

BÀI 7: CÂY 2-3-4

1. Giới thiệu về cây 2-3-4

Chúng ta sẽ xem xét các đặc tính của cây 2-3-4 và mối quan hệ khá gần gũi giữa cây 2-3-4 và cây đỏ-đen.

Hình 1 trình bày một cây 2-3-4 đơn giản. Mỗi node có thể lưu trữ 1, 2 hoặc 3 mục dữ liệu.

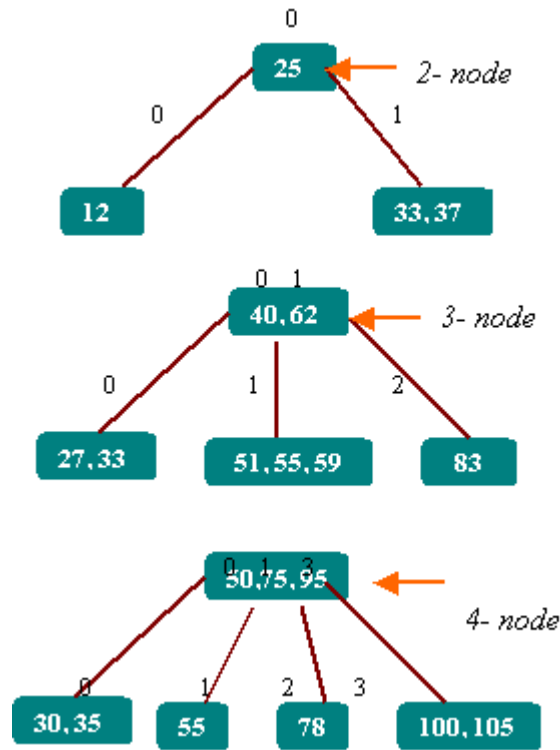


Hình 1 cây 2-3-4

Các số 2, 3 và 4 trong cụm từ *cây 2-3-4* có ý nghĩa là khả năng có bao nhiêu liên kết đến các node con có thể có được trong một node cho trước. Đối với các node không phải là lá, có thể có 3 cách sắp xếp sau:

- Một node với một mục dữ liệu thì luôn luôn có 2 con.
- Một node với hai mục dữ liệu thì luôn luôn có 3 con.
- Một node với ba mục dữ liệu thì luôn luôn có 4 con.

Như vậy, một node không phải là lá phải luôn luôn có số node con nhiều hơn 1 so với số mục dữ liệu của nó. Nói cách khác, đối với mọi node với số con là k và số mục dữ liệu là d , thì : $k = d + 1$



Hình 2. các trường hợp của cây 2-3-4

Với mọi node lá thì không có node con nhưng có thể chứa 1, 2 hoặc 3 mục dữ liệu, không có node rỗng.

Một cây 2-3-4 có thể có đến 4 cây con nên được gọi là *cây nhiều nhánh bậc 4*.

Trong cây 2-3-4 mỗi node có ít nhất là 2 liên kết, trừ node lá (node không có liên kết nào).

Hình 2 trình bày các trường hợp của cây 2-3-4. Một node với 2 liên kết gọi là một 2-node, một node với 3 liên kết gọi là một 3-node, và một node với 4 liên kết gọi là một 4-node, nhưng ở đây không có node là 1-node.

2. Tổ chức cây 2-3-4

Các mục dữ liệu trong mỗi node được sắp xếp theo thứ tự tăng dần từ trái sang phải (sắp xếp từ thấp đến cao).

