

Cơ sở dữ liệu nâng cao

Giới thiệu

Đỗ Thanh Nghị
dtnghi@cit.ctu.edu.vn

Cần Thơ
11-10-2016

Hệ quản trị CSDL

- **Phần mềm**

- quản lý lưu trữ, truy cập CSDL

- **Tập hợp các công cụ**

- mô tả dữ liệu
 - thao tác dữ liệu
 - chia sẻ dữ liệu

- **Hệ quản trị CDSL dựa trên mô hình dữ liệu mà nó hỗ trợ**

Mô hình dữ liệu

■ Định nghĩa

- tập hợp các luật, khái niệm cho phép mô tả thế giới thực

■ Ví dụ

- mô hình mạng
- mô hình quan hệ
- mô hình ngữ nghĩa
- mô hình đối tượng

Chức năng của HQTCSDL

- **Quản lý ổn định, cạnh tranh từ bộ lưu trữ thứ cấp**
 - lưu trữ hiệu quả dữ liệu trên bộ nhớ thứ cấp
 - đảm bảo sự ổn định về dữ liệu (chịu đựng được sự hỏng hóc)
 - cho phép chia sẻ dữ liệu giữa nhiều người dùng

Chức năng của HQTCSDL

- **Cung cấp các ngôn ngữ định nghĩa dữ liệu và thao tác dữ liệu cho những người dùng khác nhau**
 - ngôn ngữ định nghĩa dữ liệu cho nhà quản trị
 - ngôn ngữ thao tác dữ liệu cho người dùng đầu cuối
 - ngôn ngữ và công cụ cho người phát triển ứng dụng

Chức năng của HQTCSDL

- **Xử lý hiệu quả câu truy vấn**
 - cơ chế tối ưu câu truy vấn
 - quản lý sự toàn vẹn ngữ nghĩa cho dữ liệu
 - sử dụng các phương pháp truy cập

A thick red line forming an L-shape, starting with a vertical segment on the left and a horizontal segment extending to the right.

Lịch sử

Ba thể hệ

■ Thể hệ I: HQTCSDL phân cấp, mạng

- mô hình dữ liệu dựa trên sự tồn tại con trỏ vật lý giữa các dữ liệu
- mô hình phân cấp giới hạn số liên kết phân cấp 1-n giữa dữ liệu
- mô hình mạng cho phép mô tả khá tốt thể giới thực

HQTCSDL phân cấp, mạng

■ Ngôn ngữ

- ngôn ngữ thao tác dữ liệu cho phép duyệt qua dữ liệu theo phân cấp hay mạng
- ngôn ngữ lập trình (Cobol, etc.)

■ Sản phẩm: IMS, DBMS 32...

- thương mại từ năm 60-70
- vẫn còn tồn tại đến ngày nay

HQTCSDL quan hệ

■ Mô hình quan hệ

- dựa trên sự định nghĩa về miền và quan hệ giữa các miền
- tất cả dữ liệu được mô tả dạng bảng
- đại số quan hệ cho phép thao tác nội dung bảng

HQTCSDL quan hệ

■ Ngôn ngữ thao tác

- người dùng đặc tả cái gì muốn nhận, cái gì không
- ngôn ngữ định nghĩa và thao tác dữ liệu toàn vẹn và chuẩn hóa: SQL
- ngôn ngữ tương tác cho định nghĩa và thao tác

■ Sản phẩm: Oracle, Sybase, IBM/UDB...

- thương mại từ năm 80
- rất phổ biến trong thực tế

HQTCSDL nâng cao

■ mô hình dữ liệu

- giàu ngữ nghĩa hơn (mô hình đối tượng)
- phân biệt giữa khái niệm thực thể và sự liên kết
- phân biệt sự khác nhau giữa các liên kết

■ ngôn ngữ

- ngôn ngữ thao tác kết hợp với dạng thức của truy vấn
- tích hợp ngôn ngữ lập trình với ngôn ngữ CSDL
- tích hợp và tương hợp với ngôn ngữ lập trình hướng đối tượng

HQTCSDL nâng cao

■ Cách tiếp cận

- mở rộng mô hình quan hệ
- giới thiệu HQTCSDL hướng đối tượng
- phát triển các hệ thống chuyên dụng

