

# CHƯƠNG 1. TỔNG QUAN VỀ CSDL PHÂN TÁN

---

## NỘI DUNG

1.1 Sơ lược về mạng máy tính

1.2 Các hình thức tổ chức hệ thống phân tán

1.3 Các đặc trưng của các hệ thống phân tán

1.4 Cơ sở dữ liệu phân tán

1.5 Các loại truy xuất CSDL phân tán

Kết luận\_

# CHƯƠNG 1. TỔNG QUAN VỀ CSDL PHÂN TÁN

---

## *Tại sao phải phân tán? - Phân tán cái gì?*

- Nhu cầu dùng chung một dữ liệu hợp nhất
- 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 tổ chức
- Một hệ thống phân tán bao gồm hai phần: mạng máy tính và csdl phân tán

# 1.1 SƠ LƯỢC VỀ MẠNG MÁY TÍNH

---

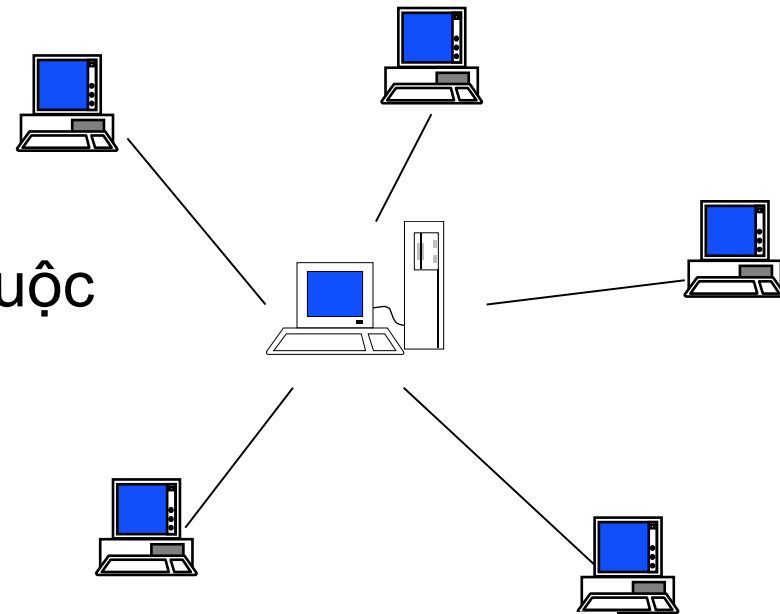
- Một mạng máy tính là một tập các máy tính tự vận hành, được kết nối lại và có khả năng trao đổi thông tin giữa chúng\_
- Các máy tính trên một mạng thường được gọi là các nút hay các trạm, chúng tạo ra các phần cứng cơ bản của mạng và được kết nối lại với nhau bởi một đường truyền.
- Một mạng máy tính là một trường hợp đặc biệt của môi trường xử lý phân tán, trong đó các máy tính là các thiết bị được kết nối vào kênh truyền dữ liệu.

# 1.1 SƠ LƯỢC VỀ MẠNG MÁY TÍNH

## *Những loại mạng máy tính cơ bản \_*

### **Mạng hình sao (star)**

- Thường được dùng trong các tổ chức có nhiều chi nhánh nằm ở nhiều vùng khác nhau
- Máy tính trung tâm được đặt tại văn phòng chính hoặc tại trung tâm vùng
- Nhược điểm:
  - độ tin cậy thấp\_
  - giao tiếp của hai máy tính phụ thuộc vào máy tính trung tâm\_
  - tải trọng quá cao trên máy tính trung tâm

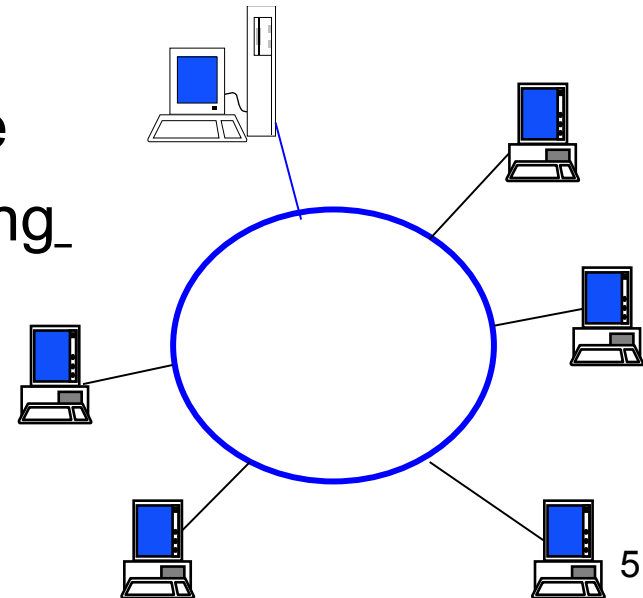


Mạng hình sao

# 1.1 SƠ LƯỢC VỀ MẠNG MÁY TÍNH

## Mạng vòng (circle)

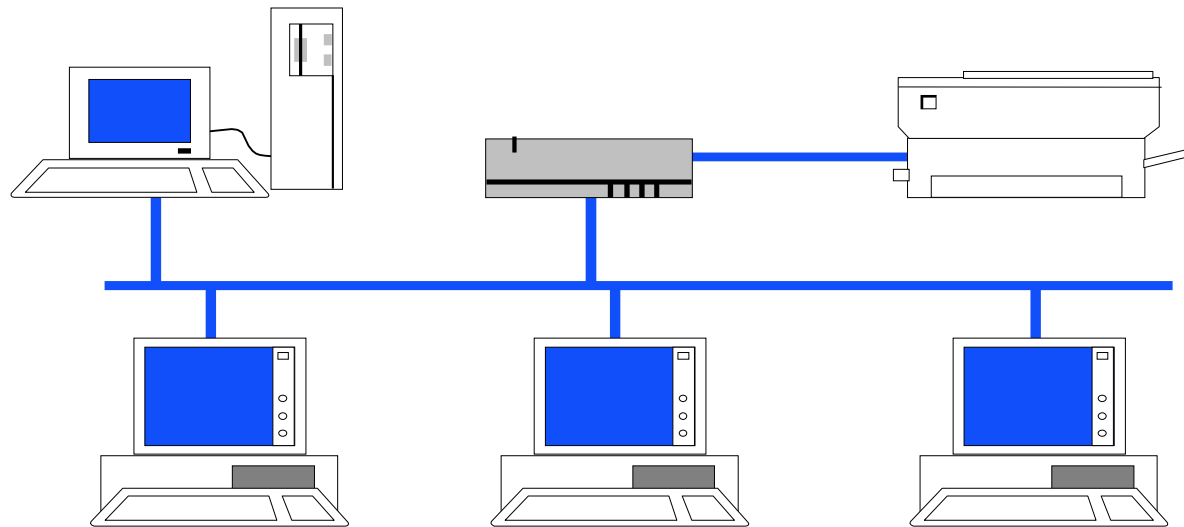
- Các máy tính được nối với nhau theo vòng khép kín\_
- Mỗi trạm đóng vai trò là một bộ chuyển tiếp
- Truyền dữ liệu quanh vòng thường theo một chiều
- Việc điều khiển truyền tin trên mạng xoay vòng thường được thực hiện bằng thẻ điều khiển
- Nhược điểm:
  - nếu đường nối chỉ bị cắt đứt thì có thể làm ngừng toàn bộ hoạt động của mạng\_
  - Lượng thông tin di chuyển trên đường truyền lớn



# 1.1 SƠ LƯỢC VỀ MẠNG MÁY TÍNH

## Mạng bus:

- Có một kênh chung để chuyển dữ liệu, các máy tính và các thiết bị đầu cuối sẽ được gắn vào đó.



