

A. TRẮC NGHIỆM (8đ)

Mỗi câu hỏi **có thể có nhiều đáp án đúng**. Sinh viên cần điền **tất cả** những đáp án đó vào bảng trả lời.

Ví dụ nếu Câu 1 có các đáp án là A và D thì điền vào ô tương ứng trong bảng là: **AD**.

Nếu câu hỏi có yêu cầu giải thích thì sinh viên phải giải thích ngắn gọn cho lựa chọn của mình.

1. E	11. A D	21. A D	31. B	41. C
2. B C D	12. C	22. B D	32. A	42. A C
3. B (gt)	13. D	23. A	33. C	43. B D
4. D	14. B D	24. B C	34. B	44. A C
5. C	15. A B D	25. A B C	35. A C	45. C D
6. D	16. D	26. B C D	36. C	46. A B C
7. B	17. A D	27. C D hoặc D	37. B	47. C
8. B C	18. B C D	28. A C	38. A (gt)	48. C
9. A	19. A C	29. A D	39. C	49. B C
10. C	20. B C	30. B	40. A	50. A

1. Những ràng buộc toàn vẹn được định nghĩa cho *specializations* trong ERD là:

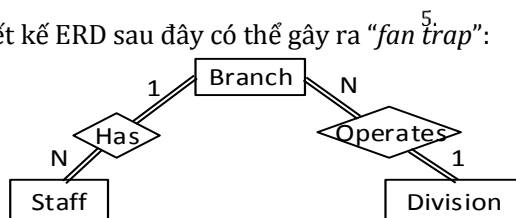
- A. Participation và completeness.
- B. Participation và disjointness.
- C. Key constraint và disjointness.
- D. Cardinality và participation.
- E. **Completeness và disjointness.**

4.

2. Trong cách tiếp cận *shared file approach* cho các ứng dụng thì phát biểu nào sau đây là **SAI**:

- A. Vấn đề cấu trúc dữ liệu không uyển chuyển (*rigid data structure*) xuất hiện và cần phải giải quyết.
- B. **Vấn đề dư thừa dữ liệu (*Data redundancy*) được loại bỏ hoàn toàn.**
- C. **Vấn đề truy xuất đồng thời (*concurrency control*) được hỗ trợ.**
- D. **Vấn đề phụ thuộc dữ liệu có tính chất vật lý (*physical data dependency*) được loại bỏ hoàn toàn.**

3. Thiết kế ERD sau đây có thể gây ra "fan trap"⁵:



A. Đúng, giải thích:

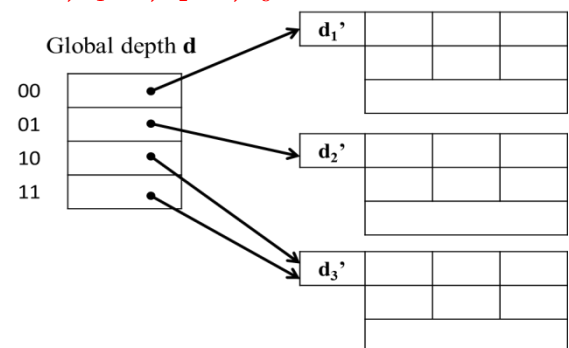
.....

B. Sai, giải thích:

.....

Cho tập tin dữ liệu được tổ chức dạng băm động dùng **extendible hashing** như hình dưới. Chọn phát biểu đúng về *global depth* và *local depth*:

- A. $d = 4, d_1' = 3, d_2' = 3, d_3' = 3$
- B. $d = 4, d_1' = 2, d_2' = 2, d_3' = 2$
- C. $d = 2, d_1' = 2, d_2' = 2, d_3' = 2$
- D. **$d = 2, d_1' = 2, d_2' = 2, d_3' = 1$**



Một câu lệnh SELECT của SQL bắt buộc phải có ít nhất:

- A. Các mệnh đề SELECT, FROM, WHERE.
- B. Các mệnh đề SELECT, WHERE.
- C. **Các mệnh đề SELECT, FROM.**
- D. Mệnh đề SELECT.

