

PGS.TS. NGUYỄN VĂN ĐỊNH

# BÀI GIẢNG

## CƠ SỞ DỮ LIỆU 2

CHƯƠNG TRÌNH ĐỊNH HƯỚNG HỌC THUẬT



HÀ NỘI 2017

# **Chương 1.**

## **Tổng quan về Cơ sở dữ liệu phân tán.**

Trong chương này, chúng ta đề cập đến một số khái niệm và kết quả cơ bản liên quan đến việc thiết kế một CSDL phân tán.

- 1.1 Cấu trúc của hệ CSDL phân tán
  - 1.1.1 Lịch sử hình thành công nghệ CSDL phân tán.
  - 1.1.2 Cấu trúc của hệ CSDLPT
- 1.2 Các mô hình CSDL phân tán
  - 1.2.1 Tính thuần nhất và không thuần nhất của một hệ CSDL phân tán
  - 1.2.2 Cơ sở dữ liệu từ xa
  - 1.2.3 Cơ sở dữ liệu liên hiệp
  - 1.2.4 Cơ sở dữ liệu song song
- 1.3. Các mục tiêu của một hệ quản trị CSDL phân tán
  - 1.3.1 Tính độc lập đối với sự phân bố dữ liệu
  - 1.3.2 *Tính độc lập đối với sự phân đoạn của các quan hệ*
  - 1.3.3 Tính độc lập đối với sự nhân bản
  - 1.3.4 Tính độc lập đối với các hệ quản trị CSDL cục bộ
  - 1.3.5 Tính tự trị của các trạm
  - 1.3.6 Tính mở rộng
  - 1.3.7 *Cải thiện hiệu năng*
- 1.4 Kiến trúc lược đồ và kiến trúc chức năng của một hệ CSDL phân tán
  - 1.4.1 Kiến trúc lược đồ của một hệ CSDL phân tán
  - 1.4.2 Kiến trúc chức năng của một hệ CSDL phân tán

Bài tập chương 1.

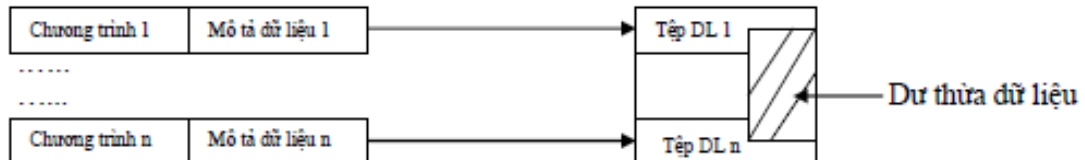
## 1.1. CẤU TRÚC CỦA HỆ CƠ SỞ DỮ LIỆU PHÂN TÁN

Dữ liệu là một thành phần không thể thiếu trong các ứng dụng công nghệ thông tin (CNTT). Quá trình phát triển của CNTT cũng kéo theo quá trình phát triển của công nghệ CSDL.

### 1.1.1. Sự phát triển của công nghệ CSDL

Công nghệ lưu trữ và xử lý dữ liệu cho đến nay đã trải qua các giai đoạn phát triển chính như sau:

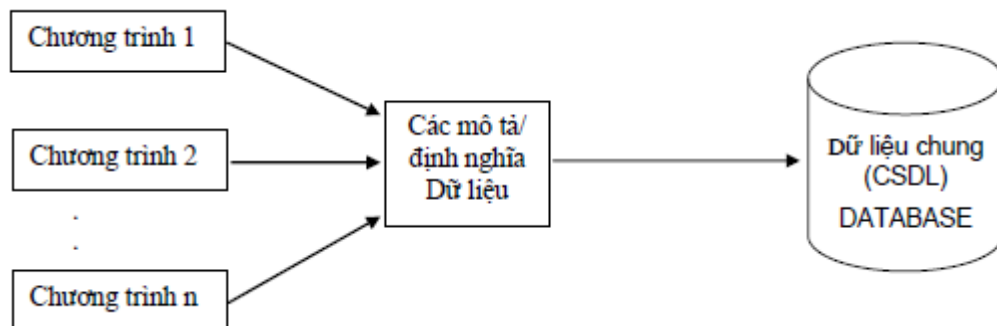
- *Xử lý tệp truyền thống:* Trong mô hình này, dữ liệu lưu trữ thành các tệp riêng rẽ, được mô tả và gắn kết với từng chương trình ứng dụng.



**Hình 1.1.** Sơ đồ xử lý tệp truyền thống

Cách xử lý này sẽ phát sinh vấn đề dư thừa dữ liệu và các dị thường trong cập nhật. Chẳng hạn, có thể có 2 chương trình ứng dụng với 2 tệp dữ liệu riêng, nhưng mỗi tệp dữ liệu này lại có những thông tin giống nhau, vì vậy tạo nên sự dư thừa dữ liệu, gây lãng phí cho lưu trữ và quản lý. Nguy hại hơn là khi cập nhật thay đổi thông tin về một thực thể trong tệp dữ liệu này thì cũng phải thay đổi trong tệp kia với đối tượng đó. Điều này rất khó thực hiện triệt để khi có nhiều tệp dữ liệu có chứa các thông tin trùng lặp, và sẽ gây ra các dị thường trong cập nhật (thông tin không nhất quán, thông tin mâu thuẫn...).

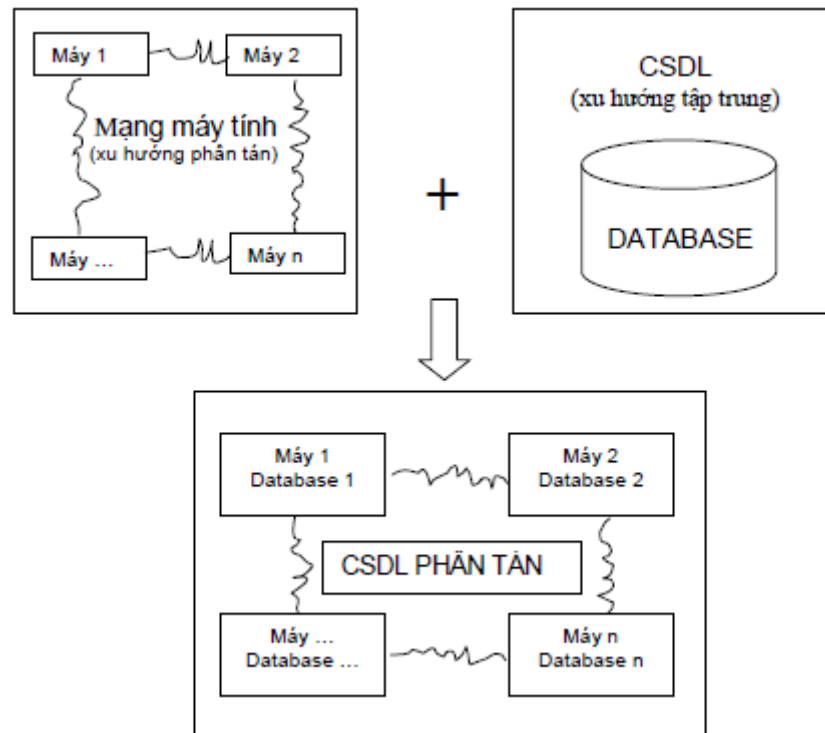
- *Xử lý CSDL (Database-DB):* Để khắc phục các nhược điểm của mô hình xử lý tệp, dữ liệu của các ứng dụng được lưu trữ tập trung, được mô tả và xử lý thống nhất cho mọi chương trình ứng dụng. Các dữ liệu tập trung lại thành một CSDL, được xử lý theo cách thống nhất cho tất cả các ứng dụng có liên quan.



