Trường Đại Học Bách Khoa - Tp. HCM
Khoa Khoa Học & Kỹ Thuật Máy Tính
***

Mã đề: **11**

Đề thi cuối kỳ

Môn: **Hệ Quản Trị Cơ Sở Dữ Liệu (503004)**

Ngành: Khoa Học Máy Tính – HK2 – 2011-2012

Thời gian làm bài: **120** phút

(*Bài thi gồm **45** câu hỏi. Sinh viên được **tham khảo ghi chú** trong **2 tờ giấy A4**.*)

Sinh viên chọn **1** câu trả lời đúng nhất. Nếu chọn câu (e) 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.

Câu 1. Hệ quản trị cơ sở dữ liệu quan hệ (*relational database management system*) ***khác*** hệ quản trị cơ sở dữ liệu XML (*XML database management system*) ở đặc điểm gì sau đây?

a. Hệ quản trị cơ sở dữ liệu quan hệ là phần mềm chuyên dụng hỗ trợ tạo, quản lý, bảo trì bền vững cho nhiều cơ sở dữ liệu có kích thước lớn; trong khi đó, hệ quản trị cơ sở dữ liệu XML hỗ trợ cho những cơ sở dữ liệu có kích thước tương đối nhỏ.

b. Hệ quản trị cơ sở dữ liệu quan hệ là hệ quản trị cơ sở dữ liệu hỗ trợ các cơ sở dữ liệu quan hệ; trong khi đó, hệ quản trị cơ sở dữ liệu XML hỗ trợ các cơ sở dữ liệu XML.

c. Hệ quản trị cơ sở dữ liệu quan hệ cung cấp nhiều tiện ích cho việc xử lý và tối ưu hóa truy vấn cơ sở dữ liệu quan hệ; trong khi đó, hệ quản trị cơ sở dữ liệu XML chủ yếu hỗ trợ việc trao đổi dữ liệu với định dạng XML.

d. Hệ quản trị cơ sở dữ liệu quan hệ giúp quản lý và xử lý giao tác của nhiều người dùng khác nhau trong môi trường đa người dùng; trong khi đó, hệ quản trị cơ sở dữ liệu XML giúp quản lý và xử lý giao tác trong môi trường đơn người dùng.

e. Ý kiến khác.

Câu 2. Các câu lệnh về định nghĩa dữ liệu (*data definition statement*) do ***ai*** gởi đến hệ quản trị cơ sở dữ liệu?

a. Người quản trị cơ sở dữ liệu (*database administrator* – DBA)

b. Các chương trình ứng dụng (*application program*) chạy trên hệ quản trị cơ sở dữ liệu

c. Người sở hữu dữ liệu (*data owner*)

d. Người thiết kế cơ sở dữ liệu (*database designer*)

e. Ý kiến khác.

Câu 3. Trong hệ quản trị cơ sở dữ liệu Oracle, người sử dụng có thể sử dụng **hints** (/*+ …. */) để dẫn hướng cho bộ tối ưu hóa truy vấn (*query optimization*) trong quá trình xử lý truy vấn. Khi đó, Oracle ***giả định điều gì*** sau đây?

a. Bộ tối ưu hóa truy vấn của Oracle không có đủ thông tin về các phương thức truy đạt trên dữ liệu của người sử dụng.

b. Quá trình xử lý truy vấn của Oracle được chuyển sang chế độ do người dùng chỉ định (*user-specified mode*).

c. Câu truy vấn có hints chưa được định nghĩa trước trên dữ liệu của người sử dụng.

d. Người sử dụng có nhiều thông tin hơn về tổ chức dữ liệu (*data organization*) và các phương thức truy đạt (*access method*) trên dữ liệu của người sử dụng so với bộ tối ưu hóa truy vấn của Oracle.

e. Ý kiến khác.

Câu 4. Việc tìm kiếm các bản ghi trong một tập tin dữ liệu với điều kiện tìm kiếm ">=" trên một vùng tin A1 của tập tin sẽ hiệu quả hơn khi tập tin được ***tổ chức ở dạng nào***?

a. Tập tin không có thứ tự theo các giá trị của vùng tin A1.

b. Tập tin có thứ tự theo các giá trị của vùng tin A1.

c. Tập tin có thứ tự theo các giá trị của vùng tin A2. A2 là một trong số những vùng tin khác của tập tin.

d. Tập tin băm theo các giá trị của vùng tin A1.

e. Ý kiến khác.

Câu 5. Xác định ***chi phí truy đạt khối trung bình*** trên tập tin dữ liệu băm (*hashed file*) khi điều kiện tìm kiếm trên vùng tin băm (*hashing field*) là "=". Cho biết **b** là số khối dữ liệu hiện có trong tập tin.

a. O(1)

b. O($\log_2$**b**)

c. O(ln**b**)

d. O(**b**)

e. Ý kiến khác.

Câu 6. Với kích thước khối (*block size*) B = 512 bytes, số lượng con trỏ cây p ở mỗi nút trong cây chỉ mục B-tree là 24 và số lượng con trỏ cây p ở mỗi nút nội trong cây chỉ mục B+-tree là 34 và ở mỗi nút lá trong cây chỉ mục B+-tree là 31. Chọn phát biểu **SAI** về sự ***khác biệt*** giữa B-tree và B+-tree?

a. Khả năng chỉ mục của B+-tree cao (hiệu quả) hơn B-tree với hệ số rẽ nhánh (*fan-out*) lớn hơn: 34 > 24.

b. Khả năng chỉ mục của B+-tree cao (hiệu quả) hơn B-tree với số mục (*entry*) tối đa của mỗi nút nội lớn hơn: 33 > 23.

c. Nếu hai cây B-tree và B+-tree này có cùng chiều cao thì tổng số con trỏ dữ liệu (*data pointer*) trong B+-tree lớn hơn nhiều so với trong B-tree.

d. Không thể so sánh được hai cây B-tree và B+-tree với phần mô tả ở trên.

e. Ý kiến khác.

Câu 7. Cho tập tin dữ liệu gồm 6 khối dữ liệu chứa các bản ghi được sắp thứ tự vật lý theo các giá trị của vùng tin khóa ID (*key field*) với hệ số phân khối (*blocking factor*) bfr = 2 bản ghi/khối. Một chỉ mục sơ cấp đa mức (*multilevel primary index*) dùng cấu trúc chỉ mục B-tree với bậc p = 3 được định nghĩa trên vùng tin ID. Xác định ***chỉ mục*** B-tree **ĐÚNG** trên tập tin dữ liệu này.

a. **Hình 1**

b. **Hình 2**

c. **Hình 3**

d. **Hình 4**

e. Ý kiến khác.

ID	....
2	
5	
8	
10	
11	
15	
18	
19	
21	
23	
28	

Tập tin dữ liệu gồm **6 blocks** với hệ số phân khối **bfr = 2 bản ghi/khối** và nội dung vùng tin khóa ID.

Câu 8. 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 9. Cho một chỉ mục thứ cấp đa mức (*multilevel secondary index*) trên thuộc tính khóa mã nhân viên EID của tập tin dữ liệu Employee dùng cấu trúc B+-tree trong **Hình 5**. Các bản ghi của tập tin được tổ chức theo cách phân khối không phủ (*unspanned blocking*) với hệ số phân khối (*blocking factor*) **bfr** = 3. Xác định ***số khối dữ liệu hiện có*** (*the current number of blocks*) trong tập tin dữ liệu Employee.

a. 3 khối

b. 4 khối

c. 6 khối

d. 11 khối

e. Ý kiến khác.

Câu 10. Xác định ***số khối dữ liệu hiện có*** trong chỉ mục thứ cấp đa mức B+-tree ở **Hình 5**.

a. 11 khối

b. 17 khối

c. 22 khối

d. 41 khối

e. Ý kiến khác.

Câu 11. A là nhân viên mới của tổ chức và có mã nhân viên là **2**. Một bản ghi về A được thêm vào tập tin Employee. Xác định ***đặc điểm thay đổi*** **ĐÚNG** của chỉ mục thứ cấp đa mức B+-tree ở **Hình 5** khi cập nhận thông tin chỉ mục của bản ghi về A.

a. Chỉ mục B+-tree này không thay đổi.

b. Chiều cao cây không đổi, nhưng số nút nội tăng 1 và số nút lá tăng 1.

c. Chiều cao cây, số nút nội, và số nút lá đều tăng 1.

d. Chiều cao cây không đổi, số nút nội không đổi, nhưng số nút lá tăng 1.

e. Ý kiến khác.

Câu 12. Cho công thức tính chi phí cho phép toán chọn (*selection*) dùng chỉ mục thứ cấp với cấu trúc chỉ mục đa mức động B+-tree trong trường hợp điều kiện chọn là điều kiện "=" trên thuộc tính được chỉ mục: C = x + s; trong đó, **x** là số mức của B+-tree, **s** là số lượng bản ghi trả về trong kết quả của phép chọn. ***Ý nghĩa*** của **x** và **s** trong công thức là gì?

a. **x** là chi phí duyệt một khối ở từng mức trên chỉ mục để đến được khối chỉ mục chứa các con trỏ dữ liệu tính theo số truy đạt khối (*block access*); **s** là chi phí chọn từ tập tin dữ liệu ra các khối dữ liệu chứa các bản ghi trong kết quả từ các con trỏ dữ liệu trong chỉ mục tính theo số truy đạt khối.

b. **x** là chi phí duyệt hết các khối ở từng mức trên chỉ mục để đến được khối chỉ mục chứa các con trỏ dữ liệu tính theo số truy đạt khối; **s** là chi phí chọn từ tập tin dữ liệu ra các khối dữ liệu chứa các bản ghi trong kết quả tính theo số truy đạt khối.

c. **x** là chi phí duyệt một số khối ở từng mức trên chỉ mục để đến được khối chỉ mục chứa các con trỏ dữ liệu tính theo số truy đạt khối; **s** là chi phí chọn từ tập tin dữ liệu ra các khối dữ liệu chứa các bản ghi trong kết quả tính theo số truy đạt khối.

d. **x** là chi phí duyệt các khối trên chỉ mục để đến được các khối chỉ mục chứa các con trỏ dữ liệu ở từng mức tính theo số truy đạt khối; **s** là chi phí chọn từ tập tin dữ liệu ra các khối dữ liệu chứa các bản ghi trong kết quả từ các con trỏ dữ liệu trong chỉ mục tính theo số truy đạt khối.

e. Ý kiến khác.

Câu 13. Cho công thức tính chi phí cho phép toán chọn (*selection*) dùng cây chỉ mục trong trường hợp điều kiện chọn là điều kiện ">=" trên thuộc tính khóa được chỉ mục: C = x + b/2; trong đó, **x** là số mức của cây chỉ mục, **b** là số khối trong tập tin dữ liệu. ***Giả định*** (*assumption*) gì được sử dụng trong công thức này?

a. Phân nửa số nút trong chỉ mục chứa các con trỏ dữ liệu chỉ đến các bản ghi trong kết quả của phép chọn.

b. Phân nửa số khối chỉ mục chính là số khối chứa các bản ghi trong kết quả của phép chọn.

c. Phân nửa số khối trong tập tin chứa các bản ghi thỏa điều kiện chọn.

d. Chỉ phân nửa số khối dữ liệu trong tập tin cần được kiểm tra điều kiện chọn.

e. Ý kiến khác.

Câu 14. Cho kích thước vùng đệm (*buffer*) là bf = 5 khối (*blocks*). Tập tin dữ liệu Employee gồm 16 khối dữ liệu (b_E = 16 blocks). Tập tin dữ liệu Department gồm 7 khối dữ liệu (b_D = 7 blocks). Thực hiện phép kết với điều kiện kết bằng giữa thuộc tính DNO của Employee và DNUMBER của Department: Employee $\bowtie_{DNO = DNUMBER}$ Department. Xác định ***tổng số truy đạt khối*** (*block accesses*) trên các tập tin dữ liệu nếu phép kết này được xử lý bằng phương pháp hai vòng lặp lồng (*nested-loop join*).

a. 112 block accesses nếu Department ở vòng lặp ngoài (*outer loop*).

b. 53 block accesses nếu Department ở vòng lặp ngoài.

c. 32 block accesses nếu Employee ở vòng lặp ngoài.

d. 23 block accesses nếu Employee ở vòng lặp ngoài.

e. Ý kiến khác.

Phần giả thiết sau được sử dụng cho các câu 15-17.

Cho lược đồ cơ sở dữ liệu gồm hai bảng EMPLOYEE (ứng với tập tin dữ liệu EMPLOYEE) và DEPARTMENT (ứng với tập tin dữ liệu DEPARTMENT) được định nghĩa dưới đây. Các thông tin cho việc truy đạt dữ liệu trong các tập tin dữ liệu EMPLOYEE và DEPARTMENT cũng được cho tương ứng.

```
CREATE TABLE EMPLOYEE
    ( FNAME        VARCHAR(15)      NOT NULL ,
      MINIT        CHAR ,
      LNAME        VARCHAR(15)      NOT NULL ,
      SSN          CHAR(9)          NOT NULL ,
      BDATE        DATE,
      ADDRESS      VARCHAR(30) ,
      SEX          CHAR ,
      SALARY       DECIMAL(10,2) ,
      SUPERSSN     CHAR(9) ,
      DNO          INT              NOT NULL ,
  PRIMARY KEY (SSN) ,
  FOREIGN KEY (SUPERSSN) REFERENCES EMPLOYEE(SSN) ,
  FOREIGN KEY (DNO) REFERENCES DEPARTMENT(DNUMBER) ) ;

CREATE TABLE DEPARTMENT
    ( DNAME        VARCHAR(15)      NOT NULL ,
      DNUMBER      INT              NOT NULL ,
      MGRSSN       CHAR(9)          NOT NULL ,
      MGRSTARTDATE DATE ,
  PRIMARY KEY (DNUMBER) ,
  UNIQUE (DNAME) ,
  FOREIGN KEY (MGRSSN) REFERENCES EMPLOYEE(SSN) ) ;
```

Tập tin EMPLOYEE có:

Số lượng bản ghi r_E = 5000 records.
Số lượng khối b_E = 1000 blocks.
Hệ số phân khối bfr_E = 5 records/block.

