

Kiểm tra giữa kỳ #1 – 19/09/2016**Môn: Hệ Quản Trị Cơ Sở Dữ Liệu (503004)**Thời gian làm bài: **60** phút**PHẦN TRẮC NGHIỆM (5đ)**

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. Các câu lệnh về định nghĩa dữ liệu (*data definition statement*) do ai gọi đến hệ quản trị cơ sở dữ liệu?

- a. Người thiết kế cơ sở dữ liệu (*database designer*)
- b. Người sở hữu dữ liệu (*data owner*)
- c. **Người quản trị cơ sở dữ liệu (*database administrator – DBA*)**
- d. Các chương trình ứng dụng (*application program*) chạy trên hệ quản trị cơ sở dữ liệu
- e. Ý kiến khác.

Câu 2. Đơn vị dữ liệu đọc/ghi giữa bộ nhớ sơ cấp (*primary memory*) bộ nhớ thứ cấp (*secondary memory*) là gì?

- a. Phần tử/ biến dữ liệu (*data item/ variable*)
- b. Bản ghi (*record*)
- c. **Khối (*block*)**
- d. Tập tin (*file*)
- e. Ý kiến khác

Câu 3. Tập tin có thứ tự (*ordered file*) **không hiệu quả** cho trường hợp nào sau đây?

- a. Thêm mới các bản ghi (*record*) vào tập tin.
- b. Tìm kiếm một bản ghi trong tập tin dựa trên giá trị của vùng tin sắp thứ tự (*ordering field*).
- c. Cập nhật giá trị của vùng tin được sắp thứ tự.
- d. **Cả 2 trường hợp a và c.**
- e. Ý kiến khác.

Câu 4. Phân khối phủ (*spanned blocking*) **bắt buộc** phải sử dụng cho trường hợp nào sau đây?

- a. Tập tin gồm những bản ghi có kích thước thay đổi.
- b. **Tập tin gồm những bản ghi có kích thước cố định với kích thước bản ghi lớn hơn kích thước khối.**
- c. Cả hai trường hợp (a, b) đều bắt buộc dùng spanned blocking.
- d. Cả hai trường hợp (a, b) đều không thể dùng spanned blocking.
- e. Ý kiến khác.

Câu 5. Kỹ thuật bộ đệm đôi (*double buffering*) được sử dụng hiệu quả khi các khối dữ liệu cần được đọc/ghi được **tổ chức** dạng nào?

- a. **Liên tục (*consecutive*)**
- b. Liên kết (*linked*)
- c. Cụm (*clustered*)
- d. Chỉ mục (*indexed*)
- e. Ý kiến khác.

Câu 6. Chọn phát biểu **ĐÚNG** về chỉ mục.

- a. Một tập tin chỉ có thể có một loại chỉ mục: chỉ mục sơ cấp, chỉ mục cụm, hoặc chỉ mục thứ cấp.
- b. **Một tập tin có nhiều nhất một chỉ mục sơ cấp (*primary index*) hoặc nhiều nhất một chỉ mục cụm (*clustering index*), nhưng không thể có cùng lúc cả hai loại chỉ mục này.**
- c. Một tập tin có nhiều nhất một chỉ mục sơ cấp (*primary index*) hoặc nhiều chỉ mục thứ cấp (*secondary index*), nhưng không thể có cùng lúc cả hai loại chỉ mục này.
- d. Một tập tin có nhiều nhất một chỉ mục cụm (*clustering index*) hoặc nhiều chỉ mục thứ cấp (*secondary index*), nhưng không thể có cùng lúc cả hai loại chỉ mục này.
- e. Ý kiến khác.

Câu 7. Cây chỉ mục **B** (*B-Tree*) và cây chỉ mục **B⁺** (*B⁺-Tree*) có những đặc điểm gì sau đây **khác nhau**?

- a. B-Tree dùng làm chỉ mục đa mức tĩnh và B⁺-Tree dùng làm chỉ mục đa mức động.
- b. B-Tree dùng làm chỉ mục sơ cấp và B⁺-Tree dùng làm chỉ mục thứ cấp.
- c. B-Tree chỉ hỗ trợ toán tử =, <, và >; trong khi đó, B⁺-Tree hỗ trợ toán tử =, <, <=, >, >=, và between.
- d. **Cấu trúc của node nội và node lá trong B-Tree giống nhau; trong khi đó, cấu trúc của node nội và node lá trong B⁺-Tree khác nhau và do đó, khả năng chỉ mục của B⁺-Tree thường hiệu quả hơn.**
- e. Ý kiến khác.

