

## 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

19

## Triển vọng của hệ quản trị CSDL phân tán

- ❶ Quản lý trong suốt dữ liệu phân tán, phân mảnh và nhân bản
- ❷ Cải thiện độ tin cậy/tính sẵn sàng qua các giao dịch phân tán
- ❸ Cải tiến hiệu năng
- ❹ Mở rộng hệ thống dễ dàng hơn và tiết kiệm hơn

20

## Tính trong suốt

- Tính trong suốt là việc tách biệt về mặt ngữ nghĩa ở mức cao hơn của hệ thống ra khỏi các vấn đề cài đặt ở mức thấp hơn.
- Vấn đề cơ bản là cung cấp **tính độc lập dữ liệu** trong môi trường phân tán
  - ❑ Trong suốt (phân tán) mạng
  - ❑ Trong suốt nhân bản
  - ❑ Trong suốt phân mảnh
    - Phân mảnh ngang: phép chọn
    - Phân mảnh dọc: phép chiếu
    - Lai ghép

21

## Ví dụ

EMP			ASG			
ENO	ENAME	TITLE	ENO	PNO	RESP	DUR
E1	J. Doe	Elect. Eng	E1	P1	Manager	12
E2	M. Smith	Syst. Anal.	E2	P1	Analyst	24
E3	A. Lee	Mech. Eng.	E2	P2	Analyst	6
E4	J. Miller	Programmer	E3	P3	Consultant	10
E5	B. Casey	Syst. Anal.	E3	P4	Engineer	48
E6	L. Chu	Elect. Eng.	E4	P2	Programmer	18
E7	R. Davis	Mech. Eng.	E5	P2	Manager	24
E8	J. Jones	Syst. Anal.	E6	P4	Manager	48
			E7	P3	Engineer	36
			E8	P3	Manager	40

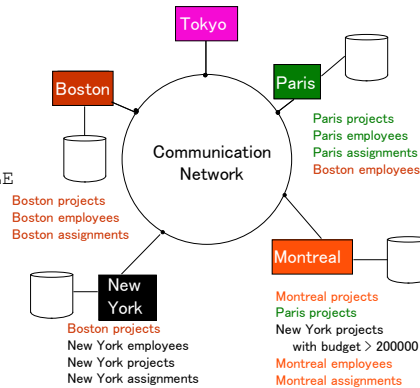
  

PROJ			PAY	
PNO	PNAME	BUDGET	TITLE	SAL
P1	Instrumentation	150000	Elect. Eng.	40000
P2	Database Develop.	135000	Syst. Anal.	34000
P3	CAD/CAM	250000	Mech. Eng.	27000
P4	Maintenance	310000	Programmer	24000

22

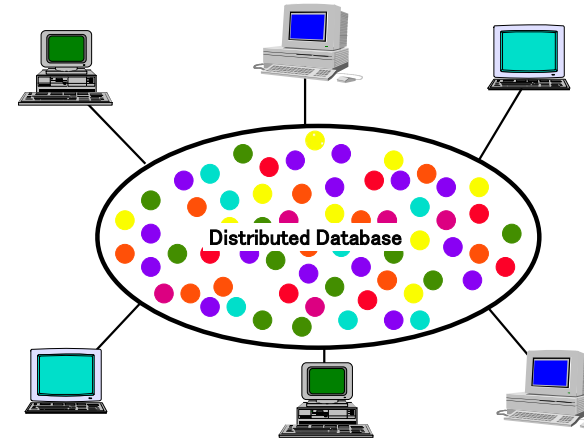
## Truy nhập trong suốt

```
SELECT ENAME, SAL
FROM EMP, ASG, PAY
WHERE DUR > 12
AND EMP.ENO = ASG.ENO
AND PAY.TITLE = EMP.TITLE
```



