**Câu hỏi ôn tập chương 1:**

1. Định nghĩa Cơ sở dữ liệu (database)?

Cơ sở dữ liệu (database) là một tập hợp các dữ liệu có liên quan với nhau trong một lĩnh vực cụ thể.

2. Các chức năng của hệ quản trị CSDL

- Cung cấp môi trường tạo lập và duy trì cơ sở dữ liệu.

- Cung cấp môi trường cập nhật và khai thác dữ liệu.

- Cung cấp công cụ điều khiển truy cập vào CSDL.

3. Ngôn ngữ định nghĩa dữ liệu là gì? Cho ví dụ.

- Ngôn ngữ định nghĩa dữ liệu (DDL) là ngôn ngữ máy tính được sử dụng để tạo và sửa đổi cấu trúc của các đối tượng cơ sở dữ liệu trong cơ sở dữ liệu.

Các đối tượng CSDL này bao gồm các khung nhìn, lược đồ, bảng, chỉ mục, ... Ngoài ra, DDL còn cung cấp cơ sở để xác định một số ràng buộc sẽ duy trì nhất quán của dữ liệu.

- Ví dụ: JSON Schema, XML Schema, YAML Schema, Avro, SQL DDL,...

4. Ngôn ngữ thao tác dữ liệu là gì? Cho ví dụ

- Ngôn ngữ thao tác dữ liệu (DML) là ngôn ngữ để người dùng diễn tả yêu cầu cập nhật hay khai thác thông tin. Trong đó, thao tác dữ liệu bao gồm: Cập nhật (thêm, sửa, xóa dữ liệu) và khai thác (sắp xếp, tìm kiếm, kết xuất báo cáo, ...)

- Ví dụ: SQL, Python, R, Java, SQLAlchemy,...

5. Ngôn ngữ kiểm soát dữ liệu và ngôn ngữ kiểm soát transaction? cho ví dụ.

- Ngôn ngữ kiểm soát dữ liệu (DCL) là tập hợp các lệnh chủ yếu các vấn đề liên quan đến quyền, cụ thể là cấp hoặc thu hồi lại quyền của người dùng với cơ sở dữ liệu. Ví dụ: SQL, Python, Cucumber, Junit,...

- Ngôn ngữ kiểm soát transaction (TCL) được dùng để thay đổi trạng thái của một số dữ liệu.

- Ví dụ: SQL, JTA, 2PC,...

6. Nhiệm vụ của công cụ kiểm soát, điều khiển truy cập vào CSDL?

- Duy trì tính nhất quán của CSDL

- Khôi phục CSDL khi có sự cố

- Phát hiện và ngăn chặn sự truy cập không được phép.

7. Các loại đối tượng sử dụng CSDL?

Các loại đối tượng sử dụng CSDL:

- Người quản trị CSDL (DBA-DataBase Administrator)

- Người thiết kế CSDL (DataBase Designer)

- Người dùng cuối (End user)

- Phân tích viên hệ thống (System analyst) và lập trình viên ứng dụng

(Application Programmer)

8. Hệ quản trị CSDL (DBMS) là gì?

Hệ quản trị CSDL (DBMS) là một gói/hệ thống các phần mềm giúp cho việc tạo và duy trì cơ sở dữ liệu trên máy tính một cách thuận lợi.

9. Kể tên một vài DBMS thông dụng

SQL-Server, DB2, Paradox, Infomix, Oracle, ...

10.Một hệ CSDL?

Hệ CSDL bao gồm một DBMS cùng với một CSDl. Đôi khi còn có cả chương trình ứng dụng.

11.Siêu dữ liệu (meta-data)?

Hệ CSDL không chỉ lưu CSDL mà còn lưu trữ định nghĩa về cấu trúc dữ liệu và các ràng buộc trong catalog. Thông tin được lưu trữ trong catalog được gọi là siêu dữ liệu.

12.Ưu điểm của DBMS

- Đảm bảo quản lý thông tin và chuyển đổi dữ liệu tốt hơn.

- Giúp người dùng có thể thực hiện các thao tác như: thêm, sửa, xóa, truy vấn dữ liệu.

13.Mô hình dữ liệu (data model) là gì?

Mô hình dữ liệu (data model) là sơ đồ về cách thức tổ chức, lưu trữ dữ liệu trong doanh nghiệp và các mối liên kết giữa các thông tin đó.

14.Database designer có nhiệm vụ?

- Hiểu yêu cầu của người dùng và tạo một thiết kế đáp ứng yêu cầu.

- Xác định dữ liệu cần lưu trữ trong CSDL và cấu trúc CSDL.

15.System Analyst có nhiệm vụ?

Phân tích viên hệ thống xác định những yêu cầu của người dùng và xây dựng bản đặc tả cho những giao tác đáp ứng yêu cầu người dùng.

16.Application Programmer có nhiệm vụ?

Lập trình viên cài đặt đặc tả thành chương trình, kiểm tra, dò lỗi, ghi sưu liệu và bảo trì những giao tác này.

17. Lược đồ (schema) là gì?

Lược đồ quan hệ là tập tất cả các thuộc tính có trong quan hệ.

18. Phân loại mô hình dữ liệu:

- Mô hình dữ liệu mức cao (mức khái niệm): cung cấp các khái niệm gần với cách người dùng cảm nhận về dữ liệu.

- Mô hình dữ liệu mức thấp (mức vật lý): mô tả cách dữ liệu được lưu trữ trên máy tính.

- Mô hình dữ liệu mức logic: mức trung gian giữa 2 mức trên để người dùng có thể hiểu nhưng cũng gần với cách dữ liệu được tổ chức trên đĩa cứng. Mô hình này giấu đi chi tiết lưu trữ đữ liệu trên đĩa nhưng có thể cài đặt trên máy tính.

19.Mô hình dữ liệu mức khái niệm?

Cung cấp các khái niệm gần với cách người dùng cảm nhận về dữ liệu.

20.Mô hình dữ liệu mức vật lý ?

Mô tả cách dữ liệu được lưu trữ trên máy tính.

21.Mô hình dữ liệu mức logic ?

Mức trung gian giữa 2 mức trên để người dùng có thể hiểu nhưng cũng gần với cách dữ liệu được tổ chức trên đĩa cứng. Mô hình này giấu đi chi tiết lưu trữ đữ liệu trên đĩa nhưng có thể cài đặt trên máy tính.

22.Kiến trúc 3 mức trừu tượng của một hệ CSDL? Công dụng của nó?

Kiến trúc 3 mức trừu tượng của một hệ CSDL là: Physical Level, Conceptual Level, và View Level, giúp tách biệt các ứng dụng người dùng với CSDL vật lý.

23.Mức vật lý trong kiến trúc 3 mức của 1 hệ CSDL?

Sử dụng mô hình vật lý để mô tả cấu trúc lưu trữ vật lý của CSDL.

24.Mức logic trong kiến trúc 3 mức trừu tượng của 1 hệ CSDL?

Giấu đi chi tiết về cấu trúc lưu trữ vật lý. Dùng mô hình dữ liệu logic để mô tả cái gì được lưu trữ trong CSDL và mối quan hệ giữa các dữ liệu đó.

25.Mức view trong kiến trúc 3 mức trừu tượng của 1 hệ CSDL?

Mô tả một phần của CSDL cho một nhóm người dùng quan tâm và giấu đi phần còn lại khỏi nhóm người dùng đó.

26. Phân loại Tính độc lập dữ liệu ?

Tính độc lập dữ liệu có thể định nghĩa như khả năng thay đổi lược đồ tại một mức của một hệ CSDL mà không phải thay đổi lược đồ tại mức cao hơn kế tiếp.

Có hai tính độc lập dữ liệu:

- Tính độc lập vật lý.

- Tính độc lập logic.

27. Tính độc lập vật lý?

Tính độc lập vật lý là khả năng thay đổi lược đồ vật lý mà không phải lược đồ quan niệm. Vì vậy lược đồ ngoài cũng không cần phải thay đổi.

28.Tính độc lập logic là gì?

Tính độc lập logic là khả năng thay đổi lược đồ quan niệm mà không phải thay đổi lược đồ ngoài và chương trình ứng dụng.

**Assigment:**

1. Why would you choose a database system instead of simply storing data in operating system files? When would it make sense not to use a database system?

- Chọn hệ thống CSDL thay vì đơn giản lưu trữ dữ liệu trong các tệp trên hệ điều hành vì có những lợi ích quan trọng như:

- Tính toàn vẹn dữ liệu

- Khả năng tìm kiếm và truy vấn

- Quản lý đa người dùng

- Bảo mật dữ liệu

- Xử lý đồng thời và hiệu quả

+ Tuy nhiên, cũng có những trường hợp không nên sử dụng hệ thống CSDL, chằng hạn như khi dự án yêu cầu tính nhất quán thấp, hoặc khi dữ liệu không cần truy xuất và quản lý phức tạp.

2. What is logical data independence and why is it important?

- Độc lập dữ liệu là khả năng thay đổi cấu trúc lưu trữ và tổ chức của CSDL mà không ảnh hưởng đến các ứng dụng sử dụng dữ liệu của nó.

- Độc lập dữ liệu quan trọng vì:

+ Tính linh hoạt và dễ dàng bảo trì

+ Quản lý thay đổi dễ dàng

+ Tách biệt giữa dữ liệu và ứng dụng

+ Tiết kiệm thời gian và tài nguyên

3. Explain the difference between logical and physical data independence.

- Tính độc lập vật lý là khả năng thay đổi lược đồ vật lý mà không phải lược đồ quan niệm. Vì vậy lược đồ ngoài cũng không cần phải thay đổi.

- Tính độc lập logic là khả năng thay đổi lược đồ quan niệm mà không phải thay đổi lược đồ ngoài và chương trình ứng dụng.

4. What are the responsibilities of a DBA? If we assume that the DBA is never interested in running his or her own queries, does the DBA still need to understand query optimization? Why?

- Trách nhiệm của người quản trị CSDL:

+ Phát triển, quản trị và duy trì cơ sở dữ liệu

+ Xác định mục đích và loại cơ sở dữ liệu cần thiết của một công ty

+ Hiểu các yêu cầu về lưu trữ dữ liệu và khả năng truy cập

+ Thiết kế kiến trúc cơ sở dữ liệu

- Nếu giả định rằng người quản trị CSDL không quan tâm đến các truy vấn riêng của mình thì vẫn cần phải hiểu về tối ưu hóa truy vấn. Vì người quản trị CSDL cần phải hiểu để có khả năng tư vấn, hỗ trợ các nhóm phát triển, đám bảo ứng dụng hoạt động hiệu quả.

5. Scrooge McNugget wants to store information (names, addresses, descriptions of embarrassing moments, etc.) about the many ducks on his payroll. Not surprisingly, the volume of data compels him to buy a database system. To save money, he wants to buy one with the fewest possible features, and he plans to run it as a stand-alone application on his PC clone.

Of course, Scrooge does not plan to share his list with anyone. Indicate which of the following DBMS features Scrooge should pay for; in each case,also indicate why Scrooge should (or should not) pay for that feature in the system he buys.

1. A security facility:

Bảo mật dữ liệu là quan trọng để đảm bảo rằng dữ liệu không bị truy cập trái phép.

2. Concurrency control:

Kiểm soát đồng thời giúp chạy nhiều tác vụ cùng lúc (có thể không cần thiết nếu muốn tiết kiệm chi phí).

3. Crash recovery:

Nên sử dụng để có thể khôi phục dữ liệu sau khi gặp sự cố.

4. A view mechanism:

Không cần phải có nếu người dùng không cần thiết tạo các views tùy chỉnh dữ liệu.

5. A query language:

Nên sử dụng để có thể truy vấn, tìm kiếm cũng như thực hiện các tác vụ khác một cách dễ dàng hơn.

6. Describe the structure of a DBMS. If your operating system is upgraded to support some new functions on OS files (e.g., the ability to force some sequence of bytes to disk), which layer(s) of the DBMS would you have to rewrite to take advantage of these new functions?

- Hệ quản trị cơ sở dữ liệu gồm 3 tầng:

+ Tầng vật lý (physical layer): Tầng này liên quan đến việc lưu trữ thực tế của dữ liệu trên đĩa và cách dữ liệu được truy cập. Nó quản lý cách dữ liệu được lưu trữ trong các tập tin và block trên đĩa cứng.

+ Tầng logic (Logical layer): Tầng này định nghĩa cấu trúc dữ liệu, quan hệ giữa các bảng, các ràng buộc và các quy tắc dữ liệu. Nó tạo ra giao diện truy vấn mà người dùng và ứng dụng có thể sử dụng để truy cập và thay đổi dữ liệu

+ Tầng ứng dụng (Application layer): Tầng này là tầng cao nhất và liênquan đến ứng dụng sử dụng dữ liệu. Nó bao gồm các ứng dụng và chương trình mà người dùng sử dụng để làm việc với dữ liệu