

CƠ SỞ DỮ LIỆU PHÂN TÁN

NGUYỄN THỊ THANH THỦY

1

Nội dung môn học

- Giới thiệu
- Thiết kế CSDL phân tán
- Điều khiển dữ liệu phân tán
- Xử lý truy vấn phân tán
- Xử lý giao dịch phân tán

- Nhân bản dữ liệu
- Tích hợp CSDL – Các hệ thống đa CSDL
- Các hệ thống CSDL song song
- Quản lý dữ liệu ngang hàng (Peer-to-Peer)
- Xử lý dữ liệu lớn
- NoSQL, NewSQL và Polystores
- Quản lý dữ liệu Web

2

Nội dung

- Giới thiệu
 - Thế nào là một hệ quản trị CSDL phân tán?
 - Lịch sử
 - Triển vọng của hệ quản trị CSDL phân tán
 - Các vấn đề về thiết kế
 - Kiến trúc của hệ quản trị CSDL phân tán

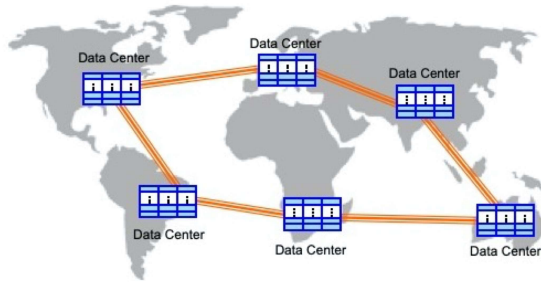
3

Tính toán phân tán

- Một số phần tử xử lý tự động (không nhất thiết phải đồng nhất) được kết nối với nhau bởi mạng máy tính và kết hợp để thực hiện một số nhiệm vụ được giao.
- Những gì được phân tán?
 - Logic xử lý
 - Chức năng
 - Dữ liệu
 - Điều khiển

4

Phân tán hiện tại – Các trung tâm dữ liệu phân tán theo địa lý



5

Hệ cơ sở dữ liệu phân tán là gì?

Cơ sở dữ liệu phân tán là tập hợp nhiều cơ sở dữ liệu **có liên quan với nhau một cách logic** được phân tán trên **mạng máy tính**.

Hệ quản trị cơ sở dữ liệu phân tán (Distributed DBMS) là phần mềm quản lý cơ sở dữ liệu phân tán (DDB) và cung cấp cơ chế truy nhập, giúp việc phân tán là **trong suốt** với người dùng.

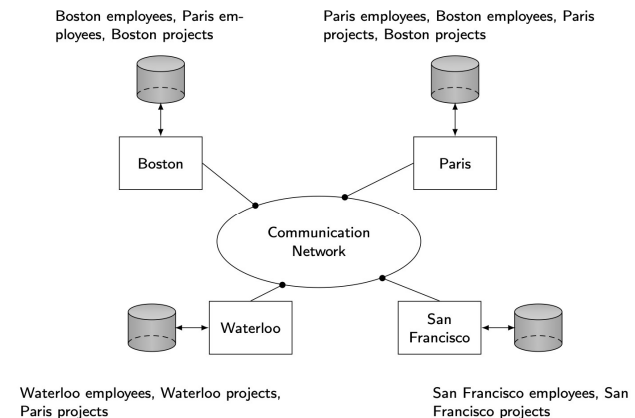
6

Những gì không phải là một hệ CSDL phân tán?

- Hệ thống máy tính chia sẻ thời gian
- Hệ đa xử lý (multiprocessor system)
- Hệ CSDL được đặt tại một nút trên mạng máy tính – đây là một hệ CSDL tập trung tại một nút mạng.

7

Môi trường hệ quản trị CSDL phân tán



8

Các giả thiết ngầm định

- Dữ liệu được lưu trữ tại một số trạm (vị trí) → mỗi trạm về *mặt logic* bao gồm một bộ xử lý đơn nhất.
- Các bộ xử lý tại các trạm khác nhau được kết nối với nhau qua một mạng máy tính → không phải là một hệ đa xử lý
 - Các hệ CSDL song song
- CSDL phân tán là một CSDL, không phải là một tập các file → dữ liệu có liên quan một cách logic như được thể hiện trong các mẫu truy nhập của người dùng
 - Mô hình dữ liệu quan hệ
- Hệ quản trị CSDL phân tán là một hệ quản trị CSDL chính thức
 - Không phải hệ thống file từ xa, không phải hệ thống xử lý giao dịch