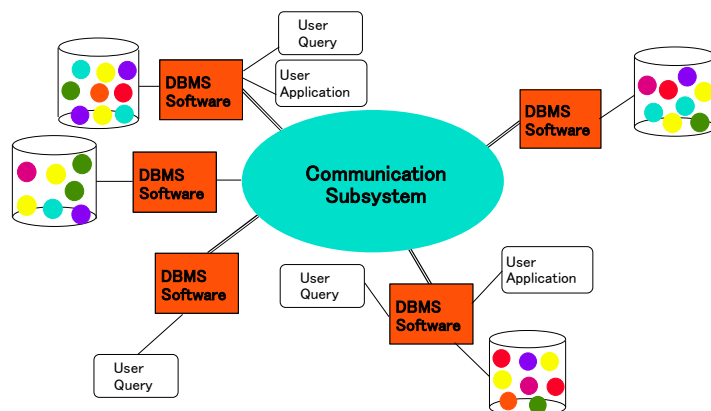
23

## CSDL phân tán – Khung nhìn người dùng



24

## Hệ quản trị CSDL phân tán – Thực tế



25

## Các loại trong suốt

- Tính độc lập dữ liệu
- Trong suốt mạng (hay trong suốt phân tán)
  - Trong suốt vị trí
  - Trong suốt phân mảnh
- Trong suốt phân mảnh
- Trong suốt nhân bản

26

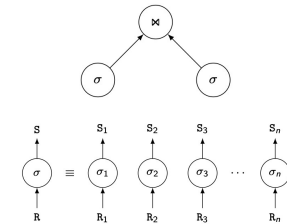
## Độ tin cậy qua các giao dịch

- Các thành phần và dữ liệu được nhân bản sẽ làm cho hệ quản trị CSDL phân tán tin cậy hơn.
- Giao dịch phân tán cung cấp
  - Tính trong suốt đồng thời
  - Tính nguyên tử khi có lỗi
- Để hỗ trợ giao dịch phân tán, yêu cầu cài đặt
  - Các giao thức điều khiển đồng thời phân tán
  - Các giao thức cam kết
- Nhân bản dữ liệu
  - Rất tốt khi có nhiều thao tác đọc, nhưng lại có vấn đề khi cập nhật
  - Các giao thức nhân bản

27

## Khả năng cải thiện hiệu năng

- Khoảng cách gần từ dữ liệu đến điểm cần sử dụng
  - Yêu cầu một số hỗ trợ cho việc phân mảnh và nhân bản
- Tính song song khi thực thi
  - Song song truy vấn ngoài
  - Song song truy vấn trong



28

## Khả năng mở rộng

- Vấn đề là mở rộng quy mô CSDL và quy mô khối lượng công việc
- Bổ sung khả năng xử lý và lưu trữ
- Mở rộng quy mô: thêm nhiều server
  - Tăng dung lượng của một server → có giới hạn

29

## 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

30

## Các vấn đề của hệ quản trị CSDL phân tán

### ■ Thiết kế CSDL phân tán

- Phân tán CSDL như thế nào
- Phân tán CSDL nhân bản và không nhân bản
- Vấn đề liên quan đến quản lý thư mục

### ■ Xử lý truy vấn phân tán

- Chuyển giao dịch của người dùng thành các chỉ dẫn thao tác dữ liệu
- Vấn đề tối ưu hóa
  - $\min\{\text{chi phí} = \text{truyền dữ liệu} + \text{xử lý cục bộ}\}$
- Bài toán tổng quát là NP-hard

31

## Các vấn đề của hệ quản trị CSDL phân tán

### ■ Điều khiển đồng thời phân tán

- Đồng bộ hóa các truy nhập đồng thời
- Tính nhất quán và cô lập các hiệu ứng của giao dịch
- Quản lý bế tắc (deadlock)

### ■ Độ tin cậy

- Làm thế nào để hệ thống có khả năng phục hồi lỗi
- Tính nguyên tử và tính bền vững

32

## Các vấn đề của hệ quản trị CSDL phân tán

### ■ Nhân bản

- Sự nhất quán giữa các bản sao
- Độ mới của bản sao
- Mức độ thường xuyên hay không thường xuyên
- Tập trung và phân tán

### ■ Hệ quản trị CSDL phân tán song song

- Mục tiêu: khả năng mở rộng và hiệu suất cao
- Không phân tán theo địa lý
- Tính toán phân cụm

33

## Các vấn đề liên quan

### ■ Các cách tiếp cận phân tán

- P2P hiện đại
- World Wide Web (WWW hay Web)

### ■ Xử lý dữ liệu lớn

- 4V: độ lớn (volume), tính đa dạng (variety), tốc độ (velocity), tính chính xác (veracity)
- MapReduce & Spark
- Truyền dữ liệu
- Phân tích đồ họa
- NoSQL
- NewSQL
- Polystores

34