CÁC PHƯƠNG PHÁP INDEX CĂN BẢN

Cuối những năm 1970, index của các cơ sở dữ liệu quan hệ, phân cấp và CODASYL ra đời với nhiều loại: tuần tự, tuần tự có chỉ mục, hash, cây tìm kiếm nhị phân, cây B, cấu trúc TRIE, file đa danh sách, file đảo và cây đôi kết nối (doubly chained tree). Có nhiều lựa chọn cho lập trình viên , vì vậy mà chọn lựa một loại cây thích hợp là một quyết định khó khăn. Đối với các loại cơ sở dữ liệu – hướng đối tượng, không gian, thời gian, …, danh sách các chiến lược chỉ mục tiềm năng sẽ còn tiếp diễn. May mắn thay, đối với cơ sở dữ liệu quan hệ, cây B+ trở thành phương pháp chỉ mục thực tế cho hầu hết cơ sở dữ liệu quan hệ ngày nay.

# 2.1 Chỉ mục cây B+

Cây B+ là một phương pháp chỉ mục cơ bản được hỗ trợ trong DB2, Oracle, SQL server. Nó không chỉ đề cao truy vấn nhanh mà còn bảo trì động nghĩa là có thể loại bỏ các vấn đề tràn xảy ra trên các phương pháp chỉ mục tuần tự và hash cũ. Hãy xem xét một cây B+ điển hình trên hình 2.1. Mỗi nút chỉ mục không lá, có thể bao gồm một dãy kết hợp các khóa và con trỏ dữ liệu trỏ đến mỗi dòng. Nút chỉ mục lá được liên kết logic bằng con trỏ block sao cho mỗi dòng theo thứ tự có thể được tìm kiếm nhanh nhất. Biến p đại diện cho thứ tự của cây B+, sự phân nhánh của các con trỏ từ một node đến node thấp hơn trong cây.

Độ cao của cây B+ là bậc của nút lá, nút gốc là bậc 1.

Nút ở giữa giúp cơ sở dữ liệu tìm kiếm nút lá mong muốn với ít I/O. Tuy nhiên, những gì lưu trong nút lá mới quan trọng. nút lá sẽ lưu 3 thứ rất quan trọng: khóa, một mã record (record identifier – RID) và một con trỏ đến nút lá kế tiếp. Ví dụ, nếu một index được định nghĩa trên cột CITY, khóa của index có thể là NEW YORK, BOSTON, TORONTO. Với mỗi thành phố, chỉ mục sẽ bao gồm một định đanh của một nơi có thể tìm thấy mỗi record trong bảng trùng khớp với khóa. Nếu bảng có 400 entries cho NEW YORK, nút lá sẽ chứa một khóa entry của NEW YORK và định danh cho nơi có thể tìm thấy mỗi 400 record của bảng gốc. Những định danh này được gọi là định danh bản ghi, hay RID. Khóa và RID là hai phần thiết yếu của nội dung lá.

Khóa được lưu trong nút lá theo thứ tự sắp xếp. Kết quả là khi chỉ mục cây + có thể được dùng trong DBMS để tìm dòng hoặc một tập hợp các dòng trùng khớp với khóa đơn, thì chúng còn có thể được dùng để tìm một miền khóa (ví dụ một miền số học hoặc alphabet) một cách dễ dàng cũng như hoàn toàn trả về một record theo thứ tự sắp xếp.