Mạng bus

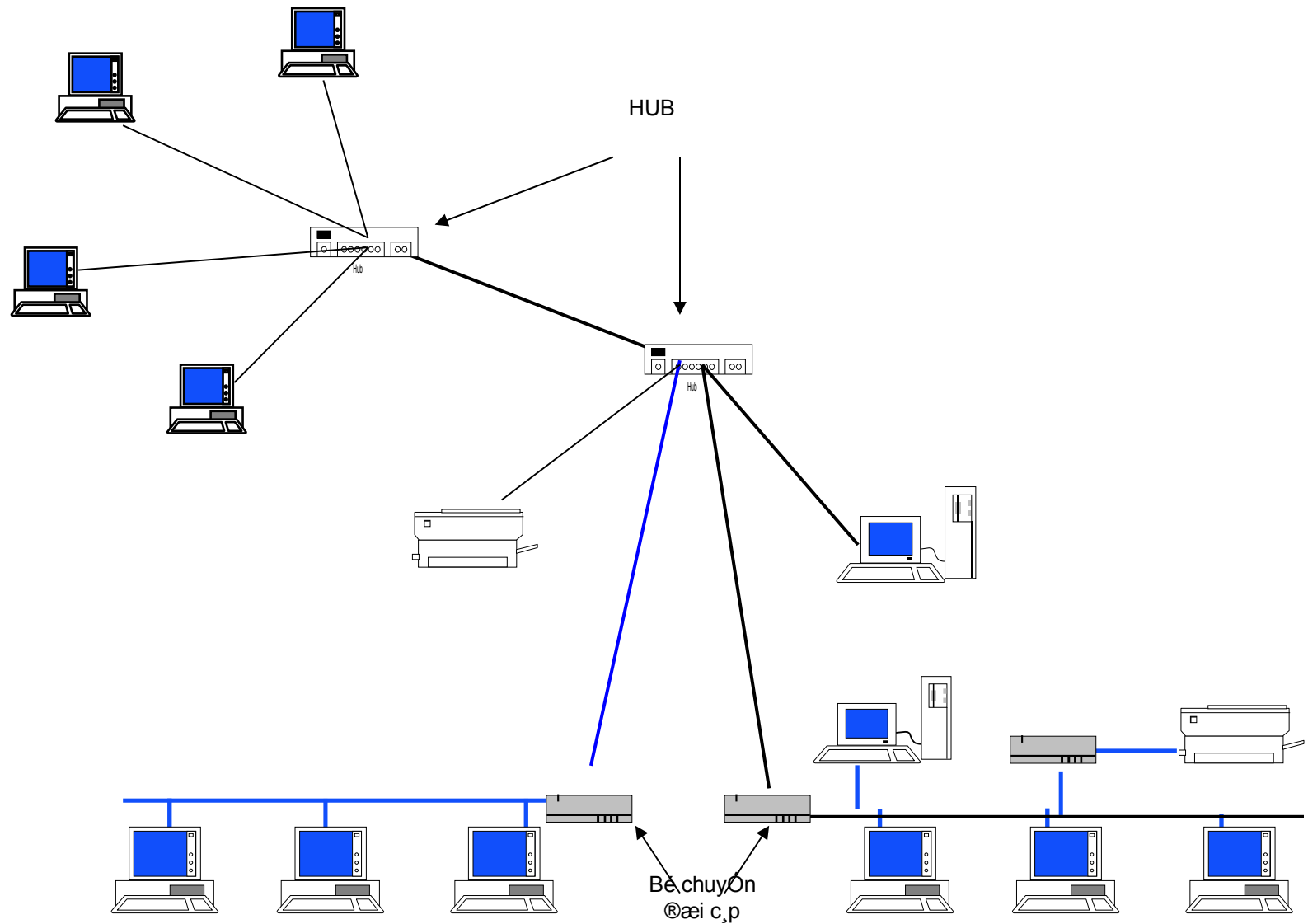
## 1.1 SƠ LƯỢC VỀ MẠNG MÁY TÍNH

---

### **Mạng hỗn hợp (hybrid):**

- Các mạng truyền thông thường có các đường nối vô định.
- Các đường nối không có tính hệ thống cũng không tuân theo một khuôn mẫu nào.
- Có thể gặp một nút chỉ nối với một nút khác và cả những nút nối với nhiều nút khác.
- Các nối kết giữa các máy tính trên mạng Internet thuộc loại này.\_

# 1.1 SƠ LƯỢC VỀ MẠNG MÁY TÍNH



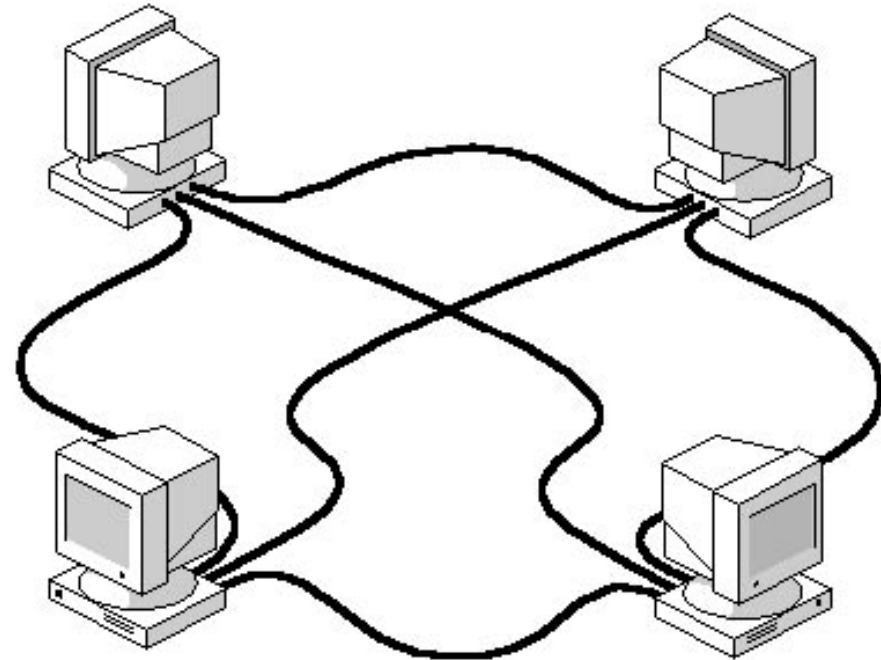
Mạng hỗn hợp



## 1.1 SƠ LƯỢC VỀ MẠNG MÁY TÍNH

### **Mạng thảm (Mesh):**

- Mỗi nút đều được nối kết với các nút còn lại.
- Có độ tin cậy cao hơn và khả năng hoạt động tốt hơn những cấu trúc đã nói ở trên.
- Cấu trúc này có chi phí cao.



## 1.2 Các hình thức tổ chức hệ thống phân tán

---

□ peer-to-peer

□ File server

□ Client/Server

## 1.2 Các hình thức tổ chức hệ thống phân tán

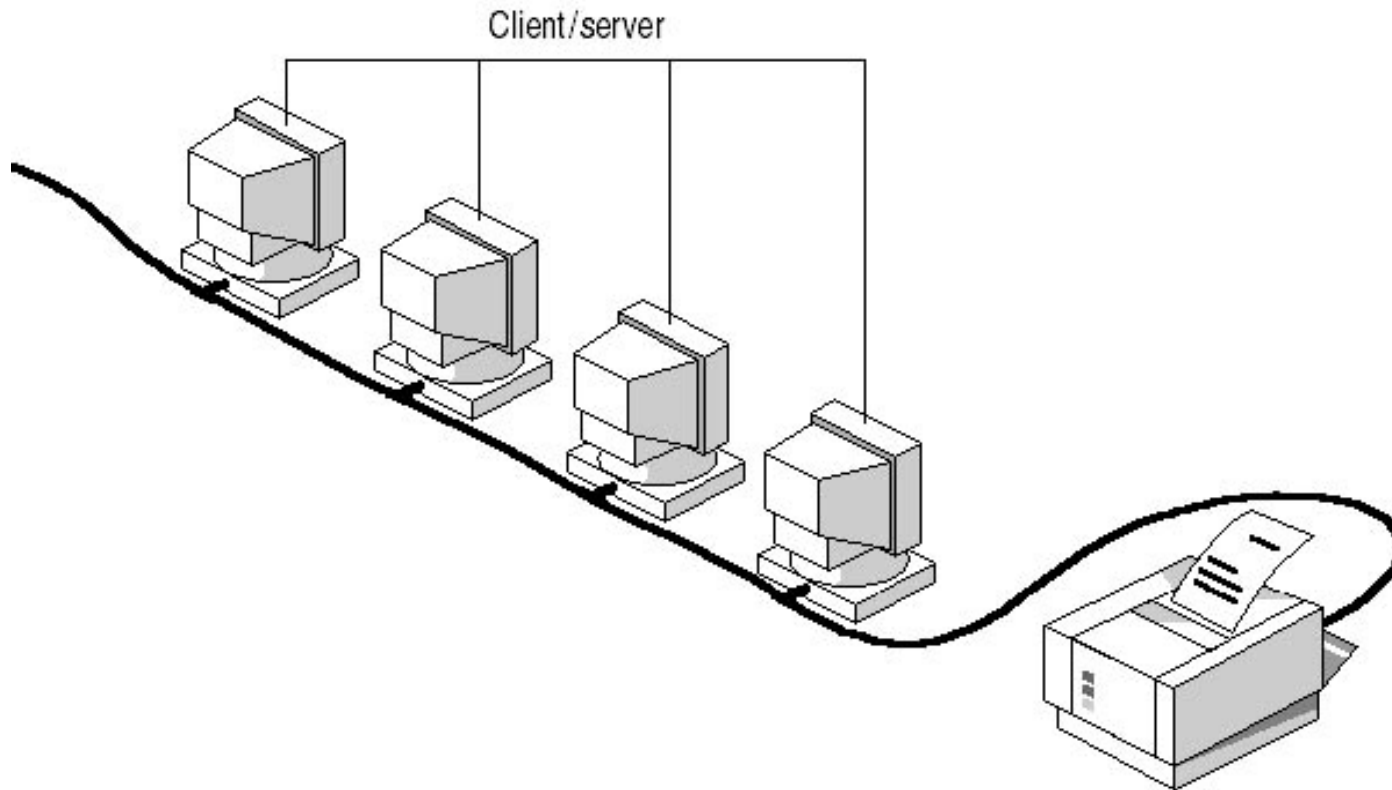
---

### **peer-to-peer (mô hình nhóm làm việc):**

- ❑ Các máy tính cá nhân và máy trạm có thể được sử dụng như một hệ thống độc lập trợ giúp các ứng dụng địa phương.
- ❑ Mỗi thành viên trong mạng có vai trò ngang nhau, tự quản lý tài nguyên của chính mình và chia sẻ tài nguyên cho các máy tính khác trên mạng
- ❑ Mỗi một máy tính trên mạng vừa đóng vai trò máy chủ (Server), vừa đóng vai trò là máy khách (Client).
- ❑ Đây là mô hình mạng đơn giản, phù hợp với những hệ thống mạng nhỏ không có yêu cầu cao về bảo mật.

## 1.2 Các hình thức tổ chức hệ thống phân tán

peer-to-peer (mô hình nhóm làm việc):



## 1.2 Các hình thức tổ chức hệ thống phân tán

### File server (máy dịch vụ file):\_

- ❑ **File server** Một số máy dịch vụ file được gán trực tiếp vào mạng LAN\_
- ❑ **File server** là một thiết bị quản lý các hoạt động file và phục vụ các máy tính cá nhân được kết nối trong mạng LAN
- ❑ Mỗi máy cá nhân được phân chia một dung lượng cố định trên ổ cứng của **File server**, chương trình ở các máy tính cá nhân có thể tham chiếu đến các file trên phần đĩa tương ứng của nó bằng một đặc tả đường dẫn.
- ❑ Những hạn chế của **File server**
  - Dữ liệu di chuyển trên mạng quá nhiều
  - Việc kiểm soát dữ liệu là phi tập trung
  - Các máy trạm phải đủ mạnh

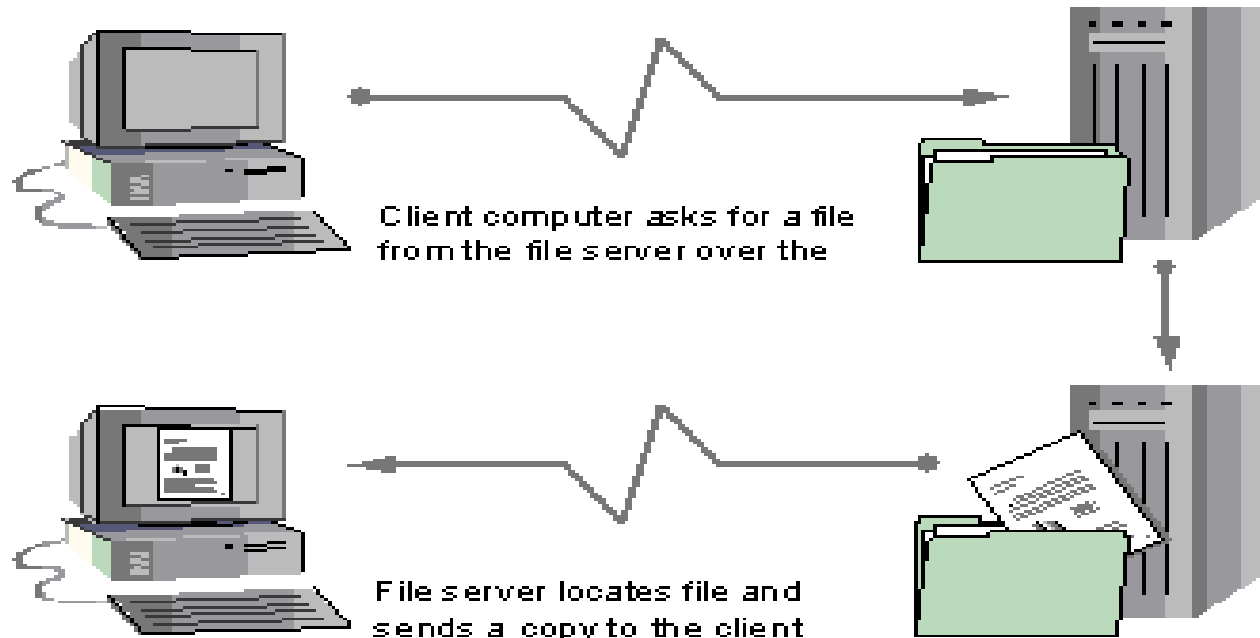
## 1.2 Các hình thức tổ chức hệ thống phân tán

### Client/server (mô hình khách/chủ)

Một số máy dịch vụ file được gán trực tiếp vào mạng LAN\_

❑ **\_server** có chức năng điều khiển, lưu trữ CSDL, xử lý các truy vấn và quản lý việc khai thác tài nguyên trên mạng của các máy tính khác.

❑ Thuật ngữ client được sử dụng để chỉ người khai thác tài nguyên mạng



## 1.3 Các đặc trưng của các loại hình hệ thống phân tán

---

Việc chuyển các ứng dụng trên máy tính cá nhân cũng như các ứng dụng trên các hệ thống máy lớn trung tâm sang mô hình phân tán là một xu hướng phát triển mạnh.

Vấn đề đặt ra là, cần lựa chọn hình thức phân tán thích hợp nào cho mỗi mô hình được chuyển đổi.

## 1.3 Các đặc trưng của các hệ thống phân tán

### 1.3.1 Đặc trưng của hệ thống File server và kiến trúc Client/Server

<b>Đặc trưng</b>	<b>File Server</b>	<b>Client/Server</b>
<b>Xử lý</b>	Chỉ ở máy khách	Cả máy khách và máy chủ
<b>Truy nhập dữ liệu đồng thời</b>	Thấp, mỗi máy khách th.hiện	Cao, Server đảm nhiệm
<b>An toàn và toàn vẹn CSDL</b>	Thấp, máy khách quản lý	Cao, Server đảm nhiệm
<b>Sử dụng mạng</b>	File lớn, chuyển cả file	Truyền dữ liệu nhiều mức
<b>Bảo trì phần mềm</b>	Thấp, chỉ ở máy Server	Hỗn hợp một số phần mềm có thể gửi đến máy khách
<b>Phần cứng và hệ thống Sự mềm dẻo của phần mềm</b>	Ghép nối máy khách và Server để có thể phối hợp	Ghép nối máy khách và Server để có thể phối hợp



## 1.3 Các đặc trưng của các hệ thống phân tán

### 1.3.2 Các chức năng của kiến trúc client/server

#### 1.3.2.1 Trình diễn thông tin phân tán

**Mục đích:** làm mới các ứng dụng trên các máy khách và để định dạng lại dữ liệu do server quản lý.

<b>Chức năng</b>	<b>Client</b>	<b>Server</b>
<b>Quản lý dữ liệu</b>		Quản lý mọi dữ liệu
<b>Phân tích dữ liệu</b>		Phân tích mọi dữ liệu
<b>Trình diễn dữ liệu</b>	Dữ liệu trình diễn trên server được định dạng để trình diễn cho các clients	Sử dụng công nghệ trình diễn của server để gửi dữ liệu gửi cho các clients

## 1.3 Các đặc trưng của các hệ thống phân tán

### 1.3.2.2. Trình diễn từ xa

**Mục đích:** giúp USERS thay đổi các biểu mẫu, báo cáo hoặc nội dung mới thì chỉ cần bảo trì phần mềm trên máy khách.

Kiểu trình diễn này được cài đặt các chức năng trình diễn dữ liệu trên các clients, nên phần mềm trên các client có khả năng trình diễn những dữ liệu được định dạng theo ý của client.

Chức năng	Client	Server
Quản lý dữ liệu		Quản lý mọi dữ liệu
Phân tích dữ liệu		Phân tích mọi dữ liệu
Trình diễn dữ liệu	Dữ liệu phân tích trên server được định dạng để trình diễn cho người dùng	

## 1.3 Các đặc trưng của các hệ thống phân tán

### 1.3.2.3. Quản lý dữ liệu từ xa

**Mục đích:** giúp client phân tích được dữ liệu thô lấy từ Server

Chức năng	Client	Server
Quản lý dữ liệu		Quản lý mọi dữ liệu
Phân tích dữ liệu	Dữ liệu thô được lấy từ server và được phân tích	
Trình diễn dữ liệu	Trình diễn tất cả dữ liệu	

## 1.3 Các đặc trưng của các hệ thống phân tán

### 1.3.2.4. Phân tán chức năng

- Các chức năng phân tán được đặt trên cả máy khách lẫn máy chủ.
- Toàn bộ chức năng **trình diễn dữ liệu** được đặt trên máy khách và toàn bộ chức năng **quản lý dữ liệu** được đặt trên máy chủ

<b>Chức năng</b>	<b>Client</b>	<b>Server</b>
<b>Quản lý dữ liệu</b>		Quản lý mọi dữ liệu
<b>Phân tích dữ liệu</b>	Các dữ liệu được lấy và phân tích từ server	Các dữ liệu được lấy và phân tích từ server sau đó được truyền cho các clients
<b>Trình diễn dữ liệu</b>	Tất cả dữ liệu (được phân tích trên cả server và clients)	

## 1.3 Các đặc trưng của các hệ thống phân tán

### 1.3.3 Xử lý phân tán (*Distributed Processing*):

- Cho phép phối hợp tốt nhất các đặc trưng của hệ phân tán như liên kết dữ liệu và chương trình trên máy chủ và chỉ để lại chức năng trình diễn trên máy khách.
- Cho phép định vị một cách mềm dẻo cả chức năng phân tích và xử lý dữ liệu ở nơi mà chúng hoạt động tốt nhất.

Chức năng	Client	Server
Quản lý dữ liệu	Quản lý dữ liệu địa phương	Chia sẻ quản lý dữ liệu trên server
Phân tích dữ liệu	Dữ liệu được lấy từ cả server và các clients để phân tích	Dữ liệu được lấy từ server để phân tích. Sau đó gửi cho máy khách để phân tích tiếp và trình diễn
Trình diễn dữ liệu	Tất cả dữ liệu	

## 1.4 Cơ sở dữ liệu phân tán (*Distributed Database*)

### 1.4.1 Định nghĩa:

Một cơ sở dữ liệu (CSDL) phân tán là một tập hợp dữ liệu, mà về mặt logic tập hợp này thuộc cùng một hệ thống, nhưng về mặt vật lý dữ liệu đó được phân tán trên các vị trí khác nhau của một mạng máy tính.

Có hai điểm quan trọng được nêu ra trong định nghĩa:

***Phân tán:*** Dữ liệu không cư trú trên một vị trí mà được phân bố rộng khắp trên nhiều máy tính đặt tại nhiều vị trí khác nhau, đây là điểm phân biệt một cơ sở dữ liệu phân tán với một cơ sở dữ liệu tập trung.

***Tương quan logic:*** Dữ liệu trong hệ phân tán có một số thuộc tính ràng buộc chúng với nhau. Điều này giúp chúng ta có thể phân biệt một cơ sở dữ liệu phân tán với một tập hợp cơ sở dữ liệu tập trung, các file dữ liệu được lưu trữ tại nhiều vị trí khác nhau, điều này thường thấy trong các ứng dụng mà hệ thống sẽ phân quyền truy nhập dữ liệu trong môi trường mạng.

## 1.4 Cơ sở dữ liệu phân tán (*Distributed Database*)

---

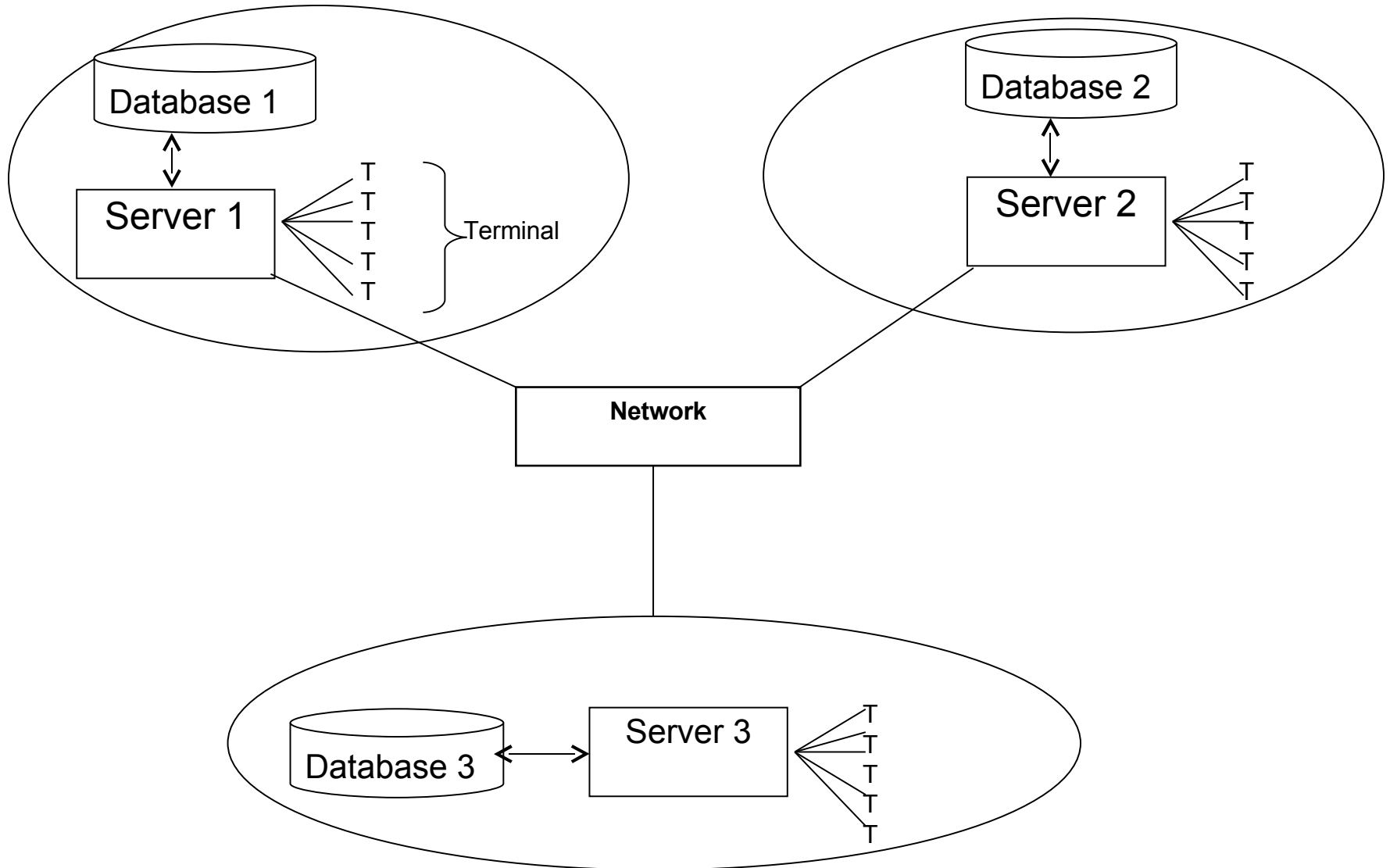
Ví dụ:

Một ngân hàng có ba chi nhánh đặt tại các vị trí khác nhau.

Tại mỗi chi nhánh có một máy tính điều khiển một số máy kế toán cuối cùng (teller terminal) và cơ sở dữ liệu thống kê của chi nhánh đó.

Mỗi máy tính với cơ sở dữ liệu thống kê địa phương của nó tại mỗi chi nhánh được đặt ở một vị trí của cơ sở dữ liệu phân tán. Các máy tính được nối với nhau bởi một mạng truyền thông. Các nút trong một mạng phân tán một mặt xử lý thông tin tại vị trí mà nó quản lý, mặt khác nó cũng tham gia vào việc xử lý các yêu cầu về thông tin cần truy cập qua nhiều địa điểm. Ví dụ như việc lên danh sách tất cả nhân viên của ngân hàng. Yêu cầu này đòi hỏi tất cả các máy tính ở các chi nhánh của công ty đều phải hoạt động để cung cấp thông tin.

