghĩa 2

- ▶ **Sự phân tán dữ liệu** (*data distribution*): dữ liệu phải được phân tán ở nhiều nơi.
- ▶ **Ứng dụng cục bộ** (*local application*): ứng dụng được chạy hoàn thành tại một nơi và chỉ sử dụng dữ liệu cục bộ của nơi này.
- ▶ **Ứng dụng toàn cục** (hoặc ứng dụng phân tán) (*global application / distributed application*): ứng dụng được chạy hoàn thành và sử dụng dữ liệu của ít nhất hai nơi.



Các đặc điểm của CSDL phân tán so với CSDL tập trung

❖ Điều khiển tập trung

- ▶ Điều khiển tập trung (*centralized control*)
- ▶ Người quản trị CSDL cục bộ (*local DBA*)
- ▶ Người quản trị CSDL toàn cục (*global DBA*)
- ▶ Tính tự trị vị trí (*site autonomy*)



Các đặc điểm của CSDL phân tán so với CSDL tập trung

❖ Độc lập dữ liệu

- ▶ Độc lập dữ liệu (*data independence*)
- ▶ Tính trong suốt dữ liệu (*data transparency*)
- ▶ **Trong suốt phân mảnh** (*fragmentation transparency*):
 - Không nhìn thấy các mảnh.
 - Nhìn thấy các quan hệ toàn cục (*global relation*).
 - Lược đồ toàn cục (*global schema*).
- ▶ **Trong suốt vị trí** (*location transparency*)
 - Không nhìn thấy các quan hệ cục bộ.
 - Nhìn thấy các mảnh (*fragment*).
 - Lược đồ phân mảnh (*fragmentation schema*).

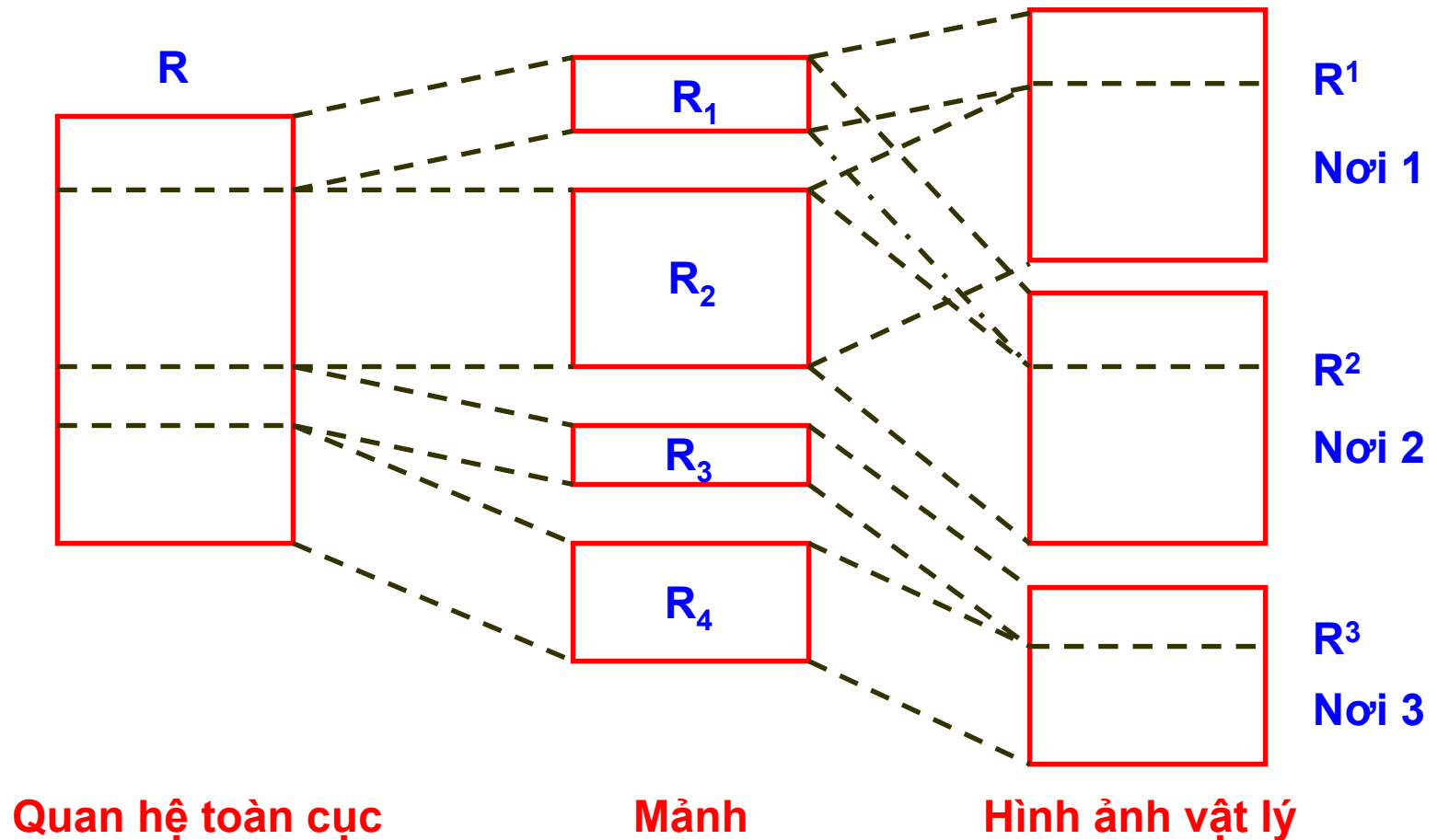


Các đặc điểm của CSDL phân tán so với CSDL tập trung

❖ Độc lập dữ liệu

- ▶ **Trong suốt nhân bản** (*replication transparency*)
 - Nhìn thấy các mảnh.
 - Không nhìn thấy sự nhân bản của các mảnh.
- ▶ **Trong suốt ánh xạ cục bộ** (*local mapping transparency*)
 - Nhìn thấy các quan hệ cục bộ (*local relation*).
 - Không nhìn thấy CSDL vật lý.
- ▶ **Trong suốt phân tán** (*distribution transparency*) gồm bốn tính trong suốt trên.

Các đặc điểm của CSDL phân tán so với CSDL tập trung



Hình 1.4. Các mảnh và các hình ảnh vật lý của một quan hệ toàn cục.

Các đặc điểm của CSDL phân tán so với CSDL tập trung

❖ Giảm dư thừa dữ liệu

- ▶ Dư thừa dữ liệu (*data redundancy*)
- ▶ **Nhược điểm của dư thừa dữ liệu**
 - Không nhất quán dữ liệu (*data inconsistency*).
 - Tốn nhiều vùng nhớ lưu trữ.
- ▶ **Ưu điểm của dư thừa dữ liệu**
 - Tính cục bộ (*locality*) của ứng dụng cao.
 - Tính sẵn sàng của dữ liệu (*data availability*) cao.
- ▶ **Nhân bản dữ liệu (*data replication*):** dữ liệu được lưu trữ thành nhiều bản.
 - Ứng dụng chỉ đọc (*read-only application*)
 - Ứng dụng cập nhật (*update application*)



Các đặc điểm của CSDL phân tán so với CSDL tập trung

- ❖ Các cấu trúc vật lý phức tạp và truy xuất hiệu quả
 - ▶ Cấu trúc vật lý phức tạp để truy xuất hiệu quả.
 - ▶ **Tối ưu hóa** (*optimization*)
 - Tối ưu hóa toàn cục (*global optimization*)
 - Tối ưu hóa cục bộ (*local optimization*)



Các đặc điểm của CSDL phân tán so với CSDL tập trung

❖ Tính toàn vẹn

▶ *integrity*

▶ **Giao tác** (*transaction*)

- Giao tác là một đơn vị thực hiện nguyên tố.
- Một chuỗi các tác vụ mà tất cả các tác vụ này đều được thực hiện hoặc đều không được thực hiện.

▶ **Giao tác toàn cục** (*global transaction*)

- Giao tác toàn cục là một ứng dụng toàn cục.

▶ **Tính nguyên tố** (*atomicity*)

- Sự hư hỏng
- Tính đồng thời



Các đặc điểm của CSDL phân tán so với CSDL tập trung

❖ Tính riêng biệt và tính bảo mật

- ▶ Thực hiện truy xuất dữ liệu có thẩm quyền.
- ▶ Bảo mật CSDL cục bộ.
- ▶ Bảo mật mạng truyền thông.

Tại sao sử dụng cơ sở dữ liệu phân tán

- ❖ **Các lý do về tổ chức và về kinh tế**
 - ▶ Nhiều tổ chức không được tập trung hóa.
- ❖ **Các CSDL hiện tại cần kết nối với nhau**
 - ▶ Nhiều CSDL đã tồn tại trong một công ty và cần phải thực hiện nhiều ứng dụng toàn cục hơn.
- ❖ **Sự lớn mạnh gia tăng**
 - ▶ Có thêm các đơn vị tổ chức tương đối độc lập.
- ❖ **Giảm chi phí truyền thông**
 - ▶ Nhiều ứng dụng cục bộ làm giảm chi phí truyền thông so với CSDL tập trung.

Tại sao sử dụng cơ sở dữ liệu phân tán

❖ Các nghiên cứu về hiệu suất

- ▶ Hiệu suất được nâng cao bằng một cơ chế song song hóa.
- ▶ Phân mảnh dữ liệu theo ứng dụng, làm cực đại hóa tính cục bộ của ứng dụng.

❖ Độ tin cậy và tính sẵn sàng

- ▶ Vì dư thừa dữ liệu, tính sẵn sàng của dữ liệu (*data availability*) cao.
- ▶ Cần phải bảo đảm độ tin cậy của dữ liệu (*data reliability*).



Hệ quản trị CSDL phân tán (DDBMS)

❖ Các thành phần của DDBMS

▶ *Truyền thông dữ liệu*

- *DC—Data Communication*
- Nhận yêu cầu truy xuất dữ liệu của ứng dụng chạy tại thiết bị đầu cuối.
- Trả kết quả về cho ứng dụng.

▶ *Quản trị CSDL*

- *DB – DataBase management*
- Quản lý CSDL.
- Thực hiện các yêu cầu của ứng dụng: xử lý dữ liệu (*data processing*).



Hệ quản trị CSDL phân tán (DDBMS)

❖ Các thành phần của DDBMS

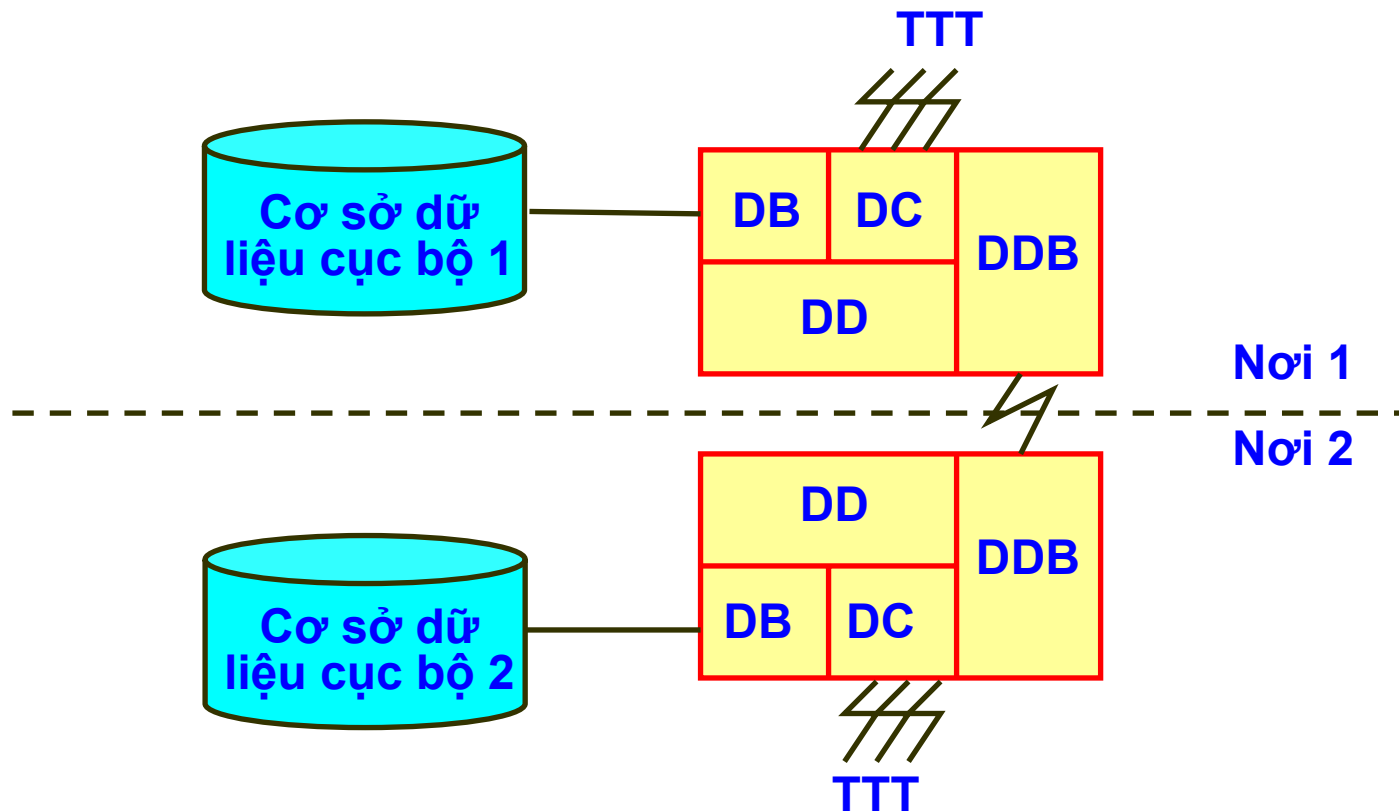
▶ *Từ điển dữ liệu*

- DD – *Data Dictionary*
- Lưu trữ thông tin về các đối tượng dữ liệu trong CSDL.
- Lưu trữ thông tin về sự phân tán dữ liệu tại các nơi.

▶ *CSDL phân tán*

- DDB – *Distributed DataBase*
- Liên lạc giữa các nơi: gửi yêu cầu và nhận kết quả.

Hệ quản trị CSDL phân tán (DDBMS)



Hình 1.6. Các thành phần của DDBMS thương mại.

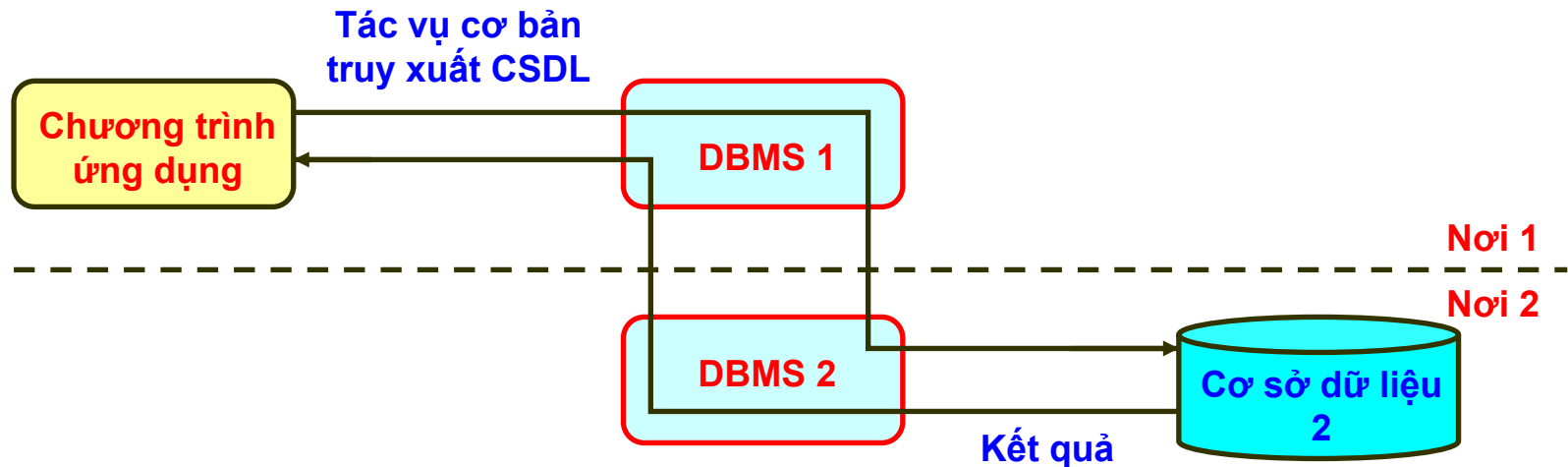


Hệ quản trị CSDL phân tán (DDBMS)

❖ Các chức năng tiêu biểu của DDBMS

- ▶ Truy xuất CSDL từ xa.
- ▶ Hỗ trợ một số mức trong suốt phân tán.
- ▶ Hỗ trợ cho việc quản trị CSDL phân tán.
- ▶ Hỗ trợ cho việc điều khiển tương tranh và phục hồi các giao tác phân tán.

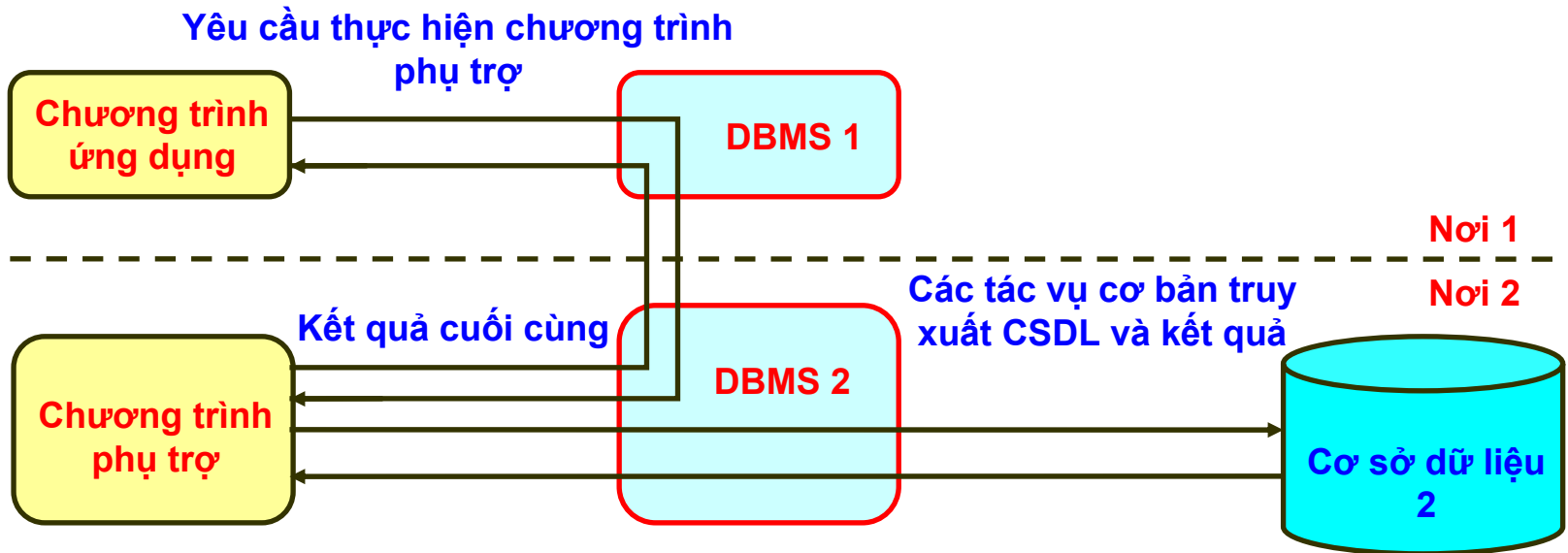
Hệ quản trị CSDL phân tán (DDBMS)



a. Truy xuất từ xa thông qua các tác vụ cơ bản của DBMS

Hình 1.7. Các loại truy xuất cơ sở dữ liệu phân tán.

Hệ quản trị CSDL phân tán (DDBMS)



b. Truy xuất từ xa thông qua chương trình phụ trợ

Hình 1.7. Các loại truy xuất cơ sở dữ liệu phân tán.



Hệ quản trị CSDL phân tán (DDBMS)

- ❖ **Tính đồng nhất và tính không đồng nhất**
 - ▶ *homogeneity, heterogeneity*
 - ▶ Phần cứng (*hardware*)
 - ▶ Hệ điều hành (*operating system*)
 - ▶ Các DBMS cục bộ
- ❖ **DDBMS đồng nhất**
 - ▶ Các DBMS cục bộ giống nhau.
- ❖ **DDBMS không đồng nhất**
 - ▶ Có ít nhất hai DBMS cục bộ khác nhau.
 - ▶ Chuyển đổi các mô hình dữ liệu khác nhau.