

Kiểm tra giữa kỳ

Môn: **Hệ Quản Trị Cơ Sở Dữ Liệu (CO3021)** - Ngành: **Khoa Học Máy Tính** - HK: 2, 2018-2019

Ngày kiểm tra: **27.03.2019** - Thời gian làm bài: **60** phút

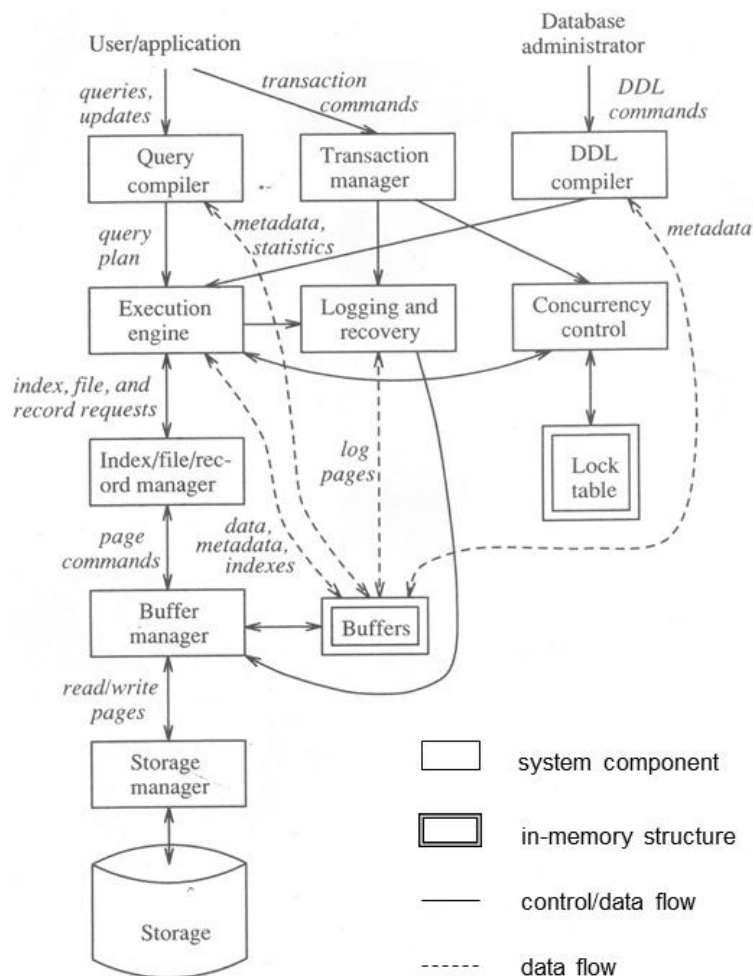
(Bài kiểm tra gồm 12 câu hỏi trắc nghiệm và 3 câu tự luận. Sinh viên được **tham khảo tài liệu**.)

Sinh viên chọn 1 câu trả lời đúng nhất cho các câu hỏi trắc nghiệm và trình bày lời giải cho các câu tự luận. Nếu chọn câu trả lời (e) cho câu hỏi trắc nghiệm thì sinh viên cần trình bày đáp án khác so với đáp án ở các câu (a), (b), (c), và (d) và/hoặc giải thích lựa chọn (e) của mình.

I. Phần trắc nghiệm:

Câu 1. Trong kiến trúc của hệ quản trị cơ sở dữ liệu (*database management system, DBMS*) ở **Hình 1**, nhà quản trị cơ sở dữ liệu (*database administrator, DBA*) chịu trách nhiệm cho công việc nào sau đây?

- a. Định nghĩa cơ sở dữ liệu và các đối tượng liên quan
- b. Định nghĩa các câu lệnh sử dụng cơ sở dữ liệu
- c. Định nghĩa các giao tác trong các chương trình ứng dụng
- d. Định nghĩa các hàm của các mô-đun như *Query compiler, Transaction manager, Logging and recovery, Concurrency control, ...*
- e. Ý kiến khác.



Hình 1 – Kiến trúc hệ quản trị cơ sở dữ liệu (*architecture of a database management system*)

Câu 2. Trong kiến trúc của DBMS ở **Hình 1**, mô-đun *Query compiler* chịu trách nhiệm chính cho công việc gì ?

