

Kiểm tra giữa kỳ

Môn: **Hệ Quản Trị Cơ Sở Dữ Liệu (CO3021)** - Ngành: **Khoa Học Máy Tính** - HK: 1, 2018-2019

Ngày kiểm tra: **17.10.2018** - Thời gian làm bài: **60** phút

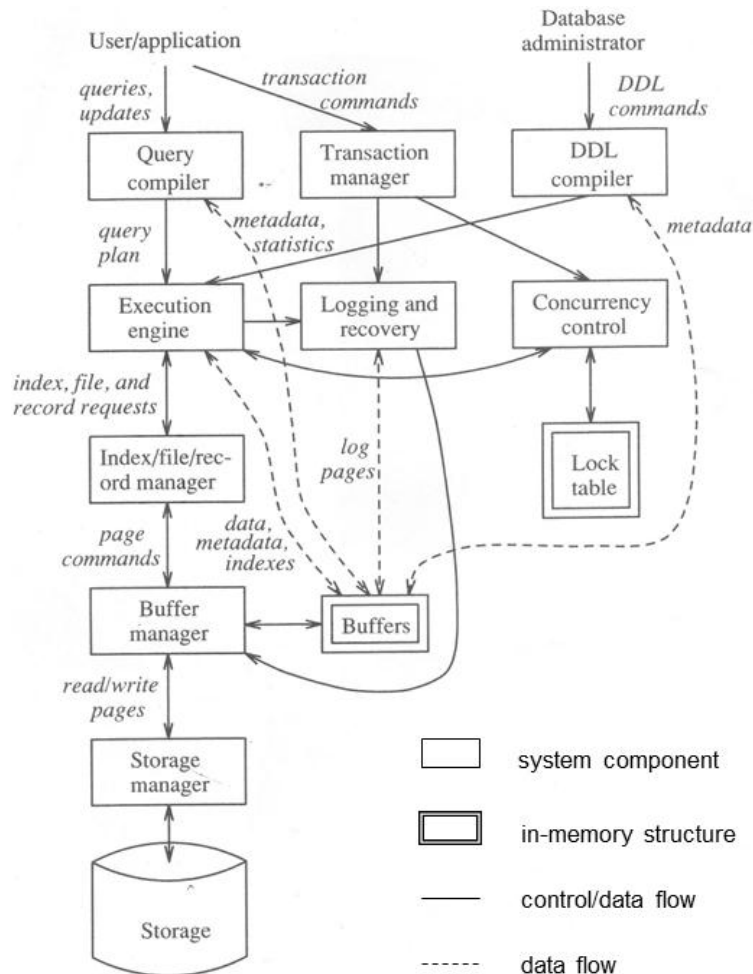
(Bài kiểm tra gồm 12 câu hỏi trắc nghiệm và 3 câu tự luận. Sinh viên được **tham khảo tài liệu**.)

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 các câu lệnh sau, câu lệnh nào được gửi đến hệ quản trị cơ sở dữ liệu (*database management system, DBMS*) bởi nhà quản trị cơ sở dữ liệu (*database administrator, DBA*)?

- a. **SELECT * FROM Nhân_Viên WHERE Tên = "A";**
- b. **UPDATE Nhân_Viên SET Tên = "AA" WHERE Tên = "A";**
- c. **CREATE INDEX index_Tên ON Nhân_Viên (Tên);**
- d. **COMMIT;**
- e. Ý kiến khác.



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

Câu 2. Cho kiến trúc của DBMS trong **Hình 1**. Mô-đun *Concurrency Control* tương tác với mô-đun *Execution Engine* nhằm :

- a. xác định trình tự truy cập dữ liệu cho các câu lệnh truy vấn và thay đổi dữ liệu
- b. xác định trình tự truy cập dữ liệu cho các câu lệnh truy vấn dữ liệu
- c. xác định trình tự thực thi cho các câu lệnh truy vấn dữ liệu
- d. xác định trình tự thực thi cho các câu lệnh thay đổi dữ liệu
- e. Ý kiến khác

Câu 3. Cho kiến trúc của một DBMS trong **Hình 1**. Mô-đun *Logging and Recovery Manager* đảm bảo đặc tính gì của các giao tác ?

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

Câu 4. Cho tập tin dữ liệu Nhân_Viên gồm 18 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. 264 bytes
- b. 378 bytes
- c. 396 bytes
- d. 756 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	Hạ	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

Câu 5. Giả sử tập tin dữ liệu Nhân_Viên được cập nhật và hiện tại có 27 khối dữ liệu (27 blocks) trong **Bảng 1** với các giả thiết được mô tả ở **Câu 4**. Các khối dữ liệu này được định vị liên tục nhau trên cùng rãnh dữ liệu (track). Kỹ thuật bộ đệm đôi (double buffering) giúp tiết kiệm thời gian gì sau đây?

- a. Xử lý của CPU b. Tìm kiếm và trở quay c. Tìm kiếm, trở quay, và truyền d. Tìm kiếm (seek)
e. Ý kiến khác.

Câu 6. Giả sử tập tin dữ liệu Nhân_Viên được cập nhật và hiện tại có 27 khối dữ liệu (27 blocks) trong **Bảng 1** với các giả thiết được mô tả ở **Câu 4**. Xác định **số truy đạt khối trung bình** (the average number of block accesses) có thể để thực hiện việc tìm kiếm bản ghi của nhân viên có ID $\neq 70$ trên tập tin dữ liệu Nhân_Viên.

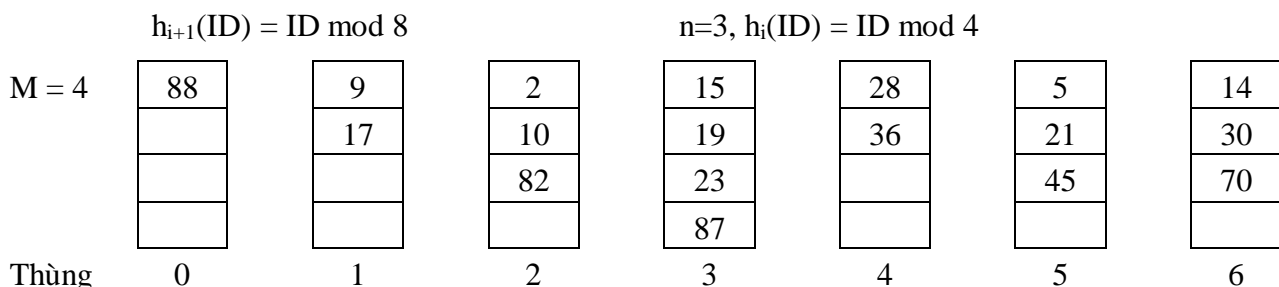
- a. 27 b. 14 c. 9 d. 5
e. Ý kiến khác.

Câu 7. Cho tập tin dữ liệu Nhân_Viên ở **Bảng 1** được lưu trữ như phần giả thiết ở **Câu 4**. Các khối dữ liệu này được định vị liên tục nhau trên cùng rãnh dữ liệu (track). Giả sử có nhân viên mới với ID = 25, Tên = Mới, Lương = 2000, ... Bản ghi tương ứng của nhân viên này sẽ được đặt ở đâu sao cho **đúng và hiệu quả nhất**?

- a. Tập tin tràn tương ứng của tập tin dữ liệu Nhân_Viên.
b. Khối sau cùng, ngay sau khối của các bản ghi với ID = 87 và ID = 88.
c. Khối giữa, ngay sau khối của các bản ghi với ID = 21 và ID = 23, và ngay trước khối của các bản ghi với ID = 28 và ID = 30.
d. Khối giữa, ngay sau khối của các bản ghi với ID = 21 và ID = 23, và chung khối của bản ghi với ID = 28.
e. Ý kiến khác.

Câu 8. Cho tập tin dữ liệu Nhân_Viên ở **Bảng 1** với các bản ghi như phần giả thiết ở **Câu 4**. Giả sử **Hình 2** minh họa các bản ghi của tập tin này được tổ chức theo kỹ thuật băm tuyến tính (linear 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 $h(ID) = ID \bmod 2^k$ với $k = 1, 2, 3, 4, 5, 6, \dots$. Nếu tìm kiếm bản ghi có ID = 20 thì hàm băm nào được sử dụng với bao nhiêu truy đạt khối?

- a. $h_i(ID) = ID \bmod 4$, $h_{i+1}(ID) = ID \bmod 8$, 2 truy đạt khối
b. $h_{i+1}(ID) = ID \bmod 8$, 2 truy đạt khối
c. $h_{i+1}(ID) = ID \bmod 8$, 1 truy đạt khối
d. $h_i(ID) = ID \bmod 4$, $h_{i+1}(ID) = ID \bmod 8$, 1 truy đạt khối
e. Ý kiến khác.



Hình 2 – Tổ chức băm tuyến tính của tập tin dữ liệu Nhân_Viên

Câu 9. Cho tập tin dữ liệu Nhân_Viên ở **Bảng 1** với các bản ghi như phần giả thiết ở **Câu 4**. Giả sử các bản ghi của tập tin này được tổ chức theo kỹ thuật băm tuyến tính (linear 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 $h(ID) = ID \bmod 2^k$ với $k = 1, 2, 3, 4, 5, 6, \dots$. Nếu tìm kiếm các bản ghi có ID > 20 thì hàm băm nào được sử dụng?

- a. Không vì tìm kiếm nhị phân cho điều kiện này
- b. $h_i(ID) = ID \bmod 4$
- c. Không vì tìm kiếm tuần tự cho điều kiện này
- d. $h_i(ID) = ID \bmod 4$ và $h_{i+1}(ID) = ID \bmod 8$
- e. Ý kiến khác.

Câu 10. Giả sử tập tin dữ liệu Nhân_Viên được cập nhật và hiện tại có 1580 khối dữ liệu (1580 *blocks*) với mỗi khối chứa đầy 2 bản ghi 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 kích thước của con trỏ khối $P_B = 6$ bytes. Xác định mức độ hiệu quả của chỉ mục khi tìm kiếm bản ghi với Tên = Tâm trên tập tin này thông qua chỉ mục so với tìm kiếm tuần tự trực tiếp trên tập tin này.

- a. Giảm chi phí khoảng 47 lần
- b. Giảm chi phí khoảng 99 lần
- c. Giảm chi phí khoảng 113 lần
- d. Giảm chi phí khoảng 226 lần
- e. Ý kiến khác.

Câu 11. Chọn phát biểu **đúng** về sự khác biệt giữa cấu trúc chỉ mục đa mức động B-tree và B+-tree.

- a. Nút lá của B+-tree tốn nhiều không gian lưu trữ hơn nút lá của B-tree.
- b. Nút nội của B-tree tốn nhiều không gian lưu trữ hơn nút nội của B+-tree.
- c. Số lượng giá trị được chỉ mục của B+-tree nhiều hơn số lượng giá trị được chỉ mục của B-tree trong cùng điều kiện chỉ mục.
- d. Khả năng chỉ mục của B+-tree cao hơn khả năng chỉ mục của B-tree trong cùng điều kiện chỉ mục.
- e. Ý kiến khác.

Câu 12. Cho tập tin dữ liệu Nhân_Viên ở **Bảng 1** với các bản ghi như phần giả thiết ở **Câu 4**. Giả sử các bản ghi của tập tin này được tổ chức theo kỹ thuật băm tuyến tính (*linear 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 $h(ID) = ID \bmod 2^k$ với $k = 1, 2, 3, 4, \dots$. Chọn phát biểu **đúng** về đặc điểm chỉ mục trên tập tin này.

- a. Có thể định nghĩa duy nhất 1 chỉ mục sơ cấp trên vùng tin ID và nhiều chỉ mục thứ cấp trên các vùng tin còn lại.
- b. Có thể định nghĩa duy nhất 1 chỉ mục cụm trên vùng tin Lương và nhiều chỉ mục thứ cấp trên các vùng tin còn lại.
- c. Không thể định nghĩa chỉ mục sơ cấp hay chỉ mục cụm vì không có vùng tin sắp thứ tự; nhưng có thể định nghĩa chỉ mục thứ cấp trên bất kỳ vùng tin nào.
- d. Không thể định nghĩa chỉ mục sơ cấp vì vùng tin ID không là vùng tin sắp thứ tự; nhưng có thể định nghĩa chỉ mục cụm hay chỉ mục thứ cấp trên bất kỳ vùng tin nào.
- e. Ý kiến khác.

II. Phần tự luận:

Câu 13. Giả sử tập tin dữ liệu Nhân_Viên có 9 khối dữ liệu 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 khối dữ liệu (*block size*) là $B = 512$ bytes. 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.

13.a. Nếu các khối dữ liệu của tập tin dữ liệu này được tổ chức trên cùng một *cylinder* thì **số truy đạt khối là bao nhiêu** và **tổng thời gian trung bình** thực hiện chuyển các khối dữ liệu vào vùng nhớ chính để tìm thấy nhân viên có ID = 10 là **bao nhiêu msec?** (1.25 điểm)

13.b. Nếu các khối dữ liệu của tập tin dữ liệu này được tổ chức trên cùng một *cylinder* thì **số truy đạt khối là bao nhiêu** và **tổng thời gian trung bình** thực hiện chuyển các khối dữ liệu vào vùng nhớ chính để

tìm thấy nhân viên có $ID \geq 10$ là ***bao nhiêu msec***? Giả sử có phân nửa số bản ghi của tập tin thỏa điều kiện $ID \geq 10$. (1.25 điểm)

Câu 14. Với tập tin dữ liệu Nhân_Viên có 18 bản ghi ở **Bảng 1** với các giả thiết được mô tả ở **Câu 4**, xây dựng chỉ mục đa mức động B+-tree trên vùng tin ID.

14.a. Gọi tên và mô tả các đặc điểm của chỉ mục này. (0.75 điểm)

14.b. Giả sử B+-tree này có bậc của nút nội $p = 3$ và bậc của nút lá $p_{leaf} = 2$. Trình bày các bước cập nhật chỉ mục lần lượt theo thứ tự của vùng tin ID. (1.5 điểm)

14.c. Tính **số truy đạt khối** và **thời gian** để tìm thấy nhân viên có $ID = 10$ thông qua chỉ mục B+-tree? Giả sử chỉ mục được lưu trữ trên cùng một *cylinder* với tập tin dữ liệu với bộ đĩa ở **câu 13** (1.25 điểm)

14.d. Giả sử câu lệnh SQL sau được thực thi trên tập tin dữ liệu Nhân_Viên:

```
INSERT INTO Nhân_Viên
VALUES (40, 'NhânViênMới', 3000, 1, 'Quá trình công tác');
```

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

Ngày:/10/2018

Bộ môn/Khoa	Giảng viên ra đề

Mã Số Sinh Viên:

Họ - Tên:

Mã Số Sinh Viên:

