Tìm hiểu cấu trúc B - Tree

B - Tree là một cấu trúc cân bằng mà lưu trữ dữ liệu đã được sắp xếp và cho phép tìm kiếm, truy cập, chèn, xóa với độ phức tạp thời gian logarithmic. Thực chất B - Tree là một thế hệ sau của cây tìm kiếm nhị phân với một nút có nhiều hơn 2 cây con. Khác với cây tìm kiếm nhị phân cân bằng, B-Tree tối ưu hóa hệ thống bằng cách ghi và đọc một lượng lớn dữ liệu. Ý tưởng chính của B-Tree là giảm thời gian truy cập ổ cứng, hầu hết các hoạt động của cây (như tìm kiếm, chèn, xóa, tìm max, tìm min,…) đều yêu cầu truy cập O(h) với h là chiều cao của cây. B-Tree là một cây mở rộng, chiều cao cây của B-Tree được giảm bớt đi nhiều nhờ vào việc lưu trữ một số lượng lớn các khóa trong một nút B-Tree, hay còn gọi là một trang. Kích cỡ của một trang được lưu như một block trong ổ cứng. Vì thế chiều cao của B-Tree giảm, các hoạt động đọc từ ổ đĩa giảm hơn hẳn so với cây nhị phân tìm kiếm cân bằng như cây AVL, cây đỏ đen,…

Tính chất của B-Tree:

* Tất cả các trang lá phải cùng mức.
* Mỗi trang có tối đa 2n khóa.
* Mỗi trang có tối thiểu n khóa, trừ trang gốc.
* Mỗi trang, trừ trang lá, nếu có m khóa thì có m+1 trang con.
* Các khóa trong mỗi trang được sắp xếp tăng dần từ trái qua phải.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| P0 | K1 | P1 | K2 | … | Pm-1 | Km | Pm |

Yếu tố tìm kiếm:

* Ki<x<Ki+1, với 1<I<m: thì tìm đến trang trỏ bởi pi.
* Km <x: thì tìm đến trang trỏ bởi pm.
* X<K1: thì tìm đến trang trỏ bởi p0.
* Nếu con trỏ bằng NULL, tức là không tìm thấy.

Chèn phần tử vào B-Tree:

* Tìm trang thích hợp và chèn phần tử vào.
* Nếu trang đầy thì tiến hành tách trang: phần tử đứng giữa sẽ được đấy lên trang cha.
* Nếu trang cha cũng đầy thì tiếp tục tiến hành tách trang cha theo tiến trình đi từ lá lên gốc.

Xóa phần tử từ B-Tree:

* Giai đoạn 1: tìm và xóa
* Nếu là trang lá: xóa phần tử.
* Nếu là trang giữa:
* Tìm phần tử thay thế ở trang lá: phải nhất bên trái hoặc trái nhất bên phải.
* Sao chép dữ liệu của phần tử thay thế với phần tử cần xóa.
* Xóa phần tử thay thế.
* Giai đoạn 2: Cân bằng lại

Xóa đi 1 phần tử ở trang lá (m phần tử) thì làm trang lá giảm đi 1 đơn vị.

* Xét m>=n: thì dừng.
* Nếu m<n: xét trang kế bên
* m2 > n: cân bằng số lượng phần tử cho 2 trang.
* m2 = n: Kéo phần tử trung gian ở cha xuống và ghép 2 trang lại -> trang đầy.
* Quá trình cứ thế lan từ lá đến gốc và có thể giảm chiều cao cây.