**Hình 1.2.** Sơ đồ xử lý CSDL

Như vậy vấn đề tập trung dữ liệu là mấu chốt của xử lý CSDL. Công nghệ CSDL là một bước tiến lớn của công nghệ xử lý dữ liệu. Rõ ràng là nhờ sự tập trung dữ liệu, CSDL đã khắc phục được sự dư thừa dữ liệu, hạn chế các dị thường trong cập nhật. Các hệ CSDL đầu tiên đã xuất hiện từ những năm 1960, và từ đó đến nay chúng ta đã chứng kiến nhiều hệ quản trị CSDL ra đời như FoxBase, FoxPro, MS Access, MySQL, SQL Server, Firebird, Oracle.

▪ **Công nghệ CSDL phân tán (Distributed Database-DDB)** : Khi công nghệ mạng máy tính ra đời, người ta phân chia CSDL tập trung thành các CSDL cục bộ đặt trên các máy tính khác nhau (các trạm / sites) và được kết nối bởi mạng truyền thông. Các CSDL cục bộ này tích hợp được với nhau qua mạng máy tính mà không cần có sự tập trung dữ liệu. Do đó có thể xây dựng được một hệ thống CSDL, trong đó dữ liệu phân tán trên một mạng máy tính. Đó là cách tiếp cận của CSDL phân tán.



**Hình 1.3.** Sơ đồ hình thành CSDL phân tán

Có thể nhận thấy một điều thú vị là: Công nghệ mạng máy tính xúc tiến một kiểu làm việc đi ngược lại nỗ lực tập trung hóa của Công nghệ CSDL, tuy vậy hai xu hướng trái ngược nhau này lại có thể được tổng hợp lại để cho ra đời một công nghệ mạnh mẽ hơn nhiều so với từng công nghệ riêng lẻ: đó là công nghệ CSDL phân tán.

Do đó có thể thấy rằng mục tiêu quan trọng nhất của công nghệ CSDL là sự tích hợp chứ không phải là sự tập trung hóa, có nghĩa là vẫn có thể tích hợp dữ liệu mà không cần tập trung hóa, và đó là bản chất của CSDL phân tán.

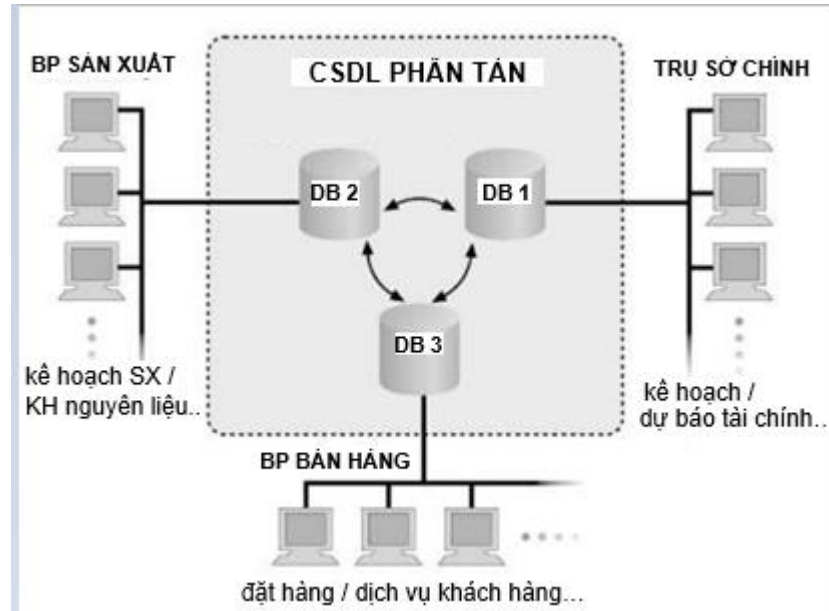
**Một số ứng dụng của CSDL phân tán.** Có rất nhiều lý do để chúng ta cần phải có nhiều CSDL trong một hệ thống phân tán như:

- Nhiều database có thể gắn với nhiều chức năng của một hệ thống như: Bộ phận sản xuất, bộ phận quản lý nhân sự, bộ phận bán hàng... và các chức năng này lại là những hệ thống độc lập có nhu cầu lấy thông tin lẫn nhau.
- Đơn vị sử dụng hệ thống database có nhiều trụ sở đặt ở nhiều vị trí địa lý khác nhau và cách xa nhau, ví dụ như hệ thống quản lý tài khoản của mạng điện thoại di động, các server game...
- Những ứng dụng đòi hỏi dữ liệu có tốc độ truy xuất rất cao được đặt các bản copy ở nhiều server, việc truy xuất được phân tải cho các server này để đảm bảo hiệu năng, nhưng dữ liệu trên các phần copy này phải đảm bảo đồng bộ tuyệt đối với nhau.

- Nhiều ứng dụng khác nhau có thể đòi hỏi phải truy xuất vào cùng một dữ liệu theo cách khác nhau, khi đó đòi hỏi phải có cơ chế truy xuất khác nhau.

Hình vẽ dưới đây mô tả một ứng dụng CSDL phân tán trong một doanh nghiệp gồm các bộ phận chính: *Trụ sở chính* là nơi điều hành mọi hoạt động của doanh nghiệp, cân đối dự báo tài chính..., *Bộ phận sản xuất* là nơi sản xuất, lập kế hoạch SX và kế hoạch nguyên liệu... và *Bộ phận bán hàng* nhận các đơn đặt hàng, phân phối hàng hóa, dịch vụ sau bán hàng...

Các bộ phận này của doanh nghiệp đặt ở 3 vị trí địa lý khác nhau, và CSDL của doanh nghiệp được kết nối theo mô hình phân tán.



**Hình. 2.1. Một ứng dụng của CSDL phân tán**

### 1.1.2. Cấu trúc của hệ CSDL phân tán

#### 1.1.2.1 Định nghĩa CSDL phân tán.

Có thể đưa ra định nghĩa cho CSDL phân theo cách mô tả như sau:

**Định nghĩa 1.1** CSDL phân tán là một tập hợp các CSDL cùng hợp tác hoạt động, được lưu trữ trên các máy tính khác nhau (gọi là các trạm/sites) được kết nối với nhau bởi một mạng truyền thông, và được quản lý bởi một hệ quản trị CSDL thống nhất.

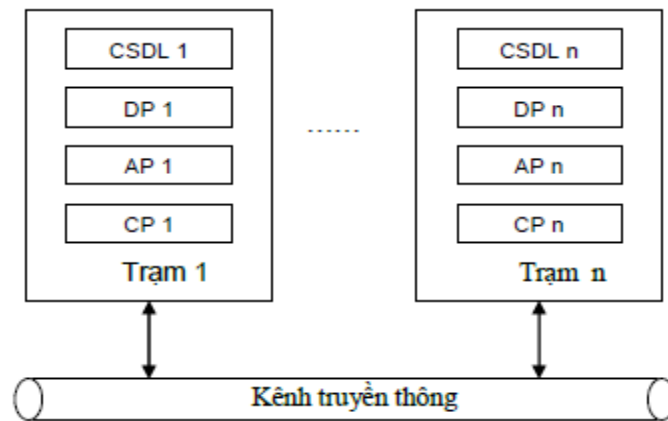
Một cách lý tưởng, người dùng tại một trạm bất kỳ có thể truy cập và thao tác dữ liệu trên bất cứ một trạm nào, các CSDL đó được người dùng nhìn thấy và xử lý chỉ như là một CSDL tập trung, người dùng có thể không biết và không cần biết là có sự phân tán của dữ liệu.

Tính chất này còn gọi là “sự phân tán của dữ liệu là trong suốt đối với người dùng”

#### 1.1.2.2. Các thành phần của CSDL phân tán:

a/. Về mặt vật lý: Yếu tố chính để phân biệt một cơ sở dữ liệu phân tán với một cơ sở dữ liệu tập trung là:

- Cơ sở dữ liệu phân tán phải có nhiều máy tính gọi là các trạm (sites hay gọi là các nút-Nodes).
- Các trạm phải được kết nối bởi một kênh truyền thông nào đó để truyền dữ liệu (trao đổi dữ liệu) và các lệnh.



**Hình 1.4** Các thành phần của CSDL phân tán

b/. Về mặt logic (phần mềm) gồm 3 module chính sau:

- Xử lý dữ liệu (DP-Database Processing): Quản lý dữ liệu cục bộ tại các trạm giống như với cơ sở dữ liệu tập trung.
- Xử lý ứng dụng (AP-Application Processing) cơ sở dữ liệu trên từng trạm: Xử lý các chức năng phân tán và truy cập các thông tin phân tán.
- Xử lý Truyền thông (CP-Communication Processing): Liên kết (kết nối) với kênh truyền thông.

c/. Ngoài ra hệ quản trị cơ sở dữ liệu phân tán còn có thêm một số chức năng:

- Quản lý từ điển dữ liệu tổng thể lưu trữ thông tin liên quan đến các dữ liệu phân tán (siêu dữ liệu- Super Data).
- Định nghĩa các dữ liệu phân tán.
- Kiểm tra các ngữ nghĩa của các dữ liệu.
- Định giá (trả lời) các câu hỏi phân tán.
- Quản lý các giao dịch phân tán

d/. Một số ưu điểm tiềm năng của hệ CSDLPT:

- Thích hợp cho bản chất phân tán của nhiều ứng dụng cơ sở dữ liệu.
- Cho phép dùng chung dữ liệu trong khi vẫn duy trì được các điều khiển cục bộ.
- Nâng cao hiệu năng hệ thống: do CSDL phân tán bao gồm nhiều CSDL cục bộ tại nhiều trạm, do đó cho phép tăng khả năng truy cập, thực hiện được nhiều xử lý song song do đó tăng được hiệu suất, tốc độ xử lý dữ liệu và ứng dụng
- Tăng độ tin cậy và an toàn dữ liệu: do dữ liệu được lưu trữ trên nhiều bản sao đồng bộ, do đó có khả năng khôi phục dữ liệu khi xảy ra thảm họa ở một vài trạm.

**Kết luận:** Như vậy, có thể xem CSDLPT là một bộ sưu tập các dữ liệu mà về mặt logic thuộc cùng một hệ thống CSDL nhưng về mặt vật lý được phân bố trên các trạm của một mạng máy tính. Hệ quản trị CSDL PT (DDBMS) là hệ thống phần mềm cho phép quản lý CSDLPT, làm cho việc phân tán trở nên trong suốt với người dùng. [1]

## 1.2. CÁC MÔ HÌNH PHÂN TÁN DỮ LIỆU

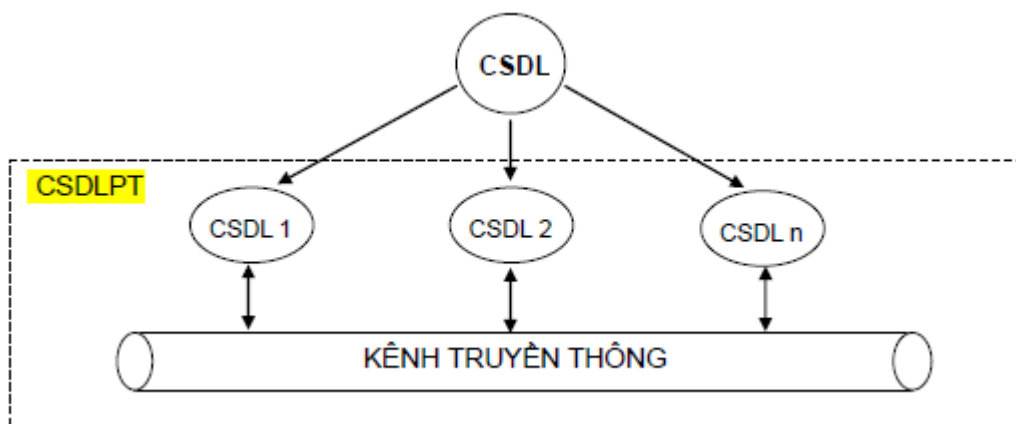
### 1.2.1 Tính thuần nhất và không thuần nhất của một hệ CSDL phân tán.

Để xây dựng CSDL phân tán, người ta phải phân bố dữ liệu trên các trạm khác nhau. Mỗi cách tiếp cận cho việc phân bố dữ liệu sẽ nhận được một CSDL phân tán với những tính chất khác nhau.

#### 1.2.1.1 CSDLPT thuần nhất (Homogeneous DDB)

Tính thuần nhất của một CSDLPT thể hiện ở các tính chất cơ bản:

- (i) Các CSDL cục bộ đều có cùng mô hình CSDL và có cùng cấu trúc dữ liệu.
- (ii) Các CSDL cục bộ được quản trị bởi cùng một hệ QTCSDL.



**Hình 1.5. CSDL PT thuần nhất nhận được từ việc chia CSDL thành các CSDL cục bộ**

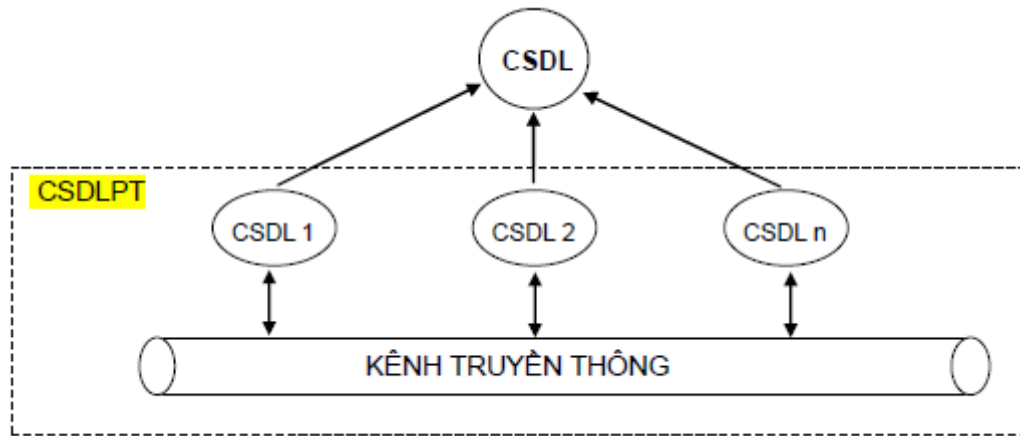
Nói chung điều kiện (ii) thường bao gồm cả điều kiện (i). Một CSDL phân tán thuần nhất thường có được từ việc chia một CSDL tập trung thành một tập các CSDL cục bộ. Mỗi cơ sở dữ liệu cục bộ này được quản trị bởi cùng một hệ quản trị cơ sở dữ liệu.

Như vậy, CSDLPT thuần nhất thường là kết quả của một cách thiết kế “từ trên xuống”, trong đó ta thiết kế một CSDLPT từ một CSDL tập trung, nó tạo thuận lợi cho sự tăng trưởng của một tổ chức (cơ quan, xí nghiệp...) khi cần thêm một trạm mới vào hệ thống CSDL. Đồng thời còn cải thiện hiệu năng của việc quản lý và khai thác dữ liệu nhờ khả năng xử lý song song được cung cấp bởi sự có mặt của nhiều trạm.

#### 1.2.1.2 CSDLPT không thuần nhất (Heterogeneous DDB)

Một CSDL PT không thuần nhất sẽ không có những tính chất cơ bản trên, và thường được xây dựng bằng việc tích hợp các CSDL cục bộ đã có thành một hệ thống CSDL duy nhất. Các CSDL cục bộ này có thể được thiết kế theo nhiều mô hình dữ liệu khác nhau, vì vậy có thể có cấu trúc dữ liệu khác nhau và được quản trị bởi các hệ quản trị dữ liệu khác nhau.

Như vậy, CSDL phân tán không thuần nhất thường là kết quả của các thiết kế “từ dưới lên”, cho phép tích hợp các CSDL hiện có và độc lập với nhau, và như vậy, không phải tạo ra một CSDL hoàn toàn mới. Như vậy, ngoài các chức năng chung, các hệ quản trị CSDL PT không thuần nhất phải cung cấp các giao diện để các hệ QTCSDL cục bộ có thể giao tiếp được với nhau, việc này thường được giải quyết bằng cách xây dựng một “mô hình trụ cột” chung cho tất cả các CSDL cục bộ, mọi thao tác của các CSDL cục bộ đều thông qua mô hình trụ cột này



**Hình. 1.6. CSDL phân tán không thuần nhất nhận được từ sự tích hợp các CSDL hiện có.**

Rõ ràng các tiếp cận không thuần nhất cho việc phân tán dữ liệu là phức tạp hơn nhiều so với cách tiếp cận thuần nhất.

Oracle hỗ trợ việc liên kết dữ liệu trong một CSDL phân tán không thuần nhất

## 1.2.2 Các mô hình phân tán dữ liệu

### 1.2.2.1 Cơ sở dữ liệu từ xa (Remote Database).

- **Khái niệm:** Cơ sở dữ liệu từ xa (*remote database*) là một CSDL đặt trên một máy tính khác với máy tính của người dùng và được truy cập nhờ vào các lệnh truyền thông được xác định bởi người dùng.
- **Các đặc điểm của remote DB:**
  - Máy tính của người dùng duy trì một CSDL riêng được nạp từ CSDL từ xa trên máy chủ.
  - Việc quản trị CSDL từ xa được thực hiện bởi một DBMS trên máy chủ.
  - Việc quản trị CSDL người dùng thực hiện trên máy trạm của người dùng, bởi phiên bản micro của DBMS trên máy chủ, và người dùng phải đảm bảo sự nhất quán dữ liệu riêng với dữ liệu máy chủ.
  - Nói chung việc truy cập CSDL từ xa thường được hạn chế trong việc tra cứu thông tin, các phép cập nhật được làm theo cách tập trung trên máy chủ.

Cách tiếp cận CSDL từ xa cho một chức năng hạn chế, nhưng không vấp phải những vấn đề khó khăn nhất của CSDL phân tán (quản lý từ điển dữ liệu, xử lý các giao dịch phân tán...).

**Thí dụ 1.1:** Phòng đào tạo có một CSDL đặt trên một máy chủ gồm danh sách SV mỗi khóa của mỗi môn học, điểm kiểm tra, điểm chuyên cần, điểm thi. Giáo viên và Sinh viên có thể truy cập và tải về để sử dụng. Đây là một mô hình CSDL từ xa.

### 1.2.2.2 Cơ sở dữ liệu liên hiệp.

Cơ sở dữ liệu liên hiệp (*Federated database*) là một mô hình phân tán dữ liệu bao gồm các CSDL cục bộ, nhưng các CSDL cục bộ này không tích hợp thành một CSDL duy nhất như với CSDL phân tán. mà chúng tạo thành các liên tác, trong phạm vi chúng có thể trao đổi dữ

liệu. Cách tiếp cận CSDL liên hiệp nhằm tránh gặp phải những khó khăn rất lớn của việc quản trị tập trung các CSDL phân tán, nhất là khi có một số lớn các CSDL cục bộ tham gia, đó là:

- Việc quản trị tập trung của CSDL phân tán là rất phức tạp nhất là khi số trạm là rất lớn.
- Mọi truy cập tới CSDL phân tán đều thông qua hệ quản trị CSDL.
- Khi truy cập tới một đơn vị dữ liệu dù trên trạm cục bộ nhưng vẫn thông qua hệ quản trị CSDL phân tán làm cho tốc độ xử lý dữ liệu cục bộ chậm lại bởi các chi phí của việc quản lý các dữ liệu phân tán.

Trong CSDL liên hiệp, người ta nói lỏng các ràng buộc chặt giữa các CSDL cục bộ và cung cấp cho người dùng một ngôn ngữ để định nghĩa các mối quan hệ giữa các CSDL khác nhau và thao tác trên nhiều CSDL cùng một lúc nhờ một ngôn ngữ đa cơ sở dữ liệu (multidatabase language) Thường là không có quá nhiều máy trạm trong một CSDL liên hiệp..

*Một số đặc điểm của CSDL liên hiệp:*

- Nhằm ưu tiên người dùng thao tác CSDL cục bộ và xem việc truy cập tới một CSDL khác là không thường xuyên.
- Sự phân tán là không trong suốt đối với người dùng.

*Sự khác nhau giữa hệ quản trị CSDL liên hiệp và hệ quản trị CSDL phân tán là:*

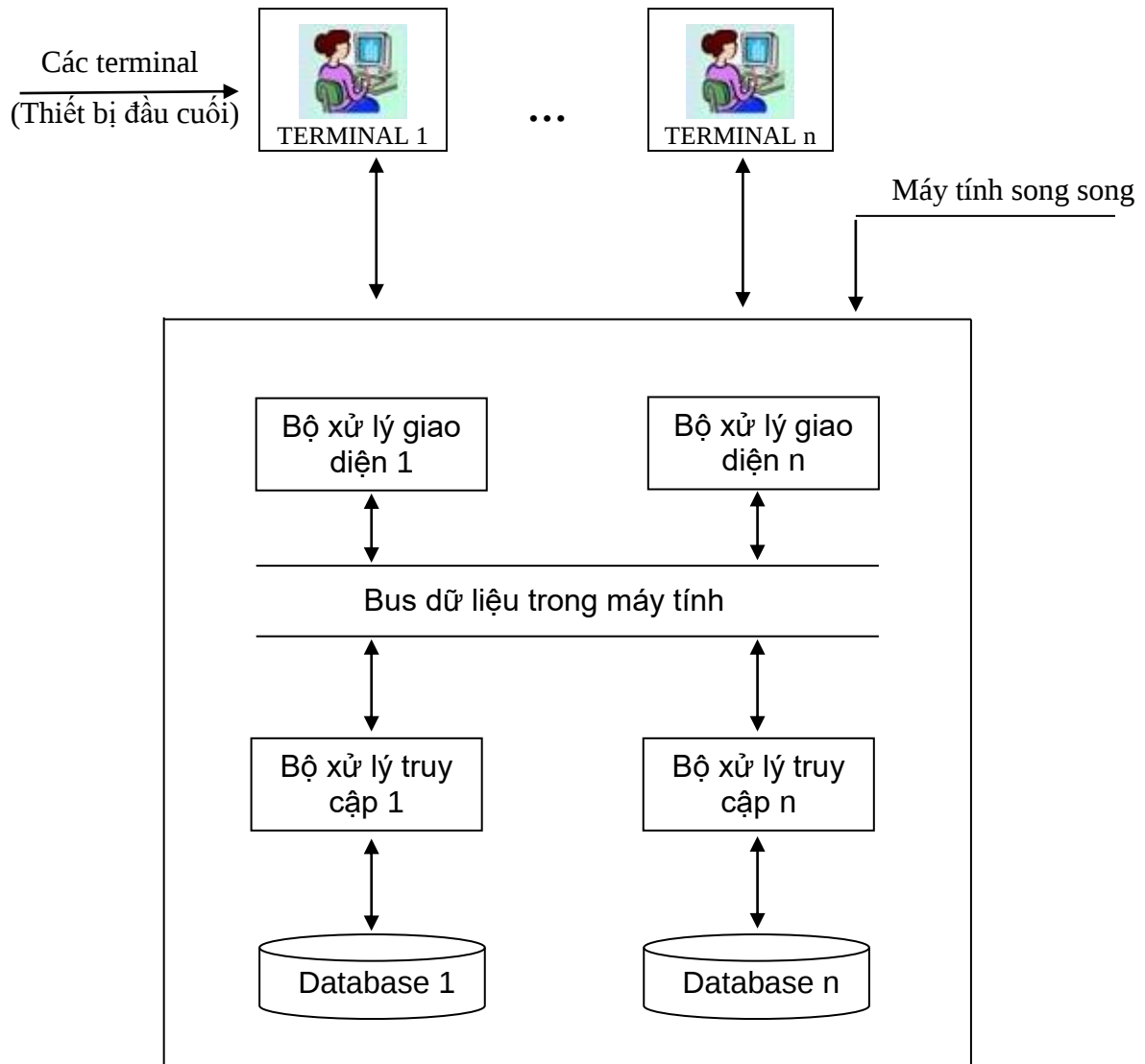
- Nó không quản lý từ điển dữ liệu tổng thể.
- Nó hỗ trợ một ngôn ngữ để định nghĩa các phụ thuộc có thể có giữa các CSDL trong liên hiệp. (multidatabase language)
- Nó hỗ trợ một ngôn ngữ để định nghĩa và thao tác trên các CSDL thuộc liên hiệp (multidatabase language)

### **1.2.2.3 Cơ sở dữ liệu song song.**

Một mô hình phân tán dữ liệu được khai thác nhiều trong thời gian gần đây là việc phân bố dữ liệu trên một tập các nút (nodes) của một máy tính song song (siêu máy tính với bộ đa xử lý) để tăng hiệu quả của việc quản lý dữ liệu. Mô hình này còn được gọi là cơ sở dữ liệu song song (parallel database)

Như vậy có thể mô tả CSDL song song như sau:

- Phần cứng: Là một máy tính song song (siêu máy tính với nhiều bộ vi xử lý), mỗi một node của máy tính song song sẽ xử lý một CSDL cục bộ. Các liên hệ giữa các node sẽ thông qua bus dữ liệu của máy tính.
- Phần mềm: Dùng chung một phần mềm QTCSDL duy nhất.
- CSDL song song được xem như một hộp đen xử lý dữ liệu và được gọi là “máy CSDL” (database machine).



**Hình 1.7. Cơ sở dữ liệu song song (Máy cơ sở dữ liệu)**

**So sánh CSDL phân tán và CSDL song song:**

- Giống nhau:
  - Cùng có nhiều CSDL cục bộ
  - Chỉ có chung một phần mềm quản lý CSDL.
  - Hoạt động như một CSDL phân tán thuần nhất.
- Khác nhau:
  - Chỉ có một máy đặt tại một trạm
  - Không dùng kênh truyền thông.

### 1.3 CÁC MỤC TIÊU CỦA MỘT HỆ CƠ SỞ DỮ LIỆU PHÂN TÁN

Các hệ QTCSDL PT phải làm cho người sử dụng trong hệ thống không biết và không cần biết tới sự phân tán của dữ liệu và việc sử dụng CSDLPT phải không khó hơn CSDL tập trung. Như vậy các hệ QTCSDLPT phải cung cấp các mức độc lập khác nhau đối với dữ liệu và đối với người dùng. Đó là các mục tiêu của một hệ QTCSDL PT.

#### 1.3.1 Tính độc lập đối với sự phân bố dữ liệu:

Nhờ tính chất này, người dùng sẽ không cần quan tâm tới sự phân tán của dữ liệu; và có thể không biết là có sự phân tán. Thông tin sự phân tán dữ liệu được lưu trữ trong *từ điển dữ liệu tổng thể* và được hệ quản trị CSDL phân tán tham khảo để xác định vị trí của các quan hệ có liên quan đến một thông tin nào đó. Lưu trữ từ điển dữ liệu tổng thể có 3 cách:

- Tập trung tại một trạm xem như máy chủ : từ điển DL được tổ chức tập trung
- Mỗi một trạm được cung cấp một từ điển dữ liệu tổng thể: từ điển DL được tổ chức có nhân bản
- Từ điển dữ liệu được phân đoạn và lưu trữ trên các trạm : Từ điển được tổ chức phân tán.

Tính độc lập với sự phân tán dữ liệu sẽ đảm bảo sự định vị dữ liệu sẽ độc lập với sự phân bố dữ liệu vật lý: Khi thay đổi dữ liệu về mặt vật lý thì chỉ có từ điển dữ liệu là thay đổi; không làm thay đổi phân mềm ứng dụng.

#### 1.3.2 Tính độc lập đối với sự phân đoạn của các quan hệ:

Tính độc lập đối với sự phân đoạn sẽ che giấu người dùng về việc các dữ liệu đã bị phân đoạn (sự phân đoạn là trong suốt đối với người dùng).

Trong CSDL phân tán, các quan hệ được phân đoạn và có thể lưu trữ ở các trạm khác nhau, việc phân đoạn là nhằm tăng tính hiệu quả của một CSDL phân tán và nó cho phép đơn giản hóa các truy cập địa phương

Khi người dùng truy cập tới các quan hệ tổng thể, hệ quản trị CSDL phân tán sẽ tham chiếu tới các đoạn một cách phù hợp, thông tin về sự phân đoạn được lưu trữ trong từ điển, vì vậy sự phân đoạn là độc lập với người dùng.

Tính độc lập đối với sự phân đoạn được đảm bảo bởi hệ quản trị CSDL phân tán.

#### 1.3.3 Tính độc lập đối với sự nhân bản:

Sự nhân bản là một giải pháp cho sự an toàn dữ liệu: Là sự dư thừa cho các bản sao. Một đoạn được nhân bản khi tồn tại hai hay nhiều hơn bản sao của nó. Mỗi bản sao được lưu trữ trên một hay nhiều trạm khác nhau. Ngoài ra sự nhân bản còn làm tăng tính sẵn sàng truy cập và tăng xử lý địa phương.

- Việc duy trì và đồng bộ các bản sao là khá phức tạp và tốn kém. Tuy nhiên do vấn đề an toàn dữ liệu thì mọi CSDL phân tán đều phải thực hiện nhân bản.

Tính độc lập đối với sự nhân bản là làm cho người dùng không biết là có sự nhân bản mà chỉ nhìn thấy các quan hệ không nhân bản. Như vậy sự nhân bản là trong suốt đối với người dùng.

### 1.3.4 Tính độc lập đối với các hệ quản trị CSDL địa phương (cục bộ) :

- Tính độc lập này cho phép che giấu một sự kiện là các hệ quản trị CSDL địa phương có thể khác nhau. Mục tiêu này rất khó thực hiện hoàn toàn, nhất là đối với hệ CSDL phân tán không thuần nhất, người ta đưa ra một giải pháp gần đây nhất, là dùng một mô hình trụ cột (thường chọn là mô hình quan hệ) và xác định một bộ chương trình dịch giữa các mô hình cục bộ và mô hình trụ cột.
- Người dùng ở các trạm không cần quan tâm đến các hệ quản trị CSDL ở các trạm khác, và anh ta cứ việc dùng chương trình quen thuộc của mình. Việc dịch các lệnh, câu hỏi từ các mô hình cục bộ về mô hình trụ cột do hệ quản trị CSDL phân tán đảm nhận.
- Tính độc lập này được đảm bảo bởi hệ quản trị CSDL phân tán.

### 1.3.5 Tính tự trị của các trạm:

- Đây là một mục tiêu cho phép các trạm điều khiển và thao tác dữ liệu địa phương của mình độc lập với các trạm khác.  
Ưu điểm: Việc quản trị CSDL phân tán hoàn toàn phi tập trung.
- Tính tự trị được đảm bảo bởi hệ quản trị CSDL phân tán.

### 1.3.6 Tính mở rộng:

- Là một khả năng tăng trưởng bằng việc đưa thêm vào các trạm mới với các tác động tối thiểu lên các CSDL địa phương và các chương trình ứng dụng hiện có.

### 1.3.7. Tính ổn định

- Sự cố trên một phần nhỏ của hệ thống rất ít hoặc hoàn toàn không ảnh hưởng đến hoạt động của toàn bộ hệ thống

### 1.3.8 Cải thiện hiệu năng:

- Đây là tính chất sống còn của cách tiếp cận CSDL phân tán.
- Vấn đề cốt yếu là phân đoạn hợp lý và nhân bản dữ liệu để có thể khai thác song song.

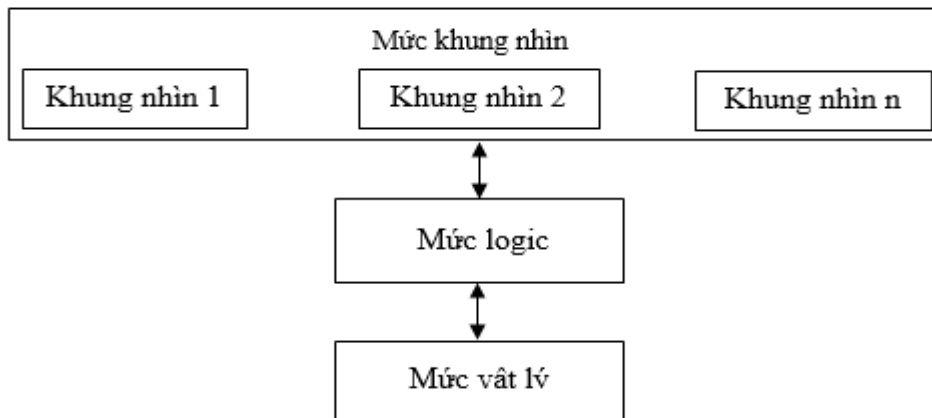
## 1.4 KIẾN TRÚC THAM CHIẾU CỦA MỘT HỆ CƠ SỞ DỮ LIỆU PHÂN TÁN

### 1.4.1 Kiến trúc lược đồ của CSDLPT

Kiến trúc lược đồ nhằm đề xuất việc tổ chức lý tưởng nhất một CSDLPT hỗ trợ các mức độc lập đối với dữ liệu và người dùng. Kiến trúc lược đồ của CSDL phân tán là sự mở rộng kiến trúc 3 mức cho CSDL tập trung,

#### 1.4.1.1 Nhắc lại kiến trúc 3 mức trong CSDL tập trung

Kiến trúc 3 mức theo tiêu chuẩn ANSI/SPARC (American National Standards Institute: Viện Tiêu chuẩn quốc gia Mỹ/Planning and Requirement Committee: Ủy ban nhu cầu và kế hoạch Mỹ) được mô tả như sau: (Hình 1.8)



**Hình 1.8. Ba mức của kiến trúc lược đồ trong các CSDL cục bộ (CSDL tập trung)**

1. *Mức khung nhìn*: gồm các lược đồ ngoài, mỗi lược đồ ngoài mô tả chỉ 1 phần của CSDL, thích hợp với 1 nhóm người sử dụng nhất định. Mỗi người sử dụng có thể không quan tâm đến toàn bộ CSDL mà chỉ cần một phần thông tin nào đó.
2. *Mức logic*: gồm các lược đồ khái niệm, các lược đồ này mô tả những dữ liệu nào được lưu trữ trong CSDL, và mối quan hệ giữa chúng. Chẳng hạn lược đồ khái niệm chỉ quan tâm đến các quan hệ được lưu trữ trong CSDL và các mối liên hệ giữa chúng chứ không quan tâm đến cách thức lưu trữ các quan hệ này.
3. *Mức vật lý*: gồm các lược đồ trong, các lược đồ này mô tả dữ liệu được lưu trữ như thế nào trong bộ nhớ thứ cấp. Mức này phản ánh cấu trúc dữ liệu, cách tổ chức tệp, các kỹ thuật nén dữ liệu, cấp phát vùng nhớ...

#### **1.4.1.2 Kiến trúc lược đồ của CSDL phân tán.**

Kiến trúc lược đồ của CSDL phân tán là sự mở rộng kiến trúc 3 mức cho CSDL tập trung. Với kiến trúc lược đồ này, tính độc lập dữ liệu dễ dàng được đảm bảo. Mọi thông tin liên quan đến các lược đồ và các phép biến đổi giữa các lược đồ (các ánh xạ) được lưu giữ trong từ điển dữ liệu.

Trước hết ta đưa ra các khái niệm sau, trong đó các mức của lược đồ mô tả CSDLPT mà độc lập với mọi CSDL cục bộ thì được gọi là mức tổng thể, các mức lược đồ mô tả một CSDL cục bộ được là mức cục bộ.

#### **Định nghĩa 1.2.**

*Lược đồ tổng thể (Global Scheme) là một mô tả tổng thể và thống nhất của tất cả các dữ liệu của CSDL phân tán và độc lập với môi trường phân tán.*

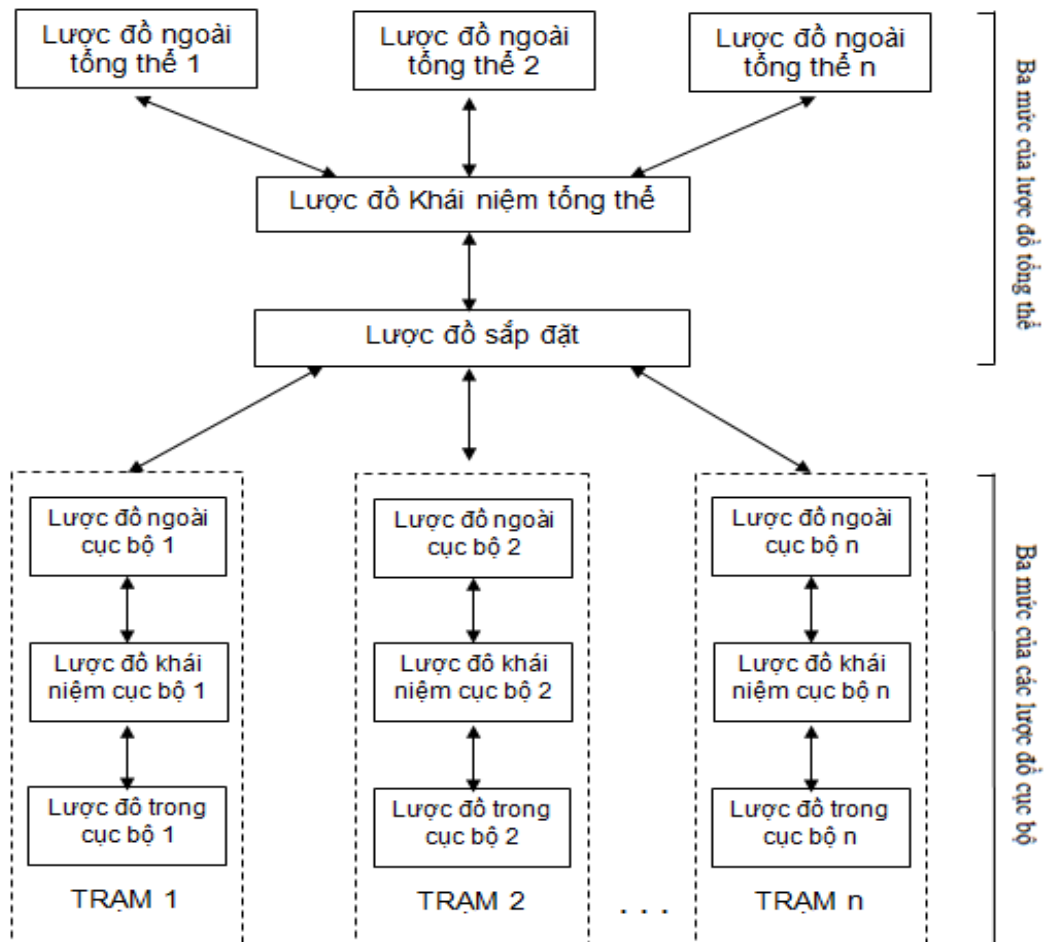
#### **Định nghĩa 1.3.**

*Câu hỏi phân tán (Distributed Query) là câu hỏi được diễn tả phù hợp với các dữ liệu được mô tả bởi lược đồ tổng thể của CSDL phân tán*

#### **Định nghĩa 1.4.**

*Câu hỏi cục bộ (Local Query) là câu hỏi diễn tả phù hợp với các dữ liệu được mô tả bởi lược đồ của một CSDL cục bộ*

Trong kiến trúc lược đồ của một CSDL phân tán (hình 1.9) có phân biệt 3 mức của lược đồ tổng thể và 3 mức của các lược đồ cục bộ :



**Hình 1.9.** Kiến trúc lược đồ của CSDL phân tán

**Ba mức của lược đồ tổng thể :**

- Lược đồ khái niệm tổng thể : định nghĩa tất cả các quan hệ có trong CSDL phân tán và cung cấp tính độc lập dữ liệu với môi trường phân tán.
- Các lược đồ ngoài tổng thể : mỗi lược đồ ngoài tổng thể xác định một khung nhìn của 1 lớp người dùng nhất định. Các quy tắc định nghĩa khung nhìn được dùng để biến đổi dữ liệu ở mức lược đồ ngoài tổng thể thành dữ liệu ở mức lược đồ khái niệm tổng thể. Nhờ các quy tắc này mà tính độc lập logic của dữ liệu được đảm bảo.
- Lược đồ sắp đặt xác định chính xác cách thức và vị trí các qua hệ được sắp đặt trên các trạm khác nhau trong mạng. Nó chứa tất cả các thông tin liên quan đến sự định vị, sự phân đoạn và nhân bản dữ liệu, cùng với các quy tắc biến đổi dữ liệu ở mức lược đồ khái niệm tổng thể thành dữ liệu cục bộ có phân đoạn và nhân bản.

**Ba mức của các lược đồ cục bộ :**

- Các lược đồ trong cục bộ : là mức vật lý của các lược đồ CSDL cục bộ, ý nghĩa và hoạt động giống như các lược đồ trong của CSDL tập trung (xem 1.4.1.1).
- Các lược đồ khái niệm cục bộ : là mức logic của các lược đồ CSDL cục bộ, ý nghĩa và hoạt động giống như các lược đồ khái niệm của CSDL tập trung (xem 1.4.1.1).
- Các lược đồ ngoài cục bộ : có vai trò biểu diễn các đoạn của quan hệ như đã mô tả trong lược đồ sắp đặt thành các đối tượng của lược đồ ngoài, được xem như các khung nhìn của

CSDL cục bộ. Các lược đồ ngoài cục bộ có nhiệm vụ kết nối lược đồ sắp đặt của các lược đồ tổng thể với các CSDL cục bộ. Các lược đồ ngoài cục bộ còn đảm bảo tính độc lập của các hệ quản trị CSDL và phải có nếu môi trường là không thuần nhất.

#### **1.4.1.3 Kiến trúc chức năng của CSDL phân tán.**

Các chức năng trong một hệ thống CSDL được chia thành 2 nhóm chính : *Các chức năng xử lý ứng dụng* (trình quản lý ứng dụng : Khách – Client) và *Các chức năng xử lý dữ liệu* (trình quản lý dữ liệu : Chủ - Server) có mặt tại mỗi trạm, như vậy mỗi trạm sẽ vừa có vai trò là trạm khách và vừa có vai trò là trạm chủ. Nói chung, một câu hỏi của người dùng sẽ việi dẫn tới phần mềm quản lý ứng dụng của trạm khách và phần mềm quản lý dữ liệu của một hay nhiều trạm chủ. Các chức năng xử lý ứng dụng và chức năng xử lý dữ liệu đều được tích hợp trong phần mềm quản trị CSDL phân tán.

- Các chức năng xử lý ứng dụng :

a/. Giao diện người dùng : kiểm tra các dữ liệu trong câu hỏi người dùng có phù hợp với lược đồ ngoài tổng thể, sau khi thực thi câu hỏi thì trả về kết quả cho người dùng.

b/. Kiểm tra ngữ nghĩa : Biến đổi câu hỏi được diễn đạt trên các khung nhìn trong lược đồ ngoài tổng thể thành câu hỏi diễn đạt trên các quan hệ trong lược đồ khái niệm tổng thể (còn gọi là câu hỏi chính tắc, được biểu diễn bởi các biểu thức đại số quan hệ).

c/. Định giá câu hỏi phân tán : biến đổi các câu hỏi chính tắc (là các biểu thức đại số quan hệ trên các quan hệ tổng thể trong các quan hệ ở lược đồ khái niệm tổng thể) thành một phương án thực thi phân tán trên các đoạn trong các CSDL cục bộ (nhờ các thông tin trong lược đồ sắp đặt). Chức năng này cũng cho phép tối ưu hóa các câu hỏi phân tán để làm cực tiểu một hàm chi phí (theo thời gian thực hiện hoặc dung lượng dữ liệu cần phải truyền tải giữa các trạm).

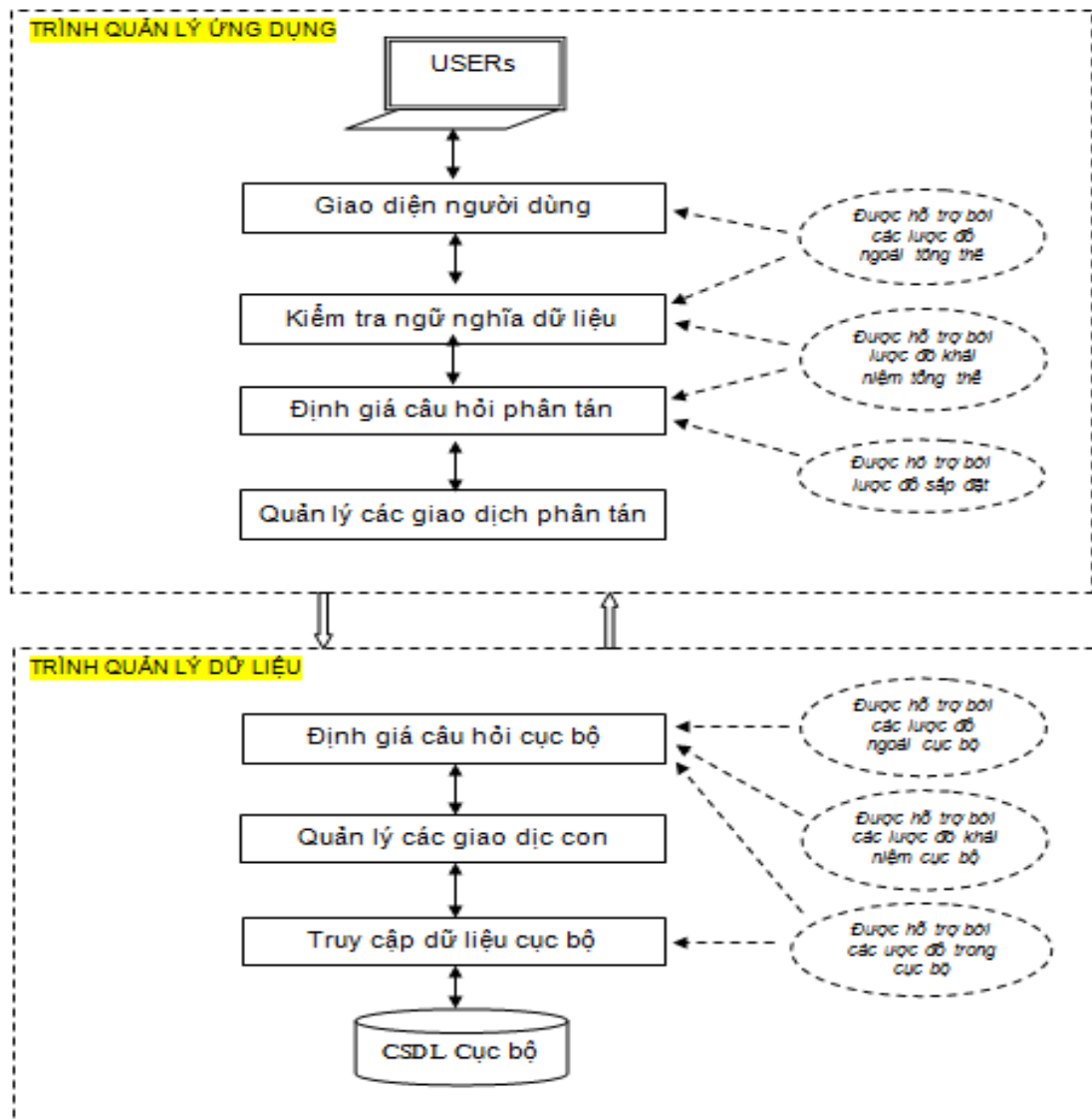
d/. Quản lý các giao dịch phân tán : chức năng này đảm bảo việc điều khiển tương tranh và tính nguyên tử của các giao dịch phân tán.

- Các chức năng xử lý dữ liệu : thực hiện trên các CSDL cục bộ, bao gồm :

a/. Định giá các câu hỏi cục bộ : các câu hỏi cục bộ này nhận được từ các câu hỏi được diễn đạt theo lược đồ sắp đặt thành các câu hỏi diễn đạt theo lược đồ ngoài cục bộ của CSDL cục bộ.

b/. Quản lý các giao dịch con trên mỗi CSDL cục bộ.

c/. Truy cập các CSDL cục bộ.



**Hình 1.10.** Kiến trúc chức năng của một hệ CSDL phân tán

## CÂU HỎI VÀ BÀI TẬP CHƯƠNG 1

1. Hãy trình bày những thành phần chính của một hệ Cơ sở dữ liệu phân tán (DDB). Mục đích và yêu cầu của một hệ CSDL phân tán?
2. Thế nào là một CSDL phân tán thuần nhất và không thuần nhất? Các đặc điểm chính của các mô hình phân tán dữ liệu: CSDL từ xa, CSDL song song, CSDL liên hiệp. So sánh mỗi mô hình này với CSDL phân tán.
3. Một CSDL phân tán phải đảm bảo các tính độc lập gì đối với dữ liệu và hệ QTCSDL?
4. Hãy phân biệt các khái niệm: Dữ liệu cục bộ, dữ liệu phân tán, câu hỏi cục bộ, câu hỏi phân tán.
5. Phân tích kiến trúc lược đồ và kiến trúc chức năng của một hệ CSDL phân tán. Đối với các lược đồ tổng thể, theo anh/chị thì lược đồ nào là quan trọng nhất đối với một hệ CSDL phân tán