**Chương 1.**

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

CSDL là một tập hợp các dữ liệu có liên quan với nhau trong một lĩnh vực cụ thể nào đó. Một CSDL có thể có kích thước bất kỳ. Một CSDL có thể được tạo lập và bảo trì một cách thủ công hoặc được vi tính hoá.

1. Các chức năng của hệ quản trị CSDL

Cung cấp cách tạo lập CSDL, cập nhật dữ liệu, tìm kiếm và kết xuất thông tin và Cung cấp công cụ kiểm soát điều khiển việc truy cập vào CSDL.

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

Ngôn ngữ định nghĩa dữ liệu thật chất là hệ thống các kí hiệu để mô tả cơ sở dữ liệu trong hệ quản trị Cơ sở dữ liệu. Ví du: các khung nhìn, lược đồ, bảng, chỉ mục

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

Ngôn ngữ thao tác dữ liệu là ngôn ngữ để người dùng diễn tả yêu cầu cập nhật hay khai thác thông tin. Ví dụ: thao tác thêm , sửa, xoá, cập nhật dữ liệu

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

Ngôn ngữ kiểm soát dữ liệu (DCL) cho phép các quản trị viên cơ sở dữ liệu cấu hình truy cập bảo mật cho các cơ sở dữ liệu quan hệ. Ví dụ: lệnh GRANT thêm quyền cho một đối tượng, lệnh REVOKE thu hồi quyền truy cập CSDL

1. 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

- Tổ chức và điều khiển các truy cập đồng thời

- Quản lí các mô tả dữ liệu.

1. Các loại đối tượng sử dụng CSDL?

- Quản trị viên CSDL (DBA)

- Người thiết kế CSDL (Database designer)

- Phân tích viên hệ thống (System Analyst)

- Lập trình viên ứng dụng (Application Programmer)

1. Hệ quản trị CSDL (DBMS) là gì?

Hệ quản trị CSDL (DBMS) là một gói/hệ thống phần mềm giúp cho việc tạo lập và duy trì cơ sở dữ liệu trên máy tính một cách thuận lợi. Nó là một phần mềm với mục đích tổng quát là cho phép dễ dàng thực hiện các tiến trình định nghĩa, xây dựng, khai thác và chia sẻ CSDL giữa những người dùng và ứng dụng khác nhau

1. Kể tên một vài DBMS thông dụng

SQL-Server, Oracle, DB2, Paradox, Informix…

1. Một hệ CSDL?

Bao gồm một DBMS cùng với một CSDL. Đôi khi có thể bao gồm cả chương trình ứng dụng

1. 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.

1. Ưu điểm của DBMS

Cung cấp nhiều phương thức để lưu trữ và truy xuất các dữ liệu có nhiều định dạng khác nhau bằng cách sử dụng ngôn ngữ truy vấn (Query Language).

Dễ dàng được duy trì.

DBMS hỗ trợ nhiều ứng dụng có thể sử dụng trong cùng một hệ cơ sở dữ liệu với thời gian phát triển và duy trì được tối ưu nhất.

Bảo mật và toàn vẹn dữ liệu với hoạt động cập nhật và sao lưu dữ liệu tối thiểu.

DBMS còn tương thích với các ngôn ngữ lập trình ứng dụng như Java và Python nhằm mục đích hỗ trợ người dùng kết nối cơ sở dữ liệu với bất kỳ ứng dụng hay một trang web nào.

Có chức năng sao lưu tự động và phục hồi khi cần.

Cho phép người dùng có thể truy cập và chia sẻ dữ liệu.

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

Một tập hợp các khái niệm dùng để mô tả cấu trúc của cơ sở dữ liệu

1. Database designer có nhiệm vụ?

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

Xác định dữ liệu cần lưu trong CSDL và cấu trúc CSDL

1. System Analyst có nhiệm vụ?

Cấp quyền khai thác CSDL

Xem xét việc sử dụng CSDL và thu hồi tài nguyên

Bảo mật và thời gian đáp ứng nhu cầu của hệ thống

1. Application Programmer có nhiệm vụ?

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

1. Lược đồ (schema) là gì?

Lược đồ là một mô tả về một tập hợp dữ liệu cụ thể, sử dụng một mô hình dữ liệu nhất định

1. 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 dấu đi chi tiết lưu trữ dữ liệu trên đĩa nhưng có thể cài đặt trên máy tính.

1. Mô hình dữ liệu mức khái niệm?

Cung cấp các khái niệm gần với cách người dùng cảm nhận về dữ liệu. Ví dụ: Mô hình thực thể kết hợp là mô hình dữ liệu mức khái niệm.

1. Mô hình dữ liệu mức vật lý ?

* Cung cấp những khái niệm để biểu diễn chi tiết cách thức dữ liệu được lưu trữ trong máy tính. Mô hình này chỉ ra định dạng bản ghi (record formats), thứ tự sắp xếp các bản ghi (record ordering) và đường dẫn để truy cập dữ liệu (access paths).

1. Mô hình dữ liệu mức logic ?

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

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

* Có 3 mức biểu diễn
* Kiến trúc 3 mức giúp tách biệt các ứng dụng người dùng với CSDL vật lý

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

* Mức vật lý: Sử dụng mô hình dữ liệu vật lý để mô tả cấu trúc lưu trữ vật lý của CSDL

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

* Mức logic/quan niệm: Dấu 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

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

* Mức ngoài/ view: Mô tả một phần của CSDL cho một nhóm người dùng quan tâm và dấu đi phần còn lại của CSDL khỏi nhóm người dùng đó.

1. 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 2 tính độc lập dữ liệu: vật lý và logic.

1. Tính độc lập vật lý?

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

1. 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 hay chương trình ứng dụng.

**Assigment:**

**Câu 1**. Tại sao bạn lựa chọn hệ thống CSDL thay vì lưu dữ liệu trong file quản lý bởi hệ điều hành? Khi nào bạn không nên dùng hệ CSDL?

* Một số vấn đề khi lưu trữ trên file:
* Bộ nhớ không đủ để lưu trữ dữ liệu lớn chỉ thích hợp với các ứng dụng nhỏ 1 người dùng
* Dữ liệu liên tục bị chuyển qua lại giữa bộ nhớ chính và phụ
* Mỗi câu truy vấn phải viết 1 chương trình đặc biệt
* Không đảm bảo tính nhất quán
* Khó khôi phục dữ liệu
* Các vấn đề bảo mật và kiểm soát truy cập
* Những hạn chế của file là các ưu điểm của CSDL
* Độc lập dữ liệu và truy cập hiệu quả
* Giảm thời gian phát triển ứng dụng
* Toàn vẹn và bảo mật dữ liệu
* Quản trị dữ liệu thống nhất
* Truy cập đồng thời và khôi phục khi có sự cố

**Câu 2**. Độc lập dữ liệu mức logic (Logical data independence) là gì? Tại sao nó quan trọng?

* Là khả năng thay đổi lược đồ quan niệm mà không phải thay đổi lược đồ ngoài hay chương trình ứng dụng.
* Nó quan trọng vì có thể thay đổi các thông tin của các thực thể hoặc xóa bớt các thông tin về các thực thể đang tồn tại trong CSDL mà việc làm này không ảnh hưởng đến các lược đồ con đang tồn tại, do đó nó quan trọng.

**Câu 3**. Giải thích sự khác nhau giữa độc lập dữ liệu mức logic và độc lập dữ liệu mức vật lý (physical data independence)? Hãy cho ví dụ minh họa.

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

Ví dụ từ dữ liệu Sinhvien(MaSV: string, Hoten: char(20), tenKhoa: char(20)) ta thêm chỉ mục Hoten để tìm kiếm theo họ tên hiệu quả hơn sẽ làm thay đổi cấu trúc lưu trữ nhưng không phải thay đổi lược đồ quan niệm.

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

Ví dụ Với lược đồ khái niệm về Sinh viên ta có thể mở rộng CSDL bằng cách thêm dữ liệu về DiemTB: float sẽ không làm thay đổi lược đồ ngoài hay chương trình ứng dụng đang có.

**Câu 4**. Giải thích sự khác biệt giữa lược đồ ý niệm/logic (conceptual/logical schema), lược đồ vật lý/bên trong (Physical/internal schema) và lược đồ ngoài (external schema).

Lược đồ ngoài (external schema) là những bản ghi cụ thể, người dùng cũng như người quản trị CSDL sẽ thao tác với mô hình ngoài dùng ngôn ngữ truy vấn dữ liệu. Lược đồ ý niệm/logic(conceptual/logical schema) là định nghĩa các mô hình quan hệ( các bảng). Còn lược đồ trong(Physical/internal schema) là mô hình, cách thức lưu dữ liệu vật lý của hệ thống CSDL.

**Câu 5**. Trách nhiệm của DBA. Giả sử là DBA không cần quan tâm đến việc thực hiện các câu truy vấn của chính DBA, thì DBA có cần hiểu về tối ưu hóa câu truy vấn không? Tại sao?

Trách nhiệm của DBA:

* Thiết kế lược đồ logic và lược đồ vật lý.
* Điều khiển bảo mật và phân quyền, dữ liệu luôn được toàn vẹn.
* Đảm bảo cho dữ liệu luôn trong tình trạng sẵn sang và phục hồi lại khi có sự cố xảy ra.
* Sự điều chỉnh dữ liệu cần phải tối ưu hơn.

DBA cần phải hiểu về tối ưu hóa câu truy vấn vì tầm quan trọng của DBA là rất lớn. Cần phải am hiểu trong nhiều mảng liên quan đến lập trình để đảm bảo cho hệ thống CSDL luôn được hoạt động tốt và am hiểu về tối ưu câu truy vấn để có thể tối ưu hệ thống CSDL 1 cách tốt nhất khi cần thiết.

**Câu 6**. Ông A cần mua một hệ CSDL. Để tiết kiệm chi phí, ông A chỉ mua một hệ CSDL với số tính năng ít nhất có thể. Ông ta lập kế hoạch chỉ chạy nó một mình trên máy PC của ông ấy và không share thông tin với ai cả. Hãy cho biết tính năng nào trong các tính năng dưới đây của DBMS ông A mua nên có và tại sao:

+ Tiện ích bảo mật

+ Kiểm soát đồng thời

+ Khôi phục dữ liệu sau sự cố

+ Cơ chế khung nhìn

+ Ngôn ngữ truy vấn

**Trả lời:**

Ông A nên mua DBMS có các tính năng sau:

* Tiện ích bảo mật(nên có vì không có chức năng này dữ liệu của ông A có thể sẽ bị tấn công).
* Khôi phục dữ liệu sau sự cố( Cần thiết để tránh mất dữ liệu khi gặp sự cố).
* Cơ chế khung nhìn(Phải có để ông A dễ dàng sử dụng).

Kiểm soát đồng thời với ngôn ngữ truy vấn có thể không cần vì chỉ có 1 mình ông A sử dụng và ngôn ngữ truy vấn ông A cũng không cần biết vì đó là công việc của DBA quản lý hệ CSDL đó.

**Câu 7**. Mô tả cấu trúc của một DBMS. Giả sử hệ điều hành của bạn được nâng cấp để hỗ trợ thêm một số chức năng về file (ví dụ khả năng cho phép lưu một chuỗi các bytes lên đĩa). Hãy cho biết lớp nào của DBMS bạn cần phải viết lại để có thể tận dụng ưu điểm của các tính năng mới đó.

Cấu trúc của một DBMS gồm 3 phần sau:

- Mức trong (internal level): mô tả cấu trúc lưu trữ vật lý của CSDL

- Mức ý niệm (conceptual level):

o Mô tả toàn bộ cấu trúc của cơ sở dữ liệu

o Che dấu chi tiết vật lý

o Tập trung vào mô tả dữ liệu, mối quan hệ, các ràng buộc

- Mức ngoài (external level):

o Mô tả một phần của CSDL ứng với góc nhìn của một nhóm người dùng

Để có thể tận dụng ưu điểm của tính năng cho phép lưu một chuỗi các byte lên đĩa chúng ta cần phải viết lại mức ngoài.

**Câu 8**. Trả lời các câu hỏi sau:

1. Giao tác (transaction) là gì?

Transaction (giao tác) là một tập các hành động thực hiện cùng một chức năng và chúng chỉ có thể cùng thành công hoặc thất bại

1. Tại sao một DBMS thực hiện xen kẽ các hành động của các giao dịch khác nhau thay vì thực hiện lần lượt từng giao dịch một ?

DBMS thực hiện xen kẽ các transaction nhằm mục đích tận dụng tối đa thời gian trống của các chương trình trong hệ thống. Nếu chỉ thực hiện từng giao dịch một thì một số chương trình sẽ rảnh thời gian, từ đó hiệu suất kém.

1. Một user phải chắc chắn điều gì để đảm bảo tính nhất quán giữa một giao dịch và CSDL ? Một DBMS nên chắc chắn điều gì để đảm bảo tính nhất quán giữa thực hiện đồng thời nhiều giao dịch và CSDL.

DBMS phải đảm bảo rằng giao dịch được thực thi toàn bộ và độc lập với các giao dịch khác. Một tính chất quan trọng của giao dịch trong DBMS là tính nguyên tử và có tính cô lập. Các giao dịch hoặc là thành công hoặc là thất bại hoàn toàn. Điều này đảm bảo tính nhất quán của dữ liệu.

1. Giải thích về nghi thức khóa 2 giai đoạn nghiêm ngặt (the strict two-phase locking protocol).

Khóa hai giai đoạn nghiêm ngặt sử dụng các khóa chia sẻ và khóa độc quyền để bảo vệ dữ liệu. Một giao dịch phải nắm giữ tất cả các khóa cần thiết trước khi thực thi và không giải phóng bất cứ khóa nào trước khi giao dịch kết thúc hoàn toàn.

1. Tính chất WAL là gì và tại sao nó quan trọng ?

Đặc tính WAL ảnh hưởng đến chiến lược logging trong DBMS. WAL là viết tắt của Write- Ahead Log, có nghĩa là mỗi một thao tác viết lên cơ sở dữ liệu đều phải được ghi vào log (trên đĩa) trước khi nó thực sự xảy ra trong cơ sở dữ liệu. Điều này bảo vệ được cơ sở dữ liệu nếu có sự cố xảy ra trong quá trình giao dịch đang thực hiện. Bằng việc ghi lại những thay đổi vào trong log trước khi những thay đổi này thực sự diễn ra, cơ sở dữ liệu biết được phải làm lại những thao tác nào để khôi phục sự cố. Ngược lại, nếu việc ghi vào log diễn ra sau thì cơ sở dữ liệu sẽ không thể phát hiện được có những thay đổi nào đã xảy ra trong quá trình khôi phục sự cố.