Cho lược đồ quan hệ $R(A1, A2, A3)$; $P1, P2$ chứa kết quả của các phép truy xuất dữ liệu như mô tả dưới đây:
 $P1 \leftarrow \Pi_{A2, A3} (\sigma_{A2="a"} (R))$; $P2 \leftarrow \text{SELECT } A2, A3 \text{ FROM } R \text{ WHERE } A2="a"$. Chọn **MỘT** phát biểu **ĐÚNG NHẤT**:

6. A. $|P1| = |P2|$
 B. $|P1| > |P2|$
 C. $|P1| < |P2|$
 D. Tất cả các câu trên đều chưa chính xác.

Phát biểu nào sau đây **ĐÚNG**:

7. A. Cơ sở dữ liệu bao gồm hệ quản trị cơ sở dữ liệu và hệ cơ sở dữ liệu.
 B. Hệ cơ sở dữ liệu bao gồm hệ quản trị cơ sở dữ liệu và cơ sở dữ liệu.
 C. Trong kiến trúc lược đồ 3 lớp (*three-schema architecture*), dữ liệu thật sự được lưu ở cấp *internal level*.
 D. Sự độc lập dữ liệu vật lý (*Physical Data Independence*) là khi ta thay đổi DBMS mà không cần thay đổi lược đồ quan hệ.

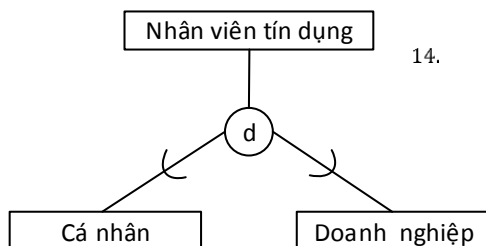
Chức năng nào sau đây **KHÔNG PHẢI** là của một DBA:

8. A. Security and authorization.
 B. Database design.
 C. Transaction management.
 D. Database tuning.
 E. Back up.
 F. Performance monitoring.

Chức năng nào sau đây **KHÔNG THUỘC** về mô hình dữ liệu mức ý niệm (*conceptual*):

9. A. Cho phép CSDL được hiện thực trong một DBMS.
 B. Cho phép người dùng CSDL hiểu cấu trúc của CSDL.
 C. Đóng vai trò như một công cụ giao tiếp giữa người dùng và người thiết kế CSDL.
 D. Biểu diễn các loại dữ liệu được lưu trữ trong CSDL.
 E. Biểu diễn những mối quan hệ của dữ liệu.

Cho ERRD như bên dưới:



Phát biểu nào sau đây là **ĐÚNG**:

- A. Nhân viên tín dụng chỉ bao gồm tín dụng cá nhân và tín dụng doanh nghiệp.
 B. Một nhân viên tín dụng có thể vừa làm tín dụng cá nhân có thể vừa làm tín dụng doanh nghiệp.

- C. Nhân viên tín dụng cá nhân sẽ có tất cả những mối quan hệ mà nhân viên tín dụng có.
 D. Tất cả đều sai.

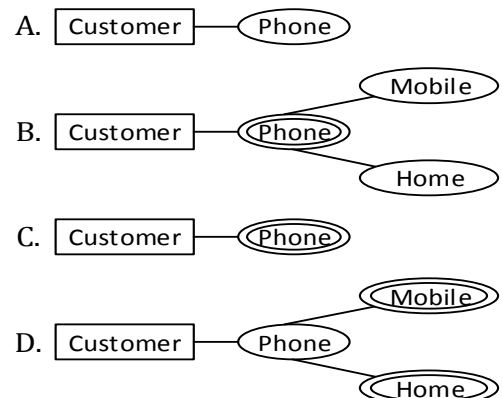
Chọn phát biểu **ĐÚNG** về chỉ mục.

- A. Một tập tin có nhiều nhất một *primary index* hoặc nhiều nhất một *clustering index*, nhưng không thể có cùng lúc cả hai loại chỉ mục này.
 B. Một tập tin có nhiều nhất một *primary index* hoặc nhiều *secondary index*, nhưng không thể có cùng lúc cả hai loại chỉ mục này.
 C. Một tập tin có nhiều nhất một *clustering index* hoặc nhiều *secondary index*, nhưng không thể có cùng lúc cả hai loại chỉ mục này.
 D. Một tập tin có thể có một *primary index* và nhiều *secondary index*.

Một đối tượng nộp thuế có thể là công ty hoặc cá nhân. Cá nhân và công ty là 2 thực thể khác nhau, không có nguồn gốc chung. Chọn **MỘT** cách biểu diễn ERD/EERD **thích hợp nhất** để mô hình hóa cho tình huống trên:

- A. Dùng quan hệ 3 ngôi giữa đối tượng nộp thuế, công ty và cá nhân.
 B. Dùng quan hệ cha con: *Superclass* (đối tượng nộp thuế) và *Subclass* (công ty và cá nhân).
 C. Dùng *Union type* (hay còn gọi là *category*): đối tượng nộp thuế là *Union type*.
 D. Dùng *shared subclass* (đối tượng nộp thuế).