Trong cây tìm kiếm nhị phân, tất cả node của cây con bên trái có khoá nhỏ hơn khoá của node đang xét và tất cả node của cây con bên phải có khoá lớn hơn hoặc bằng khoá của

node đang xét. Trong cây 2-3-4 thì nguyên tắc cũng giống như trên, nhưng có thêm một số điểm sau:

- Tất cả các node con của cây con có gốc tại node con thứ 0 thì có các giá trị khoá nhỏ hơn khoá 0.
- Tất cả các node con của cây con có gốc tại node con thứ 1 thì có các giá trị khoá lớn hơn khoá 0 và nhỏ hơn khoá 1.
- Tất cả các node con của cây con có gốc tại node con thứ 2 thì có các giá trị khoá lớn hơn khoá 1 và nhỏ hơn khoá 2.
- Tất cả các node con của cây con có gốc tại node con thứ 3 thì có các giá trị khoá lớn hơn khoá 2.

Trong cây 2-3-4, các nút lá đều nằm trên cùng một mức. Các node ở mức trên thường không đầy đủ, nghĩa là chúng có thể chứa chỉ 1 hoặc 2 mục dữ liệu thay vì 3 mục.

Lưu ý rằng cây 2-3-4 là cây cân bằng. Nó vẫn giữ được sự cân bằng khi thêm vào các phần tử có thứ tự (tăng dần hoặc giảm dần).

3. Tìm kiếm

Thao tác tìm kiếm trong cây 2-3-4 tương tự như thủ tục tìm kiếm trong cây nhị phân. Việc tìm kiếm bắt đầu từ node gốc và chọn liên kết dẫn đến cây con với phạm vi giá trị phù hợp.

Ví dụ, để tìm kiếm mục dữ liệu với khoá là 64 trên cây ở hình 1, bạn bắt đầu từ gốc. Tại node gốc không tìm thấy mục khoá này. Bởi vì 64 lớn 50, chúng ta đi đến node con 1, (60/70/80)(lưu ý node con 1 nằm bên phải, bởi vì việc đánh số của các node con và các liên kết bắt đầu tại 0 từ bên trái). Tại vị trí này vẫn không tìm thấy mục dữ liệu, vì thế phải đi đến node con tiếp theo. Tại đây bởi vì 64 lớn hơn 60 nhưng nhỏ hơn 70 nên đi tiếp đến node con 1. Tại thời điểm chúng ta tìm được mục dữ liệu đã cho với liên kết là 62/64/66.



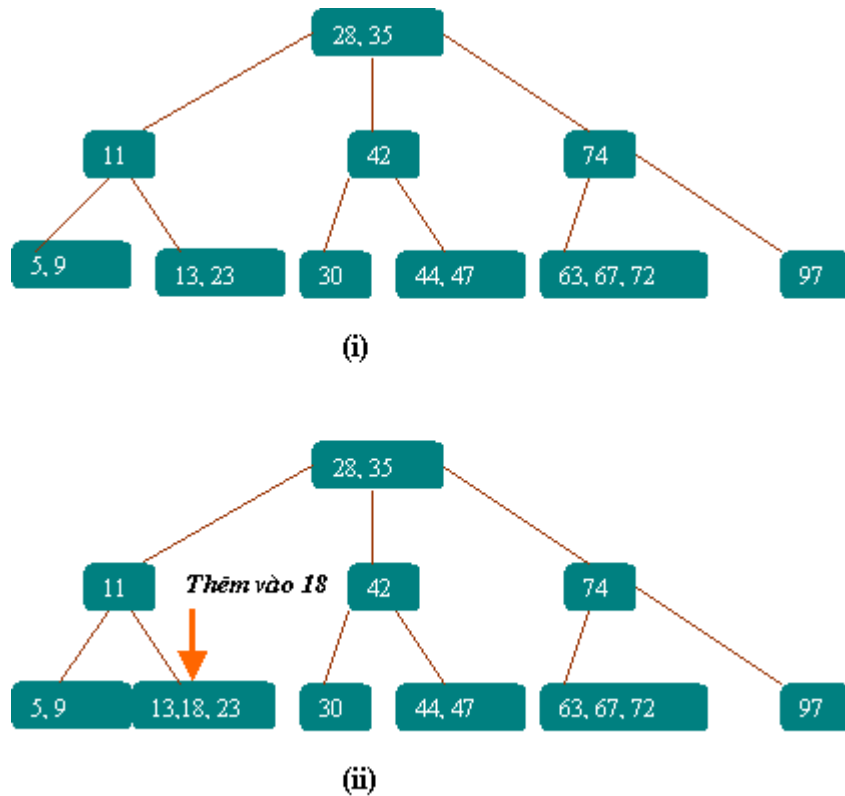
4. Thêm vào

Các mục dữ liệu mới luôn luôn được chèn vào tại các node lá . Nếu mục dữ liệu được thêm vào node mà có node con, thì số lượng của các node con cần thiết phải được biến đổi để duy trì cấu trúc cho cây, đây là lý do tại sao phải có số node con nhiều hơn 1 so với các mục dữ liệu trong một nút.

Việc thêm vào cây 2-3-4 trong bất cứ trường hợp nào thì quá trình cũng bắt đầu bằng cách tìm kiếm node lá phù hợp.

Nếu không có node đầy nào (node có đủ 3 mục dữ liệu) được bắt gặp trong quá trình tìm kiếm, việc chèn vào khá là dễ dàng. Khi node lá phù hợp được tìm thấy, mục dữ liệu mới đơn giản là thêm vào nó. Hình 3 trình bày một mục dữ liệu với khoá 18 được thêm vào cây 2-3-4.

Việc chèn vào có thể dẫn đến phải thay đổi vị trí của một hoặc hai mục dữ liệu trong node vì thế các khoá sẽ nằm với trật tự đúng sau khi mục dữ liệu mới được thêm vào. Trong ví dụ này số 23 phải được đẩy sang phải để nhường chỗ cho 18.



Hình 3 Chèn vào không làm tách cây
 (i) trước khi chèn vào
 (ii) sau khi chèn vào

Tách nút

Việc thêm vào sẽ trở nên phức tạp hơn nếu gặp phải một node đầy (node có số mục dữ liệu đầy đủ) trên nhánh dẫn đến điểm thêm vào. Khi điều này xảy ra, node này cần thiết phải được tách ra. Quá trình tách nhằm giữ cho cây cân bằng. Loại cây 2-3-4 mà chúng ta đề cập ở đây thường được gọi là cây 2-3-4 top-down bởi vì các node được tách ra theo hướng đi xuống điểm chèn.

Giả sử ta đặt tên các mục dữ liệu trên node bị phân chia là A, B và C. Sau đây là tiến trình tách (chúng ta giả sử rằng node bị tách không phải là node gốc; chúng ta sẽ kiểm tra việc tách node gốc sau này):

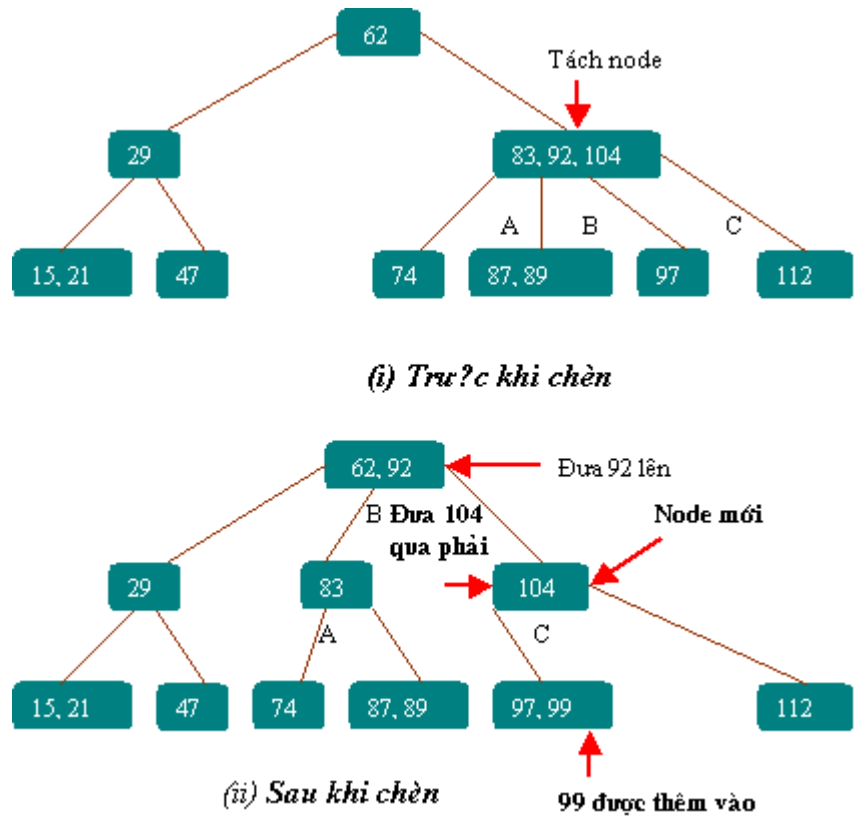
- Một node mới (rỗng) được tạo. Nó là anh em với node sẽ được tách và được đưa vào bên phải của nó.
- Mục dữ liệu C được đưa vào node mới.
- Mục dữ liệu B được đưa vào node cha của node được tách.

- Mục dữ liệu A không thay đổi.

- Hai node con bên phải nhất bị hủy kết nối từ node được tách và kết nối đến node mới.

(Một cách khác để mô tả sự tách node là một 4-node được tách thành hai 2-nút)

Một ví dụ về việc tách node trình bày trên hình 4.



Hình 4: Tách một nút

(i) Trước khi chèn vào

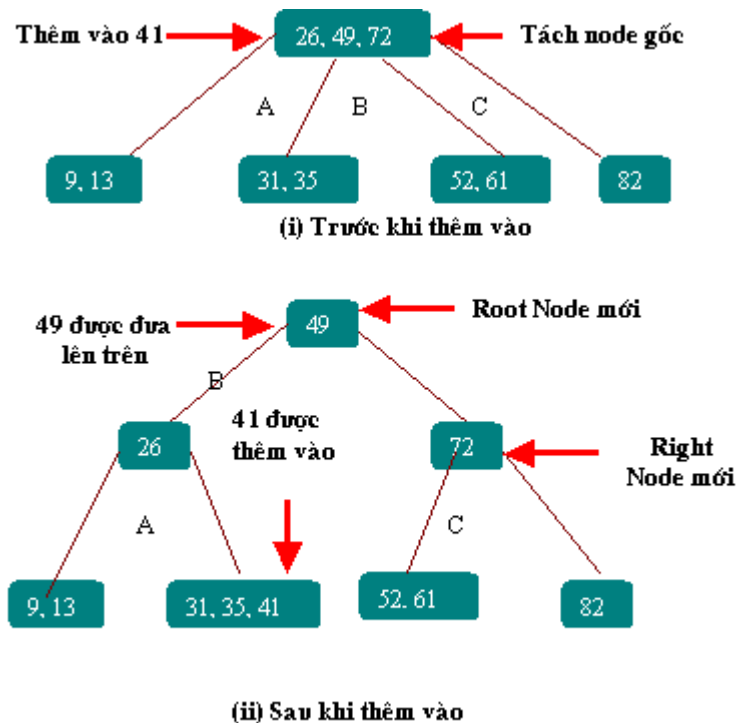
(ii) Sau khi chèn vào

Tách node gốc

Khi gặp phải node gốc đầy tại thời điểm bắt đầu tìm kiếm điểm chèn, kết quả của việc tách thực hiện như sau:

- Node mới được tạo ra để trở thành gốc mới và là cha của node được tách.
- Node mới thứ hai được tạo ra để trở thành anh em với node được tách.

- Mục dữ liệu C được dịch đưa sang node anh em mới.
- Mục dữ liệu B được dịch đưa sang node gốc mới.
- Mục dữ liệu A vẫn không đổi.
- Hai node con bên phải nhất của node được phân chia bị hủy kết nối khỏi nó và kết nối đến node mới bên phải.



Hình 4.5 Tách node gốc

i) Trước khi thêm vào

ii) Sau khi thêm vào

Hình 5 chỉ ra việc tách node gốc. Tiến trình này tạo ra một node gốc mới ở mức cao hơn mức của node gốc cũ. Kết quả là chiều cao tổng thể của cây được tăng lên 1.

Đi theo node được tách này, việc tìm kiếm điểm chèn tiếp tục đi xuống phía dưới của cây. Trong hình 5 mục dữ liệu với khoá 41 được thêm vào lá phù hợp.

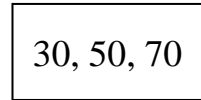
Tách theo hướng đi xuống

Chú ý rằng, bởi vì tất cả các node đầy được tách trên đường đi xuống nên việc tách node không gây ảnh hưởng gì khi phải đi ngược lên trên của cây. Node cha của bất cứ node

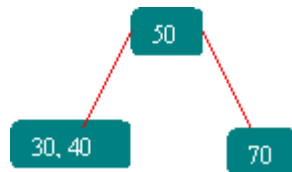
nào bị tách phải đảm bảo rằng không phải là node đầy, để đảm bảo node cha này có thể chấp nhận mục dữ liệu B mà không cần thiết nó phải tách ra. Tất nhiên nếu node cha này đã có hai con thì khi node con bị tách, nó sẽ trở thành node đầy. Tuy nhiên điều này chỉ có nghĩa là nó có thể sẽ bị tách ra khi lần tìm kiếm kế tiếp gặp nó.

Hình 6 trình bày một loạt các thao tác chèn vào một cây rỗng. Có 4 node được tách, 2 node gốc và 2 node lá.

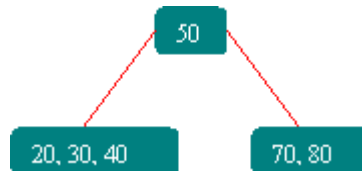
Thêm vào 70, 30, 50



Thêm 40



Thêm vào 20, 80

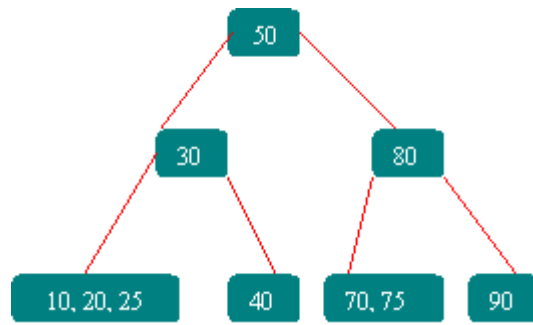


Thêm vào 25, 90



Thêm vào 75

Thêm vào 10



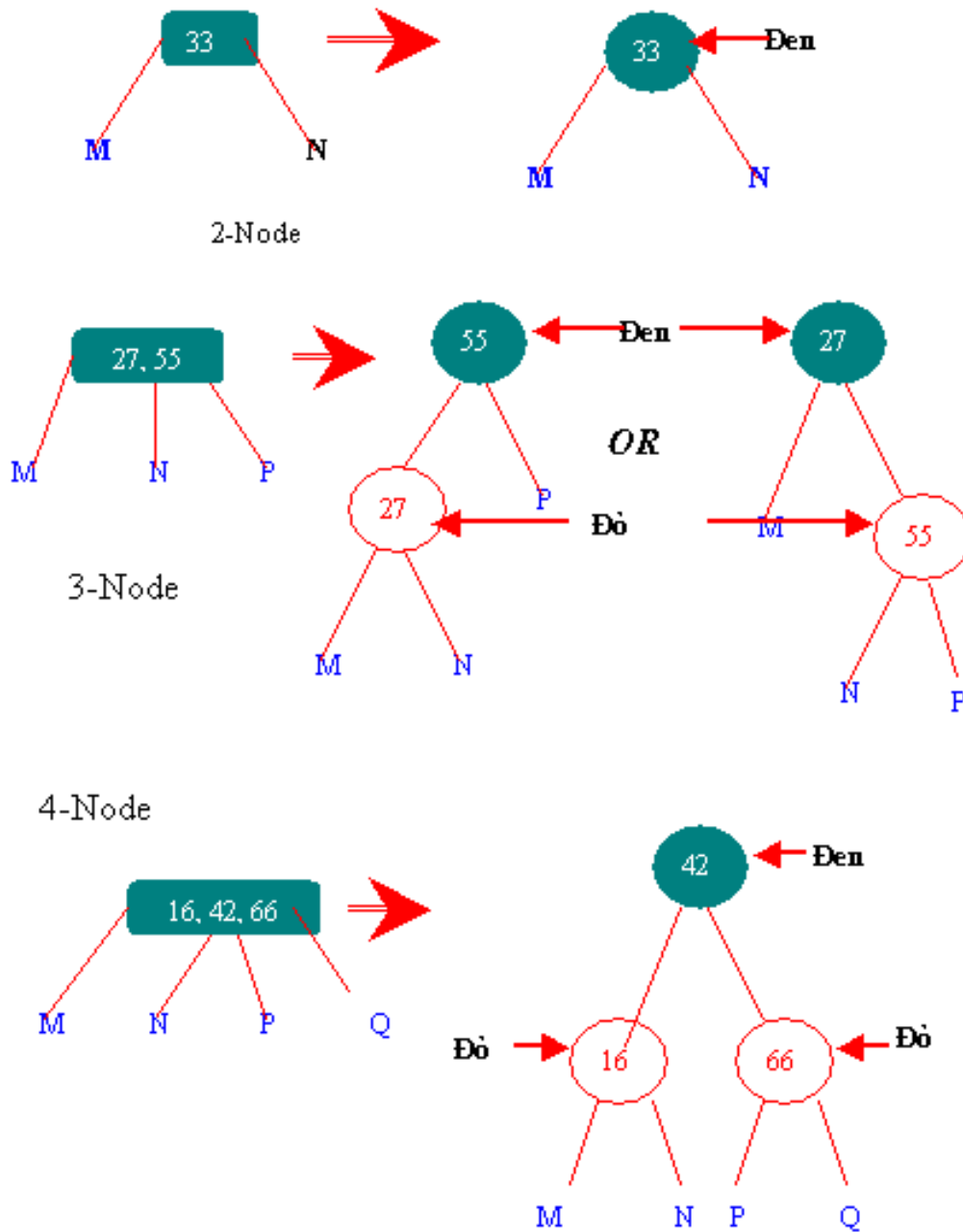
Hình 6 Minh họa thêm một node vào cây 2-3-4

5. Biến đổi cây 2-3-4 sang cây Đỏ-Đen

Một cây 2-3-4 có thể được biến đổi sang cây đỏ-đen bằng cách áp dụng các luật sau:

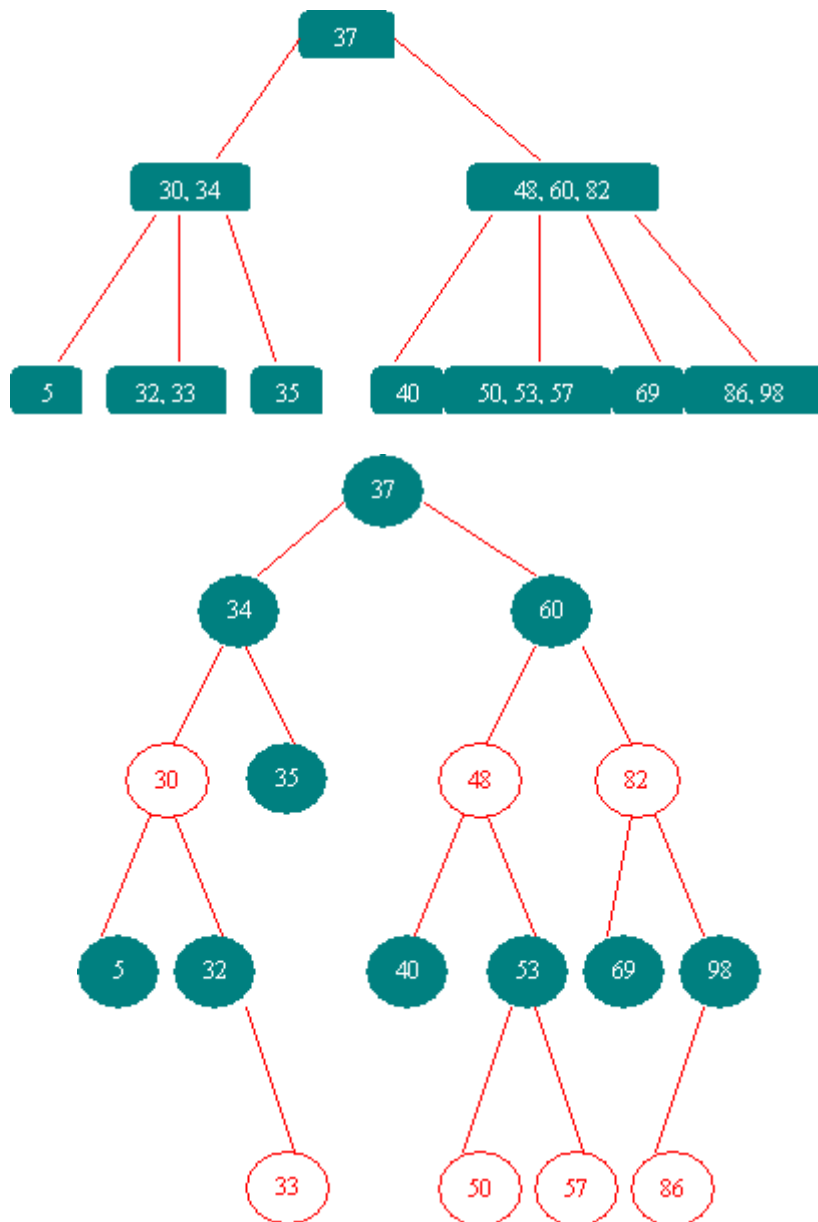
- Biến đổi bất kỳ 2-node ở cây 2-3-4 sang node đen ở cây đỏ-đen.
- Biến đổi bất kỳ 3-node sang node con C (với hai con của chính nó) và node cha P (với các node con C và node con khác). Không có vấn đề gì ở đây khi một mục trở thành node con và mục khác thành node cha. C được tô màu đỏ và P được tô màu đen.
- Biến đổi bất kỳ 4-node sang node cha P và cả hai node con C1, C2 màu đỏ.

Hình 4.7 trình bày các chuyển đổi này. Các node con trong các cây con được tô màu đỏ; tất cả các node khác được tô màu đen.



Hình 7 Chuyển đổi từ cây 2-3-4 sang cây đò-đen

Hình 4.8 trình bày cây 2-3-4 và cây đỏ-đen tương ứng với nó bằng cách áp dụng các chuyển đổi này. Các đường chấm xung quanh các cây con được tạo ra từ 3-node và 4-nút. Các luật của cây đỏ-đen tự động thoả mãn với sự chuyển đổi này. Kiểm tra rằng: Hai node đỏ không bao giờ được kết nối, và số lượng các node đen là như nhau ở mọi đường dẫn từ gốc đến lá (hoặc node con null).



Hình 4.8 Cây 2-3-4 và cây đỏ-đen tương ứng