Nguyễn Kiều Châu Anh - 20110234

Chương 1.

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

Một tập các dữ liệu có liên quan với nhau trong một lĩnh vực cụ thể. VD: CSDL trong một ứng dụng quản lí sinh viên gồm các thông tin như Họ tên, mã số SV, lớp, điểm trung bình, …

1. Các chức năng của hệ quản trị CSDL

Chức năng của hệ quản trị CSDL là cung cấp cách tạo lập CSDL, cung cấp cách 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 (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 cơ sở dữ liệu này bao gồm các khung nhìn, các lược đồ, bảng, chỉ mục,…DDL thường được kết hợp vào bất kỳ ngôn ngữ chính thức nào mô tả dữ liệu. Các DDL thường được sử dụng trong truy vấn SQL là:

* CREATE: tạo bảng mới và có cú pháp xác định trước.
* ALTER: lệnh thay đổi bảng cơ sở dữ liệu hiện có. Lệnh này có thể bổ sung, thả thêm các cột hiện có và thậm chí thay đổi kiểu dữ liệu của các cột liên quan đến bảng cơ sở dữ liệu.
* DROP: lệnh drop xóa bảng, chỉ mục hoặc khung hình.

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

Ngôn ngữ thao tác dữ liệu (DML) bao gồm các lệnh cho phép nhập, sửa, xóa dữ liệu, khai thác dữ liệu như: tìm kiếm, sắp xếp, xuất kết quả báo cáo.

VD: lệnh INSERT để chèn dữ liệu, lệnh UPDATE dùng để chỉnh sửa nội dung của một hoặc nhiều dòng dữ liệu, lệnh DELETE để có nội dung của một hoặc nhiều dòng 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) là một tập con của ngôn ngữ truy vấn dữ có cấu trúc (SQL) và cho phép các quản trị viên cơ sở dữ liệu **cấu hình truy cập** cho các cơ sở dữ liệu quan hệ. Nó **bổ sung cho ngôn ngữ định nghĩa dữ liệu (DDL)**, được sử dụng để thêm, xóa các đối tượng cơ sở dữ liệu và ngôn ngữ thao tác dữ liệu (DML) được sử dụng để truy xuất, chèn và sửa đổi nội dung của cơ sở dữ liệu. DCL là tập con đơn giản nhất của cơ sở dữ liệu vì nó chỉ gồm 3 lệnh: GRANT (cấp quyền), REVOKE(loại bỏ quyền đã cung cấp cho người dùng trước đó) và DENY(ngăn chặn người dùng nhận được một quyền cụ thể nào đó). Ba lệnh này kết hợp cung cấp cho quản trị viên sự linh hoạt để thiết lập và loại bỏ quyền cơ sở dữ liệu theo kiểu cực kì chi tiết.

**Tham khảo**: <https://vi.eyewated.com/ngon-ngu-kiem-soat-du-lieu/>

* Ngôn ngữ kiểm soát transaction (TCL) là ngôn ngữ được sử dụng để kiểm soát các transaction (giao dịch) trong cơ sở dữ liệu. Chúng được sử dụng để **kiểm soát các thay đổi được tạo bởi DML**. Chúng cũng cho phép các câu lệnh được nhóm lại với nhau thành các transaction có nghĩa.

VD: Câu lệnh **COMMIT, SAVEPOINT, ROLLBACK**

**Tham khảo:**

<https://www.geeksforgeeks.org/sql-ddl-dml-tcl-dcl/>

<https://in4tintuc.com/transaction-la-gi-sql/>

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 truy cập trái phép.

1. Các loại đối tượng sử dụng CSDL?

Có **4 nhóm** đối tượng sử dụng CSDL:

* Người quản trị CSDL (DBA-Database administrator): cấp quyền khai thác CSDL, xem xét việc sử dụng CSDL là thu hồi tài nguyên, bảo mật thời gian và thời gian đáp ứng
* Người thiết kế CSDL (Database designer): Hiểu yêu cầu người dùng và tạo thiết kế đáp ứng yêu cầu đó, xác định dữ liệu cần lưu trữ trong CSDL và cấu trúc CSDL.
* Người dùng cuối (end users): Người dùng ngẫu nhiên, người dùng thường xuyên, người dùng cho công việc phức tạp, người dùng một mình
* Phân tích viên hệ thống (System analyst) và lập trình viên ứng dụng (Application programmer): phân tích viên hệ thống xác định yêu cầu của người dùng và xây dựng bản đặc tả cho những giao tác đáp ứng yêu cầu người dùng, lập trình viên cài đặt đặc tả thành chương trình, kiểm tra, dò lỗi, ghi sưu liệu và bảo trì những giao tác này.

1. Hệ quản trị CSDL (DBMS) là gì?

Hệ quản trị cơ sở dữ liệu là một gói/ hệ thống phần mềm được dùng để việc tạo và duy trì cơ sở dữ liệu trên máy tính được thuận lợi.

1. Kể tên một vài DBMS thông dụng

SQL-Server, DB2, Paradox, Oracle, …

1. Một hệ 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 lưu trữ trong catalog được gọi là siêu dữ liệu.*

1. Ưu điểm của DBMS

* Đảm bảo tính độc lập dữ liệu và truy cập hiệu quả.
* Giảm thời gian phát triển thực thi
* Đảm bảo bảo mật và thống nhất của dữ liệu
* Nhất quán trong việc quản trị dữ liệu
* Phục hồi dữ liệu khi có sự cố xảy ra

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

Là 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 người dùng và tạo thiết kế đáp ứng yêu cầu đó, xác định dữ liệu cần lưu trữ trong CSDL và cấu trúc CSDL.

1. System Analyst có nhiệm vụ?

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

1. 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.

1. Lược đồ (schema) là gì?

Lược đồ (schema) là tập hợp các đối tượng cơ sở dữ liệu bao gồm table, view, trigger, stored procedure, index, v.v

1. Phân loại mô hình dữ liệu

* Mô hình dữ liệu mức cao (mức khái niệm)
* Mô hình dữ liệu mức thấp (mức vật lý)
* Mô hình dữ liệu mức logic

1. Mô hình dữ liệu mức khái niệm?

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

VD: Mô hình thực thể kết hợp ERD

1. Mô hình dữ liệu mức vật lý ?

Mô tả cách dữ liệu được lưu trữ trên máy tính

VD: file trên máy tính được mô tả bởi định dạng file, chỉ mục,…

1. Mô hình dữ liệu mức logic ?

Mức trung gian giữa hai 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 những chi tiết lưu trữ dữ liệu trên đĩa nhưng có thể cài đặt trên máy tính

VD: mô hình quan hệ, mô hình phân cấp, mô hình mạng, …

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

Theo kiến trúc ANSI-PARC, một cơ sở dữ liệu có 3 mức biểu diễn, nhằm tách biệt các ứng dụng người dùng với CSDL vật lý. Ba mức đó là:

* Mức vật lý
* Mức logic/ quan niệm
* Mức ngoài/ view

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

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?

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?

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 ?

Có hai tính độc lập dữ liệu:

* Tính độc lập vật lý
* Tính đọc lập logic

1. 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 thay đổi.

1. Tính độc lập logic là gì?

Là khả năng thay đổi lược đồ quan niệm mà không phải thay đổi lược đồ ngoài hay chương trình ứng dụng.

**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?

Khi lưu trữ trên file thì chúng ta gặp phải 1 số vấn đề sau:

-         Bộ nhớ không đủ để lưu trữ nếu dữ liệu quá lớn.

-         Dữ liệu phải chuyển qua lại giữa bộ nhớ chính và bộ nhớ phụ nên nếu dùng hệ điều hành 32bit thì không đủ.

-         Với lưu trữ trên file thì mỗi câu truy vấn phải viết 1 chương trình đặc biệt.

-         Không đảm bảo tính nhất quán.

-         Khi gặp sự cố thì khôi phục lại dữ liệu 1 cách khó khăn.

-         Gặp vấn đề trong bảo mật và kiểm soát truy cập

Trong khi đó chúng ta dùng hệ thống CSDL có những lợi ích sau:

-         Độc lập dữ liệu và truy cập hiệu quả.

-         Giảm thời gian phát triển ứng dụng.

-         Hợp nhất dữ liệu và bảo mật.

-         Quản trị dữ liệu thống nhất.

-         Truy cập đồng thời và khôi phục khi gặp sự cố.

Khi không cần lưu trữ thông tin hay dữ liệu gì thì không nên dùng hệ CSDL.

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

Độc lập dữ liệu logic là người quản trị có thể thay đổi các cấu trúc trong mức khái niệm mà không ảnh hưởng đến mức vật lý. Vì những thay đổi này do người quản trị thay đổi cho phù hợp với sự thay đổi của dữ liệu thực tế liên quan đến hệ thống CSDL, view ra cho người dùng xem nên nó quan trọng hơn độc lập vật lý.

**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.

Độc lập dữ liệu mức vật lý: thay đổi lược đồ vật lý mà không ảnh hưởng đến lược đồ quan niệm và lược đồ ngoài.

Độc lập dữ liệu mức logic: là khả năng thay đổi lược đồ quan niệm mà không phải thay đổi lược đồ ngoài hay chương trình ứng dụng.

**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:

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

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 4 phần sau:

-         Mức nội (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 ngoại (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 ngoại.

**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 nhưng không có công việc làm, 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-pha 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.