- a. Biến đổi câu truy vấn từ ngôn ngữ SQL thành chuỗi thực thi tối ưu của các toán tử đại số quan hệ.
- b. Biến đổi câu truy vấn từ ngôn ngữ SQL thành chuỗi thực thi của các hàm của DBMS.
- c. Biến đổi câu truy vấn từ ngôn ngữ SQL thành câu truy vấn của ngôn ngữ máy.
- d. Biến đổi câu truy vấn từ ngôn ngữ SQL thành câu truy vấn của ngôn ngữ lập trình đã được dùng để hiện thực DBMS.
- e. Ý kiến khác

Câu 3. Trong kiến trúc của DBMS ở **Hình 1**, mô-đun *Concurrency control* chịu trách nhiệm chính cho đặc tính gì của các giao tác?

- a. Đơn thể (*atomic*)
- b. Nhất quán (*consistency*)
- c. Bền vững theo thời gian (*durability*)
- d. Đơn lập (*isolation*)
- e. Ý kiến khác.

Câu 4. Cho tập tin dữ liệu Nhân_Viên gồm 19 bản ghi với hệ số phân khối là 2 bản ghi/khối (*records/block*) trong **Bảng 1**. Vùng tin ID là vùng tin khóa (*key field*) của tập tin; các vùng tin còn lại là vùng tin phi khóa (*non-key field*). Ngoài ra, các bản ghi (*record*) trong tập tin này được sắp thứ tự vật lý theo các giá trị tăng dần của vùng tin ID, nghĩa là tập tin bắt đầu với bản ghi có ID = 2. Trong đó, kích thước cố định của vùng tin ID là 9 bytes, vùng tin Tên là 50 bytes, vùng tin Lương là 16 bytes, vùng tin Phòng là 9 bytes, và vùng tin Quá_trình_công_tác là 150 bytes. *Deletion_marker* có kích thước 1 byte được sử dụng để đánh dấu xóa cho các bản ghi. Giả sử rằng tập tin này được lưu trữ vật lý trên bộ đĩa từ được định dạng với kích thước một khối dữ liệu (*block size*) là **B** = 512 bytes. Các bản ghi không phải được dùng cho tập tin này. Xác định **tổng số dung lượng lãng phí (bytes) cho tập tin** này.

- a. 798 bytes
- b. 655 bytes
- c. 420 bytes
- d. 42 bytes
- e. Ý kiến khác.

Bảng 1 - Tập tin dữ liệu Nhân_Viên tương ứng với bảng Nhân_Viên trong cơ sở dữ liệu

ID	Tên	Lương	Phòng	Quá_trình_công_tác	Deletion_marker
2	An	2000	0
5	Đông	1000	0
9	Hà	2500	0
10	Phúc	1500	0
14	Hoa	1800	0
15	Tuấn	2000	0
17	Cường	1900	0
19	Minh	2100	0
21	Mai	2500	0
23	Đào	2300	0
28	Xuân	1600	0
30	Dũng	2000	0
36	Thu	2300	0
45	Nga	2500	0
70	Tú	1800	0
82	Tây	2000	0
87	Bắc	3100	0
88	Thượng	2700	0
93	Châu	3000	0

Câu 5. Giả sử tập tin dữ liệu Nhân_Viên trong **Bảng 1** với các giả thiết ở **Câu 4**. Tập tin dữ liệu này được lưu trữ trên bộ đĩa từ có số vòng quay trong 1 phút là $p = 7200$ rpm, có thời gian tìm kiếm trung bình (*average seek time*) là $s = 20$ msec. Kích thước một rãnh dữ liệu (*track size*) là $T = 102400$ bytes. Giả sử kết quả của các phép toán trong các tính toán sau được làm tròn với 2 chữ số thập phân. Xác định **thời gian** đọc toàn bộ tập tin khi các khối dữ liệu được đặt liên tục nhau trên cùng rãnh.

a. 242.08 (ms) b. 62.08 (ms) c. 24.58 (ms) d. 24.21 (ms)

e. Ý kiến khác.

Câu 6. Giả sử tập tin dữ liệu Nhân_Viên trong **Bảng 1** với các giả thiết được mô tả ở **Câu 4**. Xác định **thời gian trung bình** đọc tập tin để tìm kiếm bản ghi của nhân viên có ID = 70.

a. 96.83 (ms) b. 72.63 (ms) c. 36.83 (ms) d. 24.22 (ms)

e. Ý kiến khác.

Câu 7. Giả sử tập tin dữ liệu Nhân_Viên trong **Bảng 1** với các giả thiết được mô tả ở **Câu 4**. Xác định các thao tác cần được thực hiện để thay đổi ID = 70 thành ID = 33 cho nhân viên tương ứng.

a. Định vị bản ghi có ID = 33, thay đổi giá trị của Deletion_marker của bản ghi này từ 1 thành 0.

b. Định vị bản ghi có ID = 70, thay đổi giá trị của Deletion_marker của bản ghi này từ 0 thành 1, thêm bản ghi tương ứng có ID = 33 và Deletion_marker = 0 vào tập tin dữ liệu Nhân_Viên.

c. Định vị bản ghi có ID = 70, thay đổi giá trị của Deletion_marker của bản ghi này từ 1 thành 0, thêm bản ghi tương ứng có ID = 33 và Deletion_marker = 1 vào tập tin tràn của tập tin dữ liệu Nhân_Viên.

d. Định vị bản ghi có ID = 70, thay đổi giá trị của Deletion_marker của bản ghi này từ 0 thành 1, thêm bản ghi tương ứng có ID = 33 và Deletion_marker = 0 vào tập tin tràn của tập tin dữ liệu Nhân_Viên.

e. Ý kiến khác.

Câu 8. Giả sử tập tin dữ liệu Nhân_Viên trong **Bảng 1** với các giả thiết được mô tả ở **Câu 4**. Các khối của tập tin này cần được phân bổ (*allocate*) như thế nào để áp dụng kỹ thuật bộ đệm đôi (*double buffering*) hiệu quả?

a. cụm (*clustered*) b. chỉ mục (*indexed*) c. liên tục (*contiguous*) d. liên kết (*linked*).

e. Ý kiến khác.

Câu 9. Giả sử tập tin dữ liệu Nhân_Viên trong **Bảng 1** với các giả thiết được mô tả ở **Câu 4**. Giả sử các bản ghi của tập tin này được tổ chức theo kỹ thuật băm tĩnh (*static hashing*) với mỗi thùng (*bucket*) gồm 2 khối liên tục nhau, mỗi khối có thể chứa 2 bản ghi tương ứng hệ số phân khối $bfr = 2$ bản ghi/khối. Vùng tin băm là ID và hàm băm là $h(ID) = ID \bmod 4$ với xâu riêng (*chaining*) dùng để giải quyết đụng độ. Nếu tìm kiếm bản ghi có ID = 45 thì bao nhiêu truy đạt khối cần được thực hiện?

a. 1 b. 2 c. 3 d. 4

e. Ý kiến khác.

Câu 10. Chọn phát biểu **đúng** về sự khác biệt giữa *băm tĩnh* và *băm động*.

a. Số lượng thùng trong băm động được thay đổi tùy theo nội dung dữ liệu của tập tin.

b. Số lượng bản ghi tối đa trong mỗi thùng của băm động thay đổi tùy theo nội dung dữ liệu của tập tin.

c. Thay đổi tập tin băm với băm tĩnh được thực hiện cục bộ và với băm động được thực hiện toàn cục.

d. Câu a, b, và c đều đúng.

e. Ý kiến khác.

Câu 11. Giả sử tập tin dữ liệu Nhân_Viên trong **Bảng 1** với các giả thiết được mô tả ở **Câu 4**. Chỉ mục được định nghĩa trên vùng tin Lương được gọi là gì?

a. Chỉ mục sơ cấp c. Chỉ mục thứ cấp

b. Chỉ mục cụm d. Vùng tin Lương không thể được dùng cho chỉ mục.

e. Ý kiến khác.