Trên thuộc tính khóa SSN, một chỉ mục thứ cấp (*secondary index*) được định nghĩa gồm 3 mức (x_{SSN} = 3) với lượng bản ghi được chọn trung bình là s_{SSN} = 1 và số lượng khối ở tầng cơ sở (tầng lá) là $bl1_{SSN}$ = 30.

Trên thuộc tính không khóa SALARY, một chỉ mục cụm (*clustering index*) được định nghĩa gồm 3 mức (x_{SALARY} = 3), số lượng khối ở tầng cơ sở (tầng lá) $bl1_{SALARY}$ = 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 thứ cấp (*secondary index*) được định nghĩa gồm 2 mức (x_{DNO} = 2), số lượng khối ở tầng cơ sở (tầng lá) $bl1_{DNO}$ = 4 blocks, 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 = 10 blocks.
Hệ số phân khối bfr_D = 8 records/block.

Trên thuộc tính khóa DNUMBER, một chỉ mục sơ cấp 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 tầng cơ sở là $bl1_{MGRSSN}$ = 2.

Cho trước các hàm tính chi phí của phép chọn và phép kết như sau:

Cost Functions for SELECT:

S1. Linear search (brute force) approach:

C_{S1a}= b

For an equality condition on a key, C_{S1b} = (b/2) if the record is found; otherwise C_{S1a}= b.

S2. Binary search:

$C_{S2} = \log_2 b + \lceil (s/bfr) \rceil - 1$

For an equality condition on a unique (key) attribute: $C_{S2} = \log_2 b$

S3. Using a primary index (S3a) or hash key (S3b) to retrieve a single record:

C_{S3a}= x + 1; C_{S3b} = 1 for static or linear hashing;

C_{S3b} = 2 for extendible hashing

S4. Using an ordering index to retrieve multiple records:

For the comparison condition on a key field with an ordering index: C_{S4}= x + (b/2)

S5. Using a clustering index to retrieve multiple records for an equality condition:

$C_{S5} = x + \lceil (s/bfr) \rceil$

S6. Using a secondary (B^+-tree) index:

For an equality comparison (=): C_{S6a}= x + s

For a comparison condition (>, <, >=, or <=):

C_{S6b}= x + (bl1/2) + (r/2)

Cost Functions for JOIN:

J1. Nested-loop join:

$C_{J1} = b_R + (b_R * b_S) + ((js * |R| * |S|)/bfr_{RS})$

(Use *R* for outer loop)

J2. Single-loop join(using an access structure to retrieve the matching record(s))

For a secondary index:

$C_{J2a} = b_R + (|R| * (x_B + s_B)) + ((js * |R| * |S|)/bfr_{RS})$

For a clustering index:

$C_{J2b} = b_R + (|R| * (x_B + (s_B/bfr_B))) + ((js * |R| * |S|)/bfr_{RS})$

For a primary index:

$C_{J2c} = b_R + (|R| * (x_B + 1)) + ((js * |R| * |S|)/bfr_{RS})$

If a hash key exists for one of the two join attributes — B of S:

$C_{J2d} = b_R + (|R| * h) + ((js * |R| * |S|)/bfr_{RS})$

(h: the average number of block accesses to retrieve a record, given its hash key value, h>=1)

J3. Sort-merge join:

$C_{J3a} = C_S + b_R + b_S + ((js * |R| * |S|)/bfr_{RS})$

(C_S: cost for sorting files)

Câu 15. Cho biểu thức truy vấn chọn ra những nhân viên có lương bằng 1000 và ở những phòng ban có mã lớn hơn 5: $\sigma_{SALARY=1000 \text{ AND } DNO>5}$(EMPLOYEE). Xác định ***chi phí của bản kế hoạch thực thi*** được chọn nếu dùng tối ưu hóa truy vấn dựa trên chi phí (*cost-based optimization*).

a. 7 block accesses

b. 20 block accesses

c. 1000 block accesses

d. 2000 block accesses

e. Ý kiến khác.

Câu 16. Cho biểu thức truy vấn chọn ra phòng ban được quản lý bởi nhân viên có mã '123456789': $\sigma_{MGRSSN='123456789'}$(DEPARTMENT). Dựa trên tối ưu hóa truy vấn dùng chi phí, ***kế hoạch thực thi*** (*execution plan*) nào sau đây là hợp lý nhất cho biểu thức truy vấn này?

a. Phương pháp tìm kiếm tuyến tính trên vùng tin MGRSSN với chi phí là 10 block accesses

b. Phương pháp chỉ mục thứ cấp trên vùng tin MGRSSN với chi phí là 2 block accesses

c. Phương pháp chỉ mục thứ cấp trên vùng tin MGRSSN với chi phí là 3 block accesses

d. Phương pháp chỉ mục thứ cấp trên vùng tin MGRSSN với chi phí là 43 block accesses

e. Ý kiến khác.

Câu 17. Cho biểu thức truy vấn kết giữa EMPLOYEE và DEPARTMENT: DEPARTMENT $\bowtie_{MGRSSN=SSN}$EMPLOYEE. Việc xử lý truy vấn kết này bằng phương pháp một vòng lặp (*single-loop join*) trên DEPARTMENT và chỉ mục thứ cấp trên vùng tin SSN của EMPLOYEE có ***tốt hơn*** việc xử lý bằng phương pháp sắp thứ tự-trộn (*sort-merge join*) hay không? Cho trước chi phí sắp thứ tự ngoại trên tập tin EMPLOYEE là 10000 block accesses và trên tập tin DEPARTMENT là 40 block accesses với vùng đệm có kích thước là 5; độ chọn lọc kết (*join selectivity*) dành cho biểu thức truy vấn trên là js = $1/r_E$ = 1/5000; hệ số phân khối của tập tin kết quả kết từ biểu thức truy vấn trên là bfr_{ED} = 4 records/block.

a. Phương pháp single-loop join tốt hơn.

b. Phương pháp sort-merge join tốt hơn.

c. Hai phương pháp có chi phí như nhau.

d. Không thể kết luận được phương pháp nào tốt hơn.

e. Ý kiến khác.

Câu 18. Một giao tác (*transaction*) ***không*** bị ngưng thực thi giữa chừng (*aborted*) trong ***tình huống*** nào?

a. Lỗi từ các tác vụ đọc/ghi dữ liệu trong giao tác xảy ra.

b. Lỗi tắt nguồn xảy ra.

c. Yêu cầu ngưng thực thi từ trình điều khiển tương tranh (concurrency controller) được thực hiện.

d. Lượng tác vụ đọc/ghi dữ liệu trong giao tác (*long transaction*) được thực hiện nhiều.

e. Ý kiến khác.

Câu 19. Cho các trình tự thực thi của các tác vụ trong các giao tác khác nhau. Xác định ***trình tự thực thi*** **ĐÚNG** là một ***lịch biểu*** (*schedule*).

Giao tác T1: $r_1(X); w_1(X); r_1(Y); w_1(Y); r_1(X)$

Giao tác T2: $r_2(Y); r_2(X); w_2(X); w_2(Y)$

a. Trình tự S1: $r_1(X); r_2(Y); w_2(Y); w_1(X); r_1(Y); r_2(X); w_2(X); w_1(Y); r_1(X)$

b. Trình tự S2: $r_1(X); w_1(X); r_2(Y); r_2(X); r_1(Y); w_1(Y); w_2(X); w_2(Y); r_1(X)$

c. Trình tự S3: $r_2(Y); r_2(X); r_1(X); r_1(Y); r_1(X); w_1(X); w_1(Y); w_2(X); w_2(Y)$

d. Trình tự S4: $r_2(Y)$; $w_2(Y)$; $r_1(X)$; $w_1(X)$; $r_1(X)$; $r_1(Y)$; $w_1(Y)$; $r_2(X)$; $w_2(X)$

e. Ý kiến khác.

Câu 20. Cho lịch biểu S5: $r_1(X)$; $r_2(Z)$; $r_1(Z)$; $r_3(X)$; $r_3(Y)$; $w_1(X)$; $w_3(Y)$; $r_2(Y)$; $w_2(Z)$; $w_2(Y)$; c_1; c_2; c_3. Xác định ***đặc điểm khả phục hồi*** (*recoverability*) của S5.

a. Khả phục hồi (*recoverable*)

b. Không dắt dây (*cascadeless*)

c. Nghiêm cách (*strict*)

d. Không khả phục hồi (*non-recoverable*)

e. Ý kiến khác.

Câu 21. Cho lịch biểu S6: $r_2(X)$; $r_1(Y)$; $w_2(X)$; $r_3(X)$; $w_1(Y)$; $w_3(X)$; $r_2(Y)$; $w_2(Y)$. Xác định ***đặc điểm khả tuần tự hóa*** (*serializability*) của S6.

a. S6 không khả tuần tự hóa (*non-serializable*)

b. S6 khả tuần tự hóa xung đột (*conflict serializable*)

c. Thông tin mô tả về lịch biểu S6 không đầy đủ nên không thể xác định được đặc điểm khả tuần tự hóa của S6.

d. S6 không có đặc điểm khả tuần tự hóa.

e. Ý kiến khác.

Câu 22. Lịch biểu S7 là một lịch biểu khả tuần tự hóa không dắt dây (*serializable and cascadeless schedule*). Chọn phát biểu **SAI** về lịch biểu S7.

a. Tồn tại một lịch biểu tuần tự tương đương (*equivalent serial schedule*) với lịch biểu S7.

b. Sổ ghi hệ thống (*system log*) không cần có mục tin (*entry*) về các tác vụ đọc (*read operation*) của các giao tác trong lịch biểu S7.

c. Khi một giao tác trong lịch biểu S7 bị ngưng thực thi (*aborted*), không có giao tác nào khác trong lịch biểu S7 cùng bị ngưng thực thi kéo theo.

d. Việc phục hồi dành cho lịch biểu S7 sẽ dễ dàng hơn so với các lịch biểu khả tuần tự hóa nghiêm cách (*serializable and strict schedule*).

e. Ý kiến khác.

Câu 23. Chọn phát biểu **ĐÚNG** về ***commit point*** của một giao tác T.

a. Điểm mà khi đó, tất cả các tác vụ của T đã được thực hiện thành công và những tác động của T lên cơ sở dữ liệu đã được ghi nhận trong sổ ghi hệ thống (*system log*).

b. Điểm mà khi đó, tất cả các tác vụ của T đã được thực hiện thành công và những tác động của T lên cơ sở dữ liệu đã được ghi nhận trong cơ sở dữ liệu.

c. Điểm mà khi đó, tất cả các tác vụ của T đã được tháo gỡ thành công và những tác động của T lên cơ sở dữ liệu đã được xóa khỏi sổ ghi hệ thống.

d. Điểm mà khi đó, tất cả các tác vụ của T đã được tháo gỡ thành công và những tác động của T lên cơ sở dữ liệu đã được xóa khỏi cơ sở dữ liệu.

e. Ý kiến khác.

Câu 24. Cho lịch biểu S8: $r_1(X)$; $r_2(Z)$; $r_1(Z)$; $r_3(X)$; $r_3(Y)$; $w_1(X)$; $w_3(Y)$; $r_2(Y)$; $w_2(Z)$; $w_2(Y)$. Xác định ***lịch biểu tuần tự tương đương*** (*equivalent serial schedule*) của S8 nếu có.

a. T3; T1; T2

b. T1; T2; T3

c. T3; T2; T1

d. S không khả tuần tự hóa. Do đó, không tồn tại lịch biểu tuần tự tương đương của S8.

e. Ý kiến khác.

Câu 25. Cho các giao tác T1 và T2 thực thi trong môi trường đa người dùng nhưng điều khiển tương tranh không được thực hiện như sau:

T1	T2
Read_item(X);	
X := X-5;	
Write_item(X);	
	Read_item(X);
	X := X*2;
Read_item(Y);	
	Write_item(X);
Abort;	
	Commit;

Vấn đề gì xảy ra với việc thực thi đồng thời này?

a. Vấn đề lost update

b. Vấn đề temporary update/dirty read

c. Vấn đề incorrect summary

d. Vấn đề unrepeatable read

e. Ý kiến khác.

Câu 26. Trong môi trường thực thi đơn người dùng (*single-user execution environment*), hệ quản trị cơ sở dữ liệu ***không*** cần kiểm tra cho đặc tính nào của giao tác?

a. Tính nguyên tố (*atomicity*)

b. Tính nhất quán (*consistency*)

c. Tính cách ly (*isolation*)

d. Tính bền vững (*durability*)

e. Ý kiến khác.

Câu 27. Xác định lịch biểu mà trong đó, kỹ thuật khóa hai pha (*two-phase locking*) được sử dụng **ĐÚNG**.

a. Lịch biểu S9

T1	T2
Read_lock(X)	
Read_item(X)	
	Read_lock(X)
	Read_item(X)
Write_lock(Y)	
	Unlock(X)
Read_item(Y)	
Unlock(X)	
Write_item(Y)	
Unlock(Y)	

b. Lịch biểu S10

T1	T2
Read_lock(X)	
Read_item(X)	
	Read_lock(X)
	Read_item(X)
	Unlock(X)
	Write_lock(Y)
Write_lock(X)	
	Write_item(Y)
	Unlock(Y)
Write_item(X)	
Unlock(X)	

c. Lịch biểu S11

T1	T2
Read_lock(X)	
Read_item(X)	
	Read_lock(X)
	Read_item(X)
Write_lock(Y)	
	Unlock(X)
Write_item(Y)	
Unlock(X)	
Write_lock(X)	
Write_item(X)	
Unlock(Y)	
Unlock(X)	

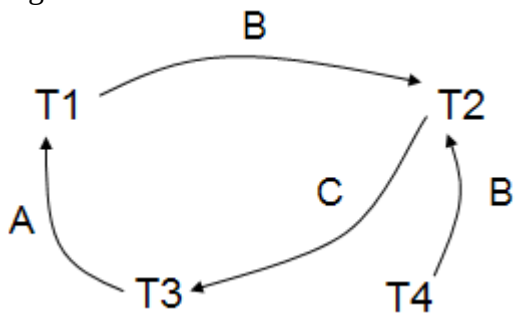
d. Lịch biểu S12

T1	T2
Read_lock(X)	
Read_item(X)	
	Read_lock(X)
	Read_item(X)
	Write_lock(Y)
	Unlock(X)
	Write_item(Y)
	Unlock(Y)
Write_item(X)	
Unlock(X)	

e. Ý kiến khác.

Câu 28. Cho đồ thị đợi (*wait-for graph*) của một lịch biểu dưới đây mà trong đó, điều khiển tương tranh của các giao tác được thực hiện với kỹ thuật khóa hai

pha. ***Khóa chết*** (**deadlock**) có xảy ra với lịch biểu này không?



a. Không đủ thông tin về các giao tác để xác định liệu khóa chết có xảy ra hay không.

b. Chỉ khi các giao tác đạt đến điểm *commit*, đồ thị đợi mới có đầy đủ thông tin và khi đó, việc xác định liệu khóa chết có xảy ra hay không mới được xác định đúng.

c. Có

d. Không

e. Ý kiến khác.

Câu 29. Điều gì luôn **ĐÚNG** với các tác vụ của các giao tác được điều khiển tương tranh bằng kỹ thuật khóa hai pha đa phiên bản với khóa chứng nhận (*multiversion two-phase locking using certify locks*)?

a. Tác vụ write(X) không bao giờ bị từ chối.

b. Tác vụ write(X) sẽ bị bỏ qua (*ignored*) nếu có giao tác khác trẻ hơn giao tác yêu cầu ghi đã đọc dữ liệu X.

c. Tác vụ read(X) không bao giờ bị từ chối (*rejected*).

d. Tác vụ read(X) sẽ bị từ chối nếu không tìm thấy được phiên bản dữ liệu đúng cho yêu cầu đọc của giao tác.

e. Ý kiến khác.

Câu 30. Cho lịch biểu S13 sau đây mà trong đó, các giao tác T1 và T2 thực thi đồng thời dưới kỹ thuật điều khiển tương tranh khóa hai pha đa phiên bản với khóa chứng nhận.

Lịch biểu S13:

T1	T2
Write_lock(Y)	
	Read_lock(Y)
Read_lock(X)	
	Read_item(Y)
Write_item(Y)	
Read_item(X)	
	Unlock(Y)
	Commit
Certify_lock(Y)	
…	

Giao tác T1 có thể có được ***certify lock*** trên dữ liệu Y trước khi T1 ***commit*** hay không?

a. T1 không thể có được certify lock trên Y do T2 đã đọc dữ liệu Y trước khi T1 cập nhật dữ liệu Y.

b. T1 có được certify lock trên Y do không có giao tác nào khác đang giữ read lock trên Y.

c. T1 có được certify lock trên Y do certify lock tương thích với read lock và xung đột loại trừ với write lock và T2 chỉ giữ read lock trên Y.

d. Không đủ thông tin mô tả về lịch biểu nên không thể xác định được liệu T1 có thể có được certify lock trên dữ liệu Y hay không.

e. Ý kiến khác.

Câu 31. Xác định ***lịch biểu tuần tự tương đương*** (*equivalent serial schedule*) với lịch biểu S14 sau đây khi các giao tác của S14 được điều khiển tương tranh bằng kỹ thuật sắp thứ tự theo nhãn thời gian (*timestamp ordering*).

Lịch biểu S14:

T1	T2	T3	T4	A
TS=150	TS=200	TS=175	TS=225	Read_TS=0; Write_TS=0
$R_1(A)$				Read_TS=150
$W_1(A)$				Write_TS=150
	$R_2(A)$			Read_TS=200
	$W_2(A)$			Write_TS=200
		$R_3(A)$		
		Abort		
			$R_4(A)$	Read_TS = 225

a. T4; T2; T3; T1

b. T3; T2; T1; T4

c. T1; T3; T2; T4

d. T2; T3; T4; T1

e. Ý kiến khác.

Câu 32. Cho lịch biểu S15 trong **Bảng 1** mà trong đó, việc điều khiển tương tranh của các giao tác được thực hiện bằng kỹ thuật đa phiên bản dựa trên sắp thứ

tự theo nhãn thời gian (*multiversion technique based on timestamp ordering*). Việc gì xảy ra khi giao tác T3 đọc A, nghĩa là $r_3(A)$ diễn ra?

a. Giao tác T3 bị ngưng thực thi (*aborted*).

b. Tác vụ đọc của giao tác T3 bị trì hoãn thực thi (*delayed*).

c. Tác vụ đọc của giao tác T3 bị bỏ qua (*ignored*).

d. Tác vụ đọc của giao tác T3 được thực hiện với phiên bản A150.

e. Ý kiến khác.

Câu 33. ***Giả định*** (*assumption*) gì được sử dụng trong kỹ thuật điều khiển tương tranh xác nhận hợp lệ lạc quan (*validation (optimistic) concurrency control technique*)?

a. Sự can thiệp lẫn nhau (*interference*) giữa các giao tác trong lịch biểu thấp.

b. Sự can thiệp lẫn nhau giữa các giao tác trong lịch biểu cao.

c. Không tồn tại sự can thiệp lẫn nhau giữa các giao tác trong lịch biểu.

d. Không có giả định gì về sự can thiệp lẫn nhau giữa các giao tác trong lịch biểu.

e. Ý kiến khác.

Câu 34. Trong kỹ thuật điều khiển tương tranh xác nhận hợp lệ lạc quan, việc cập nhật dữ liệu X của một giao tác T sẽ được thực hiện như thế nào?

a. Giá trị mới của dữ liệu X được ghi nhận vào cơ sở dữ liệu và các giao tác khác có thể thấy cập nhật này khi đọc dữ liệu X từ cơ sở dữ liệu.

b. Giá trị mới của dữ liệu X được ghi nhận thành các phiên bản cục bộ của giao tác T và các giao tác khác không thể thấy cập nhật này khi đọc dữ liệu X.

c. Giá trị mới của dữ liệu X được ghi nhận vào sổ ghi hệ thống (*system log*) và các giao tác khác có thể thấy cập nhật này từ việc đọc sổ ghi hệ thống.

d. Giá trị mới của dữ liệu X được ghi nhận thành các phiên bản cục bộ của giao tác T và các giao tác khác có thể thấy cập nhật này khi liên lạc với giao tác T.

e. Ý kiến khác.

Câu 35. Trong ma trận tương thích khóa (*lock compatibility matrix*) dưới đây của kỹ thuật khóa đa mức dữ liệu (*multiple granularity locking technique*), khóa SIX (*shared-intention-exclusive*) tương thích với ***khóa*** nào?

	IS	IX	S	SIX	X
IS	yes	yes	yes	yes	no
IX	yes	yes	no	no	no
S	yes	no	yes	no	no
SIX	yes	no	no	no	no
X	no	no	no	no	no

a. IS

b. IX

c. S

d. SIX

e. Ý kiến khác.

