

Kiểm tra giữa kỳ #2 – 17/10/2016**Môn: Hệ Quản Trị Cơ Sở Dữ Liệu (503004)**Thời gian làm bài: **60 phút****PHẦN TRẮC NGHIỆM (4đ):**

Sinh viên chọn 1 câu trả lời đúng nhất. Nếu chọn câu (e) thì sinh viên cần trình bày đáp án khác và giải thích.

Câu 1. Tối ưu hóa truy vấn (*query optimization*) là gì?

- a. Quá trình thực thi các phép toán đại số quan hệ tương ứng với câu lệnh truy vấn SQL.
- b. **Quá trình chọn phương án thực thi phù hợp nhất cho quá trình xử lý câu lệnh truy vấn SQL hiệu quả.**
- c. Quá trình phân tích câu lệnh truy vấn SQL và biểu diễn lại câu lệnh bằng cây truy vấn.
- d. Quá trình phân tích các tập tin dữ liệu (bảng dữ liệu) trong câu lệnh truy vấn SQL và sau đó, chọn được các đường truy đạt (access path) tốt nhất cho từng tập tin dữ liệu liên quan.
- e. Ý kiến khác.

Câu 2. Loại chi phí nào sau đây được ưu tiên xem xét trong trường hợp xử lý truy vấn từ cơ sở dữ liệu phân tán?

- a. **Chi phí giao tiếp (*communication cost*)**
- b. Chi phí lưu trữ (*storage cost*)
- c. Chi phí truy đạt (*access cost*) đến vùng lưu trữ thứ cấp
- d. Chi phí tính toán (*computation cost*)
- e. Ý kiến khác.

Câu 3. Sắp thứ tự ngoại (*external sorting*) được sử dụng trong việc xử lý truy vấn nào sau đây?

- a. Phép kết với phương pháp hai vòng lặp lồng (*nested loop join*)
- b. Phép giao (*intersection, \cap*)
- c. Phép kết với phương pháp sort-merge
- d. **Cả 2 trường hợp b và c**
- e. Ý kiến khác

Câu 4. Tập tin DEPARTMENT chứa các bản ghi được đánh chỉ mục sơ cấp (*primary index*) theo vùng tin DNUMBER. Quá trình xử lý câu truy vấn nào sau đây có thể sử dụng chỉ mục này?

- a. $\sigma_{DNUMBER > 5}(DEPARTMENT)$
- b. $\sigma_{DNUMBER \geq 5}(DEPARTMENT)$
- c. $\sigma_{DNUMBER = 5}(DEPARTMENT)$
- d. **Câu a, b, và c đều đúng.**
- e. Ý kiến khác.

Câu 5. Đơn vị cơ bản có thể chuyển câu truy vấn SQL thành các toán tử đại số (*algebra operator*) dùng để tối ưu hoá là:

- a. Cả câu truy vấn
- b. Mỗi dòng trong câu truy vấn (*line of query*)
- c. **Khối truy vấn (*query block*)**
- d. Từ khoá (*keyword*) trong câu truy vấn
- e. Ý kiến khác

Câu 6. Các phép toán tập hợp (*hội \cup , giao \cap , hiệu $-$*) có những tính chất nào dưới đây? (Giả sử θ đại diện cho bất kỳ phép toán nào trong 3 phép toán trên)

- a. Tính giao hoán: $R \theta S \equiv S \theta R$
- b. Tính kết hợp giữa phép tích CARTESIAN (\times) và phép toán tập hợp: $(R \theta S) \times T \equiv R \theta (S \times T)$
- c. **Tính giao hoán phép chọn với phép toán tập hợp: $\sigma_c(R \theta S) \equiv (\sigma_c(R)) \theta (\sigma_c(S))$**
- d. Câu a và c đúng
- e. Ý kiến khác

Câu 7. Chọn phát biểu ĐÚNG NHẤT về cơ chế xử lý của bộ tối ưu hóa truy vấn dùng qui tắc biến đổi heuristic (*heuristic rule-based query optimizer*).

- a. Ưu tiên thực hiện phép chọn trước phép kết.
- b. Ưu tiên xử lý tập tin dữ liệu nhỏ trước tập tin dữ liệu lớn.
- c. Ưu tiên thực hiện phép kết trước phép tích Cartesian.
- d. **Ưu tiên thực hiện các phép toán trả về kết quả (trung gian) ít hơn trước các phép toán trả về kết quả nhiều hơn.**
- e. Ý kiến khác.

Câu 8. Chọn phát biểu đúng về cây truy vấn (*query tree*) và đồ thị truy vấn (*query graph*):

- Các quan hệ (relations) của câu truy vấn có thể xuất hiện ở nốt lá và nốt nội của cây truy vấn.
- Các quan hệ của cây truy vấn được biểu diễn dưới dạng hình tròn nét đôi trong đồ thị truy vấn.
- Cây truy vấn không cho biết thứ tự thực thi của các tác vụ trong câu truy vấn trong khi đồ thị truy vấn thì cho biết thứ tự này.
- Tất cả đều sai.**
- Ý kiến khác.

PHẦN TỰ LUẬN (6đ)

Câu 1. Cần sắp xếp một file có **50.000 blocks** dùng giải thuật sắp thứ tự ngoại (*External sorting*). Không gian đệm sẵn sàng (*available buffer space*) là **11 blocks**. (1đ)

- Tính số *run* tạo ra trong giai đoạn sắp xếp (*sort phase*)
- Cần bao nhiêu lần trộn (*pass*) trong giai đoạn trộn (*merge phase*)?

Câu 2. Cho 3 quan hệ EMPLOYEE, DEPARTMENT và DEPT_LOCATIONS có lược đồ như sau:

EMPLOYEE (ENAME, SSN, ADDRESS, SALARY, DoB, DNO);

DEPARTMENT (DNAME, DNUMBER, MGRSSN)

DEPT_LOCATIONS (DNUM, LOCATION)

- Hãy viết câu lệnh SQL và biểu thức đại số quan hệ liệt kê tên (*ENAME*) và địa chỉ (*ADDRESS*) của những trưởng phòng ban tại vị trí 'Ho Chi Minh' có lương lớn hơn 10.000. (0.5đ)
- Vẽ cây truy vấn ban đầu (*initial query tree*) ứng với câu lệnh SQL ở trên. (0.5đ)
- Hãy thực hiện tối ưu hoá và cho biết kết quả từng bước của cây truy vấn dựa trên các qui tắc tối ưu kinh nghiệm (*Heuristic Optimization*). Biết rằng số lượng phòng ban tại 'Ho Chi Minh' ít hơn rất nhiều so với số lượng nhân viên có lương lớn hơn 100.000. (1đ)

Câu 3. Tập tin EMPLOYEE có:

- Số lượng bản ghi $r_E = 8000$ records, Số lượng khối $b_E = 1000$ blocks, Hệ số phân khối $bfr_E = 8$ records/block.
- Trên thuộc tính không khóa SALARY, một chỉ mục thứ cấp (*secondary index*) được định nghĩa gồm 3 mức ($x_{SALARY} = 3$), số lượng khối ở mức lá $b_{lSALARY} = 20$ blocks, và lượng bản ghi được chọn trung bình (*average selection cardinality*) $s_{SALARY} = 20$.
- Trên thuộc tính không khóa DNO, một chỉ mục cụm (*clustering index*) được định nghĩa gồm 2 mức ($x_{DNO} = 2$), số lượng giá trị phân biệt tại thuộc tính DNO là $d_{DNO} = 80$, và lượng bản ghi được chọn trung bình $s_{DNO} = r_E/d_{DNO} = 62$.

Tập tin DEPARTMENT có:

- Số lượng bản ghi $r_D = 80$ records, Số lượng khối $b_D = 8$ blocks, hệ số phân khối $bfr_D = 10$ records/block.
- Trên thuộc tính khóa DNUMBER, một chỉ mục sơ cấp (*primary index*) được định nghĩa gồm 1 mức ($x_{DNUMBER} = 1$) với lượng bản ghi được chọn trung bình là $s_{DNUMBER} = 1$.
- Trên thuộc tính không khóa MGRSSN, một chỉ mục thứ cấp được định nghĩa gồm 2 mức ($x_{MGRSSN} = 2$) với lượng bản ghi được chọn trung bình là $s_{MGRSSN} = 1$ và số lượng khối ở mức 1 là $b_{lMGRSSN} = 2$.

Dựa trên tối ưu hóa truy vấn dùng chi phí, hãy xác định kế hoạch thực thi hợp lý nhất cho biểu thức truy vấn sau:

- $\sigma_{SALARY > 5000}(EMPLOYEE)$ (1đ)
- $EMPLOYEE \bowtie_{DNO=DNUMBER} DEPARTMENT$? Cho trước độ chọn lọc kết (*join selectivity*) dành cho biểu thức truy vấn trên là $js = 1/r_D = 1/80$; hệ số phân khối của tập tin kết quả kết từ biểu thức truy vấn trên là $bfr_{ED} = 5$ records/block; vùng đệm sẵn sàng trong bộ nhớ chính để hiện thực phép kết là $n_B = 7$ blocks. (2đ)