## 1.4 Cơ sở dữ liệu phân tán (*Distributed Database*)





## 1.4 Cơ sở dữ liệu phân tán (*Distributed Database*)

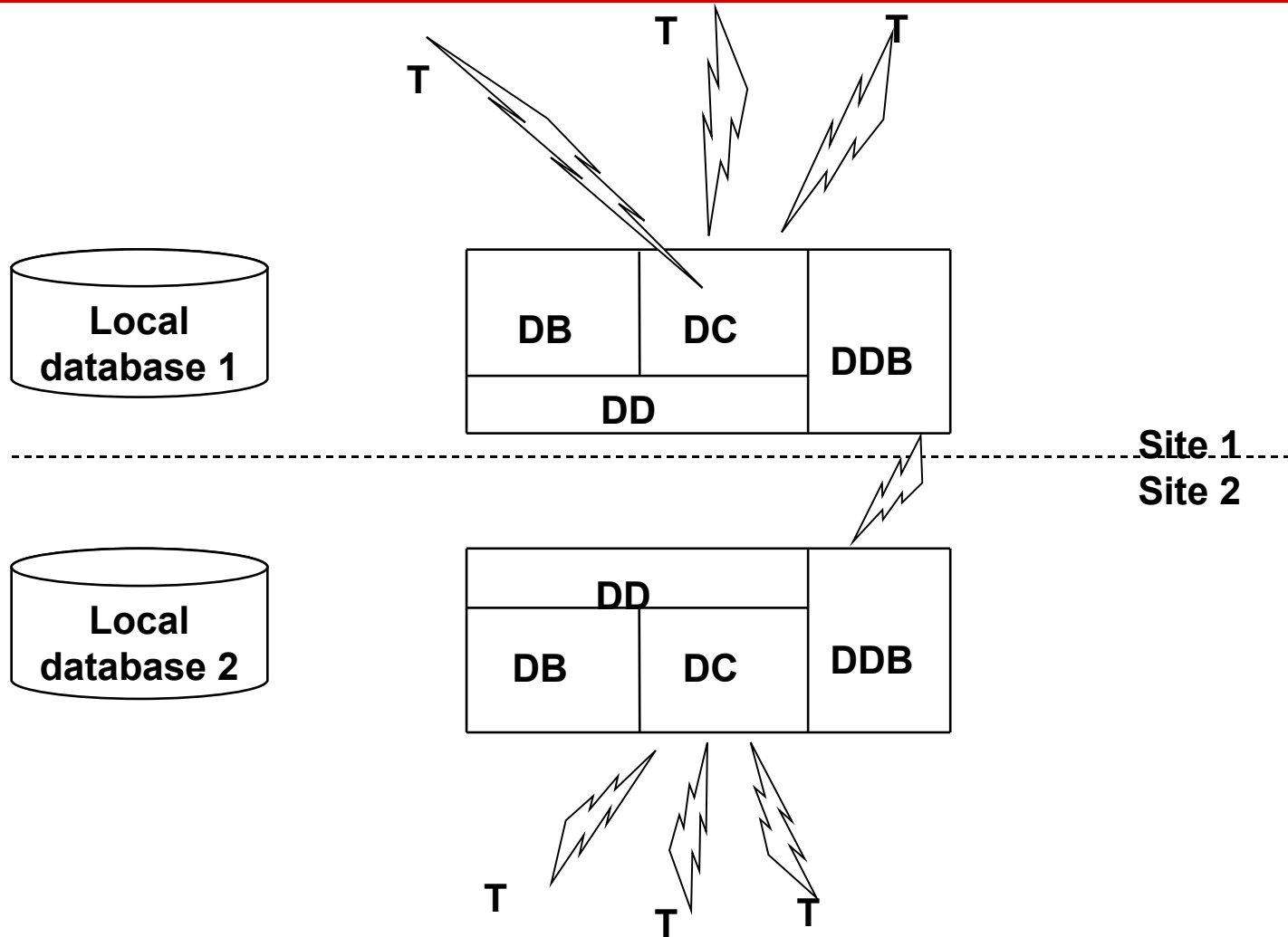
### 1.4.2 Hệ quản trị cơ sở dữ liệu phân tán (DDBMSs)

DDBMS có chức năng hỗ trợ việc tạo và bảo trì cơ sở dữ liệu phân tán, chúng có các thành phần tương tự như một hệ quản trị cơ sở dữ liệu tập trung và các thành phần hỗ trợ trong việc chuyển tải dữ liệu đến các trạm và ngược lại.

Các thành phần sau đây đòi hỏi một DDBMSs thương mại phải có:

- ☐ Quản trị dữ liệu (Database management): DB
- ☐ Truyền thông dữ liệu (Data Communication): DC
- ☐ Từ điển dữ liệu (Data Dictionary): DD dùng để mô tả thông tin về sự phân tán của dữ liệu trên mạng.
- ☐ Cơ sở dữ liệu phân tán (Distributed Database): DDB

## 1.4 Cơ sở dữ liệu phân tán (*Distributed Database*)



Các thành phần của một DDBMS thương mại

## 1.4 Cơ sở dữ liệu phân tán (*Distributed Database*)

### 1.4.3 So sánh csdl phân tán và csdl tập trung

**Nhận xét:** CSDL phân tán không đơn giản là những sự thực hiện phân tán của CSDL tập trung, bởi vì chúng cho phép thiết kế các đặc trưng khác với CSDL tập trung truyền thống.

**Các đặc điểm tiêu biểu của CSDL truyền thống:**

- điều khiển tập trung
- độc lập dữ liệu
- giảm dư thừa
- biệt lập và bảo mật dữ liệu.\_

## 1.4 Cơ sở dữ liệu phân tán (*Distributed Database*)

### a. Điều khiển tập trung

**Trong CSDL tập trung**: Khả năng điều khiển tập trung trên toàn nguồn tài nguyên thông tin của tổ chức, được xem là động cơ mạnh nhất cho việc ra đời CSDL. Chúng được phát triển như là sự tiến hoá của hệ thống thông tin, mà trong đó mỗi ứng dụng có các tập tin riêng của nó.

**Trong CSDL phân tán**, ý niệm về điều khiển tập trung ít được nhấn mạnh hơn. Điều này phụ thuộc vào kiến trúc của CSDL phân tán. Một cách tổng quát, CSDL phân tán được điều khiển với cấu trúc phân lớp dựa vào một hệ quản trị CSDL toàn cục (có trách nhiệm trên toàn bộ CSDL phân tán) và hệ quản trị CSDL địa phương (có trách nhiệm với CSDL địa phương riêng). Điều này cho thấy rằng hệ quản trị CSDL địa phương có thể có một mức tự trị cao. Các CSDL phân tán có thể rất khác nhau về mức độ tự trị: từ hoàn toàn tự trị, không có bất cứ một hệ quản trị CSDL tập trung nào, đến hầu như hoàn toàn điều khiển tập trung.

