

Ngày thi: /5/2015

Thời gian: 60 phút. Được sử dụng tài liệu giấy.

Câu 1: Hãy trình bày ý nghĩa thực tiễn và sự khác biệt giữa cơ sở dữ liệu (CSDL) phân tán và CSDL song song. Phân tích và đánh giá mức độ trong suốt (transparency) của mỗi loại CSDL đó dựa trên các tiêu chí về thiết kế lược đồ, xử lý truy vấn.

Câu 2: Gần đây, Google đã triển khai hệ CSDL phân tán toàn cầu Spanner, được xem là thế hệ sau của BigTable (theo mô hình NoSQL) và dựa theo mô hình NewSQL. Hệ quản trị này cho phép nhân bản (replicate) dữ liệu trong các trung tâm dữ liệu giữa các lục địa. Hãy phân tích ưu, nhược điểm của mô hình CSDL phân tán toàn cầu này.

Câu 3: Xét lược đồ CSDL đã giới thiệu trong khoá học, bao gồm bốn quan hệ EMP(eNo, eName, Title), ASG(eNo, pNo, Resp, Dur), Proj(pNo, pName, Budget) và Pay(Title, Sal). Giả sử size(EMP \bowtie ASG) = 2000, size(ASG \bowtie PROJ) = 1000 và CSDL này được phân mảnh ngang minh hoạ như bảng sau:

Quan hệ	Điểm A	Điểm B	Điểm C
EMP	2000 (eNo <= 10000)	1000 (1000 < eNo < 10000)	1000 (eNo >= 10000)
ASG		2000	
PROJ			1000

Xét truy vấn SELECT eName, pName FROM EMP, ASG, PROJ WHERE
(dur > 24 OR resp = "PM")
AND EMP.eNo = ASG.eNo
AND (title = "IT Eng." OR ASG.pNo < "P3")
AND (dur > 24 OR resp NOT = "PM")
AND ASG.pNo = PROJ.pNo

- Hãy đơn giản hoá truy vấn trên và biến đổi về cây truy vấn tối ưu nhất có thể.
- Giả sử truy vấn này được thi hành ở điểm D, với chỉ phí đường truyền là 1 đơn vị với mỗi bộ dữ liệu của tất cả các quan hệ trên; số dự án có dur > 24 là 100; các bộ của EMP mà (title = "IT Eng.") đều nằm ở phần mảnh tại điểm B. Hãy xác định chiến lược thi hành truy vấn trên một cách tối ưu nhất (bỏ qua chi phí tính toán cục bộ tại mỗi điểm).