Câu 36. Cho phân cấp cơ sở dữ liệu trong **Hình 6**. Giao tác T1 thực hiện các tác vụ sau: đọc trang p_{21} và cập nhật bản ghi r_{211}. Xác định ***việc khóa đa mức dữ liệu*** được thực hiện trong T1 với kỹ thuật khóa đa mức dữ liệu.

a. IX(db)
IX(f_2)
S(p_{21})
Read(p_{21})
IX(p_{21})
X(r_{211})
Write(r_{211})
Unlock(r_{211})
Unlock(p_{21})
Unlock(p_{21})
Unlock(f_2)
Unlock(db)

b. IX(db)
IX(f_2)
IX(p_{21})

X(r_{211})
Read(p_{21})
Write(r_{211})
Unlock(r_{211})
Unlock(p_{21})
Unlock(f_2)
Unlock(db)

c. IX(db)
IX(f_2)
SIX(p_{21})
Read(p_{21})
X(r_{211})
Write(r_{211})
Unlock(r_{211})
Unlock(p_{21})
Unlock(f_2)
Unlock(db)

d. IX(db)
IX(f_2)
S(p_{21})
Read(p_{21})
Unlock(p_{21})
Unlock(f_2)
Unlock(db)
IX(db)
IX(f_2)
IX(p_{21})
X(r_{211})
Write(r_{211})
Unlock(r_{211})
Unlock(p_{21})
Unlock(f_2)
Unlock(db)

e. Ý kiến khác.

Câu 37. ***Cập nhật trì hoãn*** (*deferred update*) là gì?

a. Ngay khi dữ liệu bị cập nhật trong bộ nhớ cache, thay đổi đó sẽ được thực hiện trên đĩa chỉ sau khi giao tác đạt đến điểm *commit*.

b. Ngay khi dữ liệu bị cập nhật trong bộ nhớ cache, thay đổi đó sẽ được thực hiện trên sổ ghi hệ thống (*system log*) chỉ sau khi giao tác đạt đến điểm *commit*.

c. Ngay khi dữ liệu bị cập nhật trong bộ nhớ cache, thay đổi đó sẽ được ghi nhận thành phiên bản cục bộ của giao tác chỉ sau khi giao tác đạt điểm *commit*.

d. Ngay khi dữ liệu bị cập nhật trong bộ nhớ cache, thay đổi đó sẽ được thực hiện trên vùng đĩa riêng gọi là các trang bóng âm (*shadow pages*) chỉ sau khi giao tác đạt điểm *commit*.

e. Ý kiến khác.

Câu 38. Điều gì **ĐÚNG** với giao thức ghi *log* trước (*write ahead logging* – WAL)?

a. Việc cập nhật dữ liệu nên được thực hiện trước tác vụ ghi *log*.

b. Bản ghi *log* cho một tác vụ cập nhật nên được ghi *log* trước khi dữ liệu thật sự được ghi.

c. Tấ cả các bản ghi *log* nên được ghi *log* trước khi một giao tác mới bắt đầu thực thi.

d. Sổ ghi *log* không bao giờ cần được ghi xuống đĩa.

e. Ý kiến khác.

Câu 39. Để đảm bảo việc phục hồi dữ liệu luôn đúng, các tác vụ ***REDO*** cần có đặc điểm gì?

a. Hoán vị lẫn nhau (*commutative*)

b. Kết hợp lẫn nhau (*associative*)

c. Không thay đổi giá trị (*idempotent*)

d. Phân tán (*distributive*)

e. Ý kiến khác.

Câu 40. Chọn phát biểu **ĐÚNG** về các tác vụ bảo trì tính nguyên tố (*atomicity*) của giao tác trong quá trình phục hồi dữ liệu.

a. ***UNDO*** là tác vụ tháo gỡ bằng cách loại bỏ ***BFIM*** khỏi cơ sở dữ liệu trên đĩa và ***REDO*** là tác vụ tái thực hiện bằng cách phục hồi lại ***AFIM*** đến cơ sở dữ liệu trên đĩa.

b. ***UNDO*** là tác vụ tháo gỡ bằng cách loại bỏ ***AFIM*** khỏi cơ sở dữ liệu trên đĩa và ***REDO*** là tác vụ tái thực hiện bằng cách phục hồi lại ***AFIM*** đến cơ sở dữ liệu trên đĩa.

c. ***UNDO*** là tác vụ tháo gỡ bằng cách loại bỏ ***AFIM*** khỏi cơ sở dữ liệu trên đĩa và ***REDO*** là tác vụ tái thực hiện bằng cách phục hồi lại ***BFIM*** đến cơ sở dữ liệu trên đĩa.

d. ***UNDO*** là tác vụ tháo gỡ bằng cách loại bỏ ***BFIM*** khỏi cơ sở dữ liệu trên đĩa và ***REDO*** là tác vụ tái thực hiện bằng cách phục hồi lại ***BFIM*** đến cơ sở dữ liệu trên đĩa.

e. Ý kiến khác.

Câu 41. Giải thuật phục hồi ***ARIES*** dùng cách tiếp cận nào để ghi dữ liệu xuống cơ sở dữ liệu trên đĩa?

a. Steal/Force

b. Steal/No-Force

c. No-Steal/No-Force

d. No-Steal/Force

e. Ý kiến khác.

Câu 42. *Shadow paging* là dạng kỹ thuật phục hồi gì?

a. UNDO/REDO

b. NO-UNDO/REDO

c. NO-UNDO/NO-REDO

d. UNDO/NO-REDO

e. Ý kiến khác.

Câu 43. Cho các giao tác thực thi trong môi trường đơn người dùng trong **Hình 7**. Xác định ***nội dung của sổ ghi hệ thống*** (*system log*) tại điểm thời gian 9:30 khi kỹ thuật phục hồi dùng cập nhật trì hoãn (*deferred update*) được sử dụng.

a. [start_transaction, T1]
[write_item, T1, A, 15]
[commit, T1]

b. [start_transaction, T1]
[read_item, T1, A]
[write_item, T1, A, 15]
[commit, T1]

c. [start_transaction, T1]
[read_item, T1, A]
[write_item, T1, A, 15]
[commit, T1]
[start_transaction, T2]
[write_item, T2, B, 20]
[read_item, T2, A]

d. [start_transaction, T1]
[write_item, T1, A, 15]
[commit, T1]
[start_transaction, T2]
[write_item, T2, B, 20]

e. Ý kiến khác.

Câu 44. Cho lịch biểu S17 của các giao tác thực thi trong môi trường đa người dùng trong **Hình 8**. Chọn phát biểu **SAI** về việc phục hồi dữ liệu khi hệ thống ngưng thực thi tại điểm thời gian 9:30 với kỹ thuật phục hồi dùng cập nhật trì hoãn (*deferred update*).

a. Các tác vụ ghi của giao tác T1 không cần tái thực hiện do T1 đã *commit* trước *checkpoint*.

b. Các tác vụ ghi của giao tác T2 cần tái thực hiện do T2 đã *commit* sau *checkpoint* nhưng trước 9:30.

c. Các tác vụ ghi của giao tác T3 cần được tháo gỡ do T3 bắt đầu trước *checkpoint* nhưng chưa đạt điểm *commit* trước 9:30.

d. Các tác vụ ghi của giao tác T4 được bỏ qua do T4 bắt đầu sau *checkpoint* nhưng chưa đạt điểm *commit* trước 9:30.

e. Ý kiến khác.