Biết rằng một công ty cần lưu số điện thoại liên lạc của khách hàng bao gồm số cố định và di động. Một khách hàng có thể có cùng lúc nhiều số điện thoại. Cho biết ERD/ERRD nào sau đây thỏa yêu cầu này:



Chọn phát biểu **ĐÚNG** về *shared subclass* và *category*:

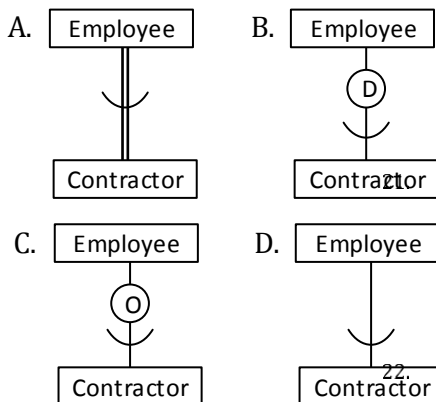
- A. *Shared subclass* và *category* thừa kế mọi thuộc tính và quan hệ từ các lớp cha (*superclass*) của nó.
 B. *Shared subclass* thừa kế mọi thuộc tính và quan hệ từ các lớp cha của nó; *category* chỉ thừa kế thuộc tính và quan hệ của một trong các lớp cha của nó.

- C. Một *category* là một tập con của giao tất cả các lớp cha của nó.
- D. Một lược đồ có tồn tại ít nhất một *shared subclass* thì được gọi là một *lattice*.

Đối tượng nào sau đây trong ERD sẽ được chuyển thành một quan hệ trong quá trình ánh xạ:

- A. Thực thể yếu. 19.
- B. Mỗi liên kết 3 ngôi *one-one-one*.
- C. Mỗi liên kết 2 ngôi *one-to-many* với phía "many" là bắt buộc (mandatory).
- D. Thuộc tính đa trị.

15. Một công ty cần quản lý nhân viên (*Employee*) của mình. Trong số nhân viên, có những người đặc biệt, làm việc theo dự án, được gọi là *contractor*. ERD/ERRD nào sau đây mô tả tình huống trên:



17. Cho kích thước của một khối (*block*) là $B = 512$ bytes. Một tập tin gồm các bản ghi (*record*) có chiều dài cố định và không có bản ghi nào nằm trên 2 blocks. Mỗi bản ghi gồm các vùng tin (*field*) sau: NAME (30 bytes), SSN (9 bytes), ADDRESS (4 bytes), JOB CODE (4 bytes), GENDER (1 byte). Chọn phát biểu **ĐÚNG**:

- A. Các bản ghi trong tập tin này được lưu trữ theo dạng không phủ (*unspanned*).
- B. Các bản ghi trong tập tin này được lưu trữ theo dạng phủ (*spanned*).
- C. Blocking factor - *bfr* của tập tin này: $bfr = 10.667$.
- D. Blocking factor - *bfr* của tập tin này: $bfr = 10$.
- E. Blocking factor - *bfr* của tập tin này: $bfr = 11$.

18. Phát biểu nào sau đây là **ĐÚNG**: 24.
- A. Bậc của một quan hệ (*relation*) là số lượng các dòng dữ liệu đang có trong quan hệ đó.
- B. Bậc (*degree*) của một mối quan hệ trong ERD là số lượng các kiểu thực thể tham gia mỗi quan hệ đó.
- C. Một thuộc tính phức hợp cũng có thể là đa trị.
- D. Ta không nên cho người sử dụng nhập dữ liệu vào thuộc tính dẫn xuất.

Trong bảo mật CSDL, với 2 kỹ thuật điều khiển truy xuất là DAC và MAC thì:

- A. Hệ thống sử dụng kỹ thuật MAC không gặp phải vấn đề *Trojan Horses*.
- B. Vấn đề *covert channel* không tồn tại trong các hệ thống sử dụng MAC.
- C. DAC thích hợp hơn cho các ứng dụng trong môi trường Web.
- D. Tất cả các câu trên đều sai.

Chọn phát biểu **ĐÚNG** về ORDER BY trong câu lệnh SELECT không có truy vấn con:

- A. Mọi thuộc tính được liệt kê trong mệnh đề ORDER BY phải xuất hiện trong mệnh đề SELECT.
- B. Nếu có thêm từ khóa DISTINCT, thì mọi thuộc tính được liệt kê trong mệnh đề ORDER BY phải xuất hiện trong mệnh đề SELECT.
- C. Có thể ORDER BY trên một hoặc nhiều thuộc tính.
- D. Có thể ORDER BY trên một thuộc tính không xuất hiện trong GROUP BY.

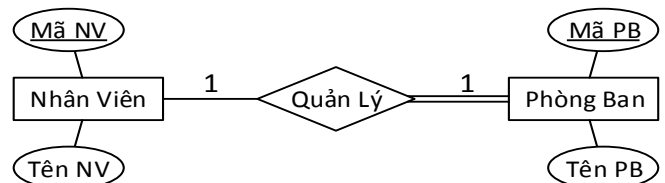
Cho hai quan hệ R và S, chọn phát biểu **ĐÚNG** sau đây (ký hiệu "x": tích Đề-các, "*": phép toán nhân):

- A. $|R \times S| = |R| * |S|$
- B. $|R \times S| < |R| * |S|$
- C. Bậc ($R \times S$) = bậc R * bậc S
- D. Bậc ($R \times S$) = bậc R + bậc S

Cho quan hệ R(A, B, C) và 3 tác vụ: (1) $\sigma_{A=c}(R(A, B, C))$; (2) $\{t \mid R(t) \text{ AND } t.A=c\}$; (3) $\{xyz \mid R(xyz) \text{ AND } z=c\}$.

- A. Cả ba tác vụ này là tương đương nhau trên R.
- B. (1) và (2) là tương đương nhau trên R.
- C. (2) và (3) là tương đương nhau trên R.
- D. (1) và (3) là không tương đương nhau trên R.
- E. Tất cả các câu trên đều sai.

Ánh xạ ERD sang mô hình dữ liệu quan hệ, ta được:



- A. NV (Mã NV, Tên NV) và PB (Mã PB, Tên PB, Mã NV Quản lý).
- B. NV (Mã NV, Tên NV) và PB (Mã PB, Mã NV Quản lý, Tên PB).
- C. NV (Mã NV, Tên NV, Mã PB) và PB (Mã PB, Tên PB).
- D. NV (Mã NV, Tên NV), PB(Mã PB, Tên PB) và Quản lý (Mã NV, Mã PB).

Chọn phát biểu **ĐÚNG** về thực thể yếu (*weak entity*):

- A. Thực thể yếu chỉ được phép có quan hệ với thực thể mạnh mà nó phụ thuộc vào.
- B. Thực thể yếu có thể không cần khoá riêng phần (*partial key*).
- C. Thực thể yếu có thể có nhiều hơn một thực thể mạnh cùng tham gia vào mỗi quan hệ định danh (*identifying relationship*) của mình.
- D. Thực thể yếu không được phép có mối quan hệ định danh với một thực thể yếu khác.

Cho quan hệ R. Phát biểu nào sau đây là **ĐÚNG**:

25. A. $\sigma_{\langle \text{cond1} \rangle}(\sigma_{\langle \text{cond2} \rangle}(R)) = \sigma_{\langle \text{cond2} \rangle}(\sigma_{\langle \text{cond1} \rangle}(R))$
- B. $\sigma_{\langle \text{cond1} \rangle}(\sigma_{\langle \text{cond2} \rangle}(R)) = \sigma_{\langle \text{cond1} \rangle \text{ AND } \langle \text{cond2} \rangle}(R)$
- C. $|\sigma_{\langle \text{cond1} \rangle}(\sigma_{\langle \text{cond2} \rangle}(R))| \leq |\sigma_{\langle \text{cond1} \rangle}(R)|$
- D. Tất cả các câu trên đều sai.

Cho lược đồ quan hệ:

LOTS(PROPERTY_ID#, COUNTY_NAME, LOT#, AREA)

26. và 3 phụ thuộc hàm:

FD1: PROPERTY_ID# → {COUNTY_NAME, LOT#, AREA};
 FD2: {COUNTY_NAME, LOT#} → {PROPERTY_ID#, AREA};
 FD3: {AREA → COUNTY_NAME}.

Chúng ta có LOTS thuộc những dạng chuẩn nào?

27. A. BCNF.
- B. **3NF.**
- C. **2NF.**
- D. **1NF.**

Cho quan hệ R như sau:

R (<u>PartNumber</u> , <u>Supplier</u> , SupplierAddress, Price, Description)
--

R có một khoá duy nhất là {PartNumber, Supplier}.

Các phụ thuộc hàm:

FD1: PartNumber → Description

FD2: Supplier → SupplierAddress

FD3: PartNumber, Supplier → Price

Phân rã R thành các quan hệ thoả **BCNF**. Quan hệ nào dưới đây **KHÔNG** xuất hiện trong kết quả phân rã:

28. A. R1 (PartNumber, Description)
- B. R2 (Supplier, SupplierAddress)
- C. R3 (PartNumber, Supplier, Price)
- D. **R4 (Description, Price)**

29. Phát biểu nào sau đây là **ĐÚNG** trong mô hình dữ liệu quan hệ (*relational data model*):

- A. **Tuple, row, record là các thuật ngữ tương đương.**
- B. Một quan hệ có thể có nhiều khóa chính.
- C. **Một quan hệ phải có ít nhất 1 khóa.**
- D. Một quan hệ phải có ít nhất một khóa ngoại.

Một quan hệ luôn luôn thuộc dạng chuẩn 2 khi:

- A. **Chỉ có một khóa và khóa này chỉ có một thuộc tính.**
- B. Tồn tại những phụ thuộc hàm bậc cầu.
- C. Chỉ có một khóa.
- D. **Tất cả những khóa dự tuyển là những khóa chỉ có một thuộc tính.**
- E. Tất cả những phụ thuộc hàm đều thuộc dạng không đầy đủ (*partial*).

Trong SQL, phát biểu "Các hàm gộp MIN, MAX, AVERAGE, SUM, COUNT(A) chỉ thao tác trên các giá trị khác NULL" là:

- A. Sai vì hàm COUNT(A) thao tác trên cả giá trị NULL.
- B. **Đúng vì các giá trị NULL sẽ bị bỏ qua đối với các hàm gộp này.**
- C. Sai vì hàm SUM sẽ coi các giá trị NULL bằng với số không (0).
- D. Các câu trên đều sai.

Giả sử kích thước các nodes (*node size*) của B-tree và B⁺-tree là như nhau và các cây này được dùng để đánh chỉ mục (*index*) trên cùng một cột của một bảng trong cơ sở dữ liệu thì B⁺-tree luôn có chiều cao nhỏ hơn hoặc bằng chiều cao của B-tree. Phát biểu này là:

- A. Sai.
- B. **Đúng.**
- C. Tùy trường hợp cụ thể.

Dữ liệu sau dùng cho **Câu 32** và **33**: Cho quan hệ *Employee* đang sử dụng cơ chế bảo mật MAC: (Quy ước: S ≥ C ≥ U)

Name		Salary		Performance		TC
Smith	U	40	C	Fair	U	C
Brown	C	80	S	Good	C	S

Một user có **level C** sẽ nhìn thấy *Employee* như sau:

A.

Name		Salary		Performance		TC
Smith	U	40	C	Fair	U	C
Brown	C	Null	C	Good	C	C

B.

Name		Salary		Performance		TC
Smith	U	40	C	Fair	U	C
Brown	C	Null	S	Good	C	S

C.

Name		Salary		Performance		TC
Smith	U	40	C	Fair	U	C
Brown	C	Null	C	Good	C	S

D.

Name		Salary		Performance		TC
------	--	--------	--	-------------	--	----

Smith	C	40	C	Fair	C	C
Brown	C	Null	C	Good	C	C

User có **level C** cập nhật thành công giá trị *Salary* của *Brown* thành **100** thì quan hệ kết quả sẽ ra sao?

A.

Name		Salary		Performance		TC
Smith	U	40	C	Fair	U	C
Brown	C	100	S	Good	C	S

B.

Name		Salary		Performance		TC
Smith	U	40	C	Fair	U	C
Brown	C	100	C	Good	C	C

C.

Name		Salary		Performance		TC
Smith	U	40	C	Fair	U	C
Brown	C	80	S	Good	C	S
Brown	C	100	C	Good	C	C

D.

Name		Salary		Performance		TC
Smith	U	40	C	Fair	U	C
Brown	C	80	S	Good	C	S
Brown	C	100	S	Good	C	S

34. Cho quan hệ R(A,B,C,D,E) và một tập các phụ thuộc hàm $F = \{AB \rightarrow CE, B \rightarrow D, D \rightarrow A\}$. Phụ thuộc hàm nào sau đây **KHÔNG THỂ** suy ra được từ F:

- A. $BC \rightarrow DE$
- B. **$AD \rightarrow CE$**
- C. $B \rightarrow E$

35. D. $AB \rightarrow A$ E. Không có phụ thuộc hàm nào ở trên.

Chọn phát biểu **SAI** sau đây:

- A. **Khóa ngoại không được NULL.**
- B. Mọi thành phần của khóa chính không được NULL.
- C. **Mỗi quan hệ chỉ có thể có nhiều nhất một khóa ngoại.**
- D. Một thành phần nào đó của khóa chính cũng có thể là khóa ngoại tham khảo đến khóa chính của quan hệ khác.

Các khái niệm cơ bản về bảo mật CSDL cho rằng:

- A. *Identification* và *authentication* là hai công việc giống nhau.
- B. Nguyên lý *no write-down* trong MAC cũng tồn tại trong RBAC.

- C. **Dùng dấu vân tay (fingerprint) là một kỹ thuật sử dụng trong việc authentication.**
- D. Tất cả các câu trên đều sai.

Chọn phát biểu **SAI** về *big data*:

- A. Các DBMS thương mại hiện nay chưa thể quản lý và xử lý câu truy vấn hiệu quả.
- B. **Dữ liệu phải được quản lý trên hạ tầng điện toán đám mây.**
- C. Lượng dữ liệu cần lưu trữ nhiều.
- D. Tốc độ dữ liệu được tạo ra cao.
- E. Dữ liệu lưu trữ đa dạng.

Phát biểu "Mọi quan hệ có hai thuộc tính đều thuộc dạng chuẩn BCNF" là:

A. **Đúng, giải thích:**

.....

B. Sai, giải thích:

.....

Hai nguyên tắc bảo mật của mô hình điều khiển truy xuất MAC (mô hình Bell-LaPadula) là:

- A. No read-down và No write-down.
- B. No read-down và No write-up.
- C. **No read-up và No write-down.**
- D. No read-up và No write-up.

Công nghệ lưu trữ **RAID** (Redundant Arrays of Inexpensive/Independent Disks) và **SAN** (Storage Area Networks) có cùng đặc điểm nào sau đây?

- A. **Hỗ trợ dư thừa dữ liệu (data redundancy) để gia tăng độ đáng tin cậy (reliability).**
- B. Hỗ trợ khoảng cách vật lý nhất định giữa các thiết bị (đĩa) lưu trữ và máy chủ truy cập dữ liệu.
- C. Cho phép chia sẻ dữ liệu để giảm tải kiểm tra lỗi của bộ điều khiển đĩa (disk controller).
- D. Hỗ trợ việc thêm vào/lấy ra mỗi thiết bị (đĩa) lưu trữ trong dãy đĩa (**RAID**) hoặc trong mạng lưu trữ (**SAN**) một cách tiện lợi.

Quan hệ bên dưới dùng cho các **câu 41, 42, 43**

ChiNhanh (<u>MaCN</u> , TenCN, DiaChi)
PhongBan (<u>MaPB</u> , TenPB, DoanhThu, MaCN)

- Chọn câu truy vấn **hợp lệ**:
41. A. **SELECT** TenPB, DoanhThu
FROM PhongBan
WHERE DoanhThu = **MAX**(DoanhThu);
- B. **SELECT** TenPB, DoanhThu
FROM PhongBan
WHERE DoanhThu =(SELECT **MAX** (DoanhThu)
FROM PhongBan
GROUP BY MaCN);
- C. **SELECT** TenPB, DoanhThu
FROM PhongBan
WHERE DoanhThu = (SELECT **MAX**(DoanhThu)
FROM PhongBan);
- D. Cả 3 câu truy vấn đều sai.
- Câu truy vấn sau bị lỗi ở các dòng lệnh nào?
42. **SELECT** P2.MaPB -- dòng 1
FROM PhongBan P1 -- dòng 2
WHERE P1.DoanhThu = -- dòng 3
(SELECT **MAX** (P2.DoanhThu) -- dòng 4
FROM PhongBan P2 46. -- dòng 5
WHERE P1.MaCN = PhongBan.MaCN); -- dòng 6
- A. Dòng 1.
B. Dòng 4.
C. Dòng 6.
D. Câu truy vấn trên không bị lỗi.
43. Câu truy vấn nào sau đây liệt kê các tên chi nhánh mà tất cả các phòng ban của nó đều có doanh thu lớn hơn 10.000?
- A. **SELECT** TenCN
FROM ChiNhanh C, PhongBan P
WHERE C.MaCN = P.MaCN
AND DoanhThu > 10.000;
- B. **SELECT** TenCN
FROM ChiNhanh C
WHERE 10.000 < **ALL** (SELECT DoanhThu
FROM PhongBan P
WHERE C.MaCN = P. MaCN);
- C. **SELECT** TenCN
FROM ChiNhanh C
WHERE 10.000 < **ANY** (SELECT DoanhThu
FROM PhongBan P
WHERE C.MaCN = P. MaCN);
- D. **SELECT** TenCN
FROM ChiNhanh C
WHERE MaCN **NOT IN** (SELECT MaCN
FROM PhongBan P
WHERE DoanhThu <= 10.000);

Loại chỉ mục nào có thể được lập trên một trường là khoá (*key field*)?

- A. **Primary index.**
B. Clustering index.
C. **Secondary index.**

Cho T là một cây R-tree lưu trữ các MBR (*Minimum Bounding Rectangles*) của các đối tượng dữ liệu không gian 3 chiều phức tạp. Giả sử $Q(x,y,z)$ là một điểm trong không gian dữ liệu của T. Để xác định xem đối tượng dữ liệu nào trong T thực sự chứa Q thì:

- A. Chỉ phải duyệt cây T theo 1 nhánh.
B. Phải duyệt cây T theo 2 nhánh.
C. **Có thể chỉ phải kiểm tra gốc (root) của cây T.**
D. **Chưa thể xác định phải duyệt cây T theo bao nhiêu nhánh.**
E. Tất cả các câu trên đều sai

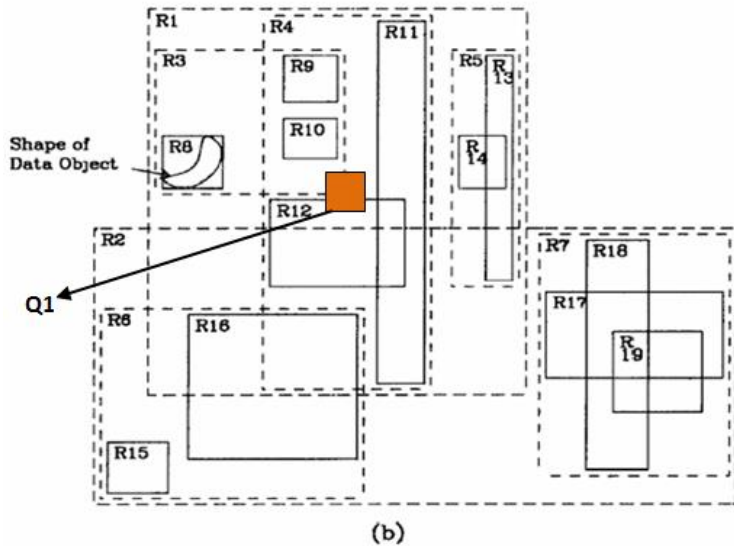
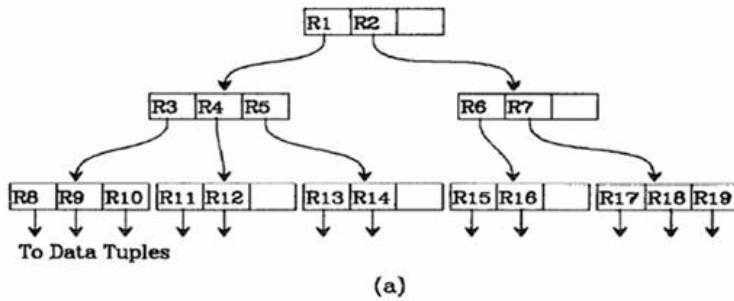
Cho quan hệ R(X, Y, Z, T) với các phụ thuộc hàm:

- $X \rightarrow T$
- $Y \rightarrow X$
- $Z, T \rightarrow Y$

Tập nào dưới đây là **một khoá** của quan hệ trên:

- A. {X, Z}
B. {Y, Z}
C. {Z, T}
D. {X, Y}

Hình sau sử dụng cho 2 câu 47 và 48:



47. Những nodes nào của cây R-tree sẽ được truy xuất để trả lời câu hỏi “Tìm các đối tượng dữ liệu CHỨA Q1”:
- Root, R1, R3.
 - Root, R2, R6.
 - Root, R1, R4.**
 - Các câu trên đều sai.

Những nodes nào của cây R-tree sẽ được truy xuất để trả lời câu hỏi “Tìm các đối tượng dữ liệu GIAO với Q1”:

- Root, R1, R3, R5, R13.
- Root, R2, R6, R7, R17.
- Root, R1, R3, R4, R12.**
- Các câu trên đều sai.

Physical database design có thể gồm:

- Thiết kế lại ERD của CSDL.
- Thiết kế chỉ mục mức vật lý.**
- “Denormalization” nhằm tăng tốc độ xử lý cho các câu truy vấn.**
- Tất cả các câu trên đều sai.

Giả sử ta có 4 account: Minh, Chi, Linh, Thu

- Minh tạo bảng Dept và thực hiện lệnh sau:
GRANT SELECT ON Dept TO Chi;
- Chi tiếp tục thực hiện lệnh sau:
GRANT SELECT ON Minh.Dept TO Linh;
- Linh tiếp tục thực hiện lệnh sau:
GRANT SELECT ON Minh.Dept TO Thu;
- Sau đó Chi lại thực hiện lệnh:
REVOKE SELECT ON Dept FROM Linh;

Chọn phát biểu **ĐÚNG** trong tình huống này:

- Linh và Thu không có quyền SELECT trên Dept, Chi có quyền này.**
- Linh không còn quyền SELECT trên Dept, Chi và Thu vẫn còn quyền này.
- Linh, Thu, Chi vẫn còn quyền SELECT trên dept.
- Tất cả các câu trên đều sai.

B. TỰ LUẬN (2đ)

Trường Đại học Quốc gia TP.HCM cần xây dựng hệ thống phục vụ nhu cầu lưu trữ dữ liệu và quản lý toàn bộ hoạt động của tất cả các trường/ trung tâm/ viện thành viên. Hệ thống này cung cấp các nhóm chức năng chính sau:

- Hoạt động học tập & giảng dạy:**
 - Cho phép sinh viên đăng kí môn học trực tuyến, tra cứu thông tin môn học, điểm thi, thời khoá biểu, lịch thi.
 - Hệ thống e-learning (học trực tuyến, thảo luận, nộp bài tập, cung cấp các tài liệu học tập đa phương tiện, khảo sát ý kiến sinh viên, thi trực tuyến...).
 - Cho phép giảng viên tra cứu các thông tin liên quan tới hoạt động giảng dạy và quản lý lớp học, sinh viên.
- Hoạt động nghiên cứu khoa học:** cho phép lưu trữ, quản lý và thống kê các đề tài, dự án nghiên cứu khoa học, chuyển giao công nghệ, hợp tác quốc tế.
- Hoạt động quản lý:**
 - Sinh viên:** Lưu trữ thông tin cá nhân (lý lịch, hình ảnh, khen thưởng, kỷ luật, hoạt động ngoại khoá...) để phục vụ cho việc cấp học bổng, trợ cấp xã hội đến từ các tổ chức, đơn vị khác nhau...
 - Cán bộ, công nhân viên:** lưu trữ thông tin cá nhân (lý lịch, hình ảnh, thu nhập, thuế, khen thưởng, kỷ luật, quá trình công tác...).
 - Các hoạt động, sự kiện:** Hội nghị, hội thảo, cuộc thi, hội chợ, chiến dịch Mùa Hè Xanh (thông tin, hình ảnh, video...)...

- **Các dịch vụ khác:** Cho phép các tổ chức bên ngoài xác minh kết quả học tập và làm việc của sinh viên, giảng viên trong trường.

Ngoài ra, mọi thông tin về lịch sử hoạt động của người dùng trên hệ thống cần được lưu lại để phục vụ cho mục đích phân tích sau này.

Toàn bộ các dữ liệu và thông tin sẽ được dùng để nâng cao hiệu suất, lập các kế hoạch tài chính, xây dựng chiến lược hoạt động cho trường.

Câu hỏi:

1. Hãy mô tả ngắn gọn các ví dụ cho thấy hệ thống trên mang 3 đặc trưng sau của **Big Data**. Mỗi đặc trưng 1 ví dụ. **(1.5 đ)**

- Volume

.....

.....

.....

.....

.....

- Variety

.....

.....

.....

.....

.....

- Velocity

.....

.....

.....

.....

.....

2. Trình bày một vấn đề bảo mật đặc trưng cho dữ liệu **Big Data** trong hệ thống này. **(0.5đ)**

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....