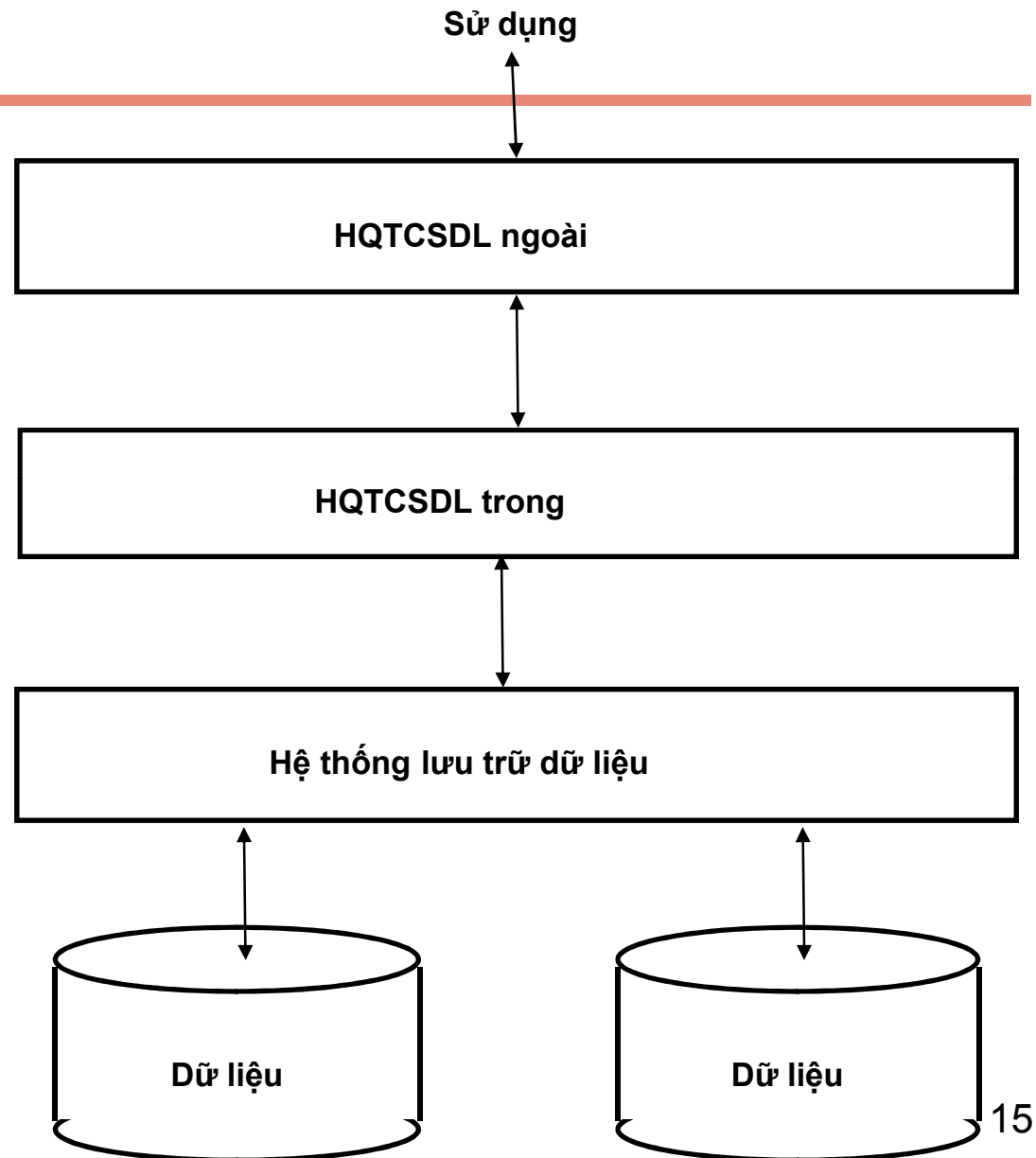
■ Sản phẩm

- thương mại vào cuối năm 80, sử dụng cho ứng dụng mới



Kiến trúc

Kiểu kiến trúc



Thành phần

■ Hệ thống lưu trữ dữ liệu

- đảm nhận chức năng quản lý dữ liệu trên bộ nhớ thứ cấp

■ HQTCSDL trong

- đảm nhận xếp đặt dữ liệu trong tập tin, tập hợp dữ liệu, liên kết giữa dữ liệu và cấu trúc

■ HQTCSDL ngoài

- đảm nhận sự phân tích, thông dịch câu truy vấn, định dạng hiển thị kết quả

A decorative L-shaped line in a reddish-brown color, consisting of a vertical segment on the left and a horizontal segment extending to the right, framing the text.