Câu 45. Cho nội dung của sổ ghi hệ thống tại điểm thời gian 9:30 khi hệ thống bị ngưng thực thi (*system crash*) trong **Bảng 2,** nội dung của các bảng ***Transaction*** và ***Dirty Page*** tại điểm *checkpoint* trong **Bảng 3,** và nội dung của các bảng ***Transaction*** và ***Dirty Page*** sau giai đoạn phân tích (*analysis phase*) của kỹ thuật phục hồi ***ARIES*** trong **Bảng** 4. ***REDO*** và ***UNDO*** cần được thực hiện cho các cập nhật dữ liệu của những giao tác nào?

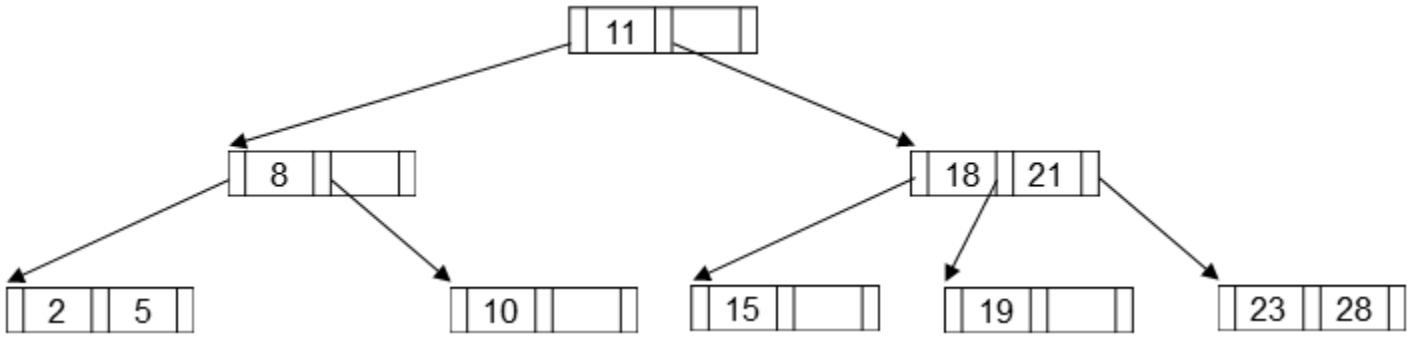
a. Không cần thực hiện ***REDO*** cho các cập nhật dữ liệu của giao tác T1. ***UNDO*** cho các cập nhật dữ liệu của giao tác T2 và T3.

b. Không đủ thông tin để xác định việc ***REDO*** và ***UNDO*** có cần được thực hiện cho các cập nhật dữ liệu của các giao tác T1, T2, và T3.

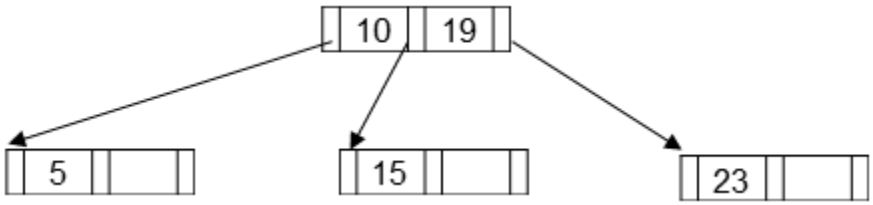
c. ***REDO*** cho các cập nhật dữ liệu của giao tác T1. Không cần thực hiện ***UNDO*** cho các cập nhật dữ liệu của giao tác T2 và T3.

d. ***REDO*** cho các cập nhật dữ liệu của giao tác T1. ***UNDO*** cho các cập nhật dữ liệu của giao tác T2 và T3.

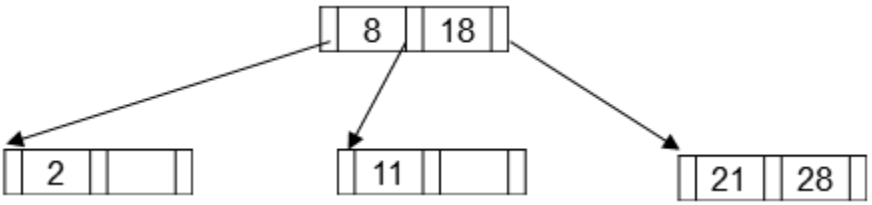
e. Ý kiến khác.



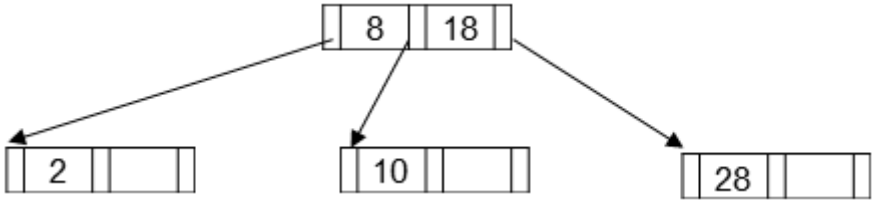
Hình 1 – Cây chỉ mục B-tree có bậc p = 3 được tạo trên vùng tin khóa ID (Câu 7. a)



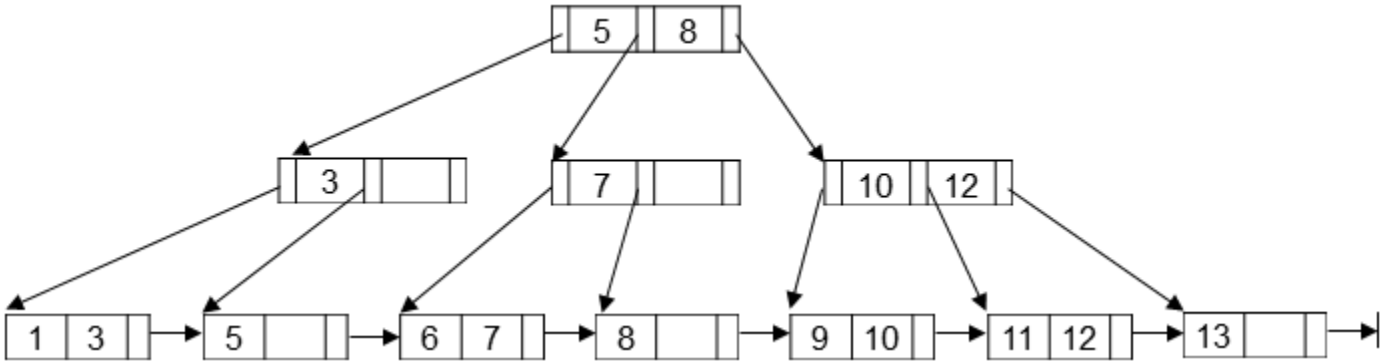
Hình 2 – Cây chỉ mục B-tree có bậc p = 3 được tạo trên vùng tin khóa ID (Câu 7. b)



