



# ĐẠI HỌC KINH TẾ ĐÀ NẴNG

## KHOA THỐNG KÊ TIN HỌC

# Lý thuyết về Distributed DataBase (Cơ sở dữ liệu phân tán)

Môn : **CƠ SỞ DỮ LIỆU**

GVHD : Ths. Nguyễn Văn Chức

SVTH : Nhóm 13 (Lớp 36H10K14.1-A)

**Đăng Diệu**

Lê Thị Thúy

Lê Thị Mạnh

Huỳnh Thị Vy

Ngô Thị Phương Thảo

- 1 Giới thiệu về CSDL phân tán
- 2 Các mô hình của CSDL phân tán
- 3 Các phương pháp thiết kế
- 4 Các mô hình phân mảnh
- 5 Các mức trong suốt
- 6 Ưu và nhược điểm



# I. GIỚI THIỆU VỀ CSDL PHÂN TÁN

## 1. CSDL phân tán là gì?

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.

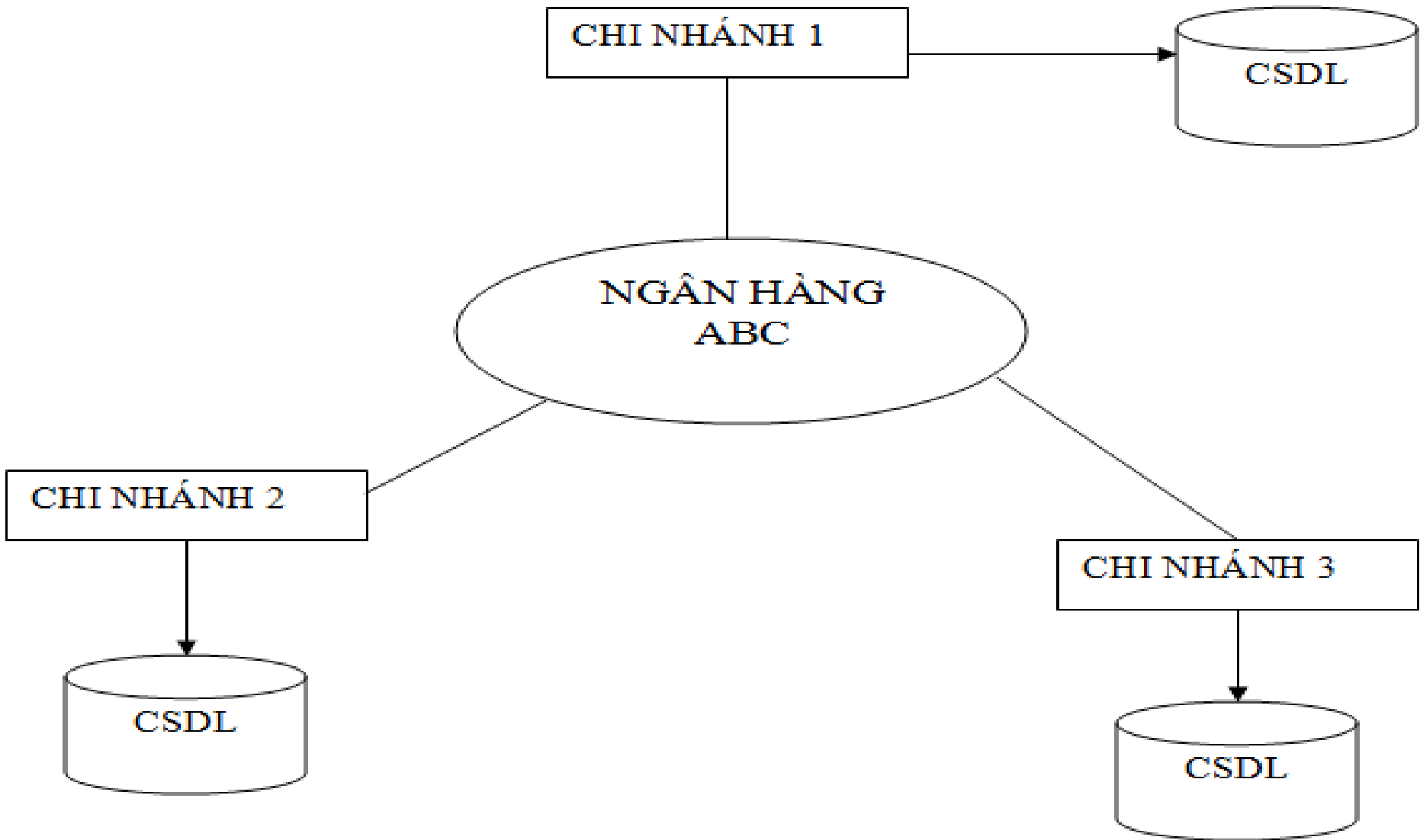
Trong hệ thống cơ sở dữ liệu phân tán gồm nhiều trạm, mỗi trạm có thể khai thác các giao tác truy nhập dữ liệu trên nhiều trạm khác.