Ứng dụng mới

Ứng dụng mới

■ Ví dụ

- Mạng lưới kinh doanh bất động sản Québec

■ Mỗi chi nhánh

- mô tả các bất động sản: biệt thự, nhà và căn hộ chung cư rao bán
- có thể thay đổi mô tả các bất động sản
- thực hiện tìm kiếm tập bất động sản của chi nhánh
- tìm kiếm bất động sản: nhà, căn hộ chung cư trong danh sách quản lý của các chi nhánh khác

Ứng dụng mới

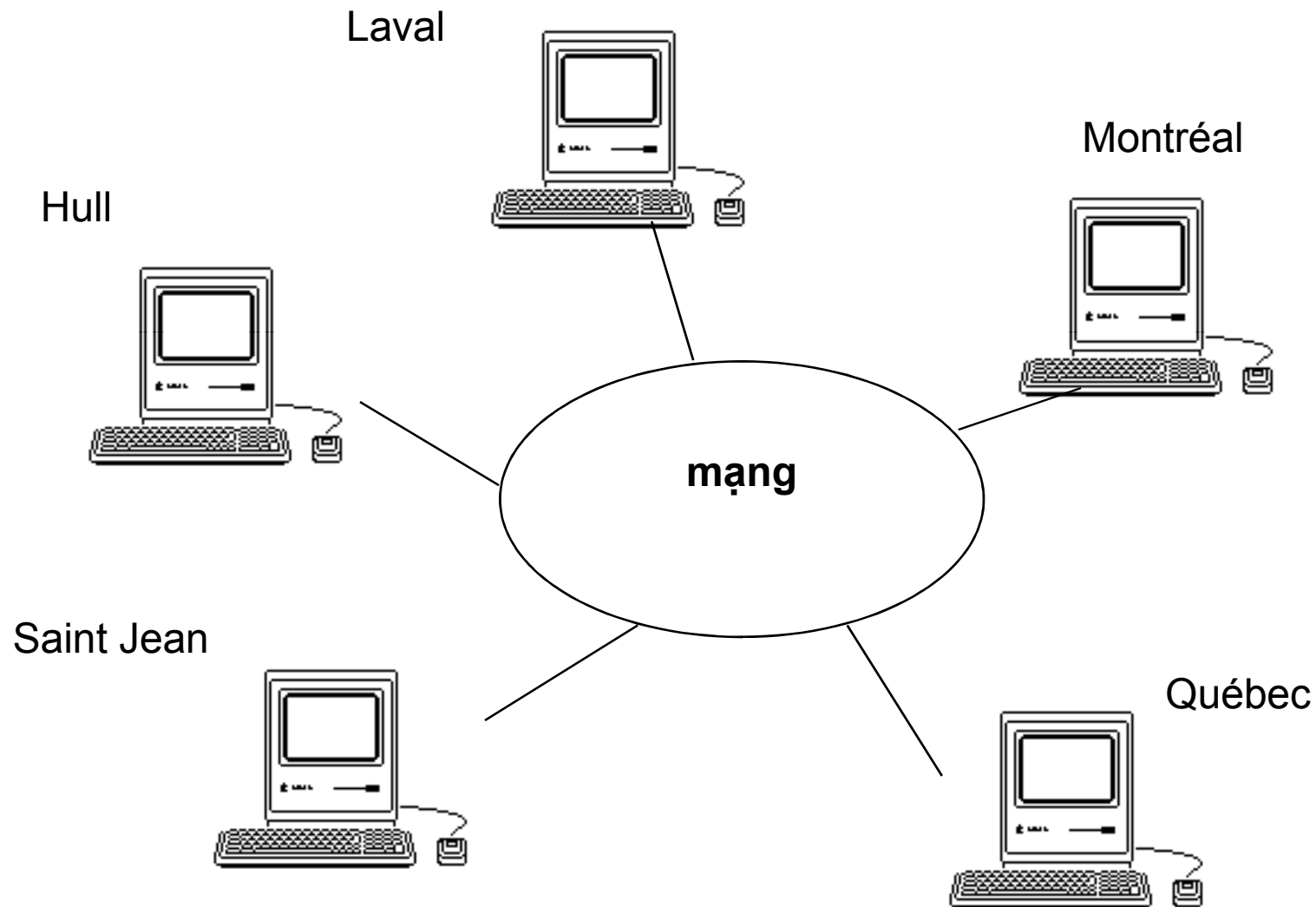
■ Tập hợp các chi nhánh

- có mô tả thuần nhất về bất động sản
- hợp tác, trao đổi thông tin
- truy xuất các mô tả của chi nhánh khác

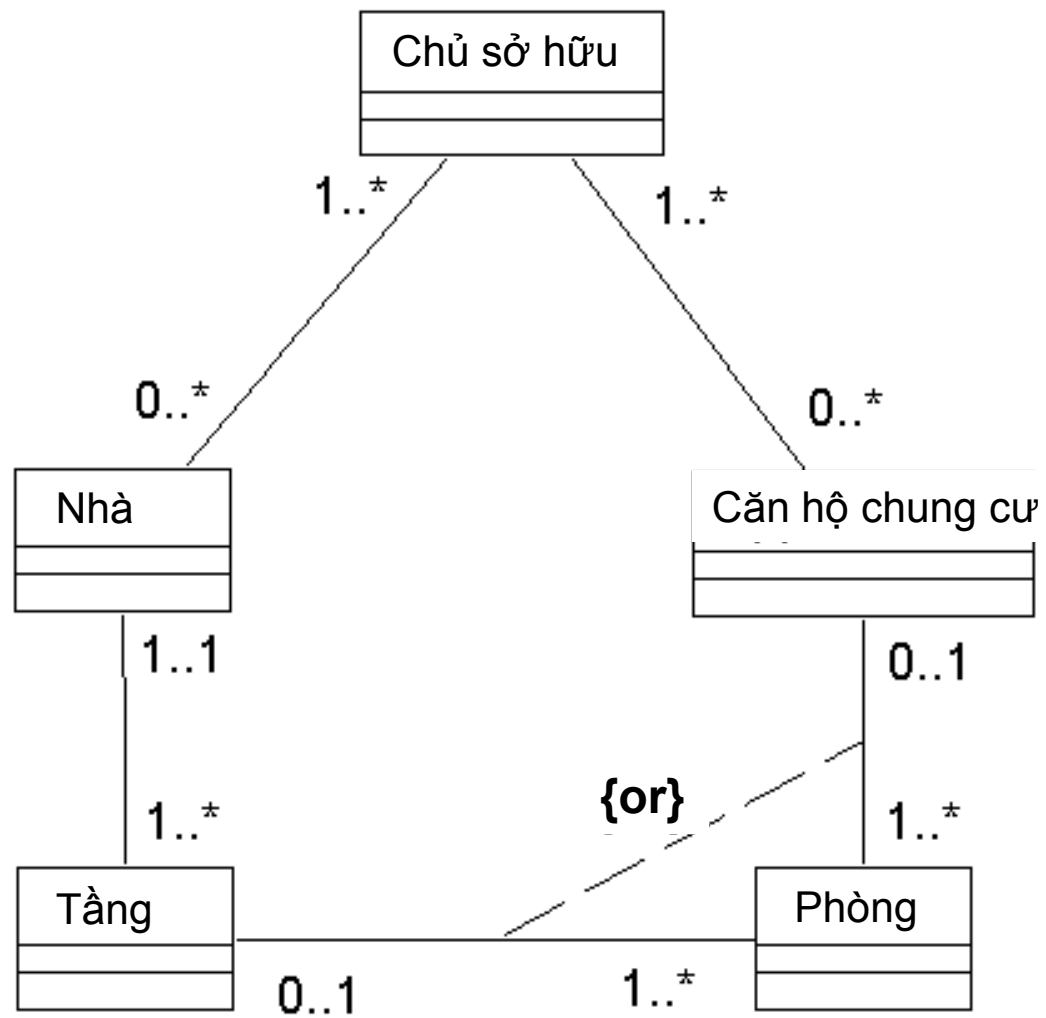
■ Đặc điểm của hệ thống thông tin

- Phân tán
- Dữ liệu phức tạp
- Dữ liệu đa phương tiện

Kiến trúc hệ thống



Mô hình dữ liệu



Ví dụ về dữ liệu

Mô tả nhà rao bán

địa chỉ

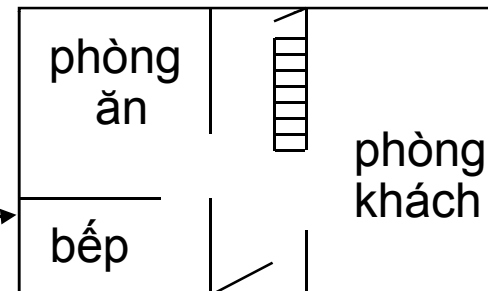
số tầng

tầng 1

mô tả các phòng

sơ đồ

hình



Lược đồ quan hệ

Căn hộ	NoAp	Địa chỉ	Tầng
	1	123 Green Montréal	1
	2	123 Green Montréal	2
	3	45 Maple Saint Lambert	1

Phòng	NoAp	NoR	Tên	Diện tích
	1	1	Khách	30
	1	2	Phòng ăn	20
	1	3	Bếp	10
	1	4	Phòng ngủ	20
	1	5	Phòng tắm	6
	2	2	Phòng ngủ	15
	2	1	Phòng ngủ	40
	2	3	Phòng tắm	4



Đặc thù riêng của ứng dụng mới

Đặc thù

■ Kiến trúc hệ thống

- phân tán
- hợp tác
- trong suốt

■ Kiểu dữ liệu

- dữ liệu truyền thống
- ảnh, bản đồ, âm thanh, video, ...

Mô hình hóa

- **Kiểu dữ liệu mới**
 - dữ liệu cấu trúc
 - đối tượng lớn
- **Phân biệt những liên kết khác nhau**
 - chuyên biệt, tổng quát, ...
- **Định nghĩa thủ tục liên kết**
 - ví dụ: tính diện tích

Xử lý và thao tác

■ **Độ quy**

- tìm kiếm bất động sản lân cận với bất động sản địa chỉ 540 Green à St Lambert