## 1.4 Cơ sở dữ liệu phân tán (*Distributed Database*)

### b. Độc lập dữ liệu

Độc lập dữ liệu cũng là một trong những động lực cho việc mở đầu sự tiếp cận dữ liệu phân tán. Thuận lợi chính của độc lập dữ liệu là các chương trình không bị ảnh hưởng bởi sự thay đổi trong cấu trúc vật lý của dữ liệu. Trong CSDL phân tán, độc lập dữ liệu cũng quan trọng giống như trong CSDL truyền thống. Tuy nhiên, một khía cạnh mới được thêm vào trong ý niệm của độc lập dữ liệu là *trong suốt phân tán*. Với trong suốt phân tán chúng ta hiểu rằng các chương trình ứng dụng có thể sử dụng CSDL như là nó không được tổ chức phân tán. Vì thế sự chính xác của chương trình không bị ảnh hưởng bởi việc dịch chuyển dữ liệu từ trạm này đến trạm khác. Tuy nhiên, tốc độ thực hiện của chúng bị ảnh hưởng.

## 1.4 Cơ sở dữ liệu phân tán (*Distributed Database*)

### c. Giảm dư thừa dữ liệu

**Trong CSDL truyền thống**, dữ liệu dư thừa được giảm đến mức tối thiểu bởi hai lý do:

- Sự không tương thích giữa nhiều bản sao của cùng một tập dữ liệu.
- Tiết kiệm không gian lưu trữ bằng cách loại bỏ các dư thừa. Việc giảm dư thừa dữ liệu có thể đạt được bằng cách chia sẻ dữ liệu, cho phép nhiều ứng dụng truy cập cùng các bản tin và bản ghi.

**Trong CSDL phân tán**, việc giảm dư thừa phức tạp hơn vì ngoài hai lý do trên, còn nhiều lý do để giảm dư thừa như:

- Hoạt động của các trình ứng dụng có thể bị tăng lên khi dữ liệu được sao lại tất cả các vị trí, nơi trình ứng dụng cần nó.
- Tính thường trực của hệ thống sẽ tăng lên, bởi vì khi có lỗi xảy ra ở một trạm nào đó sẽ không dừng việc thực hiện các ứng dụng của trạm khác nếu dữ liệu đã được sao chép lại.

## 1.4 Cơ sở dữ liệu phân tán (*Distributed Database*)

### d. Biệt lập và bảo mật

**Trong CSDL truyền thống**, hệ quản trị CSDL tập trung có thể bảo đảm chỉ truy cập đến dữ liệu đã được uỷ quyền.

**Trong CSDL phân tán**, hệ quản trị dữ liệu địa phương thực chất phải đương đầu với các vấn đề giống như hệ quản trị CSDL trong CSDL truyền thống.

Tuy nhiên, hai khía cạnh đặc biệt sau đây của CSDL phân tán cần phải được xem xét:

- Trong CSDL phân tán với một mức độ tự trị rất cao của các địa phương, người chủ dữ liệu địa phương cảm giác được bảo vệ tốt hơn vì họ có thể tự chủ thực hiện bảo vệ thay vì phụ thuộc vào người quản trị CSDL trung tâm.
- Vấn đề bảo mật là bản chất trong hệ phân tán nói chung, vì các mạng truyền thông diện rộng cho phép nhiều người cập nhật và khai thác dữ liệu nên cần được bảo vệ.

## 1.4 Cơ sở dữ liệu phân tán (*Distributed Database*)

---

### 1.4.4 Ưu và nhược điểm của hệ phân tán

#### Ưu điểm

- Đáp ứng nhanh hầu hết các ứng dụng sử dụng dữ liệu tại các trạm
- Tăng cường các đơn thể ứng dụng và CSDL mà không làm cản trở người sử dụng hiện tại.
- Kiểm soát dữ liệu địa phương theo hướng hoàn thiện sự tích hợp và quản trị dữ liệu từ xa.
- Tăng cường khả năng của hệ thống liên quan đến sự dư thừa dữ liệu.



## 1.4 Cơ sở dữ liệu phân tán (*Distributed Database*)

---

### 1.4.4 Ưu và nhược điểm của hệ phân tán

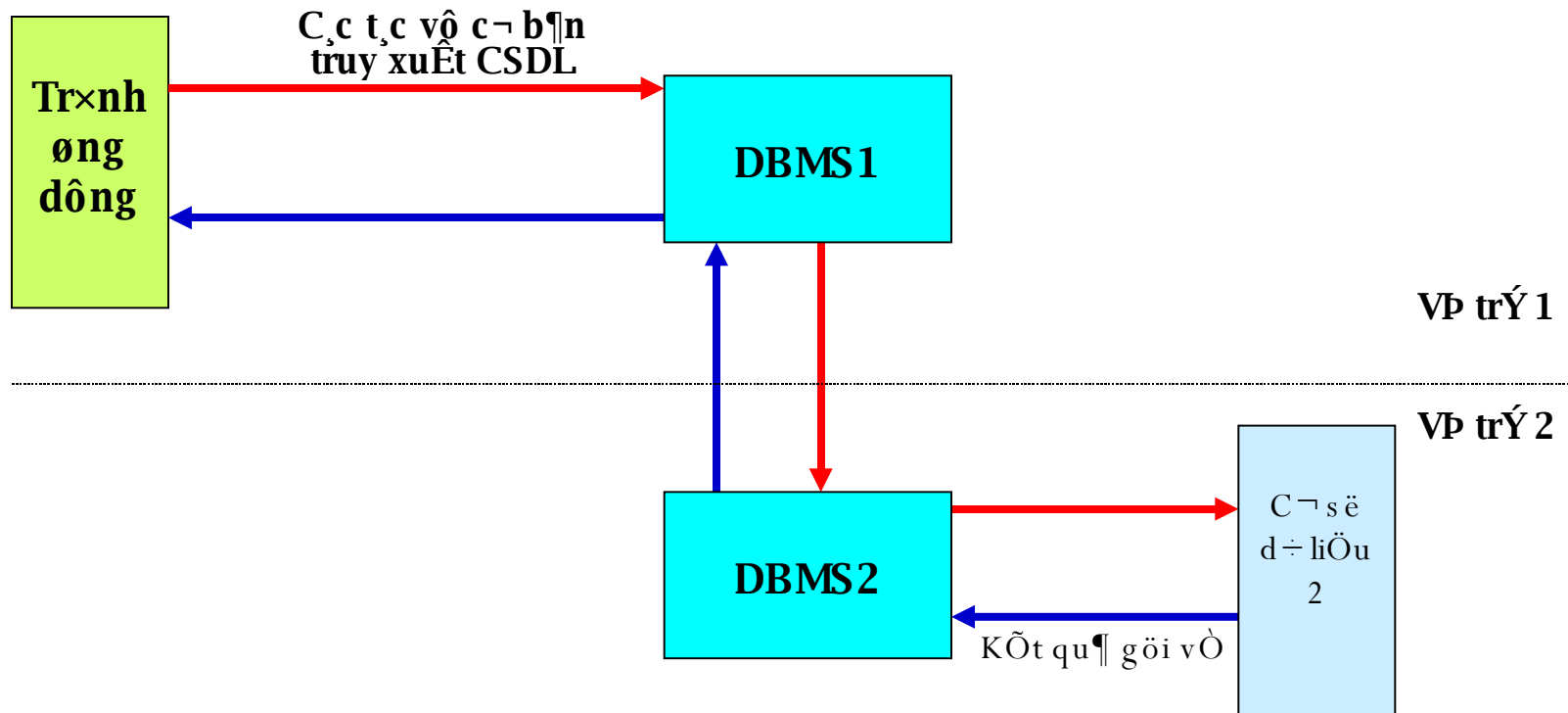
#### Nhược điểm

- Phần mềm đắt và phức tạp
- Phải xử lý các thay đổi thông báo trong mọi địa điểm
- Khó kiểm soát tính toàn vẹn dữ liệu với nhiều bản sao dữ liệu được phân bố khắp mọi nơi.
- Đáp ứng chậm nhu cầu của các trạm trong trường hợp các phần mềm ứng dụng không được phân bố phù hợp với việc sử dụng chung.

