

Kiểm tra giữa kỳ

Môn: **Hệ Quản Trị Cơ Sở Dữ Liệu (CO3021)** - Ngành: **Khoa Học Máy Tính** - HK: 1, 2019-2020

Ngày kiểm tra: **25.10.2019** - Thời gian làm bài: **60** phút

(Bài kiểm tra gồm 12 câu hỏi trắc nghiệm và 2 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 DBMS, mô-đun *Query Compiler* thực hiện thao tác nào sau đây ? DML = Data Manipulation Language, ngôn ngữ thao tác dữ liệu.

- a. Sinh mã cho các câu lệnh SQL của người dùng
- b. **Sinh kế hoạch thực thi tối ưu của câu lệnh DML**
- c. Sinh lịch biểu tối ưu của các câu lệnh DML
- d. Sinh trình kiểm tra ngữ nghĩa của câu lệnh DML
- e. Ý kiến khác.

Câu 2. Hệ quản trị cơ sở dữ liệu (*database management system*, DBMS) hỗ trợ ...

- a. Quản lý dữ liệu và mã của các câu lệnh SQL trên dữ liệu hiệu quả và bền vững theo thời gian
- b. Quản lý dữ liệu và mã của các câu lệnh của chương trình yêu cầu dữ liệu bền vững theo thời gian
- c. **Quản lý dữ liệu hiệu quả và bền vững theo thời gian**
- d. Quản lý siêu dữ liệu, các trang log, địa chỉ dữ liệu, và thông tin thống kê bền vững theo thời gian
- e. Ý kiến khác.

Câu 3. Trong kiến trúc của DBMS, mô-đun *Execution Engine* có kết nối đến mô-đun *Logging and Recovery Manager* nhằm ...

- a. **ghi nhật ký các tương tác được thực hiện trên cơ sở dữ liệu**
- b. ghi nhật ký trình tự thực thi của các câu lệnh của chương trình yêu cầu dữ liệu
- c. ghi nhật ký các kết nối đến cơ sở dữ liệu của các người dùng yêu cầu dữ liệu
- d. ghi nhật ký cho các nội dung ở câu a, câu b, và câu c.
- 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 trong **Bảng 1**. Vùng tin ID là vùng tin khóa (*key field*); các vùng tin còn lại là vùng tin phi khóa (*non-key field*). Ngoài ra, các bản ghi 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 của vùng tin ID là 9 bytes, vùng tin Tên là 50 bytes, vùng tin Ngày_sinh là 9 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 tập tin này được lưu trữ vật lý trên bộ đĩa từ với kích thước một khối dữ liệu (*block size*) là **B** = 512 bytes. Các bản ghi không phù hợp dùng cho tập tin này. Xác định **kích thước của mỗi bản ghi trong tập tin** này.

- a. 512 bytes
- b. 256 bytes
- c. **228 bytes**
- d. 227 bytes
- e. Ý kiến khác.

Câu 5. Giả sử tập tin dữ liệu Nhân_Viên được cập nhật với bản ghi mới có ID = 10. Xác định **khối dữ liệu (block)** chứa bản ghi mới này. Giả sử block đầu tiên được tính là block 1.

- a. block 3
- b. block 10
- c. **block của tập tin tràn**
- d. Không xác định được
- e. Ý kiến khác.

Câu 6. Giả sử tập tin Nhân_Viên được lưu trữ trên bộ đĩa từ có số vòng quay trong 1 phút là $p = 5600$ rpm, có thời gian tìm kiếm trung bình (*average seek time*) là $s = 10$ msec. Kích thước một rãnh dữ liệu (*track size*) là $T = 102400$ bytes. Xác định thời gian dịch chuyển 1 khối của tập tin này giữa bộ nhớ thứ cấp và bộ nhớ chính? Giả sử các tính toán được làm tròn với 2 số thập phân.

- a. 0.05 msec b. 5.36 msec c. 15.36 msec **d. 15.41 msec**
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	Ngày_sinh	Phòng	Quá_trình_công_tác	Deletion_marker
2	R&D	...	0
5	Marketing	...	0
7	Finance	...	0
9	R&D	...	0
12	Logistics	...	0
16	Marketing	...	0
18	Marketing	...	0
20	Finance	...	0
21	R&D	...	0
23	R&D	...	0
28	Logistics	...	0
30	Logistics	...	0
36	Sales	...	0
45	Marketing	...	0
70	Finance	...	0
82	Sales	...	0
87	Sales	...	0
98	Marketing	...	0

Câu 7. Cho tập tin dữ liệu Nhân_Viên ở **Bảng 1** với các bản ghi như phần giả thiết ở **Câu 4**. Nếu tìm kiếm các bản ghi có ID = 70 thì trung bình cần thực hiện bao nhiêu truy đạt khối?

- a. 4 truy đạt khối b. 5 truy đạt khối c. 9 truy đạt khối **d. 18 truy đạt khối**
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**. Nếu tìm kiếm các bản ghi có Phòng = "R&D" thì trung bình cần thực hiện bao nhiêu truy đạt khối?

- a. 4 truy đạt khối b. 5 truy đạt khối **c. 9 truy đạt khối** d. 18 truy đạt khối
e. Ý kiến khác.

Câu 9. Nếu tập tin này được tái tổ chức dạng tập tin băm tĩnh với vùng tin băm là ID, hàm băm là $h(ID) = ID \bmod 10$, kích thước mỗi thùng gồm 1 block với hệ số phân khối là 2 bản ghi/block, và dùng xâu riêng (*chaining*) để giải quyết đụng độ cho mỗi thùng, thì bản ghi mới có ID = 10 sẽ được lưu trữ ở đâu?

- a. Thùng 0 **b. Vùng tràn của thùng 0** c. Thùng 10 d. Vùng tràn của thùng 10
e. Ý kiến khác.

Câu 10. Cho tập tin dữ liệu Nhân_Viên ở **Bảng 1** với các bản ghi như phần giả thiết ở **Câu 4**. Chỉ mục được tạo trên vùng tin ID được gọi là gì?

- a. Chỉ mục sơ cấp** b. Chỉ mục cụm c. Chỉ mục thứ cấp d. Chưa thể xác định
e. Ý kiến khác.

Câu 11. Cho tập tin dữ liệu Nhân_Viên ở **Bảng 1** với các bản ghi như phần giả thiết ở **Câu 4**. Chỉ mục B-tree được tạo trên vùng tin Phòng được gọi là gì?

- a. Chỉ mục cụm động b. Chỉ mục cụm tĩnh c. Chỉ mục thứ cấp động d. Chỉ mục thứ cấp tĩnh
e. Ý kiến khác.

Câu 12. Cấu trúc chỉ mục B+-tree được cải tiến gì so với cấu trúc chỉ mục B-tree?

- a. Các nút nội không chứa các con trỏ dữ liệu. c. Các nút lá không chứa các con trỏ cây.
b. Các nút lá được liên kết với nhau. d. Câu a, b, và c đều đúng.
e. Ý kiến khác.

II. Phần tự luận:

Bảng 2. Tập tin Môn_học.

Code	Name	...	Credit
1011	Programming	...	4
1006	Computing	...	4
2007	Architectures	...	4
2013	Database	...	4
2001	Communication	...	2
3001	SE	...	4
3031	Algorithms	...	3
2003	Data Structures	...	4
2005	OOP	...	4
3055	Practice on SE	...	2

Câu 13. Giả sử tập tin Môn_học là tập tin không có thứ tự (*unordered file, heap file*) gồm 10 bản ghi có 5 khối dữ liệu trong **Bảng 2** với hệ số phân khối là 2 bản ghi/khối. Vùng tin Code là vùng tin khóa chính.

13.a. Xác định số truy đạt khối trung bình khi tìm các bản ghi có Code = 3016. Nếu các bản ghi của tập tin được sắp thứ tự theo các giá trị của vùng tin Code thì số truy đạt khối trung bình khi tìm các bản ghi có Code = 3016 là bao nhiêu? So sánh kết quả của trường hợp khi tập tin Môn_học có thứ tự với kết quả của trường hợp khi tập tin Môn_học không có thứ tự. (1.5 điểm)

13.b. Mô tả quá trình cập nhật tổ chức lưu trữ của tập tin này khi thêm bản ghi mới cho môn học DBMS có Code = 2011 với câu lệnh INSERT như sau. (1 điểm)

INSERT INTO Môn_học

VALUES (2011, 'Modeling', '...', 3);

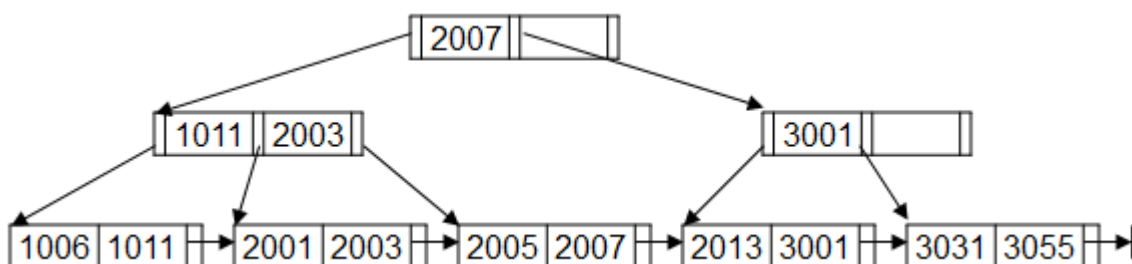
Câu 14. Giả sử chỉ mục B+-tree có bậc của nút nội $p = 3$ và bậc của nút lá $p_{leaf} = 2$ ở **Hình 1** được tạo trên vùng tin Code của tập tin Môn_học trong **Bảng 2** ở **Câu 13**.

14.a. Chỉ mục này có phải là chỉ mục dày hay không? Neo khối (block anchor) có được sử dụng trong trường hợp này không? Giải thích. (1.0 điểm)

14.b. Có bao nhiêu khối trên đĩa cứng cần dùng để lưu trữ chỉ mục này? (0.5 điểm)

14.c. Xác định số truy đạt khối khi tìm các bản ghi có Code = 3016 với chỉ mục này. So sánh kết quả của trường hợp này với kết quả của trường hợp khi tìm trực tiếp trên tập tin không có thứ tự. (1.0 điểm)

14.d. Mô tả quá trình cập nhật chỉ mục này khi **Câu 13.b** được thực thi. (1.0 điểm)



Hình 1. B+-tree có bậc $p=3$ và $p_{leaf}=2$ trên vùng tin Code của tập tin Môn_học.

Đề 14

Ngày:/10/2019

Bộ môn/Khoa	Giảng viên ra đề

Bảng ánh xạ Câu hỏi-Chuẩn đầu ra môn học:

Câu hỏi	Chuẩn đầu ra môn học
1	LO.1.2
2	LO.1.1
3	LO.1.2
4	LO.2.1
5	LO.2.1
6	LO.2.1
7	LO.2.1
8	LO.2.1
9	LO.2.1

Câu hỏi	Chuẩn đầu ra môn học
10	LO.2.2
11	LO.2.2
12	LO.2.2
13.a	LO.2.1
13.b	LO.2.1
14.a	LO.2.2
14.b	LO.2.2
14.c	LO.2.2
14.d	LO.2.2

Họ - Tên:

Mã Số Sinh Viên: Mã đề: 14

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:

Câu 13.a.

Khi tìm các bản ghi có Code = 3016, không có kết quả trả về. Đây là trường hợp không tìm thấy bằng cách tìm kiếm tuần tự (*linear search*) trên vùng tin khóa chính không được sắp thứ tự. Do đó, số truy đạt khối trung bình = $b = 5$.

Nếu các bản ghi được sắp thứ tự dựa trên các giá trị của vùng tin khóa chính thì tìm kiếm trên vùng tin này được thực hiện bằng cách tìm kiếm nhị phân (*binary search*). Do đó, số truy đạt khối trung bình = $\log_2 b = 3$.

So sánh: số truy đạt khối trung bình của trường hợp có thứ tự ít hơn khoảng 2 lần so với số truy đạt khối trung bình của trường hợp không có thứ tự. Do đó, tìm kiếm “=” trên vùng tin Code sẽ hiệu quả hơn khi tập tin có thứ tự trên vùng tin này.

Câu 13.b.

Do đây là tập tin không có thứ tự nên khi thêm bản ghi mới, bản ghi mới được thêm trực tiếp vào tập tin này. Giả sử việc thêm vào được thực hiện ở cuối tập tin.

Khi này, block cuối gồm 2 bản ghi (2005, 3055) được đưa vào vùng đệm. Do khối này đã đầy nên một khối mới được hình thành và trở thành block cuối của tập tin. Bản ghi mới được ghi vào block này. Sau đó, 2 block gồm block (2005, 3055) và block (2011) được ghi xuống đĩa cho tập tin.

Câu 14.a.

Do Code là vùng tin khóa chính và không có thứ tự nên chỉ mục được tạo trên vùng tin này là chỉ mục thứ cấp. Đây là dạng chỉ mục dày do mỗi bản ghi đều được chỉ mục.

Trong trường hợp này, neo khối không được sử dụng vì mỗi bản ghi được chỉ mục chứ không phải nhóm bản ghi được chỉ mục.

Câu 14.b. Chỉ mục gồm 8 nút. Mỗi nút được lưu trữ trong 1 block. Do đó, chỉ mục này được lưu trữ trong 8 block trên đĩa cứng.

Họ - Tên:

Mã Số Sinh Viên: Mã đề: 14

Câu 14.c. Quá trình tìm kiếm với Code = 3016 trên chỉ mục được thực hiện như sau:

Truy cập nút gốc (2007) và đi về nhánh phải.

Truy cập nút (3001) và đi về nhánh phải.

Truy cập nút lá (3031, 3055) và kiểm tra. Do trị khóa 3016 không có ở nút lá nên kết quả tìm kiếm trả về là không tìm thấy.

Khi này, số truy đạt khối = 3.

Kết quả của trường hợp này ít hơn khoảng 2 lần so với trường hợp khi tìm trực tiếp trên tập tin không có thứ tự. Do đó, tìm kiếm “=” trên vùng tin khóa không có thứ tự được chỉ mục hiệu quả hơn so với tìm trực tiếp bằng cách tìm kiếm tuần tự.

Câu 14.d. Cập nhật chỉ mục cho trị khóa 2011.

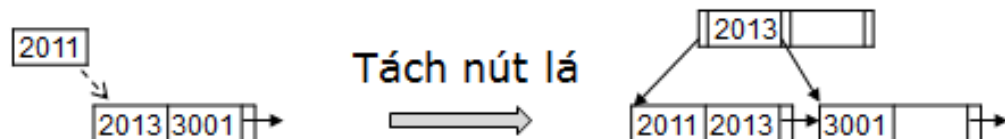
Quá trình định vị nút lá cho trị khóa 2011:

Truy cập nút gốc (2007) và đi về nhánh phải.

Truy cập nút (3001) và đi về nhánh trái.

Truy cập nút lá (2013, 3001) và kiểm tra. Trị khóa 2011 sẽ được thêm vào ở nút lá này nhưng nút này đã đầy nên hiện tượng tràn trên xảy ra.

Quá trình thêm vào:



Trị khóa 2013 được thêm vào ở nút cha (3001, _). Khi này chỉ mục được cập nhật lại như sau:

