Chương 1. Tổng quan về Hệ quản trị cơ sở dữ liệu

Nội dung

1.1. Khái niệm Hệ quản trị cơ sở dữ liệu
1.2. Các ưu điểm của Hệ quản trị cơ sở dữ liệu
1.3. Mô tả và lưu trữ dữ liệu trong Hệ quản trị cơ sở dữ liệu
1.4. Truy vấn trong Hệ quản trị cơ sở dữ liệu
1.5. Quản lý giao tác
1.6. Cấu trúc của Hệ quản trị cơ sở dữ liệu
1.7. Các đối tượng giao dịch với Cơ sở dữ liệu
1.8. Một số Hệ quản trị Cơ sở dữ liệu thông dụng

1.1. Khái niệm Hệ quản trị cơ sở dữ liệu

- Cơ sở dữ liệu
- Hệ cơ sở dữ liệu
- Hệ quản trị cơ sở dữ liệu

Cơ sở dữ liệu (Database)

- Cơ sở dữ liệu (CSDL): Là một bộ sưu tập dữ liệu, thường mô tả hoạt động của một hoặc nhiều tổ chức có liên quan. Ví dụ: CSDL của một trường đại học có thể chứa đựng các thông tin:
  - Các thực thể: sinh viên (students), giảng viên (faculty), khóa học (courses), phòng học (classrooms), …
  - Các quan hệ giữa các thực thể: sinh viên đăng ký các khóa học, giảng viên giảng dạy các khóa học, sử dụng các phòng học cho các khóa học, …
- Được tổ chức và lưu trữ trên các thiết bị nhớ thứ cấp để đáp ứng nhu cầu khai thác thông tin của nhiều người sử dụng với nhiều mục đích khác nhau

Hệ cơ sở dữ liệu

- Là một hệ thống gồm 4 thành phần:
  - Cơ sở dữ liệu
  - Những đối tượng sử dụng CSDL
  - Hệ quản trị CSDL
  - Phần cứng

Hệ quản trị cơ sở dữ liệu (Database Management System – DBMS)

- Là phần mềm cho phép tạo lập, sử dụng và quản trị các CSDL
- Các chức năng chính của DBMS:
- Cung cấp môi trường tạo lập CSDL
- Cung cấp môi trường cập nhật và khai thác dữ liệu
- Cung cấp các công cụ kiểm soát, điều khiển truy cập vào CSDL

1.2. Các ưu điểm của DBMS

- Tính độc lập dữ liệu
- Cho phép truy cập dữ liệu hiệu quả
- Tính toàn vẹn và bảo mật dữ liệu
- Quản trị dữ liệu
- Truy cập đồng thời và khả năng phục hồi sau sự cố
- Giảm thiểu thời gian phát triển ứng dụng

1.3. Mô tả và lưu trữ dữ liệu trong DBMS

- Mô hình dữ liệu
- Các mức trừu tượng hóa trong DBMS
- Tính độc lập dữ liệu

Mô hình dữ liệu (1)

- Mô hình dữ liệu (data model):
  - Là tập hợp các cấu trúc mô tả dữ liệu ở mức cao trong đó ẩn chứa nhiều chi tiết lưu trữ ở mức thấp
  - Dùng để mô tả dữ liệu, các mối quan hệ của dữ liệu, các ràng buộc trên dữ liệu của CSDL
    Ví dụ: Mô hình dữ liệu quan hệ
- DBMS cho phép người dùng định nghĩa dữ liệu được lưu trữ dưới dạng mô hình dữ liệu

Mô hình dữ liệu (2)

- Mô hình quan hệ (Realational model):
  - Kiến trúc mô tả dữ liệu trung tâm của mô hình này là *quan hệ* - có thể được xem như một tập các bản ghi
  - Một mô tả của dữ liệu về mô hình quan hệ được gọi là lược đồ. Lược đồ xác định rõ tên quan hệ, tên các thuộc tính (và có thể bao gồm kiểu của từng thuộc tính). Ví dụ: Students (sid: string, name: string, login: string, age: integer, gpa: real)

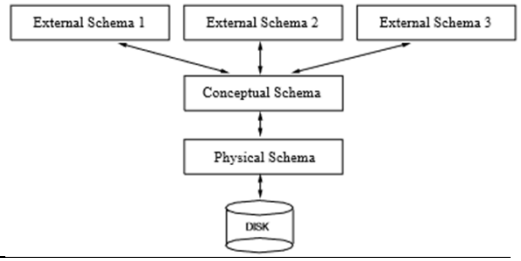
sid	*name*	*login*	*age*	*gpa*
53666	Jones	jones@cs	18	3.4
53688	Smith	smith@ee	18	3.2
53650	Smith	smith@math	19	3.8
53831	Madayan	madayan@music	11	1.8
53832	Guldu	guldu@music	12	2.0

Mô hình dữ liệu (3)

- Các mô hình dữ liệu khác:
  - Mô hình mạng (Network model)
  - Mô hình phân cấp (Hierarchical model)
  - Mô hình hướng đối tượng (Object oriented model)
  - Mô hình quan hệ đối tượng (Object-relational model)
  - …

Các mức trừu tượng hóa của DBMS

- Mô tả của một CSDL bao gồm một lược đồ ở 3 mức độ trừu tượng: mức vật lý (mức trong - physical schema), mức khái niệm (mức logic - conceptual schema), mức ngoài (mức khung nhìn – external schema)



Tính độc lập dữ liệu

- Một ưu điểm của việc sử dụng DBMS là tính độc lập dữ liệu
- Các chương trình ứng dụng được cách ly khỏi sự thay đổi trong cấu trúc và cách thức lưu trữ dữ liệu
- Sự độc lập dữ liệu được thực hiện thông qua việc sử dụng ba mức độ trừu tượng dữ liệu, đặc biệt là dựa trên việc sử dụng lược đồ khái niệm và lược đồ ngoài

1.4. Truy vấn trong DBMS (1)

- Các hệ thống CSDL quan hệ cho phép dễ dàng đặt ra một lớp câu hỏi phong phú
  Ví dụ: Với CSDL của trường đại học, có thể đặt ra một số câu hỏi
  1. Tên của sinh viên có id 123456 là gì?
  2. Lương trung bình của các giáo sư giảng dạy cho khóa học CS564 là bao nhiêu?
  3. Có bao nhiêu sinh viên đăng ký khóa học CS564?
  4. Có sinh viên nào có GPA<3.0 đăng ký khóa học CS564 không?

  …
- Những câu hỏi liên quan đến dữ liệu được lưu trữ trong DBMS được gọi là các truy vấn

1.4. Truy vấn trong DBMS (2)

- DBMS cung cấp một ngôn ngữ chuyên biệt, được gọi là ngôn ngữ truy vấn
- Tối ưu hóa và đánh giá truy vấn:
- Là một vấn đề luôn được chú trọng quan tâm
- Thông thường, hiệu quả của việc đánh giá truy vấn được xác định chủ yếu qua cách thức lưu trữ dữ liệu về mặt vật lý, các chỉ mục có thể được sử dụng để tăng tốc cho các truy vấn
- DBMS cho phép người dùng tạo, sửa đổi và truy vấn dữ liệu thông qua ngôn ngữ thao tác dữ liệu (DML)

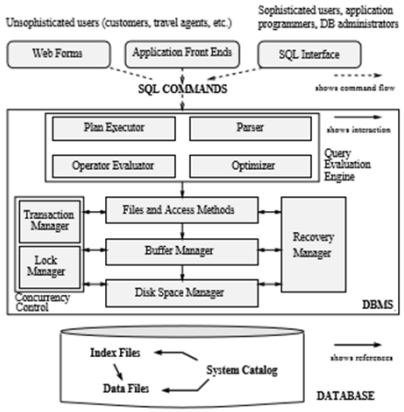
1.5. Quản lý giao tác (1)

- Giao tác (Transaction) là bất kỳ một hành động thực thi nào của chương trình người dùng trong DBMS
- Thực thi đồng thời các giao tác: DBMS lập kế hoạch truy cập đồng thời vào dữ liệu sao cho mỗi người dùng có thể bỏ qua thực tế là những người dùng khác cũng đang truy cập dữ liệu đồng thời một cách an toàn
- Giao tác không hoàn thành và các sự cố hệ thống: Các giao tác có thể bị gián đoạn trước khi thực thi xong vì nhiều lý do (ví dụ: sự cố hệ thống). DBMS phải đảm bảo tất cả các thay đổi được thực hiện bởi các giao tác không hoàn thành như vậy sẽ được gỡ bỏ khỏi cơ sở dữ liệu

1.5. Quản lý giao tác (2)

- Để điều khiển việc truy xuất đồng thời và phục hồi sau sự cố:
  - Mọi đối tượng được đọc hoặc ghi bởi một giao tác trước tiên được khóa trong chế độ độc quyền hoặc chia sẻ tương ứng
  - Để bảo trì nhật ký hiệu quả, DBMS có khả năng chọn lọc một tập các trang trong bộ nhớ chính để đẩy vào đĩa
  - Việc kiểm tra định kỳ làm giảm thời gian cần thiết để phục hồi sau sự cố

1.6. Cấu trúc của DBMS



1.7. Các đối tượng giao dịch với CSDL

- Người dùng cuối (NDC): lưu trữ, sử dụng CSDL thông qua các ứng dụng hoặc dựa trên DBMS
- Người lập trình ứng dụng CSDL: là người viết các chương trình ứng dụng (các gói hỗ trợ truy cập dữ liệu) cho phép NDC sử dụng CSDL
- Người quản trị CSDL (Database Administrator): là người thu thập dữ liệu, thiết kế và bảo trì CSDL (thiết kế lược đồ mức khái niệm và vật lý, bảo mật dữ liệu, đảm bảo dữ liệu luôn sẵn sàng và khả năng phục hồi sau sự cố, hiệu chỉnh CSDL)

1.8. Một số Hệ quản trị CSDL thông dụng

- MS Access
- MS SQL Server
- DB2
- My SQL
- PostgreSQL
- Oracle
- SQLite
- …