## 1.5 Các loại truy xuất CSDL phân tán

### 1.5.1 Truy xuất tổ xa thông qua các tập vô cùng:

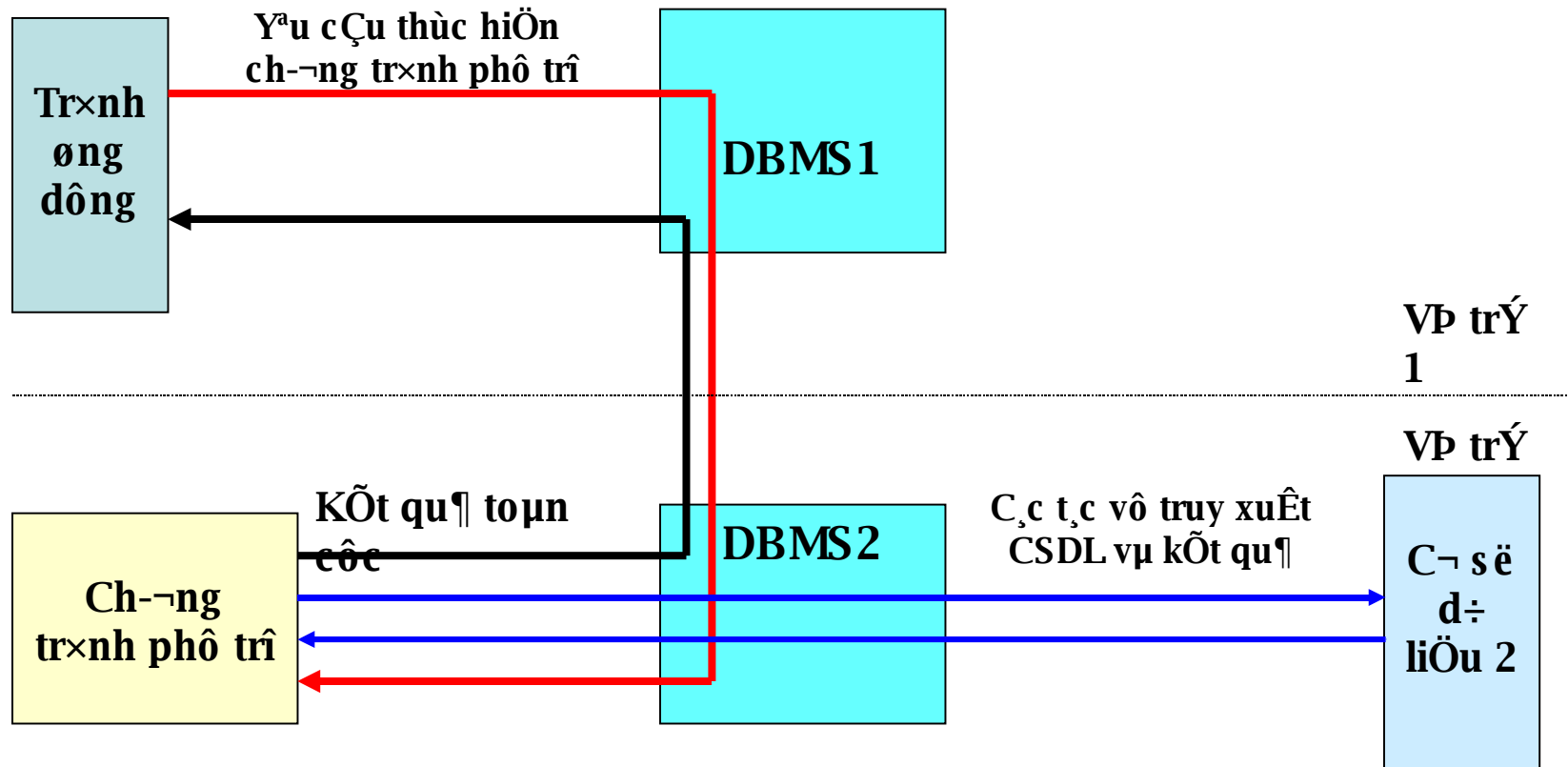
Mong đợi phát ra một yêu cầu truy xuất CSDL là một việc tự nhiên. Yêu cầu này sẽ kích hoạt quá trình truy xuất CSDL phân tán gọi là quá trình chia dữ liệu. Thực hiện xong sẽ gọi kết quả về.



## 1.5 Các loại truy xuất CSDL phân tán

### 1.5.2 Truy xuất tổ xa thông qua chương trình phân phối

Một ứng dụng yêu cầu thực hiện một chương trình phân phối trên một tập máy chủ xa. Chương trình phân phối này sẽ truy xuất CSDL tổ xa và tuân thủ các quy tắc cho ứng dụng yêu cầu.



Cơ sở dữ liệu phân tán là quan trọng trong kinh tế, tổ chức và kỹ thuật với nhiều lý do khác nhau. Chúng có thể được cài đặt trên một mạng máy tính có phạm vi rộng lớn hoặc nhỏ bé.

Hiện nay các DDBMSs thương mại đều tích hợp các ứng dụng phân tán nên rất tiện cho người sử dụng.

# Câu hỏi cuối chương

---

1. Ưu và nhược điểm của các loại mạng máy tính.
2. Trong mô hình client/Server, hãy cho biết mục đích của:
  - Trình diễn thông tin phân tán
  - Trình diễn từ xa
  - Quản lý dữ liệu từ xa
  - Phân tán chức năng
3. Định nghĩa cơ sở dữ liệu phân tán. Hệ quản trị CSDL phân tán. Các thành phần yêu cầu trong một DDBMSs thương mại. Cho ví dụ về một CSDL phân tán .
4. So sánh csdl phân tán và csdl tập trung
5. Ưu và nhược điểm của hệ phân tán
6. Các truy xuất từ xa trong hệ thống phân tán

# CHƯƠNG I TỔNG QUAN VỀ CSDL PHÂN TÁN

---



## HẾT CHƯƠNG 1