■ **Suy diễn thông tin mới**

- diện tích tổng nhà được suy ra từ diện tích các phòng

■ **Phân tích, trích thông tin**

- những nhà có giá lớn hơn 200 000\$ được mua bởi các cặp vợ chồng có nghề nghiệp

Giới hạn mô hình quan hệ

Giới hạn mô hình quan hệ

■ Mô hình hóa

- khó tích hợp kiểu dữ liệu mới
- mô hình đơn giản: phân cấp của các liên kết chặt có thể bị bể gãy
- không phân biệt giữa các khái niệm khác nhau trong khi mô hình hóa: chỉ có khái niệm quan hệ
- không mô hình hóa xử lý thông tin

Giới hạn mô hình quan hệ

■ Thao tác

- số lượng hữu hạn các phép toán thao tác: phép toán đại số quan hệ
- không hỗ trợ thao tác xử lý đệ quy
- diễn giải tường minh của các phép kết nối
- thao tác kiểu tập hợp
- không có khả năng duyệt qua các liên kết chặt

Hướng nghiên cứu

■ CSDL hướng đối tượng

- tích hợp khái niệm đối tượng vào HQTCSDL (lớp, kế thừa)

■ CSDL suy diễn

- tăng khả năng suy diễn cho HQTCSDL

■ CSDL multimedia

- tích hợp vào HQTCSDL khả năng quản lý văn bản, ảnh, âm thanh, video...

Hướng nghiên cứu

■ CSDL phân tán

- cho phép truy vấn và cập nhật dữ liệu từ các sites phân tán (địa lý)

■ Hệ hỗ trợ ra quyết định

- sinh ra các thông tin mới, tiềm ẩn, có ích từ kho dữ liệu lớn hỗ trợ ra quyết định

HQTCSDL hướng đối tượng

- **Hệ thống tích hợp chức năng của HQTCSDL và các đặc trưng của ngôn ngữ hướng đối tượng**
- **Mô hình hóa**
 - tích hợp các khái niệm của mô hình dữ liệu đối tượng
 - đối tượng phức tạp
 - khái niệm thuần nhất về dữ liệu và chương trình
 - khái niệm lớp, kế thừa

HQTCSDL hướng đối tượng

■ Khía cạnh hệ thống

- lưu trữ hiệu quả, hợp lý cho các kiểu dữ liệu khác nhau
- tìm kiếm hiệu quả
- giải thuật điều khiển truy cập cạnh tranh

HQTCSDL suy diễn

- **Tích hợp chức năng HQTCSDL và hệ chuyên gia**
 - lưu trữ
 - tìm kiếm
 - suy diễn
- **Cung cấp hệ thống hoàn chỉnh**
 - điều khiển dữ liệu
 - phát triển ứng dụng suy luận

HQTCSDL suy diễn

- **Suy diễn thông tin mới từ**
 - dữ liệu của CSDL
 - luật mô hình hóa tri thức
- **Cung cấp ngôn ngữ định nghĩa tri thức**
 - ngôn ngữ luật
- **Đảm bảo thực thi hiệu quả quá trình suy diễn**
 - lưu trữ các luật
 - tối ưu hóa chương trình luật
 - điều khiển thực thi

HQTCSDL Multimédia

■ Đáp ứng yêu cầu ứng dụng

- văn phòng
- thiết kế với sự trợ giúp máy tính
- CSDL văn bản
- CSDL âm thanh
- CSDL ảnh và video
- hệ thống thông tin địa lý

HQTCSDL Multimédia

■ Mục tiêu

- tích hợp nhiều kiểu dữ liệu
- quản lý đối tượng multimedia
- cung cấp khả năng mô hình hóa dữ liệu
- cung cấp ngôn ngữ thao tác dữ liệu thích hợp

HQTCSDL Multimédia

- **Giải pháp đề xuất ở cấp độ**
 - mô tả đối tượng multimedia
 - lưu trữ nhiều kiểu dữ liệu khác nhau
 - thao tác đối tượng multimedia
 - quản lý việc xử lý
- **Hệ thống đặc thù theo ứng dụng**
 - CSDL văn bản
 - CSDL âm thanh
 - CSDL ảnh và video
 - hệ thống thông tin địa lý

HQTCSDL phân tán

■ CSDL phân tán

- tập hợp dữ liệu có liên quan nhau, được lưu trữ trên các máy tính khác nhau trên hệ thống mạng

■ Mỗi site mạng

- có thể thực thi giao dịch cục bộ
- tham gia thực thi ít nhất 1 giao dịch toàn cục

■ HQTCSDL phân tán

- phần mềm đảm nhận việc quản lý CSDL phân tán

HQTCSDL phân tán

■ HQTCSDL phân tán

- CSDL (luận lý) duy nhất được phân tán (vật lý) trên nhiều sites
- chỉ có hệ thống quản lý CSDL luận lý

■ HQTCSDL không thuần nhất

- tập hợp các HQTCSDL cộng tác nhưng tự chủ độc lập
- khái niệm cơ bản cục bộ, toàn cục

■ HQTCSDL nhiều CSDL

- liên kết yếu từ các CSDL khác nhau mà nó không tạo thành thực thể luận lý

Hệ thống hỗ trợ ra quyết định

■ Kho dữ liệu (data warehouse)

- tập hợp dữ liệu trong lĩnh vực có mối liên quan và được lưu trữ nhiều máy tính khác nhau

■ Phân tích trực tuyến (OLAP)

- OLAP: On Line Analytic Processing
- phân tích dữ liệu và các giao dịch trên dữ liệu

■ Khai mỏ dữ liệu (Data Mining)

- khám phá tri thức hữu dụng, mới, tiềm ẩn từ kho dữ liệu lớn



Cám ơn !