

MỘT SỐ YÊU CẦU BÀI TẬP LỚN MÔN CSDL PHÂN TÁN

--o0o--

1. Định dạng

Tham khảo định dạng Khóa luận Tốt nghiệp (Môn Kỹ năng soạn thảo văn bản)

2. Cấu trúc bài tập lớn gồm những phần sau:

- Bìa quyền: (Bài tập lớn môn Cơ sở dữ liệu phân tán, Tên nhóm, Mã sinh viên, Họ tên những thành viên, Tên đề tài, Giáo viên hướng dẫn)
- Mục lục
- Chương 1. Tổng quan về cơ sở dữ liệu phân tán
 - o Xử lý dữ liệu phân tán, hệ CSDL phân tán là gì?
 - o Kiến trúc phân tán
 - o Khả năng của CSDL phân tán
 - o Các vấn đề cần quan tâm trong phân tán
(Tham khảo: Chapter 1 – Introduction trong quyển *Principles of Distributed Database Systems*)
- Chương 2. Xây dựng cơ sở dữ liệu phân tán cho XYZ (XYZ là đề tài nhóm đã chọn)
- 2.1. Mô tả bài toán:
 - o Mô tả yêu cầu dữ liệu:
 - Mô tả các thực thể dữ liệu và mối quan hệ
 - Không cần đầy đủ, sao cho khi cài đặt có khoảng 5-6 bảng
 - o Mô hình thực thể quan hệ:
 - Vẽ mô hình ER như thiết kế CSDL tập trung
 - o Lược đồ quan hệ:
 - Chuyển mô hình ER → Lược đồ quan hệ (Tên bảng và các thuộc tính)
 - Chỉ rõ khóa chính (kí hiệu: #), khóa ngoại (kí hiệu: @)
 - Ví dụ: Sinhvien(#masv, hoten, ngaysinh, quequan, @malop)
- 2.2. Thiết kế phân tán
 - o **Yêu cầu:**
 - Số trạm bằng số thành viên trong nhóm
 - Có 2 bảng phân mảnh dọc, 2 bảng phân mảnh ngang, còn lại nhân bản
 - Trong các mảnh, bảng vẫn có các khóa chính, khóa ngoại. Nhưng không có các ràng buộc tham chiếu
 - Mô tả chi tiết các mảnh, bảng đặt ở trạm nào
 - Mô tả chi tiết các mảnh, bảng nào được nhân bản ở trạm nào

- Đặt tên cho các mảnh có ý nghĩa, VD: EMP_NO_SAL; EMP_SAL hay DEPT_LA, DEPT_NY,...

○ **Kịch bản phân mảnh dọc cho 1 bảng như sau:**

Giả sử số trạm (Site), Số ứng dụng (AP) bằng số thành viên trong nhóm. VD: Nhóm 2 có 4 sinh viên: A28372, A28374, A28834, A28864 và phân mảnh dọc bảng B có 6 thuộc tính được đặt tên C, C1, C2, C3, C4, C5, với C là khóa chính.

+ Lập ma trận truy cập vào cột trong bảng B như sau: các ứng dụng lẻ: AP1, AP3,... truy cập vào các cột lẻ; các ứng dụng chẵn: AP2, AP4,... truy cập vào các cột chẵn. Ví dụ trên thì ở các dòng AP1, AP3 thì cột C1, C3, C5 có giá trị 1, còn lại giá trị 0;

+ Lập ma trận tần suất truy cập ở các trạm như sau: lấy các số cuối của các mã sinh viên (nhóm có 3 người thì 3 số cuối, 4 người thì 4 số cuối,...). Các ứng dụng AP1, AP2, AP3,... tương ứng với các mã sinh viên được sắp xếp tăng dần. Để đơn giản trong tính toán thì ma trận truy cập được lập như sau: nhóm chẵn thì các số chẵn cho về 0, nhóm lẻ thì các số lẻ cho về 0; nếu toàn số chẵn thì vị trí chẵn cho về 0; nếu toàn lẻ thì vị trí lẻ cho về 0.

Vậy ta có ma trận truy cập và tần suất cho ví dụ trên như sau:

	C1	C2	C3	C4	C5	Tần suất truy cập			
						S1	S2	S3	S4
AP1	1	0	1	0	1	0	3	7	0
AP2	0	1	0	1	0	0	3	7	0
AP3	1	0	1	0	1	0	0	3	0
AP4	0	1	0	1	0	8	0	6	0

- Sau đó dùng thuật toán năng lượng liên kết – BEA để tính toán phân mảnh dọc cho bảng B.

○ **Kịch bản phân mảnh ngang cho 1 bảng như sau:**

- Với mỗi ứng dụng (AP) hãy tự nghĩ ra 1 đến 2 vị từ. Ví dụ: với bảng B ở trên AP1 có vị từ C1>200, AP2 có vị từ C5='LA',... sau đó xây dựng các Hội vị từ
- Dùng các hội vị từ đó để phân mảnh ngang cho bảng B.

- 2.3. Kiểm soát dữ liệu

- Xây dựng các Trigger để kiểm soát dữ liệu

- Đảm bảo khóa chính là duy nhất: Trong các phân mảnh ngang, 1 bảng được phân mảnh ở nhiều trạm. Nên cần đảm bảo khi chèn giá trị khóa vào 1 mảnh thì giá trị khóa đó chưa tồn tại ở các mảnh còn lại.
- Đảm bảo tham chiếu: Chỉ được chèn các giá trị cho khóa ngoại khi giá trị đó tồn tại ở bảng tham chiếu đến. Không được xóa giá trị khóa chính khi đang tồn tại một giá trị tham chiếu đến
- 2.4. Tối ưu câu truy vấn
 - Mỗi nhóm tự nghĩ ra câu truy vấn trên 3 bảng (mỗi bảng được phân mảnh ở các trạm). Xây dựng cây toán tử và tối ưu trên cây toán tử. Gợi ý: dựa vào các dạng câu như trong bài giảng.
 - Số câu truy vấn bằng số lượng thành viên của nhóm
 - Viết lại các câu truy vấn sau khi đã tối ưu.
 - Tạo view cho các câu truy vấn đã tối ưu trên
- 2.5. Xây dựng các thao tác dữ liệu:
 - Xây dựng các thủ tục (chèn, sửa, xóa) cho tất cả các bảng
 - Chèn 1 bộ giá trị cho các bảng. Chú ý đưa các giá trị tương ứng cho các mảnh nếu bảng đó được phân mảnh
 - Sửa 1 giá trị khi biết khóa của bảng.
 - Xóa 1 bản ghi khi biết khóa của bảng.
- Kết luận
 - Nội dung đã tìm hiểu được
 - Hướng phát triển
- Tài liệu tham khảo
 - Mô tả các tài liệu tham khảo