9

Điểm quan trọng

Được tích hợp logic
nhưng
Được phân tán vật lý

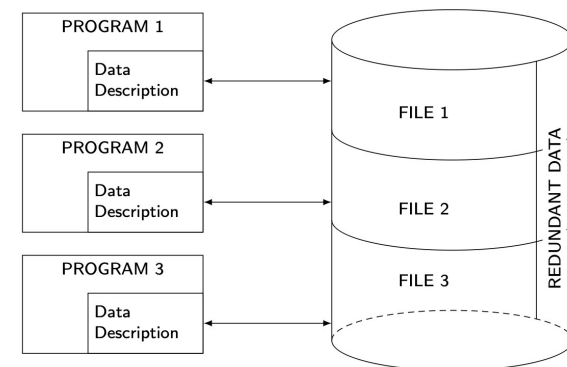
10

Nội dung

- Giới thiệu
 - Thế nào là một hệ quản trị CSDL phân tán?
 - Lịch sử
 - Triển vọng của hệ quản trị CSDL phân tán
 - Các vấn đề về thiết kế
 - Kiến trúc của hệ quản trị CSDL phân tán

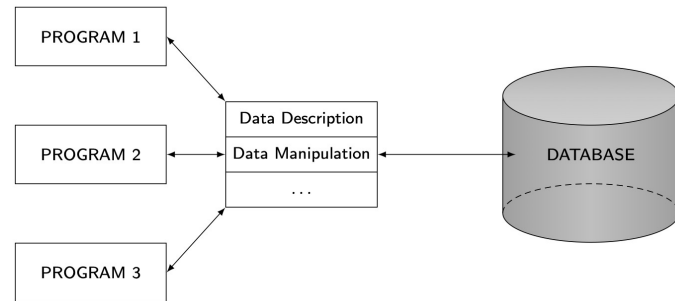
11

Lịch sử – Các hệ thống File



12

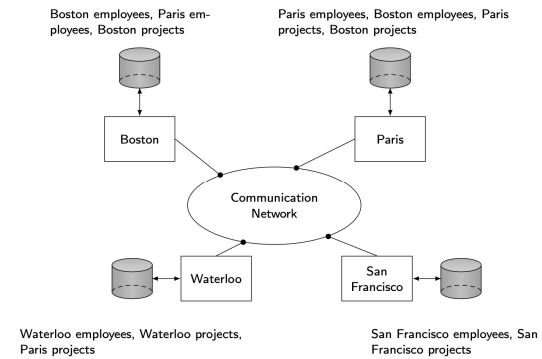
Lịch sử – Quản trị CSDL



13

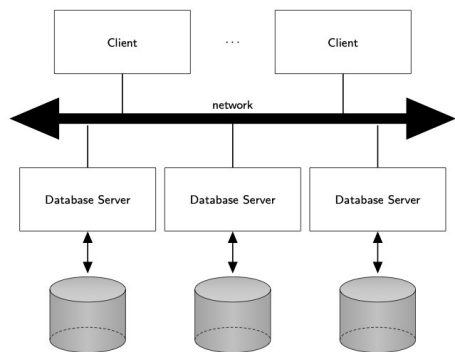
Lịch sử – Phân tán sớm

Điểm – nối – Điểm (Peer-to-Peer, P2P)



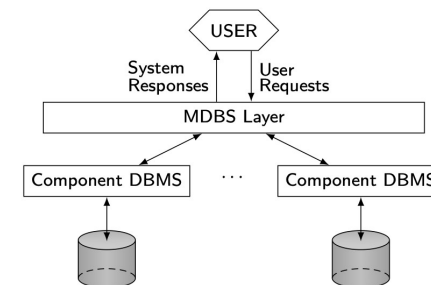
14

Lịch sử – Client/Server



15

Lịch sử – Tích hợp dữ liệu

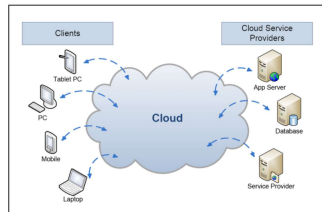


16

Lịch sử – Điện toán đám mây

Các dịch vụ tin cậy, theo nhu cầu được cung cấp qua mạng Internet với chi phí hợp lý.

- Tiết kiệm chi phí: không cần duy trì khả năng tính toán chuyên dụng
- Tính co giãn (Elasticity): có khả năng thích ứng tốt hơn với việc thay đổi khối lượng công việc



17

Các phương thức phân phối dữ liệu

- Các chế độ phân phối
 - Chỉ kéo (Pull-only)
 - Chỉ đẩy (Push-only)
 - Lai ghép (Hybrid)
- Tần suất
 - Định kỳ
 - Theo điều kiện
 - Đặc biệt (Ad-hoc) hoặc không thường xuyên
- Các phương thức truyền thông
 - Đơn hướng (Unicast)
 - Một-nhiều
- Lưu ý: không phải tất cả các kết hợp đều có ý nghĩa

18