Tổ chức dữ liệu vật lý

Nguyễn Hồng Phương
phuongnh@soict.hut.edu.vn
http://is.hut.edu.vn/~phuongnh

Bộ môn Hệ thống thông tin
Viện Công nghệ thông tin và Truyền thông
Đại học Bách Khoa Hà Nội

1

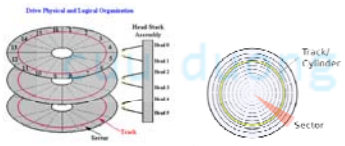
Nội dung

- 1. Mô hình tổ chức bộ nhớ ngoài
- 2. Tổ chức tệp đống
- 3. Tổ chức tệp băm
- 4. Tổ chức tệp chỉ dẫn
- 5. Cây cân bằng

2

1. Mô hình tổ chức bộ nhớ ngoài

- Bộ nhớ ngoài (bộ nhớ thứ cấp): đĩa từ, băng từ,...



- Đĩa được chia thành các khối vật lý (sector) - 512 byte đến 4096 byte được đánh địa chỉ khối gọi là địa chỉ tuyệt đối
- Mỗi tệp dữ liệu chiếm 1 hoặc nhiều khối
- Mỗi khối chứa 1 hoặc nhiều bản ghi

3

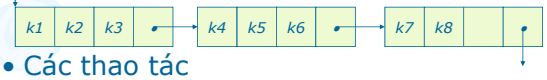
1. Mô hình tổ chức bộ nhớ ngoài

- Thao tác với dữ liệu của tệp thông qua địa chỉ tuyệt đối của các khối.
- Các bản ghi đều có địa chỉ:
  - địa chỉ tuyệt đối của byte đầu tiên
  - địa chỉ khối và số byte tính từ đầu khối đến vị trí đầu bản ghi
- Địa chỉ của các bản ghi/khối được lưu ở 1 tệp => sử dụng con trỏ (pointer) để truy cập dữ liệu của tệp.

4

2. Tổ chức tệp đống (Heap file)

- Tổ chức dữ liệu
  - Bản ghi lưu trữ kế tiếp trong các khối, không tuân theo một thứ tự đặc biệt nào.



- Các thao tác
  - Tìm kiếm một bản ghi: tìm kiếm một bản ghi có giá trị khóa cho trước => quét toàn bộ tệp.
  - Thêm một bản ghi: thêm bản ghi mới vào sau bản ghi cuối cùng

5

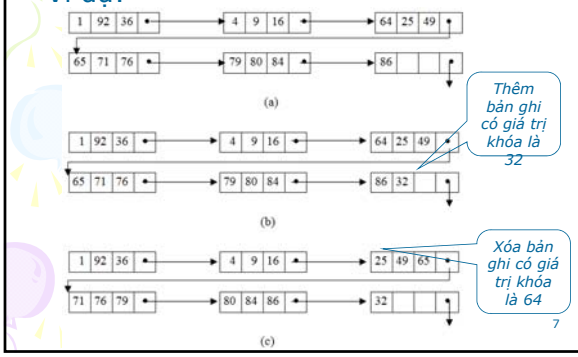
2. Tổ chức tệp đống (Heap file)

- Các thao tác (tiếp)
  - Xóa một bản ghi: thao tác xóa bao hàm thao tác tìm kiếm. Nếu có bản ghi cần xóa thì nó sẽ được đánh dấu là xóa => hệ thống cần tổ chức lại đĩa định kỳ.
  - Sửa một bản ghi: tìm bản ghi rồi sửa một hay nhiều trường.

6

2. Tổ chức tệp đống (Heap file)

- Ví dụ:



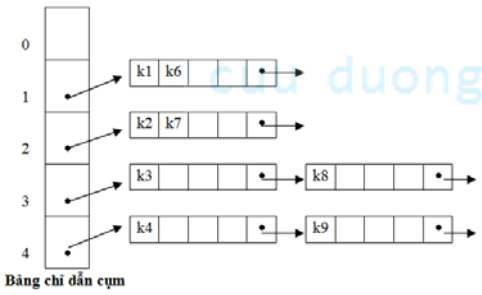
7

3. Tổ chức tệp băm (Hashed files)

- Hàm băm: h(x) nhận một giá trị trong đoạn [0,k], ví dụ: h(x)=x mod k
- Tổ chức tệp dữ liệu
  - Phân chia các bản ghi vào các cụm.
  - Mỗi cụm gồm một hoặc nhiều khối.
  - Mỗi khối chứa số lượng bản ghi cố định.
  - Tổ chức lữu trữ dữ liệu trong mỗi cụm áp dụng theo tổ chức đống
- Tiêu chí chọn hàm băm: phân bố các bản ghi tương đối đồng đều theo các cụm.

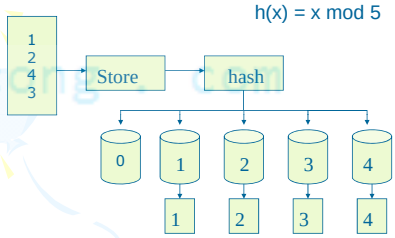
8

3. Tổ chức tệp băm (Hashed files)



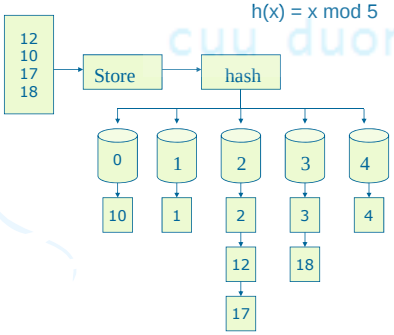
9

3. Tổ chức tệp băm (Hashed files)



10

3. Tổ chức tệp băm (Hashed files)



11

3. Tổ chức tệp băm (Hashed files)

- Các thao tác
  - Tìm kiếm một bản ghi: để tìm bản ghi có khóa x, tính h(x) sẽ được cụm chứa bản ghi, sau đó tìm kiếm theo tổ chức đống.
  - Thêm một bản ghi: thêm 1 bản ghi có giá trị khóa là x.
    - nếu trong tệp đã có một bản ghi có trùng khóa x =>bản ghi mới sai (vì khóa là duy nhất!)
    - nếu không có bản ghi trùng khóa, bản ghi được thêm vào khối còn chỗ trống đầu tiên trong cụm, nếu hết chỗ thì tạo khối mới.

12

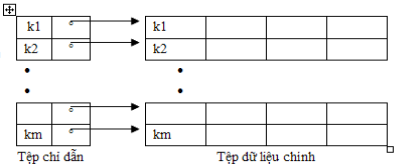
3. Tổ chức tệp băm (Hashed files)

- Xóa một bản ghi: tìm kiếm bản ghi rồi xóa
- Sửa đổi một bản ghi:
  - nếu trường cần sửa có tham gia vào trong khóa thì việc sửa sẽ là loại bỏ bản ghi này và thêm mới 1 bản ghi (bản ghi có thể thuộc vào 1 cụm khác)
  - nếu trường cần sửa không thuộc khóa: tìm kiếm rồi sửa. Nếu bản ghi không tồn tại thì xem như có lỗi.

13

4. Tổ chức tệp chỉ dẫn(Indexed Files)

- Giả sử giá trị các khóa của các bản ghi được sắp xếp tăng dần.
- Tệp chỉ dẫn được tạo bằng cách chọn các giá trị khóa trong các bản ghi
- Tệp chỉ dẫn bao gồm các cặp (k,d), trong đó k là giá trị khoá của bản ghi đầu tiên, d là địa chỉ của khối (hay con trỏ khối).



4

4. Tổ chức tệp chỉ dẫn(Indexed Files)

- Tìm kiếm trên tệp chỉ dẫn
  - Cho một giá trị khóa k_i, tìm một bản ghi (k_m,d) trong tệp chỉ dẫn sao cho $k_m <= k_i$ và:
    - hoặc (k_m,d) là bản ghi cuối cùng trong tệp chỉ dẫn
    - hoặc bản ghi tiếp theo (k_{m+1},d') thỏa mãn $k_i < k_{m+1}$
  - Khi đó, chúng ta nói k_m phủ k_i
  - Tìm kiếm này có thể là:
    - tuần tự
    - nhị phân

15

4. Tổ chức tệp chỉ dẫn(Indexed Files)

- Các thao tác
  - Tìm kiếm một bản ghi
  - Thêm một bản ghi: xác định khối i sẽ chứa bản ghi đó
    - nếu trong khối i còn chỗ thì đặt bản ghi này vào đúng chỗ theo thứ tự sắp xếp của khóa, dồn toa các bản ghi đằng sau nó.
    - nếu khối i hết chỗ thì việc thêm này sẽ đẩy bản ghi cuối cùng trong khối sang làm bản ghi đầu tiên của khối tiếp theo i+1 => sửa bản ghi chỉ dẫn tương ứng
    - nếu bản ghi mới này có giá trị khóa lớn hơn tất cả mọi khóa trong tệp dữ liệu chính và không còn chỗ thì tạo thêm một khối mới.

16

4. Tổ chức tệp chỉ dẫn(Indexed Files)

- Xóa một bản ghi: giống như thêm một bản ghi, nếu xóa mà tạo thành 1 khối rỗng, khi đó có thể loại bỏ cả khối đó.
- Sửa một bản ghi:
  - Sử dụng thủ tục tìm kiếm để xác định bản ghi cần sửa
  - nếu các trường cần sửa không phải là khóa thì sửa bình thường
  - nếu các trường cần sửa tham gia vào khóa thì quá trình sửa sẽ là quá trình thêm và xóa 1 bản ghi.

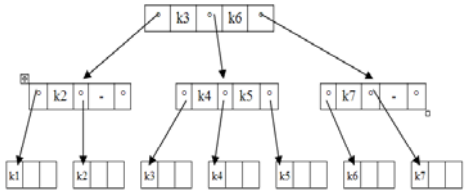
17

5. Cây cân bằng(Balanced-trees)

- B-tree được tổ chức theo cấp m, có các tính chất sau đây:
  - Gốc của cây hoặc là một nút lá hoặc ít nhất có hai con.
  - Mỗi nút (trừ nút gốc và nút lá) có từ [m/2] đến m con.
  - Mỗi đường đi từ nút gốc đến bất kỳ nút lá nào đều có độ dài như nhau.

18

5. Cây cân bằng(Balanced-trees)



- Cấu trúc của mỗi nút trong B-cây có dạng (p_0, k_1, p_1, k_2,...,k_n, p_n) với p_i (i=1..n) là con trỏ trỏ tới khối i của nút có k_i là khoá đầu tiên của khối đó. Các khoá k trong một nút được sắp xếp theo thứ tự tăng dần.

19

5. Cây cân bằng(Balanced-trees)

- Mọi khoá trong cây con, trỏ bởi con trỏ p_0 đều nhỏ hơn k_1;
- Mọi khoá trong cây con, trỏ bởi con trỏ p_i đều nhỏ hơn k_{i+1}.
- Mọi khoá trong cây con, trỏ bởi con trỏ p_n đều lớn hơn k_n.

20

5. Cây cân bằng(Balanced-trees)

- Các thao tác
  - Tìm kiếm một bản ghi: xác định đường dẫn từ nút gốc tới nút lá chứa bản ghi này
  - Thêm một bản ghi:
    - Xác định vị trí nút lá sẽ chứa bản ghi này (như tìm kiếm)
    - Nếu còn chỗ thì thêm bình thường
    - Nếu hết chỗ thì phải tạo thêm nút lá mới, chuyển nửa dữ liệu cuối của nút lá hiện tại sang nút mới, sau đó thêm bản ghi mới này vào vị trí phù hợp nút lá hiện tại hoặc nút mới tạo
    - Rất có khả năng "động chạm" đến nút cha,….nút gốc.

21

5. Cây cân bằng(Balanced-trees)

- Loại bỏ 1 bản ghi:
  - Dùng thủ tục tìm kiếm một bản ghi để xác định nút L có thể chứa bản ghi đó.
  - Rất có khả năng "động chạm" đến nút cha,…,nút gốc.

22

Kết luận

- Tổ chức tệp chỉ dẫn:
  - được áp dụng phổ biến
  - Với các ứng dụng yêu cầu cả xử lý tuần tự và truy nhập trực tiếp đến các bản ghi
  - Hiệu năng sẽ giảm khi kích thước tệp tăng =>chỉ dẫn B-cây
- Tổ chức băm:
  - Dựa trên 1 hàm băm, cho phép tìm thấy địa chỉ khoản mục dữ liệu một cách trực tiếp
  - Hàm băm tốt? Phân bố các bản ghi đồng đều trong các cụm

23

24

Lời hay ý đẹp

Bản chất của tình bạn chân thật là khoan dung với những lỗi nhỏ của bạn

David Storey

25

cuu duong than cong . com