Câu 8. Một tập tin có số bản ghi $r = 32000$ bản ghi, hệ số phân khối $bfr = 8$. Các bản ghi của tập tin được sắp thứ tự vật lý theo các giá trị của vùng tin không khóa (*nonkey field*) **A**. Vùng tin **A** có 500 giá trị phân biệt. Xác định **số lượng mục tin chỉ mục** (*index entry*) của chỉ mục được lập trên vùng tin **A**.

- a. 32000
- b. 4000
- c. **500**
- d. Không thể xác định được do thông tin mô tả về chỉ mục này chưa đầy đủ chi tiết.
- e. Ý kiến khác.

Câu 9. Cho chỉ mục B⁺-tree trên vùng tin khóa SSN (*key field*) của tập tin dữ liệu Employee. Các giá trị của vùng tin khóa SSN **không** được dùng để sắp thứ tự các bản ghi của tập tin dữ liệu Employee. Chỉ mục B⁺-tree này được gọi **tên** là gì?

- a. Chỉ mục sơ cấp (*primary index*)
- b. Chỉ mục cụm (*clustering index*)
- c. **Chỉ mục thứ cấp (*secondary index*)**
- d. Không đủ chi tiết mô tả về chỉ mục này nên không thể kết luận được dạng của chỉ mục này.
- e. Ý kiến khác.

Câu 10. Giả sử kích thước vùng tin khoá V (*search key field*) là 7 bytes, kích thước block B là 512 bytes, *record pointer* Pr chiếm 7 bytes và *block pointer* P chiếm 6 bytes. Xác định bậc nút nội (*order for internal nodes*) p và bậc nút lá (*order for leaf nodes*) p_{leaf} của cây chỉ mục B⁺-tree. Biết rằng mỗi nút của cây được chứa trọn trong 1 block.

- a. $p = 39, p_{leaf} = 38$
- b. **$p = 39, p_{leaf} = 36$**
- c. $p = 40, p_{leaf} = 39$
- d. $p = 40, p_{leaf} = 37$
- e. Ý kiến khác.

PHẦN TỰ LUẬN (5đ)

Câu 1. Một tập tin có các bản ghi được tổ chức theo kỹ thuật băm khả mở (*extendible hashing*). Giả sử: mỗi thùng (*bucket*) gồm 1 khối dữ liệu (*block*); mỗi khối dữ liệu chỉ có thể chứa 2 bản ghi (*record*); hàm băm (*hashing function*) là $h(K) = K \bmod 128$. Cho biết trạng thái của tập tin ứng với **từng giá trị** được thêm vào, mô tả ở bảng dưới. (1đ)

Thứ tự thêm vào tập tin	Giá trị vùng tin băm của mỗi bản ghi	Kết quả băm (mod 128) trong hệ thập phân	Kết quả băm (mod 128) trong hệ nhị phân
1	1096	72	1001000
2	4744	8	0001000
3	5659	27	0011011
4	1817	25	0011001

Câu 2. Cho kích thước một khối (*block*) $B = 1024$ bytes, kích thước con trỏ khối $p_b = 6$ bytes, kích thước con trỏ bản ghi $p_r = 7$ bytes. Một tập tin gồm **50.000** bản ghi (*record*) có chiều dài cố định, sử dụng cách lưu trữ không phủ (*unspanned*) và được sắp thứ tự vật lý theo giá trị **SSN**. Mỗi bản ghi gồm các vùng tin (*field*) sau: NAME (20 bytes), SSN (9 bytes), ADDRESS (10 bytes), JOBCODE (4 bytes), GENDER (1 byte), SALARY (15 bytes). Tập tin được đánh chỉ mục trên vùng tin khoá (*key*) **SSN**.

- a. Tính kích thước 1 record. (0.25đ)
- b. Tính hệ số phân khối (*blocking factor*) và số khối (*number of file blocks*) của tập tin này. (0.5đ)
- c. Tính số dòng chỉ mục (*index entry*) ở mức cuối cùng (*top level*) nếu chúng ta tổ chức chỉ mục này thành một cấu trúc chỉ mục đa mức. (0.5đ)
- d. Tính số lần truy xuất khối (*number of block access*) để đọc được các bản ghi với phép toán so sánh “=” trên vùng tin **SSN** khi dùng chỉ mục đa mức. (0.5đ)
- e. Tính số lần truy xuất khối (*number of block access*) để đọc được các bản ghi với phép toán so sánh “>” trên vùng tin **SALARY**. (0.25đ)

Câu 3. Một tập tin sử dụng cấu trúc chỉ mục đa mức động B⁺-Tree có bậc là $p = 3$ và $p_{leaf} = 2$. Giả sử giá trị tìm kiếm được lần lượt thêm vào theo thứ tự sau: 72, 55, 37, 24, 49, 8, 69, 75.

Hãy vẽ trạng thái của cây B⁺-Tree ứng với **từng giá trị** được thêm vào như trên. (2đ)