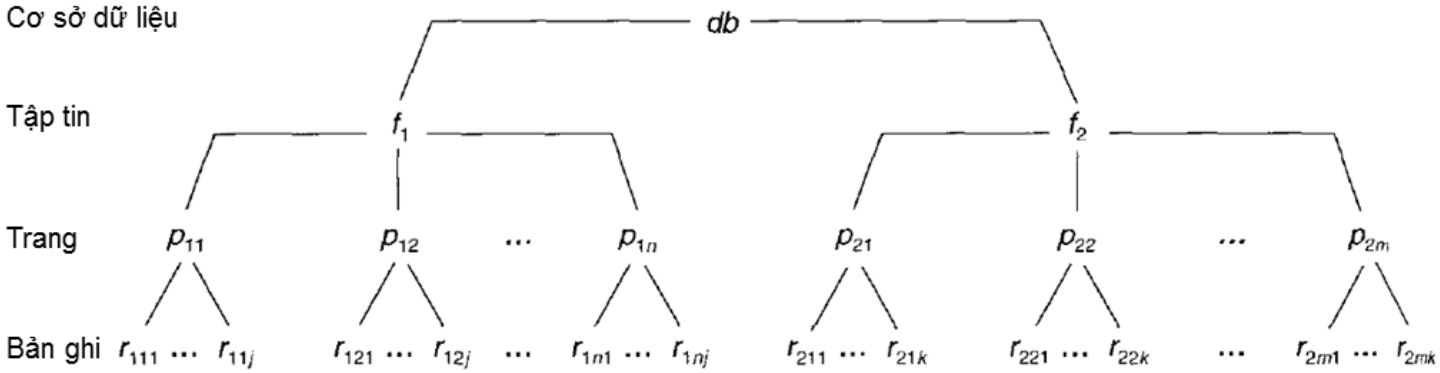
Hình 3 – Cây chỉ mục B-tree có bậc p = 3 được tạo trên vùng tin khóa ID (Câu 7. c)



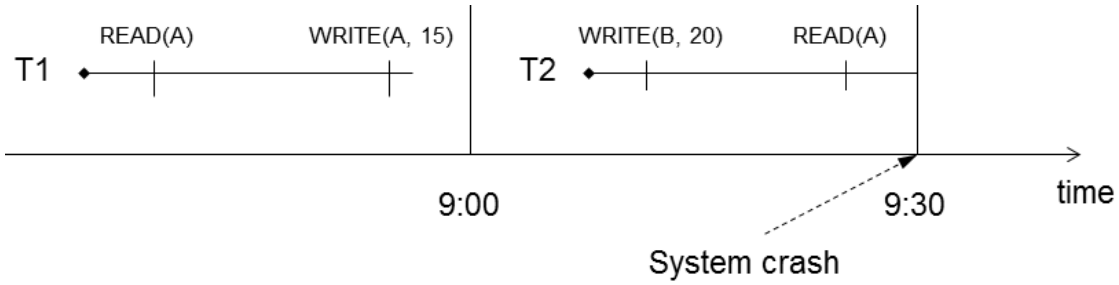
Hình 4 – Cây chỉ mục B-tree có bậc p = 3 được tạo trên vùng tin khóa ID (Câu 7. d)



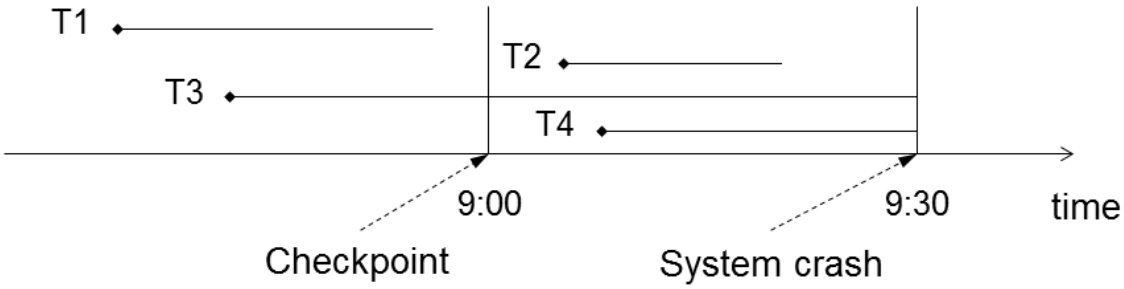
Hình 5 – Cây chỉ mục B+-tree có bậc p = 3 và pleaf = 2 (Câu 9, 10, và 11)



Hình 6 – Phân cấp dữ liệu (Câu 36)



Hình 7 – Các giao tác thực thi trong môi trường đơn người dùng với kỹ thuật phục hồi dùng cập nhật trì hoãn (Câu 43)



Hình 8 – Lịch biểu S17 của các giao tác thực thi trong môi trường đa người dùng (Câu 44)

Bảng 1 – Lịch biểu S15 với kỹ thuật điều khiển tương tranh đa phiên bản dùng sắp thứ tự theo nhãn thời gian (Câu 32)

T1	T2	T3	A0	A150	A200
TS=150	TS=200	TS=175	Read_TS=0; Write_TS=0	Read_TS=150; Write_TS=150	Read_TS=200; Write_TS=200
$r_1(A)$			Read_TS=150		
$w_1(A)$				Created	
	$r_2(A)$			Read_TS=200	
	$w_2(A)$				Created
		$r_3(A)$			

Bảng 2 – Sổ ghi hệ thống khi hệ thống bị ngưng thực thi (Câu 45)

LSN	LAST_LSN	TRAN_ID	TYPE	PAGE_ID	…
1	0	T1	update	C	…
2	1	T1	update	B	…
3	0	T2	update	C	…
4	begin_checkpoint				
5	end_checkpoint				
6	2	T1	commit		…
7	0	T3	update	A	…

Bảng 3 – Bảng Transaction và Dirty Page tại thời điểm checkpoint (Câu 45)

TRANSACTION TABLE			DIRTY PAGE TABLE	
TRANSACTION ID	LAST LSN	STATUS	PAGE ID	LSN
T1	2	in progress	C	1
T2	3	in progress	B	2

Bảng 4 – Bảng Transaction và Dirty Page sau giai đoạn phân tích (Câu 45)

TRANSACTION TABLE			DIRTY PAGE TABLE	
TRANSACTION ID	LAST LSN	STATUS	PAGE ID	LSN
T1	6	Commit	C	1
T2	3	in progress	B	2
T3	7	in progress	A	7

Họ - Tên: …….

Mã Số Sinh Viên: ……

Phần trả lời:

Câu 1 - 15:

Câu	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
a															
b															
c															
d															
e															

Câu 16 - 30:

Câu	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
a															
b															
c															
d															
e															

Câu 31 - 45:

Câu	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
a															
b															
c															
d															
e															

Phần giải thích đáp án (e) nếu có: