



Chương 3 Thiết kế cơ sở dữ liệu vật lý

Giáo viên: Nguyễn Mậu Uyên

Bộ môn: Hệ thống thông tin

Biên soạn: Đỗ Thị Mai Hương

Kiểm tra giữa kỳ

- Kiểm tra giữa kỳ về vấn đề thiết kế cơ sở dữ liệu logic

Thiết kế cơ sở dữ liệu vật lý

LOGO

- Thiết kế cơ sở dữ liệu vật lý là quá trình chọn cấu trúc lưu trữ và truy xuất dữ liệu cho cơ sở dữ liệu.
- Cần quan tâm:
 - Các vấn đề liên quan tới khối lượng dữ liệu mà CSDL đang xây dựng cần phải đáp ứng.
 - Tinh chỉnh thêm thiết kế các quan hệ.
- **Mục tiêu:**
 - Đạt được hiệu suất mong muốn.
 - Đảm bảo tính toàn vẹn, bảo mật và khả năng khôi phục dữ liệu.

Nội dung

A decorative header image featuring a close-up of a fountain pen nib and a small logo with the word "LOGO" in green capital letters.

- ❖ Phân tích kích thước và nhu cầu sử dụng dữ liệu
- ❖ Thiết kế các trường
- ❖ Tổ chức các tệp vật lý
- ❖ Lựa chọn và sử dụng các chỉ mục
- ❖ Thiết kế an toàn bảo mật
- ❖ Bài tập vận dụng và thực hành

Phân tích kích thước và nhu cầu sử dụng dữ liệu

LOGO

- ❖ Kích thước của dữ liệu đại diện cho độ phức tạp của hoạt động nghiệp vụ, nên được đánh giá trong vài năm tùy vào từng ứng dụng cụ thể.
- ❖ Nhu cầu sử dụng hay tần xuất truy cập được đánh giá dựa trên thời gian của các sự kiện, độ lớn của các giao dịch, các hoạt động liên quan đến báo cáo, truy vấn tìm kiếm dữ liệu.

Thiết kế các trường

LOGO

- ❖ Trường là đơn vị nhỏ nhất của một dữ liệu ứng dụng.
- ❖ Đối với mỗi trường dữ liệu, cần quan tâm đến các vấn đề sau:
 - Kiểu dữ liệu (hay kiểu lưu trữ).
 - Các kiểm soát toàn vẹn dữ liệu.

Thiết kế các trường

❖ Lựa chọn kiểu dữ liệu (hay kiểu lưu trữ)

Mỗi một trường dữ liệu được lưu trữ trong CSDL bắt buộc phải thuộc một kiểu dữ liệu mà HQTCSDL đó hỗ trợ. Việc lựa chọn một kiểu dữ liệu phải thỏa mãn:

- Tối thiểu hóa không gian lưu trữ.
- Biểu diễn được tất cả các khả năng của giá trị cần lưu trữ.
- Nâng cao tính toàn vẹn của dữ liệu.
- Hỗ trợ tất cả các thao tác trên dữ liệu.

Kiểm soát các toàn vẹn dữ liệu

LOGO

❖ Một số kiểm soát toàn vẹn dữ liệu:

- Kiểu dữ liệu: giới hạn kiểu dữ liệu của trường có thể nhận được.
- Độ rộng của trường dữ liệu.
- Giá trị mặc định. Giảm thời gian thời gian nhập liệu, giảm tỉ lệ mắc lỗi.
- Điều kiện khoảng.
- Điều kiện giá trị Null
- Toàn vẹn tham chiếu

Tổ chức các tệp vật lý



LOGO

❖ Dữ liệu được lưu trong các thiết bị lưu trữ ngoài như đĩa, băng từ, và chỉ được nạp vào trong bộ nhớ chính khi quá trình xử lý cần đến nó.

❖ Các thiết bị lưu trữ

- Cache
- Main memory
- Flash memory
- Magnetic-disk
- Optical storage
- Tape storage
- Main memory

Tổ chức các tệp vật lý..

- ❖ Cache: Là thiết bị lưu trữ nhanh nhất và cũng có giá đắt nhất. Bộ nhớ cache tương đối nhỏ, được quản lý bởi phần cứng của hệ thống máy tính.
- ❖ Main Memory: Dung lượng có thể lên tới vài Gigabytes trên máy tính cá nhân và hàng trăm Gigabytes trên các máy Server. Dữ liệu lưu trữ trên Main Memory sẽ mất khi mất nguồn hay lỗi hệ thống.
- ❖ Flash Memory: Khác với Main Memory, dữ liệu lưu trên thiết bị lưu trữ loại này vẫn còn ngay cả khi mất nguồn hay lỗi hệ thống. Flash Memory có giá thấp hơn Main Memory, thường sử dụng rộng rãi để lưu trữ dữ liệu trên các thiết bị như camera, máy nghe nhạc, điện thoại, USB, ...

Tổ chức các tệp vật lý..

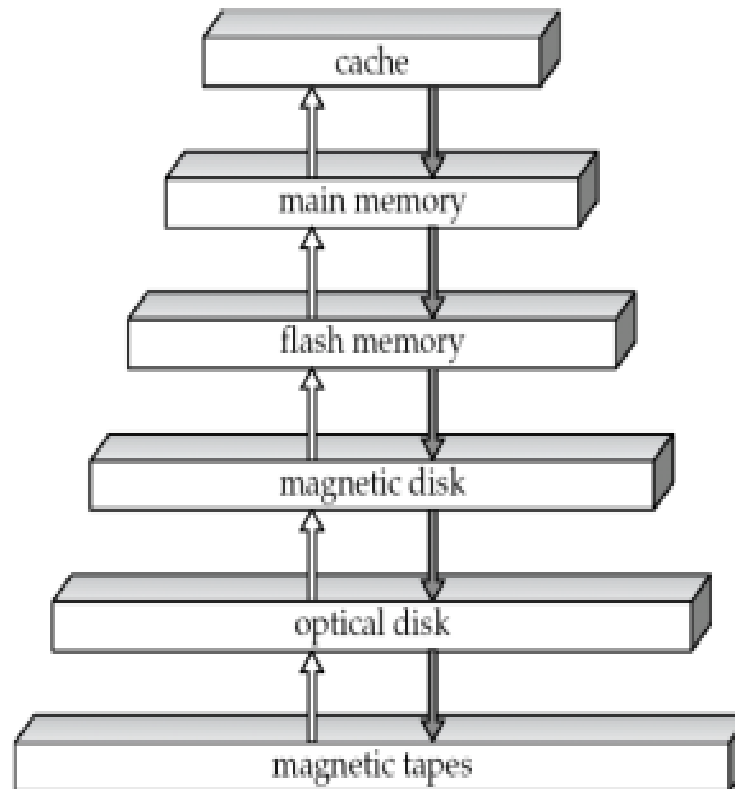
- ❖ Magnetic – disk storage: Là phương tiện chủ yếu lưu trữ dữ liệu lâu dài, thường toàn bộ cơ sở dữ liệu được lưu trữ trên đĩa từ. Để xử lý dữ liệu trên đĩa từ, dữ liệu phải được chuyển sang bộ nhớ chính (Main Memory). Hiện nay, dung lượng của đĩa từ có thể từ vài chục Gigabytes đến hàng Terabyte.
- ❖ Optical storage: Các đĩa CD (Compact Disk) với dung lượng khoảng 700 Megabytes là dạng phổ biến nhất của các thiết bị lưu trữ loại này. Và khoảng 4.7 đến 8.5 Gigabytes với đĩa DVD (Digital Video Disk). Đĩa hai mặt có dung lượng lên tới 17 Gigabytes.

Tổ chức các tệp vật lý..

- ❖ Tape storage: Là thiết bị chủ yếu dùng để sao lưu dữ liệu. Thiết bị lưu trữ loại này có giá rẻ hơn đĩa từ nhưng tốc độ truy xuất chậm do phải đọc tuần tự. Khả năng lưu trữ của băng từ là rất lớn (40 – 300 Gigabytes).
- ❖ Các thiết bị lưu trữ được tổ chức phân cấp theo tốc độ và giá cả: Ở mức cao là các thiết bị lưu trữ có khả năng truy xuất nhanh hơn nhưng có giá đắt hơn.

Tổ chức các tệp vật lý..

LOGO



Phân cấp các thiết bị lưu trữ

Tổ chức các tệp vật lý..

- ❖ Trang: là một đơn vị thông tin đọc hoặc ghi từ đĩa. Kích thước của mỗi trang là một tham biến của DBMS, có giá trị điển hình là 4KB hoặc 8KB.
- ❖ Chi phí của một trang I/O (input từ đĩa tới bộ nhớ chính và output từ bộ nhớ chính ra đĩa) có ảnh hưởng lớn đến chi phí của các phép toán cơ sở dữ liệu, và các hệ thống cơ sở dữ liệu thường tối ưu hoá rất cẩn thận chi phí này.
- ❖ Mỗi bản ghi trong một file có một định danh gọi là record id hay ngắn gọn là rid. Rid có thể được sử dụng để xác định địa chỉ của trang chứa bản ghi này.

Tổ chức các tệp vật lý..

❖ Quản lý vùng đệm:

- Dữ liệu được đọc vào bộ nhớ để xử lý, và viết vào đĩa để lưu trữ lâu dài bằng một phần mềm gọi là quản lý vùng đệm.
- Khi lớp files and access methods (lớp file) cần xử lý một trang nào đó, nó yêu cầu hệ thống quản lý vùng đệm nạp trang này, xác định rid của trang. Hệ thống quản lý vùng đệm nạp trang này từ đĩa nếu nó không sẵn sàng trong bộ nhớ.

Tổ chức các tệp vật lý..

- ❖ Những khoảng trống trên đĩa được quản lý bởi disk space manager .
- ❖ Khi lớp files and access methods cần thêm các khoảng trống để nhận thêm các bản ghi mới trong một file nào đó, nó yêu cầu disk space manager định vị một trang đĩa bổ sung cho file này, nó cũng thông báo cho disk space manager khi không cần đến một trang đĩa nào đó của nó.
- ❖ Disk space manager giữ lại dấu vết của các trang đang được lớp file sử dụng; nếu một trang nào không còn cần cho lớp file nữa, space manager sẽ ghi nhớ điều này, và tái sử dụng các vùng trống nếu lớp file yêu cầu một trang mới sau đó.

Tổ chức tệp tin

LOGO

❖ Bản ghi có độ dài cố định

```
type instructor = record
    ID varchar (5);
    name varchar(20);
    dept_name varchar (20);
    salary numeric (8,2);
end
```

Tổ chức tệp tin

❖ Bản ghi có độ dài cố định:

■ Ví dụ:

record 0	10101	Srinivasan	Comp. Sci.	65000
record 1	12121	Wu	Finance	90000
record 2	15151	Mozart	Music	40000
record 3	22222	Einstein	Physics	95000
record 4	32343	El Said	History	60000
record 5	33456	Gold	Physics	87000
record 6	45565	Katz	Comp. Sci.	75000
record 7	58583	Califieri	History	62000
record 8	76543	Singh	Finance	80000
record 9	76766	Crick	Biology	72000
record 10	83821	Brandt	Comp. Sci.	92000
record 11	98345	Kim	Elec. Eng.	80000

Tổ chức tệp tin

A decorative header image featuring a close-up of a fountain pen nib and a small logo with the word "LOGO" in green capital letters.

❖ Bản ghi có độ dài cố định:

❖ Vấn đề nảy sinh:

- Nếu kích thước của một khối không chia hết cho 53 thì một số bản ghi sẽ vượt quá một khối đĩa (một bản ghi có thể nằm trong hai khối đĩa. Phải truy xuất tới hai khối đĩa để đọc hay ghi một bản ghi.
- **Cách giải quyết:** Lưu mỗi bản ghi trong một khối. Phần khoảng trống dư lại ở cuối khối sẽ bị bỏ qua.

Tổ chức tệp tin

LOGO

❖ Bản ghi có độ dài cố định:

- ❖ Khó khăn khi xóa một bản ghi với cấu trúc lưu trữ kiểu này. Khoảng trống bị chiếm bởi bản ghi đã xóa hoặc là phải được lấp đầy bởi bản ghi khác hoặc là phải được đánh dấu xóa.

❖ Cách giải quyết:

- Dịch chuyển các bản ghi sau bản ghi bị xóa về trước.
- => Khoảng không gian còn trống luôn ở cuối khối.
- Một cách tiếp cận đơn giản hơn là di chuyển bản ghi cuối cùng.
 - Sử dụng phương pháp đánh dấu các bản ghi bị xóa: đánh dấu vị trí các bản ghi bị xóa và chờ thao tác thêm bản ghi tiếp theo sẽ sử dụng lại khoảng không gian bị trống đó

Tổ chức tệp tin

A decorative header image featuring a close-up of a fountain pen writing on a piece of paper. The image is partially obscured by a green banner at the top and a green 'LOGO' watermark in the top right corner.

❖ Bản ghi có độ dài cố định:

■ Nhận xét:

Cấu trúc bản ghi có độ dài cố định việc cài đặt các thao tác thêm hay xóa bản ghi là đơn giản do khoảng không gian để lại bởi bản ghi bị xóa cũng vừa bằng khoảng không gian cần thiết cho bản ghi thêm vào.

Tổ chức tệp tin

A decorative header image featuring a close-up of a fountain pen nib and a small logo with the word "LOGO" in green capital letters.

❖ Bản ghi có độ dài thay đổi

- Lưu trữ nhiều kiểu bản ghi trong một tệp tin.
- Các kiểu bản ghi với độ dài thay đổi trên một hay nhiều thuộc tính.
- Các kiểu bản ghi cho phép các trường được lặp lại, như mảng hay tập hợp.

Tổ chức các bản ghi trong tệp tin

LOGO

- ❖ Có một vài cách tổ chức các bản ghi trong file như sau:
 - Tổ chức tệp tin Heap: Một bản ghi được lưu ở bất kỳ đâu trong file. Không có thứ tự của các bản ghi. Thông thường một file lưu một quan hệ.
 - Tổ chức tệp tin tuần tự: Các bản ghi được tổ chức tuần tự theo giá trị của một khóa tìm kiếm (search key) trong mỗi bản ghi.
 - Tổ chức tệp tin băm: Sử dụng một hàm băm tính toán trên một số thuộc tính của bản ghi và dùng kết quả để đó để xác định khối đĩa mà bản ghi được lưu trữ.

Tổ chức các bản ghi trong tệp tin

❖ Tổ chức tệp tin Heap

- Là kiểu tổ chức file đơn giản nhất, các bản ghi không được sắp xếp thứ tự.
- Việc thêm mới một bản ghi là đơn giản: Tìm khối đĩa cuối cùng của file, sao chép khối đĩa vào bộ đệm, thêm mới bản ghi rồi ghi lại vào đĩa. Địa chỉ của khối đĩa cuối cùng được cập nhật lại vào file header.
- Khi tìm kiếm phải sử dụng phương pháp tìm kiếm tuần tự: Đọc lần lượt từng khối đĩa vào bộ nhớ chính rồi tiến hành tìm kiếm các bản ghi. Như vậy nếu file gồm n khối đĩa thì thời gian tìm kiếm trung bình sẽ là $n/2$.

Tổ chức các bản ghi trong tệp tin

LOGO

❖ Tổ chức tệp tin tuần tự

- Tổ chức file tuần tự được thiết kế để xử lý hiệu quả các bản ghi theo thứ tự được sắp dựa trên một khoá tìm kiếm nào đó.
- Một khóa tìm kiếm là một hay một nhóm thuộc tính bất kỳ, không nhất thiết phải là khóa chính hay siêu khóa.
- Sử dụng kỹ thuật con trỏ để truy xuất nhanh chóng tới bản ghi theo thứ tự khoá tìm kiếm. Con trỏ trong mỗi bản ghi sẽ trỏ tới bản ghi tiếp theo trong thứ tự khoá tìm kiếm

Tổ chức các bản ghi trong tệp tin

❖ Tổ chức tệp tin tuần tự (tt)

- Tổ chức file tuần tự cho phép đọc các bản ghi theo thứ tự được sắp thuận lợi cho mục đích hiển thị cũng như cho các thuật toán xử lý truy vấn (query – processing algorithms).
- Tuy vậy, khó khăn gặp phải khi tổ chức file tuần tự là việc duy trì thứ tự vật lý của các bản ghi trong file khi xảy ra các thao tác thêm, xóa bản ghi.

Bài tập

- ❖ Bài 1: Xác định các kiểu trường vật lý trong cơ sở dữ liệu «Hệ thống quản lý bán sách online»
- ❖ Bài 2: Xác định các kiểu trường vật lý trong cơ sở dữ liệu «Hệ thống quản lý đào tạo LQDUNI»
- ❖ Bài 3: Viết đoạn lệnh lặp cho phép insert 10000 bản ghi vào 1 bảng chính và 100000 bản ghi vào bảng phụ liên kết
- ❖ Bài 4: Viết câu lệnh kết nối lấy dữ liệu từ hai bảng đã thực hiện