Câu 12. Giả sử tập tin dữ liệu Nhân_Viên trong **Bảng 1** với các giả thiết được mô tả ở **Câu 4**. Chỉ mục đơn mức có thứ tự được định nghĩa trên vùng tin Tên. Giả sử vùng tin Tên có các giá trị phân biệt và không được dùng để sắp thứ tự các bản ghi. Xác định số lượng bản ghi chỉ mục của tập tin chỉ mục này.

- a. 10
- b. 19
- c. 20
- d. Vùng tin Tên không thể được dùng cho chỉ mục.
- e. Ý kiến khác.

II. Phần tự luận:

Câu 13. Cho tập tin dữ liệu Nhân_Viên có 10 bản ghi ở **Bảng 2**. Tập tin dữ liệu này được lưu trữ trên bộ đĩa từ dùng kỹ thuật băm khả mở (*extendible hashing*) với hàm băm là $h(K) = K \bmod 16$, mỗi thùng gồm 1 khối dữ liệu và mỗi khối dữ liệu chứa được 2 bản ghi.

13.a. Thực hiện thêm các bản ghi theo thứ tự từ trên xuống dưới của tập tin dữ liệu Nhân_Viên ở **Bảng 2**. Trình bày các bước cập nhật tập tin băm tương ứng. (1.5 điểm)

13.b. Xác định *số truy đạt khối* cần được thực hiện để xóa bản ghi tương ứng của nhân viên có ID = 70 trên tập tin băm này. Giả sử thư mục được lưu trữ trên 1 khối dữ liệu riêng, khác với các thùng. (0.5 điểm)

Câu 14. Cho tập tin dữ liệu Nhân_Viên có 10 bản ghi, không được sắp thứ tự vật lý ở **Bảng 2**, xây dựng chỉ mục đa mức động B+-tree trên vùng tin ID. Đây là vùng tin khóa của Nhân_Viên.

14.a. Giả sử B+-tree này có bậc của nút nội $p = 4$ và bậc của nút lá $p_{leaf} = 3$. Trình bày các bước cập nhật chỉ mục lần lượt theo thứ tự từ trên xuống dưới của tập tin dữ liệu Nhân_Viên. (1.5 điểm)

14.b. Trình bày cập nhật của chỉ mục B+-tree và tập tin dữ liệu Nhân_Viên tương ứng khi câu lệnh UPDATE được thực thi thành công. (1 điểm)

UPDATE Nhân_Viên
SET ID = 33
WHERE ID = 70;

14.c. Cho kích thước 1 khối là $B = 512$ bytes, kích thước của vùng tin ID là $V = 4$ bytes, kích thước của con trỏ khối là $P = 6$ bytes, kích thước của con trỏ dữ liệu chỉ đến bản ghi dữ liệu là $Pr = 7$ bytes. Mỗi nút trong chỉ mục B+-tree được giữ đầy tối đa là 75%. Nếu tập tin có 500,000 bản ghi thì *chỉ mục B+-tree tương ứng có bậc p và p_{leaf} là bao nhiêu và phải có ít nhất bao nhiêu mức (tính luôn mức gốc (root) và mức lá (leaf))* để chỉ mục được các bản ghi? Trình bày các bước tính toán. (1.5 điểm)

Bảng 2 - Tập tin dữ liệu Nhân_Viên (câu 13, 14)

ID	Tên	Lương	Phòng
2	An	2000	1
17	Cường	1900	3
21	Mai	2500	1
9	Hạ	2500	2
70	Tú	1800	4
87	Bắc	3100	3
14	Hoa	1800	2
36	Thu	2300	3
93	Châu	3000	2
18	Xuân	1600	4

Ngày:/03/2019

Bộ môn/Khoa	Giảng viên ra đề

Họ - Tên: Mã Số Sinh Viên:

Mã Đề: 102

Phần trả lời

I. Phần trắc nghiệm:

Câu	1	2	3	4	5	6	7	8	9	10	11	12
a												
b												
c												
d												
e												

II. Phần tự luận:

Họ - Tên: Mã Số Sinh Viên:

Mã Đề: **102**

Họ - Tên: Mã Số Sinh Viên:

Mã Đề: **102**

Mã Đề: 102

8