Đặng Công Tuấn - 21110709

Chương 1.

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

Một tập hợp các dữ liệu có liên quan với nhau trong một lĩnh vực cụ thể.

**Vd.** Một phần CSDL trong một ứng dụng quản lý SV của một trường đại học bao gồm các dữ liệu liên quan tới SV, các môn học và điểm thi, v.v…

1. Các chức năng của hệ quản trị CSDL

* Cung cấp môi trường tạo lập và duy trì cơ sở dữ liệu ( được cấp DDL )
* Cung cấp môi trường cập nhật và khai thác dữ liệu( tìm kiếm , truy xuất dữ liệu -được cấp DML)
* Cung cấp công cụ điều khiển truy cập vào CSDL ( đảm bảo an ninh, cấp quyền..v..v)

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

Data definition language (DDL) : language used by the database administrator (DBA) to define the database's conceptual (khái niệm ) , internal and external schemas. (tạo mới và sữa đỗi các cấu trúc dữ liệu của các đối tượng trong csdl )

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

Data manipulation language (DML ) : language used to retrive (truy xuất) , insert , delete and modify data. DML statements ( các câu lệnh ) can be entered interactively from a terminal (thiết bị đầu cuối ) or embedded in a general-purpose programming language ( được nhúng vào một ngôn ngữ lập trình có mục đích chung ).

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

+ Ngôn ngữ kiểm soát dữ liệu (DCL ): DCL (Data Control Language) là tập hợp các lệnh chủ yếu giải quyết 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:

* Grant: Lệnh được dùng để cấp quyền cho người dùng truy cập vào cơ sở dữ liệu.
* Revoke: Lệnh được dùng để thu hồi lại quyền truy cập vào cơ sở dữ liệu của người dùng.

Transaction Control Language hay còn gọi là ngôn ngữ kiểm soát giao dịch thường được viết tắt là TCL. Được sử dụng để thay đổi trạng thái của một số dữ liệu.TCL (Transaction Control Language) chỉ có thể sử dụng với các lệnh DML.

Một số câu lệnh trong TCL:

* Commit: Câu lệnh được dùng để lưu tất cả các giao dịch vào cơ sở dữ liệu.
* Rollback: lệnh được dùng để hoàn tác các giao dịch chưa được lưu vào cơ sở dữ liệu.
* Savepoint: lệnh được dùng để chuyển giao dịch trở lại một điểm nhất định mà không cần quay lại toàn bộ giao dịch.

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

-Phát hiện và ngăn chặn sự truy cập không cho phép( đảm bảo an ninh). Chức năng này đáp ứng yêu cầu an toàn và bảo mật thông tin. Ví dụ, không phải ai cũng có thể truy nhập để sửa điểm của sinh viên trong CSDL quản lý sinh viên. Chỉ có những người có thẩm quyền như giảng viên, phòng giáo vụ mới có quyền để làm việc này.

- Duy trì tính nhất quán khi cập nhật dữ liệu. Ví dụ, khi hai người cùng mua một chiếc vé máy bay nhưng chỉ còn lại một chiếc duy nhất. Nếu không có tính nhất quán khi cập nhật dữ liệu có thể cả hai người sẽ mua cùng một chiếc vé.

- điều khiển và tổ chức các hoạt động truy cập

1. Các loại đối tượng sử dụng CSDL?

+ Người quản trị CSDL (DBA – DataBase Administrator)

- Cấp quyền khai thác CSDL

- Xem xét việc sử dụng CSDL và thu hồi tài nguyên

- Bảo mật và thời gian đáp ứng yêu cầu của hệ thống

+ Người thiết kế CSDL (Database designer)

- Hiểu yêu cầu ( cần - want to ) của người dùng và tạo một thiết kế đáp ứng yêu cầu. ( action)

- Xác định dữ liệu cần lưu trong CSDL và cấu trúc CSDL ( định dạng dữ liệu và yêu được lưu trữ ra sao)

+ Phân tích viên hệ thống (System Analyst)

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



+ lập trình viên ứng dụng (Application Programmer)

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

+ Người dùng cuối (End user)

- Người dùng ngẫu nhiên

- Người dùng thường xuyên

- Người dùng cho các công việc phức tạp

- Người dùng một mình

1. Hệ quản trị CSDL (DBMS) là gì?

* A Database Management System (DBMS) là một phần mềm tạo lập và duy trì cơ sở dữ liệu

1. Kể tên một vài DBMS thông dụng

SQL-Server, DB2, Paradox, Informix, Oracle

1. Một hệ CSDL?

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

A diagram of a diagram

Description automatically generated

1. Siêu dữ liệu (meta-data)?

- Là thông tin được lưu trữ trong catalog

-Lưu trữ định nghĩa về cấu trúc dữ liệu và các ràng buộc trong catalog.

1. Ưu điểm của DBMS

* Data independence ( độc lập dữ liệu) and efficient access (DCL&TCL ).
* Recovery from system crashes(TCL)
* Concurrent access (TCL)
* Quick application development
* Data integrity (tính toàn vẹn dữ liệu)and security (TCL&DCL)
* Uniform data administration. ( quản lí dữ liệu thống nhất - DDL)

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 yêu cầu.

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

( danh sách tên nhân viên -> list < String > )

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?

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

1. Lược đồ (schema) là gì?

A *schema* is a description of a particular collection of data, using the a given data model.( gom một nhóm dữ liệu co liên quan với nhau lại )

**Relation schema :** tập tất cả các thuộc tính có trong quan hệ.

**relational database schema**: một tập các lược đồ quan hệ(**Relation schema)** có liên quan với nhau về một lĩnh vực cụ thể.

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?

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

1. Mô hình dữ liệu mức vật lý ?

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

1. Mô hình dữ liệu mức logic ?

- **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.

Vd. Mô hình quan hệ, mô hình phân cấp, mô hình mạng là các mô hình dữ liệu mức logic.

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

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

🡪 Công dụng 3 mức :

+ 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 vật lý***

***- Tính độc lập 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 đỗi lươc đồ quan niệm/ logic. 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 đỗi cấu trúc lưu trữ nhưng không phải thay đổi lược đồ quan niệm.

2. Tính độc lập logic là gì?

***- Tính độc lập logic***: là khả năng thay đổi lược đồ quan niệm không phải thay đỗi 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.

**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?

* Application must stage large datasets between main memory and secondary storage (e.g., buffering, page-oriented access, 32-bit addressing, etc.)
* Special code for different queries
* Must protect data from inconsistency(sự không nhất quán )due to multiple concurrent users
* Crash recovery ( phục hồi )
* Security and access control ( kiểm soát truy cập và bảo mật)

Khi nào không cần sử dụng đến hệ quản trị cơ sở dữ liệu :

+ Dữ liệu cần lưu trữ nhỏ

+ các quan hệ đơn giản và có các relationship đơn giản giữa các quan hệ

+ Sử dụng với ít người dùng và chưa cần đến vấn đề kiểm soát truy cập và bảo mật

2. What is logical data independence and why is it important? ( độc lập dữ liệu mức logic là gì và vì sao nó quan trọng ? )

Độc lập dữ liệu logic là khả năng bảo vệ dữ liệ u khỏi bị ảnh hưởng bởi các thay đổi trong cấu trúc logic của nó. Ví dụ nếu một quan hệ Students( id,name,gpa) được thay thế bởi Students(id,name) và Studentgpa(id,gpa) vì lí do nào đó thì các chương trình ứng dụng hoạt động trên quan hệ Students không cần phải được điều chỉnh thay đỗi để phù hợip với sự thay đỗi từ quan hệ Students

3 Explain the difference between logical and physical data independence.

Độc lập dữ liệu vật lý là các thay đổi trong mức vật lý để nâng cao hiệu năng lưu trữ mà không ảnh hưởng đến mức khái niệm (người quản trị dữ liệu không cần quan tâm nhiều đến mức vật lý cũng như sự thay đổi của mức này)

Ví dụ về các thao tác liên quan độc lập dữ liệu vật lý: Thay đổi về phương pháp nén, phương pháp băm, thiết bị lưu trữ.

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 hoặc mức ngoài .

Ví dụ về các thao tác liên quan độc lập dữ liệu logic: Thêm, Sửa đổi, Xóa một thuộc tính mới

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 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 tưu hoá câu truy vấn 🡪 vì hiểu mới thiết kế ra các lược đồ logic hay vật lí tối ưu để người thực hiện truy vấn có thể dùng lệnh truy vấn tối ưu hơn . Từ đó CSDL luôn được hoạt động tốt nhất

**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.
2. Concurrency control.
3. Crash recovery.
4. A view mechanism.
5. A query language.

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).

- ngôn ngữ truy vấn ( vẫn cần trong trường hợp ông A muốn truy vấn thông tin trong các trường hợp đặc biệt )

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?

Cấu trúc của một DBMS gồm 3 phần sau:

Quản trị không gian trên đĩa

Quản trị vùng đệm ( quản trị bộ nhớ chính khả dụng ) : đưa các trang đĩa vào bộ nhớ khi cần

Quản trị chỉ mục / tập tin/ mẫu tin

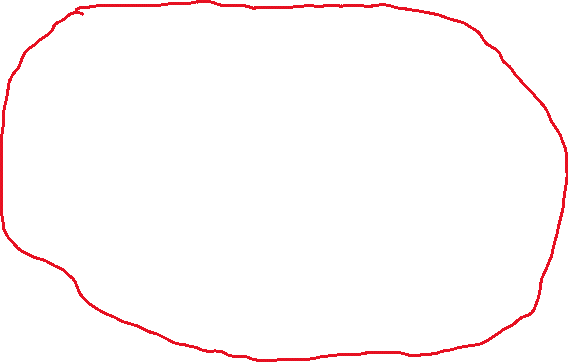
Đánh giá và thực thi câu truy vấn

Quản trị kiểm soát truy cập đồng thời

Quản trị phục hồi dữ liệu

Đánh giá và thực thi câu truy vấn





🡪 thay đỗi phần quản trị không gian trên đĩa ( lưu trữ được các bye của string ) và thay đỗi phần quản trị vùng đệm ( đưa các trang từ đĩa vào vùng nhớ đệm khi cần )