

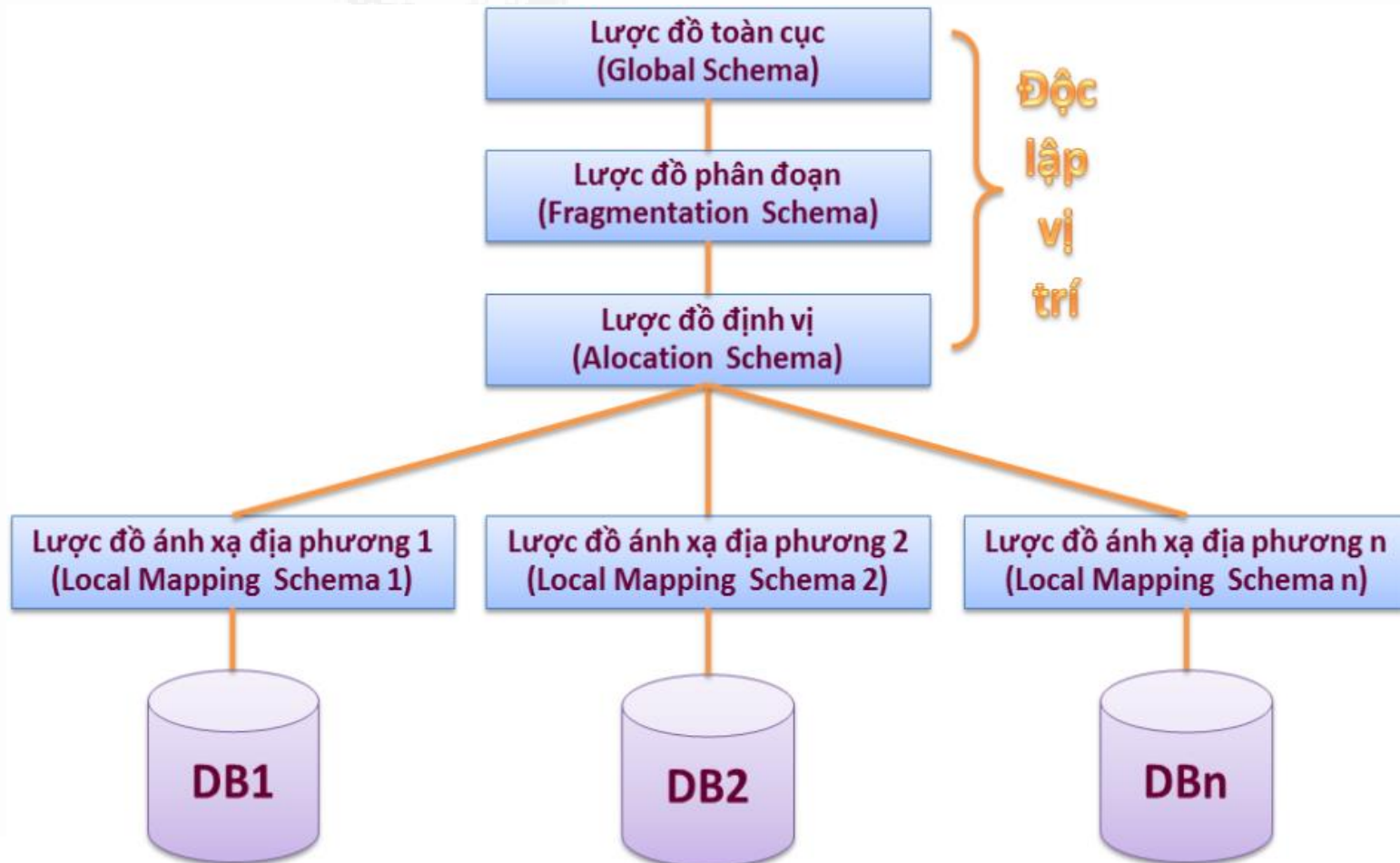
Chương 2 – Kiến trúc Hệ quản trị CSDL phân tán

Giảng viên: ThS. Nguyễn Hồ Duy Trí
tringuyen@uit.edu.vn

Nội dung

1. Kiến trúc cơ bản của một cơ sở dữ liệu phân tán
2. Các thành phần của hệ quản trị cơ sở dữ liệu phân tán
3. Kiến trúc của hệ quản trị cơ sở dữ liệu phân tán

Kiến trúc cơ bản của một CSDL phân tán



- **Lược đồ Toàn cục**

- ❖ Xác định toàn bộ cơ sở dữ liệu được lưu trữ trong cơ sở dữ liệu phân tán.
- ❖ Được định nghĩa như trong cơ sở dữ liệu tập trung.
- ❖ Trong **mô hình quan hệ** lược đồ toàn cục là **các quan hệ** và **mối liên kết** giữa chúng.

- **Lược đồ phân đoạn (phân mảnh)**

- ❖ Mỗi quan hệ tổng thể có thể được **chia** thành các phần **không giao nhau** gọi là phân đoạn (fragment).
- ❖ Có nhiều cách khác nhau để phân đoạn: Phân đoạn dọc, phân đoạn ngang, phân đoạn hỗn hợp.
- ❖ Các đoạn được mô tả bằng tên của quan hệ tổng thể cùng với chỉ mục đoạn.

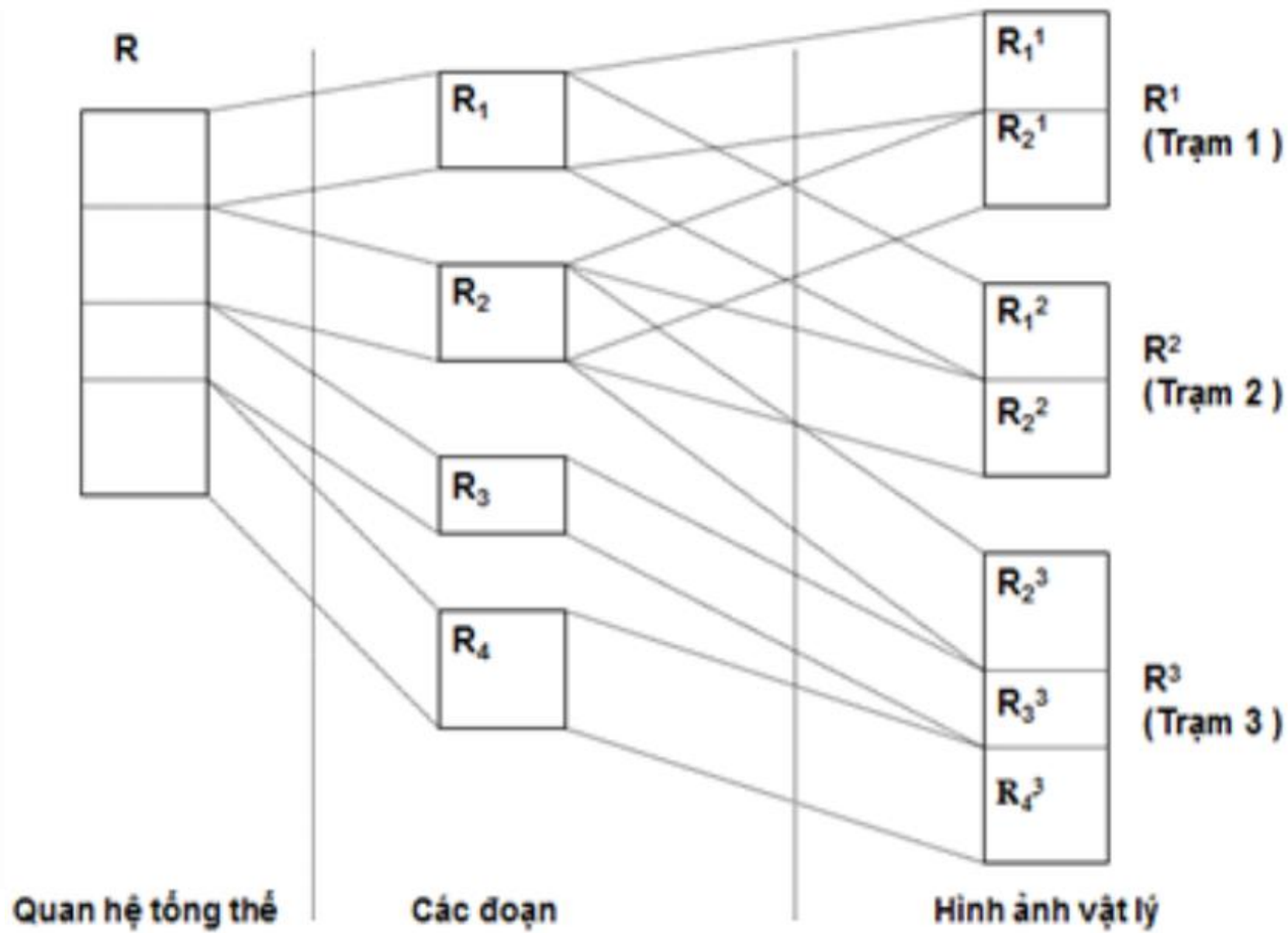
- Ví dụ: Ri là đoạn thứ i của quan hệ toàn cục R.

- **Lược đồ định vị (cấp phát):**

- ❖ Xác định đoạn dữ liệu nào được **định vị tại trạm** nào trên mạng.
- ❖ R_{ij} : Cho biết đoạn thứ i của quan hệ tổng thể R được định vị trên trạm j .

- **Lược đồ ánh xạ địa phương:**

- ❖ Ánh xạ các ảnh vật lý và các đối tượng được lưu trữ tại một trạm.



Kiến trúc cơ bản của một CSDL phân tán

- Ba **yếu tố** được suy ra từ kiểu kiến trúc này là:
 - ❖ Tách rời khái niệm phân đoạn dữ liệu với khái niệm định vị dữ liệu.
 - ❖ Biết được dữ liệu dư thừa.
 - ❖ Độc lập với các DBMS địa phương.
- Ba yếu tố này tương ứng với 3 **mức trong suốt** phân tán.

Kiến trúc cơ bản của một CSDL phân tán

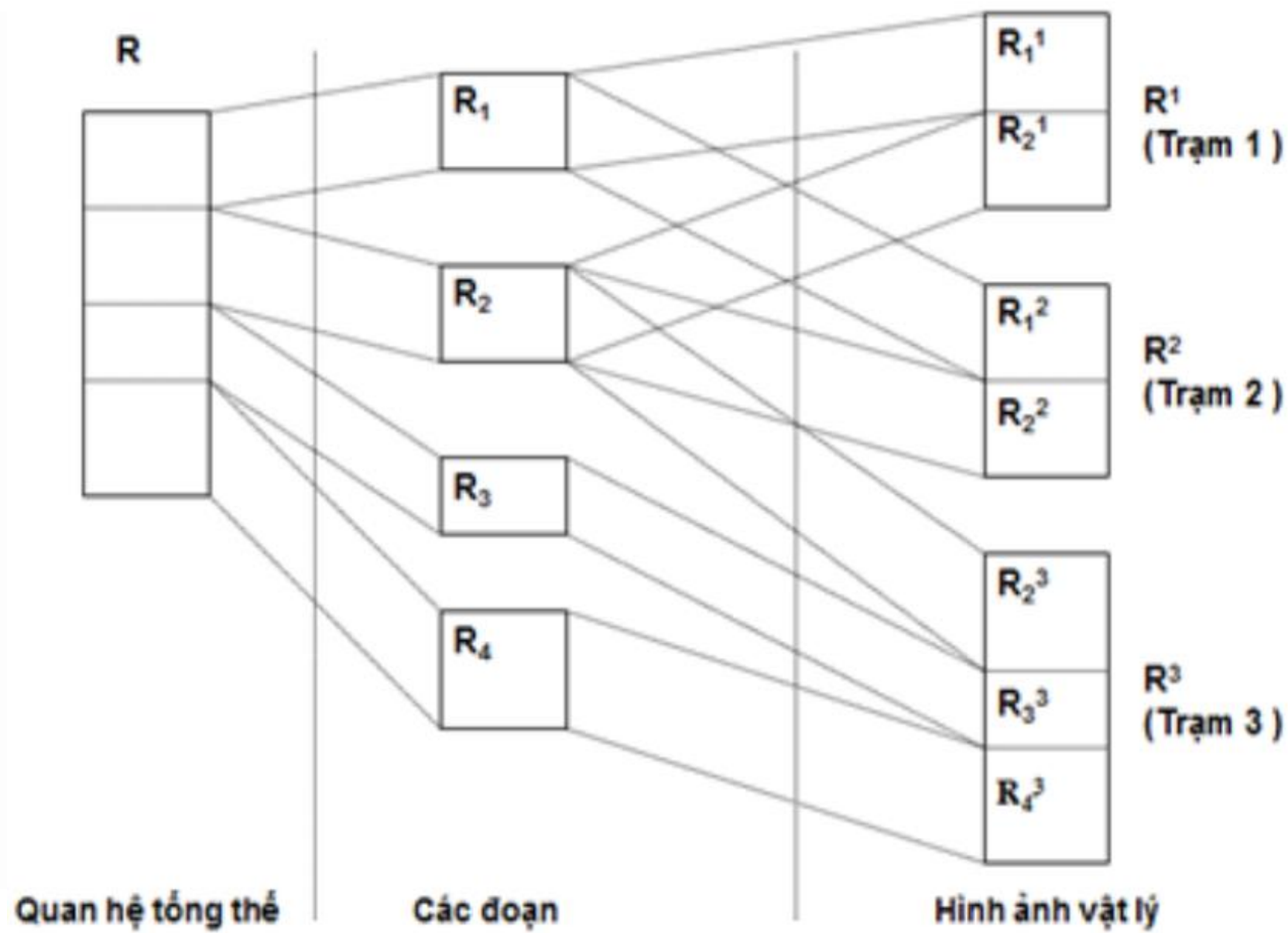
Yếu tố 1. Tách rời khái niệm phân đoạn dữ liệu với Khái niệm định vị dữ liệu

- **Phân đoạn dữ liệu**, bao gồm những công việc mà người lập trình ứng dụng làm việc với quan hệ tổng thể phân chia quan hệ tổng thể thành các đoạn.
 - Thông qua tính **trong suốt phân đoạn** (fragmentation transparency) người lập trình sẽ nhìn thấy được những đoạn dữ liệu bị phân chia như thế nào.
- **Định vị dữ liệu** lại liên quan đến các công việc của người sử dụng và người lập trình ứng dụng trên các đoạn dữ liệu được định vị tại các trạm
 - Thông qua tính **trong suốt vị trí** (location transparency) người lập trình sẽ biết được vị trí của các đoạn dữ liệu.

Kiến trúc cơ bản của một CSDL phân tán

Yếu tố 2. Biết được dữ liệu dư thừa

- Người lập trình ứng dụng có thể biết được dư thừa dữ liệu ở các trạm.
- Ở hình vẽ trên, chúng ta thấy rằng hai ảnh vật lý R_2^2 và R_2^3 có trùng lặp dữ liệu. Do đó các đoạn dữ liệu trùng nhau có thể tránh được khi xây dựng các khối ảnh vật lý.



Kiến trúc cơ bản của một CSDL phân tán

Yếu tố 3. Độc lập với các dbms địa phương

- Tính chất này còn được gọi là **trong suốt ánh xạ địa phương** (local mapping transparency).
- Cho phép chúng ta khảo sát các vấn đề về quản lý cơ sở dữ liệu phân tán mà không cần phải hiểu rõ mô hình Dữ liệu của DBMS địa phương đang sử dụng.

Các thành phần của HQT CSDL phân tán

- **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 cơ sở dữ liệu**

- DB - Database management.
 - Quản lý cơ sở dữ liệu.
 - Thực hiện các yêu cầu của ứng dụng: xử lý dữ liệu (data processing).

Các thành phần của HQT CSDL phân tán

- **Các thành phần của DDBMS (tt)**

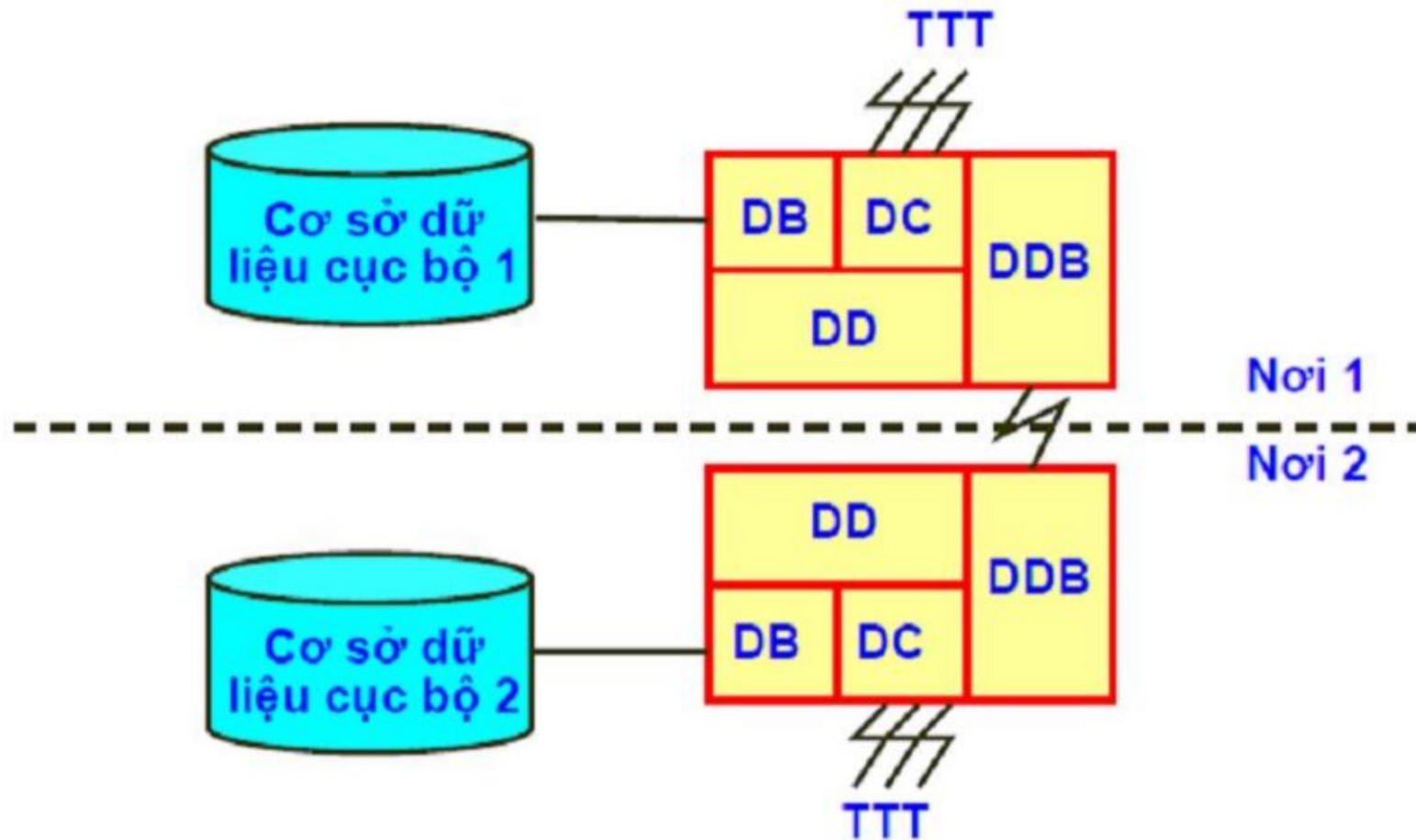
- **Từ điển dữ liệu**

- DD - Data dictionary.
 - Lưu trữ thông tin về các đối tượng dữ liệu trong cơ sở dữ liệu.
 - Lưu trữ thông tin về sự phân tán dữ liệu tại các nơi.

- **Cơ sở dữ liệu phân tán**

- DDB - Distributed database
 - Liên lạc giữa các nơi gửi yêu cầu và nhận kết quả

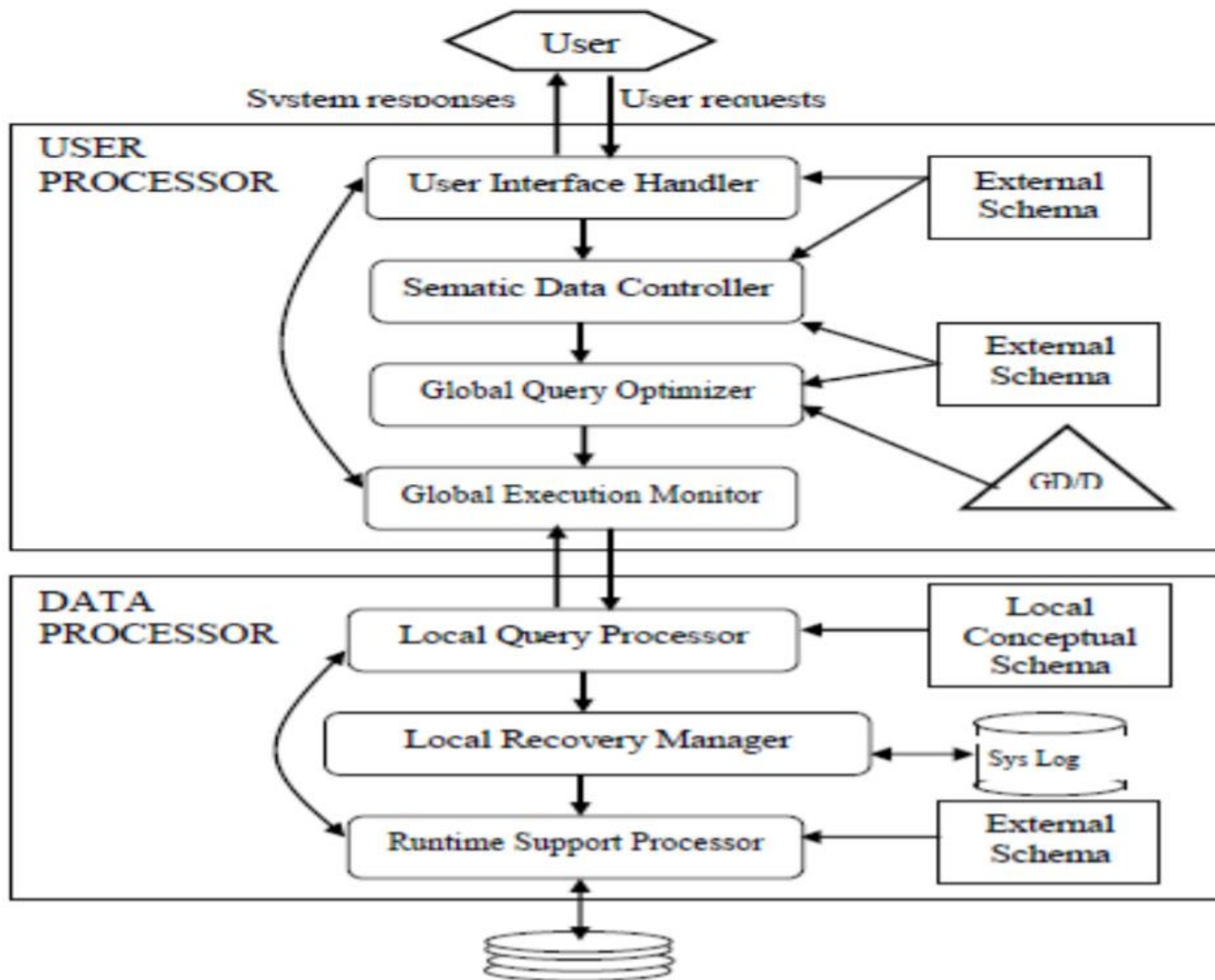
Các thành phần của HQT CSDL phân tán

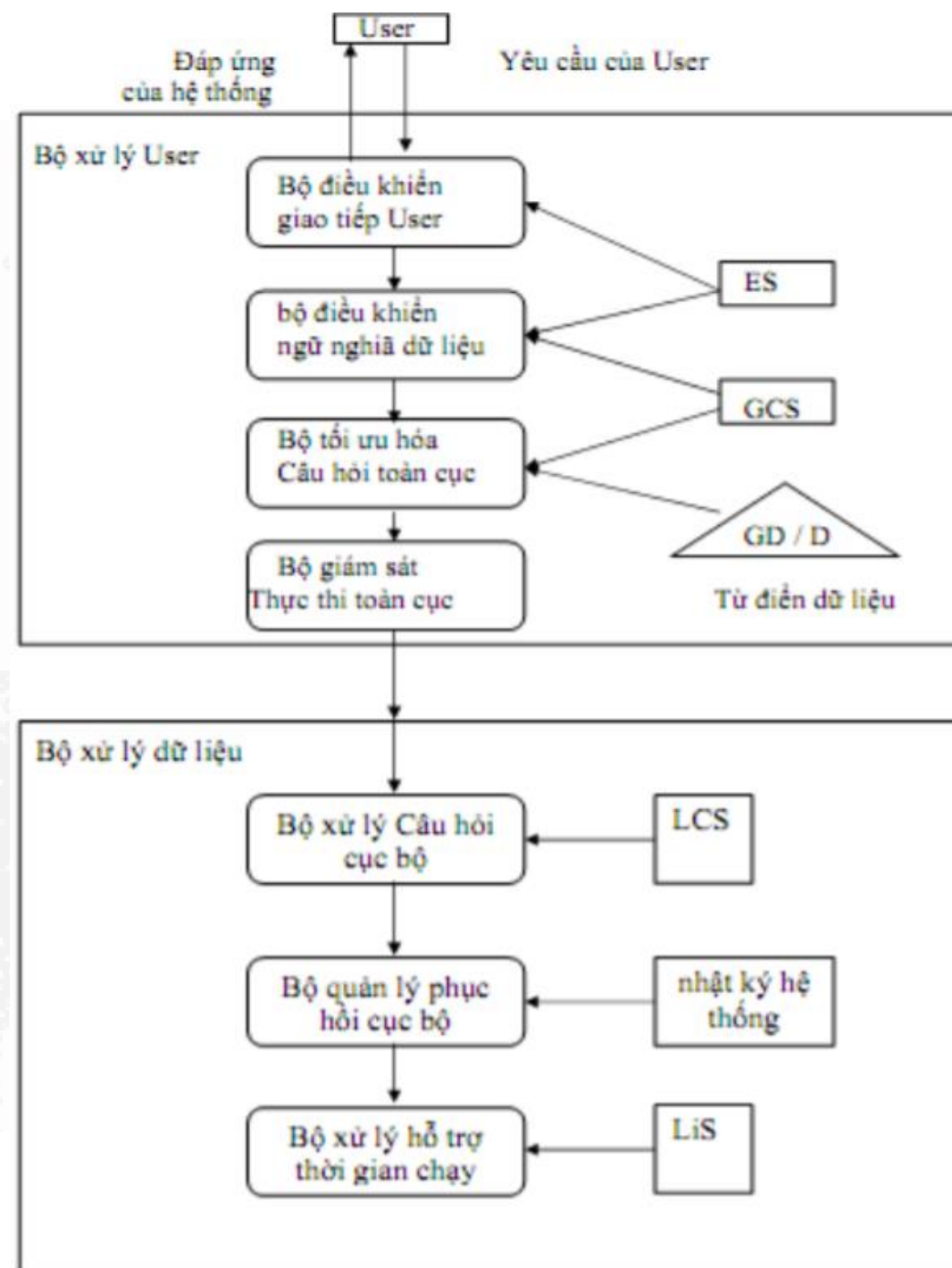


Các thành phần của DDBMS thương mại

Các thành phần của HQT CSDL phân tán

- **Bộ phận tiếp nhận người dùng (user processor)**
 - ✓ Bộ phận giao tiếp (User Interface Handler)
 - ✓ Bộ phận kiểm soát ngữ nghĩa (Semantic Data Controller)
 - ✓ Bộ phận phân rã và tối ưu vấn tin toàn cục (Global Query Optimizer and Decomposer)
 - ✓ Bộ phận theo dõi việc thực hiện phân tán (Distributed Execution Monitor)
- **Bộ xử lý (database processor)**
 - ✓ Bộ phận tối ưu hóa vấn tin cục bộ (Local Query Optimizer)
 - ✓ Bộ phận khôi phục cục bộ (Local Recovery Manager)
 - ✓ Bộ phận hỗ trợ xử lý thực thi (Runtime Support Processor)

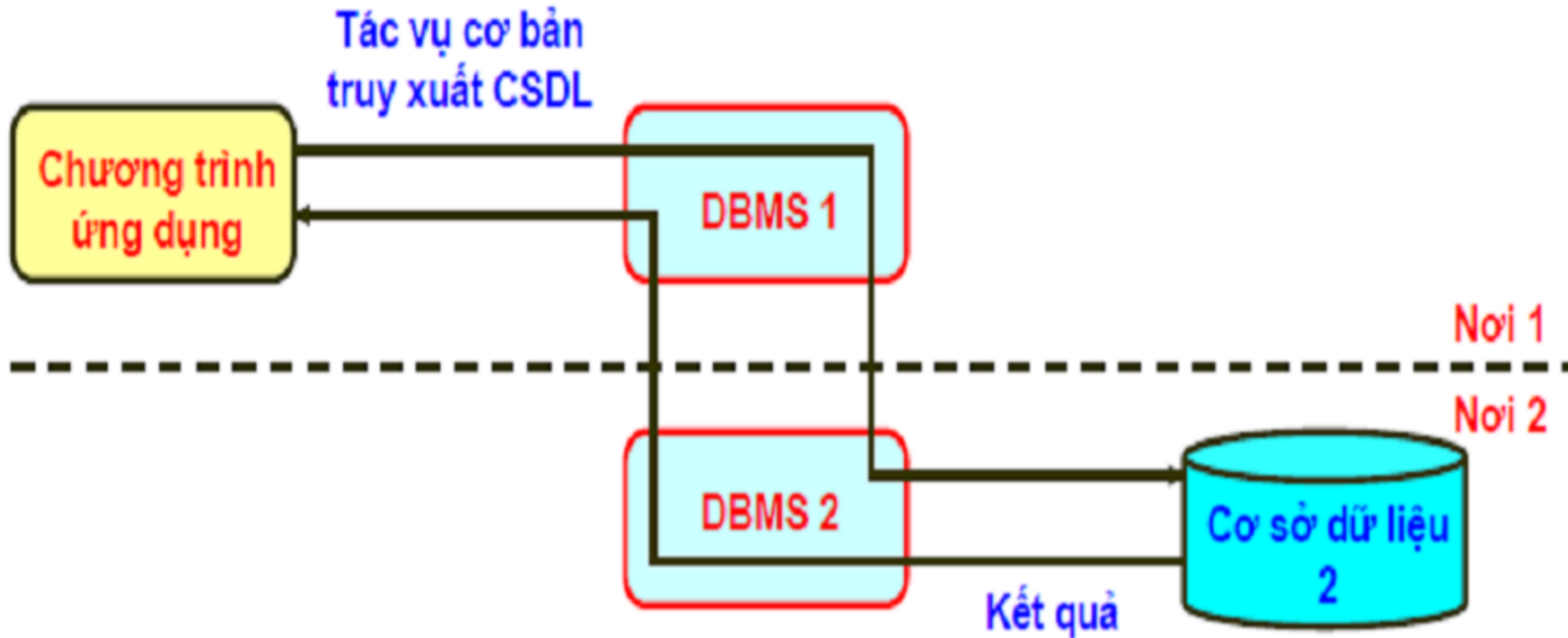




Các thành phần của HQT CSDL phân tán

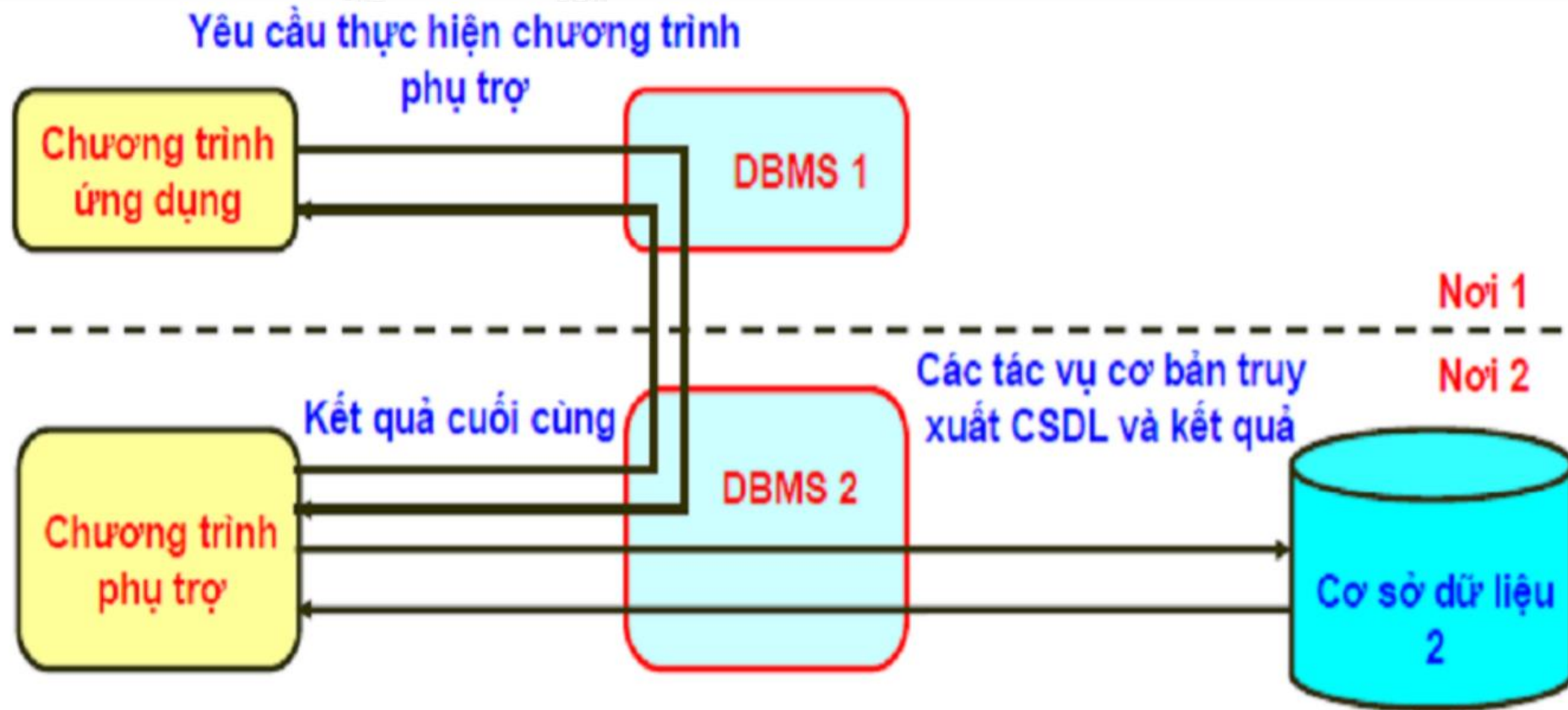
- **Các chức năng tiêu biểu của DDBMS**
 - ✓ Truy xuất cơ sở dữ liệu từ xa.
 - ✓ Hỗ trợ một số mức trong suốt phân tán.
 - ✓ Hỗ trợ cho việc quản trị cơ sở dữ liệu phân tán.
 - ✓ Hỗ trợ cho việc điều khiển tương tranh và phục hồi các giao tác phân tán.

Các thành phần của HQT CSDL phân tán



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

Các thành phần của HQT CSDL phân tán



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

Các thành phần của HQT CSDL phân tán

- **Tính đồng nhất (Homogeneity) và tính không đồng nhất (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.

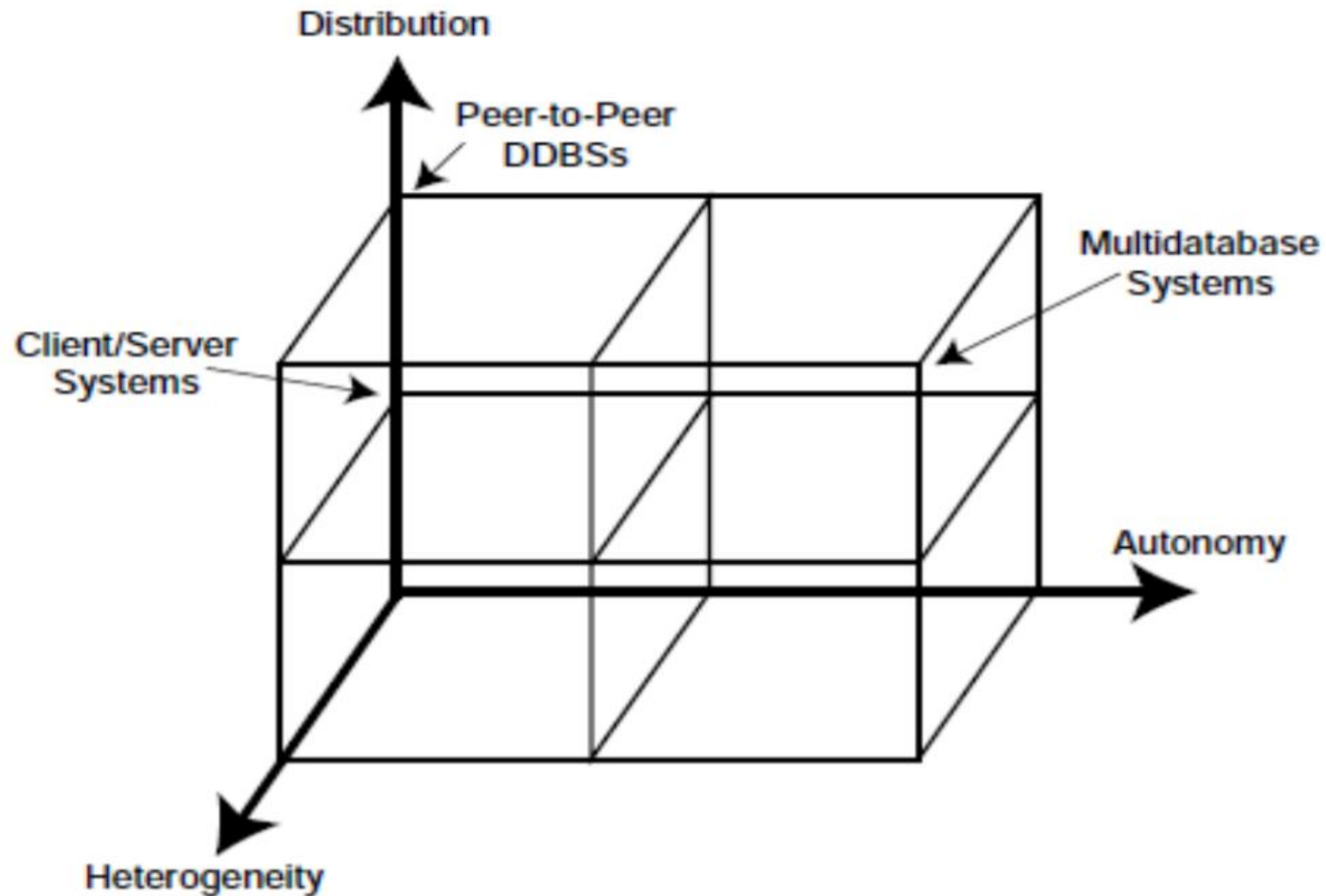
Kiến trúc hệ quản trị CSDL phân tán

- **Kiến trúc của một hệ thống là xác định cấu trúc của nó.**
 - ✓ Các thành phần của các hệ thống được xác định.
 - ✓ Chức năng của mỗi thành phần được mô tả.
 - ✓ Các mối quan hệ trung gian và tương tác giữa các thành phần này được định nghĩa.
- **Mục tiêu định nghĩa kiến trúc là xây dựng các DBMS phân tán với khả năng cung cấp được các chức năng như**
 - ✓ Vô Hình.
 - ✓ Độ tin cậy qua các giao dịch phân tán.
 - ✓ Hiệu quả và đặc tính mở rộng...

Kiến trúc hệ quản trị CSDL phân tán

- Kiến trúc hệ client/server (A_x, D_1, H_y)
- Kiến trúc hệ Peer-To-Peer (A_0, D_2, H_0)
- Kiến trúc Multi-DBMS (A_2, D_x, H_y)

Kiến trúc hệ quản trị CSDL phân tán

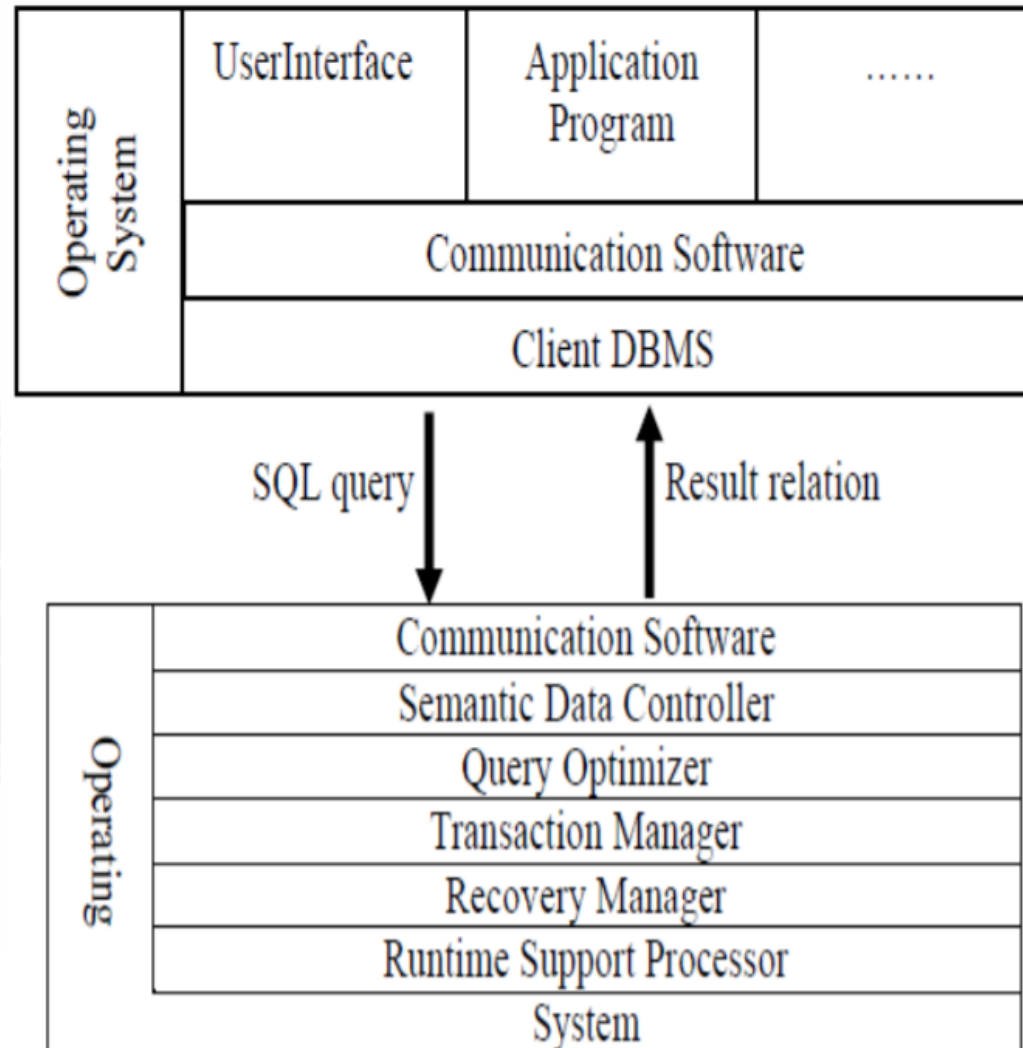


Kiến trúc hệ quản trị CSDL phân tán

Kiến trúc client/server

- Chia chức năng thành **hai lớp**: **chức năng chủ** và **chức năng khách**.
- Đây là kiến trúc hai cấp có tác dụng làm **giảm tính phức tạp** không những cho DBMS mà cả khi phân tán dữ liệu.
 - **Server** chịu trách nhiệm **thực hiện mọi xử lý** và **tối ưu hóa vấn tin, quản lý giao dịch** và **quản lý thiết bị ...**
 - Tại các **Client** ngoài các **ứng dụng giao diện** còn có riêng một DBMS chịu trách nhiệm **quản lý dữ liệu nhận từ server** và có thể **quản lý cả các phiên giao dịch**.

Kiến trúc Client/Server



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

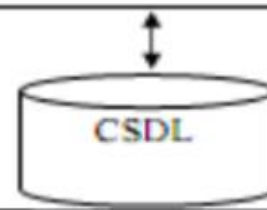
Truy vấn

SQL

Kết quả

quan hệ

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



Kiến trúc Client/Server

Các kiến trúc client/server đặc trưng nhất

1) n-Client --- 1-Server

- Kiến trúc này không có nhiều khác biệt với CSDL tập trung vì CSDL được lưu trữ trên một Server duy nhất và có một phần mềm để quản lý CSDL này.

2) n-Client --- n-Server

- Kiến trúc này có hai cách quản lý:
 - Mỗi client tự quản lý kết nối với server.
 - Mỗi client chỉ biết server trực tiếp của mình.

Kiến trúc hệ quản trị CSDL phân tán

Kiến trúc Peer-To-Peer (kiểu ngang hàng)

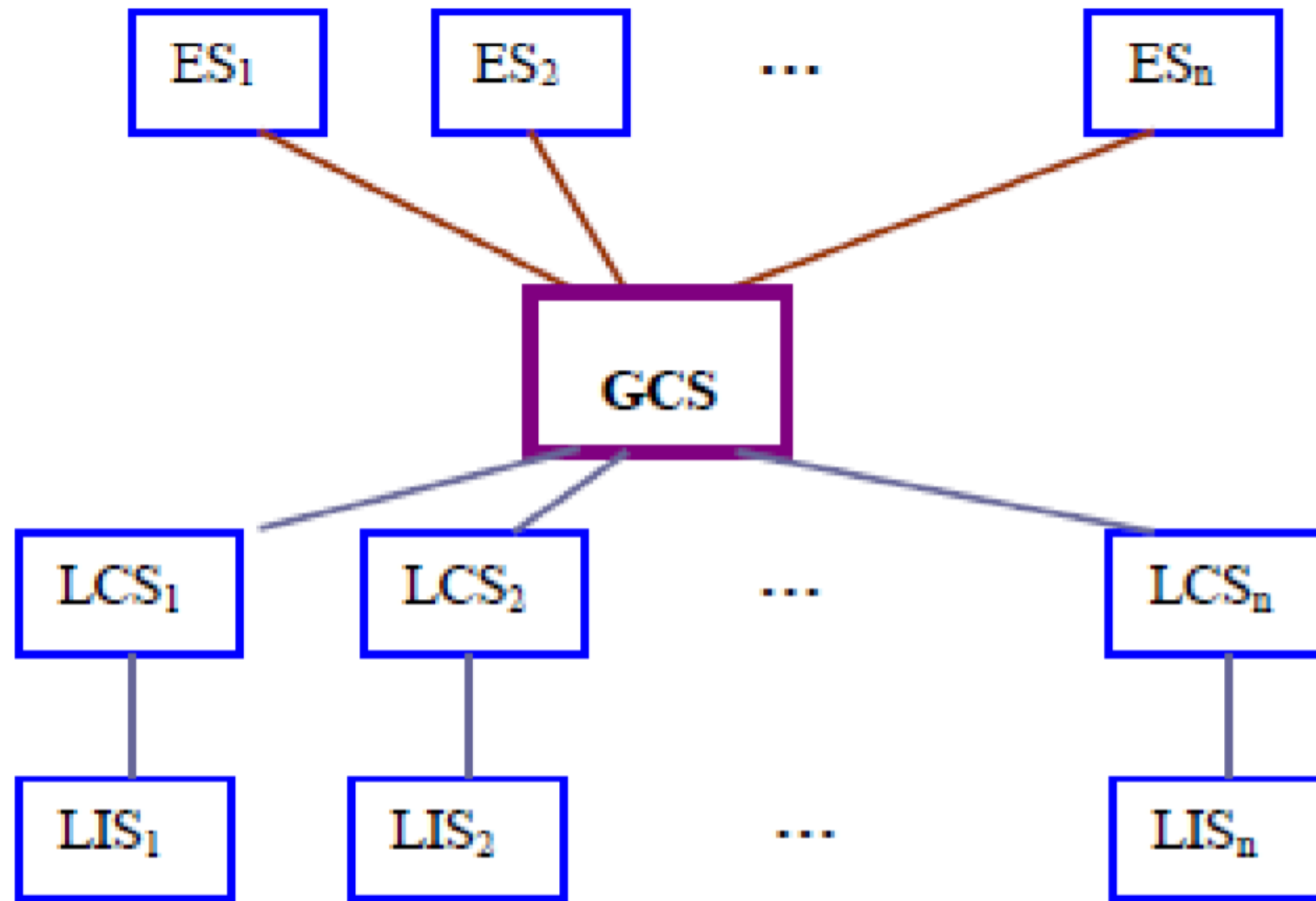
- Trong Peer-To-Peer, việc tổ chức dữ liệu vật lý trên mỗi Workstation có thể khác nhau, như vậy phải có một định nghĩa riêng cho mỗi Workstation mà ta gọi là **lược đồ cục bộ LIS** (Local Internal Schema).
- Dữ liệu trong một CSDL phân tán thường được **phân mảnh** và **nhân bản**. Để xử lý sự phân mảnh và nhân bản, chúng ta cần mô tả tổ chức logic của dữ liệu được lưu tại các Workstation nên cần phải có thêm **lược đồ khái niệm cục bộ LCS** (Local Conceptual Schema) trong kiến trúc.
- Như vậy có thể thấy GCS là hợp tác của các LCS.
- Không có sự khác biệt giữa client và server. Mỗi máy đều có đủ chức năng của 1 DBMS. Các hệ này được gọi là phân tán đầy đủ.

Kiến trúc hệ quản trị CSDL phân tán

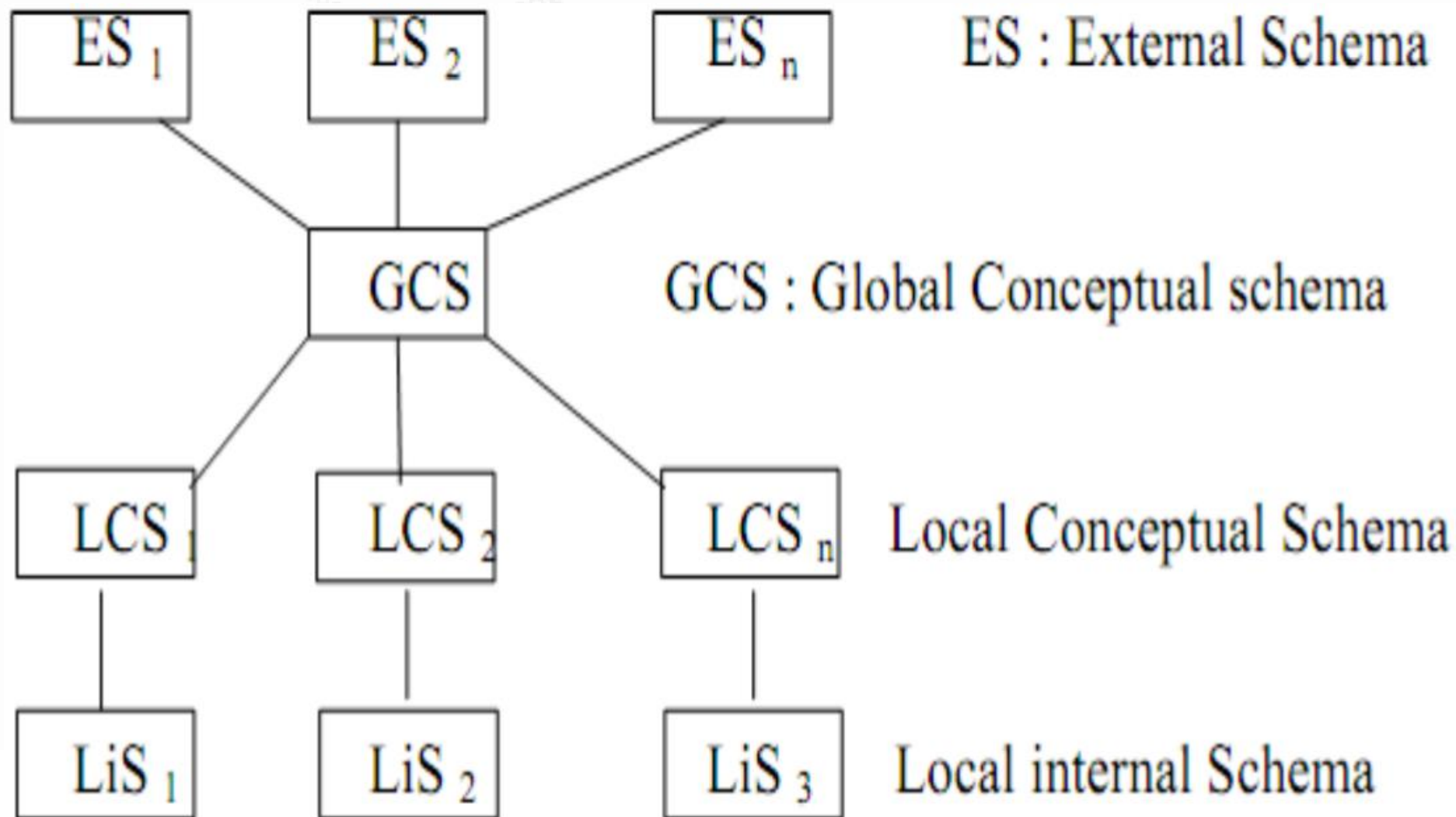
Kiến trúc Peer-To-Peer (tt)

- Bức tranh toàn cục về dữ liệu của cả công ty hay cả xí nghiệp... được mô tả bởi **lược đồ toàn cục GCS** (Global Conceptual Schema), nó được dùng để mô tả cấu trúc logic của dữ liệu được lưu tại các Workstation.
- Cuối cùng là các ứng dụng và truy cập CSDL được gọi là **lược đồ ngoài ES** (External Schema).

Kiến trúc Peer-To-Peer



Kiến trúc Peer-To-Peer



Kiến trúc hệ quản trị CSDL phân tán

Kiến trúc đa hệ cơ sở dữ liệu

- Các mô hình sử dụng lược đồ khái niệm toàn cục.
- Trong phức hệ cơ sở dữ liệu, GCS là tích hợp các lược đồ ngoài của các cơ sở dữ liệu tự vận hành cục bộ.