# I. GIỚI THIỆU VỀ CSDL PHÂN TÁN

## 2. Ví dụ:

Một ngân hàng có ba chi nhánh khác nhau. Tại mỗi chi nhánh có một máy tính điều khiển cơ sở dữ liệu thống kê của chi nhánh đó. Các máy tính được nối với nhau bởi một mạng truyền thông. Các nút trong một mạng phân tán vừa xử lý thông tin tại vị trí mà nó quản lý và vừa tham gia vào việc xử lý các yêu cầu về thông tin cần truy cập qua nhiều địa điểm. 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.

# I. GIỚI THIỆU VỀ CSDL PHÂN TÁN



# I.GIỚI THIỆU VỀ CSDL PHÂN TÁN

## 3.Đặc điểm:

- Chia sẻ tài nguyên
- Tính mở
- Khả năng song song
- Khả năng mở rộng
- Khả năng thử lỗi
- Tính trong suốt
- Đảm bảo tin cậy và nhất quán

# II. CÁC MÔ HÌNH CỦA CSDL PHÂN TÁN

## 1. Mô hình thuần nhất:

Các điều kiện sau đây có thể tồn tại:

- Các hệ điều hành máy tính tại mỗi vị trí địa lý là như nhau hay ít nhất chúng có khả năng tương thích cao.

- Các mô hình dữ liệu và các hệ quản trị CSDL được sử dụng tại mỗi vị trí địa lý là như nhau hay ít nhất chúng có khả năng tương thích cao.

- Dữ liệu tại các vị trí khác nhau có cùng các định nghĩa và khuôn dạng.

- Các hệ cơ sở dữ liệu phân tán thuần nhất đơn giản hóa việc chia sẻ dữ liệu giữa những người sử dụng khác nhau.

# CÁC MÔ HÌNH CỦA CSDL PHÂN TÁN

## 2. Mô hình không thuần nhất:

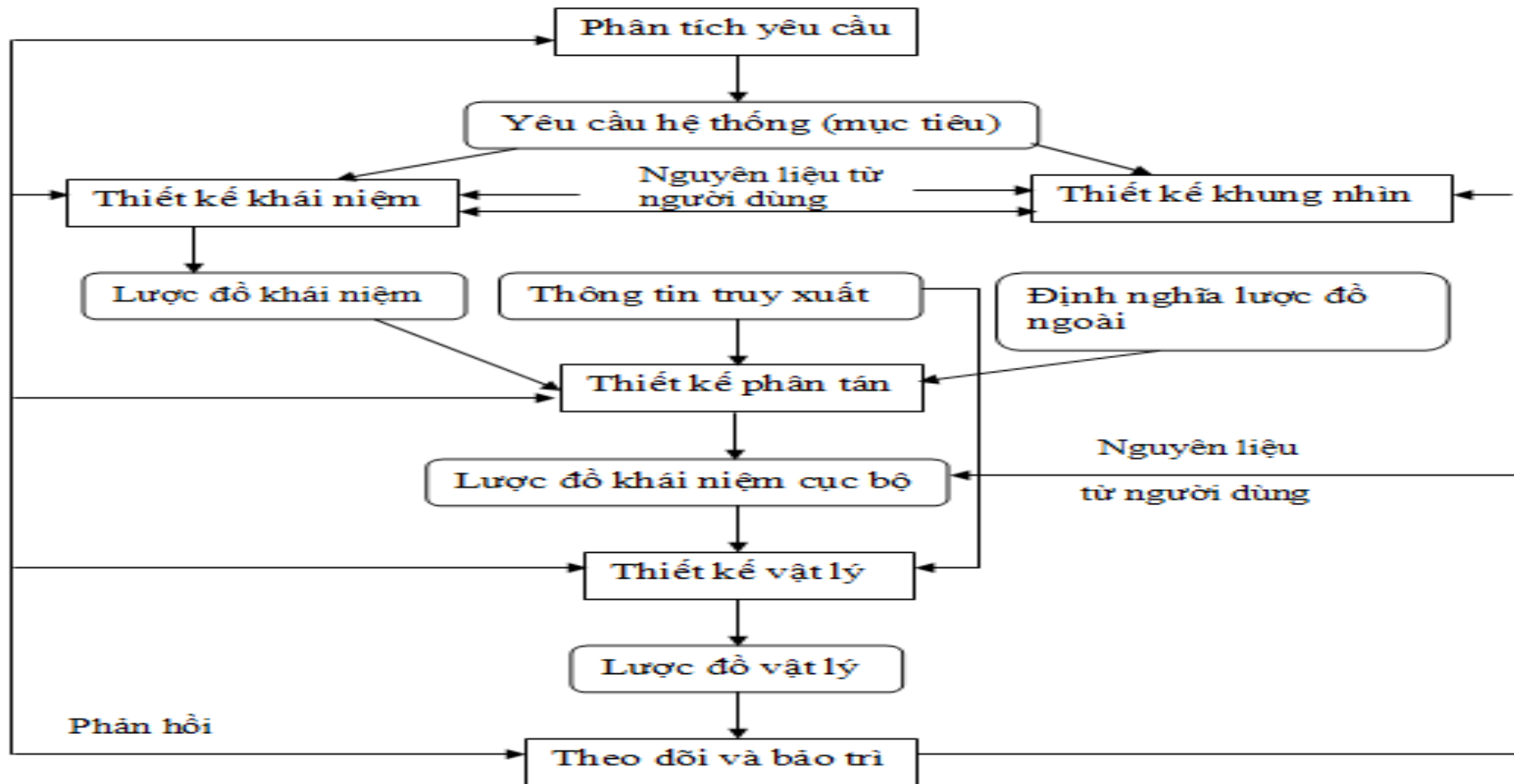
-Trong nhiều tổ chức, các hệ cơ sở dữ liệu có thể được lưu trữ trong các máy tính khác nhau và các hệ điều hành khác nhau tại mỗi vị trí địa lý. Các mô hình dữ liệu khác nhau và các hệ quản trị cơ sở dữ liệu khác nhau cũng có thể được lựa chọn sử dụng. Ví dụ, một vị trí có thể sử dụng công nghệ cơ sở dữ liệu quan hệ mới nhất, trong khi một vị trí khác có thể lưu trữ dữ liệu sử dụng các tệp truyền thống hay các cơ sở dữ liệu mạng, phân cấp cũ hơn.

-Phức tạp hơn nữa, dữ liệu trên các vị trí thường không tương thích. Ví dụ như các khác biệt về cú pháp (sự biểu diễn khác nhau của các khoản mục dữ liệu tại hai vị trí) và các khác biệt về ngữ nghĩa.



# III. CÁC PHƯƠNG PHÁP THIẾT KẾ CSDL PHÂN TÁN

## 1. Thiết kế từ trên xuống:



# III.CÁC PHƯƠNG PHÁP THIẾT KẾ CSDL PHÂN TÁN

## 2.Quá trình thiết kế từ dưới lên:

-Thiết kế từ trên xuống thích hợp với những CSDL được thiết kế từ đầu. Tuy nhiên chúng ta cũng hay gặp trong thực tế là đã có sẵn một số CSDL, nhiệm vụ là phải tích hợp thành một CSDL.Tiếp cận từ dưới lên sẽ thích hợp cho yêu cầu này. Khởi điểm của thiết kế từ dưới lên là các lược đồ khái niệm cục bộ . Quá trình này sẽ bao gồm việc tích hợp các lược đồ cục bộ thành khái niệm lược đồ toàn cục.

# IV.CÁC MÔ HÌNH PHÂN MẢNH TRONG CSDL PHÂN TÁN

## 1.Mô hình phân mảnh ngang:

-Phân mảnh ngang chia một quan hệ r theo các bộ, vì vậy mỗi mảnh là một tập con các bộ t của quan hệ r. Có hai chiến lược phân mảnh ngang cơ bản:

+Phân mảnh nguyên thủy (primary horizontal fragmentation) của một quan hệ được thực hiện dựa trên các thuộc tính được định nghĩa trên quan hệ đó.

+Phân mảnh ngang dẫn xuất (derived horizontal fragmentation ) là phân mảnh một quan hệ dựa vào các thuộc tính được định trên một quan hệ khác.

# IV. CÁC MÔ HÌNH PHÂN MẢNH TRONG CSDL PHÂN TÁN

Ví dụ: (Phân mảnh nguyên thủy)

Cho lược đồ quan hệ :

NhaCC(MaNCC, TenNCC, D/cNCC)

Chúng ta có thể phân ra thành 2 mảnh ngang như sau :

NhaCC1 =  $\sigma$  D/cNCC="HN" (NhaCC)

NhaCC2 =  $\sigma$  D/cNCC="ĐN" (NhaCC)

# IV. CÁC MÔ HÌNH PHÂN MẢNH TRONG CSDL PHÂN TÁN

Ví dụ: (Phân mảnh ngang dẫn xuất)

NhaCC( MaNCC, TenNCC , D/cNCC )

CungCap( MaNCC , MaSP, SoLuong )

-Ta muốn phân chia quan hệ này sao cho : một mảnh chứa các bộ của những nhà cung cấp ở một thành phố cho trước. Tuy nhiên, thành phố không phải là thuộc tính của CungCap mà của NhaCC :

CungCap1= CungCap >< (  $\Pi$  MaNCC (  $\sigma$  D/cNCC="HN" (NhaCC)))

CungCap2= CungCap >< (  $\Pi$  MaNCC (  $\sigma$  D/cNCC $\neq$ "HN" (NhaCC)))

# IV.CÁC MÔ HÌNH PHÂN MẢNH TRONG CSDL PHÂN TÁN

## 2.Phân mảnh dọc:

Một phân mảnh dọc của một quan hệ R sinh ra các mảnh R1, R2,..mỗi mảnh chứa một tập con thuộc tính của R và khoá của R.

Mục đích của phân mảnh dọc là phân hoạch một quan hệ thành một tập các quan hệ nhỏ hơn để nhiều ứng dụng chỉ cần chạy trên một mảnh.

Phân mảnh dọc phức tạp hơn so với phân mảnh ngang. Điều này là do tổng số chọn lựa có thể của một phân hoạch dọc rất lớn.

# IV. CÁC MÔ HÌNH PHÂN MẢNH TRONG CSDL PHÂN TÁN

Ví dụ:

Xét Quan hệ :

NhanVien(MaNV, TenNV, Luong, Thue, MaPB)

Một sự phân mảnh dọc của quan hệ này được định nghĩa như sau :

NhanVien1 =  $\Pi$  MaNV, TenNV, MaPB (NhanVien)

NhanVien2 =  $\Pi$  MaNV, Luong, Thue (NhanVien)

Phân mảnh này cho biết :

- NhanVien1 do đơn vị quản lý nhân sự lưu trữ
- NhanVien2 do đơn vị quản lý lương lưu trữ

# IV.CÁC MÔ HÌNH PHÂN MẢNH TRONG CSDL PHÂN TÁN

## 3.Phân mảnh hỗn hợp:

Đa số các trường hợp, phân mảnh ngang hoặc phân mảnh dọc đơn giản cho một lược đồ CSDL không đủ đáp ứng các yêu cầu từ ứng dụng. Trong trường hợp đó phân mảnh dọc có thể thực hiện sau một số mảnh ngang hoặc ngược lại, sinh ra một lối phân hoạch có cấu trúc cây. Bởi vì hai chiến lược này được áp dụng lần lượt, chọn lựa này được gọi là phân mảnh hỗn hợp.



# IV.CÁC MÔ HÌNH PHÂN MẢNH TRONG CSDL PHÂN TÁN

Ví dụ:

Xét quan hệ :

NhanVien(MaNV, TenNV, GioiTinh, Luong, Thue, MaPB)

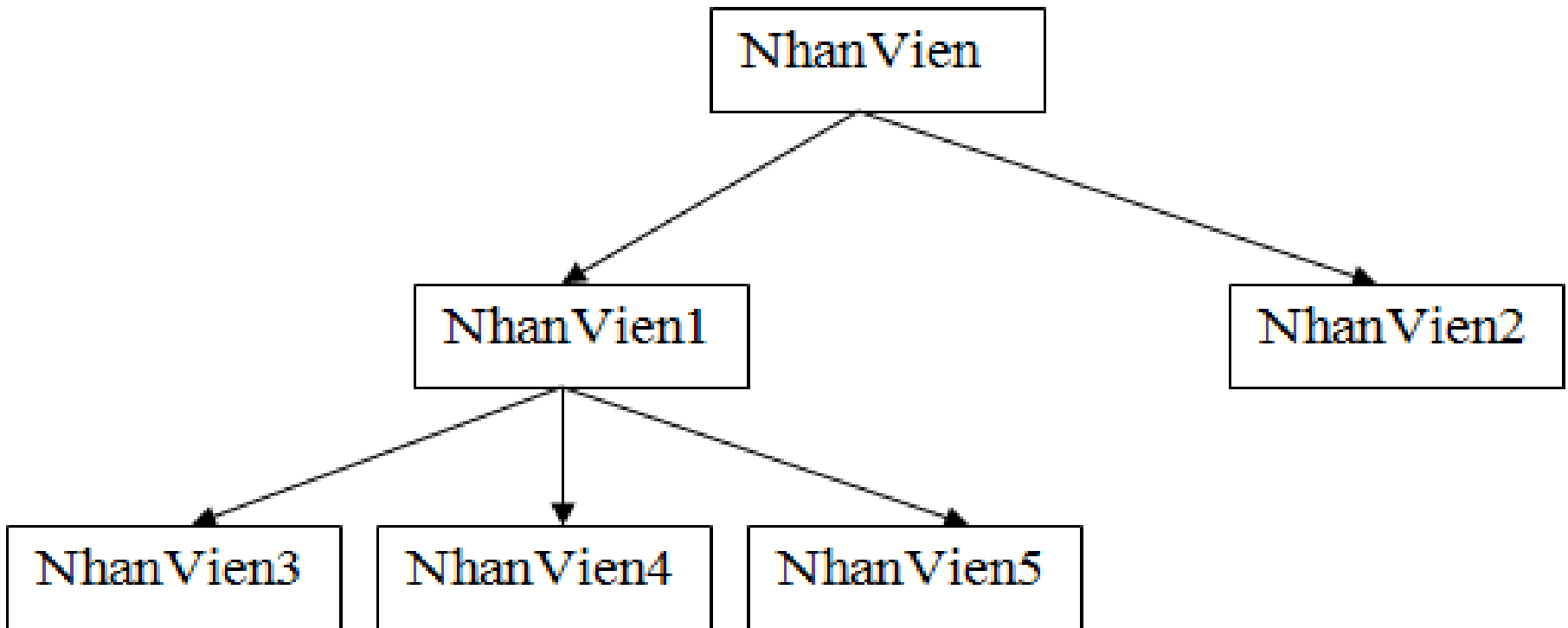
Yêu cầu phân tán :

Bộ phận quản lý quản lý các thông tin: tên, mã phòng ban, giới tính của nhân viên. Ngoài ra nhân viên còn được phân ra thành 3 vùng khác nhau : vùng 1 gồm các nhân viên thuộc phòng từ 1-10, vùng 2 gồm các nhân viên từ phòng 11-20, còn lại thuộc vùng 3.

Bộ phận kế toán quản lý các thông tin: lương, thuế của nhân viên.

# IV. CÁC MÔ HÌNH PHÂN MẢNH TRONG CSDL PHÂN TÁN

Dựa vào yêu cầu trên ta có mô hình phân tán như sau :



## V. CÁC MỨC TRONG SUỐT CỦA CSDL PHÂN TÁN

Tính trong suốt của một hệ phân tán được hiểu như là việc che khuất đi các thành phần riêng biệt của hệ đối với người sử dụng và những người lập trình ứng dụng.

Các loại trong suốt trong hệ phân tán:

- Trong suốt phân đoạn (fragmentation transparency)*
- Trong suốt về vị trí (location transparency)*\_
- Trong suốt ánh xạ địa phương (local mapping transparency)*\_

# V. CÁC MỨC TRONG SUỐT CỦA CSDL PHÂN TÁN

## **1. Trong suốt phân đoạn** (*fragmentation transparency*):

Khi dữ liệu đã được phân đoạn thì việc truy cập vào CSDL được thực hiện bình thường như là chưa bị phân tán và không ảnh hưởng tới người sử dụng.

*Ví dụ:* Xét quan hệ tổng thể NhaCC (MaNCC, TenNCC, D/c) và các phân đoạn được tách ra từ nó:

NhaCC1 (MaNCC, TenNCC, D/c)

NhaCC2 (MaNCC, TenNCC, D/c)

NhaCC3 (MaNCC, TenNCC, D/c)

Giả sử DDBMS cung cấp tính trong suốt về phân đoạn, khi đó ta có thể thấy tính trong suốt này được thể hiện như sau:

Khi muốn tìm một nhà cung cấp có MaNCC="1" thì chỉ cần tìm trên quan hệ tổng thể NCC mà không cần biết quan hệ NCC có phân tán hay không.

# V. CÁC MỨC TRONG SUỐT CỦA CSDL PHÂN TÁN

## ***2. Tính trong suốt về vị trí (location transparency):***

Người sử dụng không cần biết về vị trí vật lý của dữ liệu mà có quyền truy cập đến cơ sở dữ liệu tại bất cứ vị trí nào.

Các thao tác để lấy hoặc cập nhật một dữ liệu từ xa được tự động thực hiện bởi hệ thống tại điểm đưa ra yêu cầu.

Tính trong suốt về vị trí rất hữu ích, nó cho phép người sử dụng bỏ qua các bản sao dữ liệu đã tồn tại ở mỗi vị trí. Do đó có thể di chuyển một bản sao dữ liệu từ một vị trí này đến một vị trí khác và cho phép tạo các bản sao mới mà không ảnh hưởng đến các ứng dụng.

# V. CÁC MỨC TRONG SUỐT CỦA CSDL PHÂN TÁN

*Ví dụ:* Với quan hệ tổng thể R và các phân đoạn như đã nói ở trên

nhưng giả sử rằng DDBMS cung cấp trọng suất về vị trí nhưng không cung cấp trọng suất về phân đoạn.

Xét câu truy vấn tìm nhà cung cấp có MaNCC="1"

```
SELECT * FROM NCC1
```

```
WHERE MaNCC="1"
```

```
IF NOT FOUND THEN
```

```
SELECT * FROM NCC2
```

```
WHERE MaNCC="1"
```

```
ELSE
```

```
SELECT * FROM NCC3
```

```
WHERE MaNCC="1"
```

## ***3. Trong suốt ánh xạ địa phương (local mapping transparency):***

Là một đặc tính quan trọng trong một hệ thống DBMS không đồng nhất.

Ứng dụng tham chiếu đến các đối tượng có các tên độc lập từ các hệ thống cục bộ địa phương.

Ứng dụng được cài đặt trên một hệ thống không đồng nhất nhưng được sử dụng như một hệ thống đồng nhất.

# VI. ƯU NHƯỢC ĐIỂM CỦA CSDL PHÂN TÁN

## 1. Ưu điểm:

Sau đây là một vài lý do làm cho CSDL phân tán là cần thiết và cần được phát triển:

- 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.



# VI. ƯU NHƯỢC ĐIỂM CỦA CSDL PHÂN TÁN

## 2. Nhược điểm:

Bên cạnh những ưu điểm đã nêu ra thì cũng có một số điểm yếu cần quan tâm như:

- 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.



**THE END**

**THANK FOR YOUR LISTENING!!!**