2-3-4 Tree notes

Prof Bill, Mar 2020

I like Wikipedia here, actually:

- ➔ B-tree, en.wikipedia.org/wiki/B-tree
- ➔ 2-3-4 tree, en.wikipedia.org/wiki/2%E2%80%933%E2%80%934_tree

Read: Sedgewick Algorithms 3.3 Balanced Search Trees (btw, he covers 2-3 trees, not 2-3-4), algs4.cs.princeton.edu/33balanced

Animation: Select "B Trees"; this may be the best animation of all, www.cs.usfca.edu/~galles/visualization/Algorithms.html

---

Quickly:

In the **binary trees** we know and love, all nodes have 2 children.

B-trees are the general case, where nodes can have N children.

2-3-4 trees are B-trees that contain nodes with 2, 3, or 4 children.

B-trees

Important: "Not to be confused with Binary tree"

Definition:

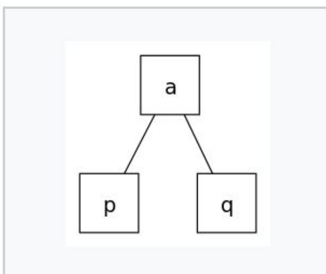
> In computer science, a B-tree is **a self-balancing tree** data structure that maintains sorted data and allows searches, sequential access, insertions, and deletions in **logarithmic time**. The B-tree generalizes the binary search tree, allowing for **nodes with more than two children**. Unlike other self-balancing binary search trees, the B-tree is well suited for storage systems that read and write relatively large blocks of data, such as discs. It is commonly used in databases and file systems.

2-3-4 trees

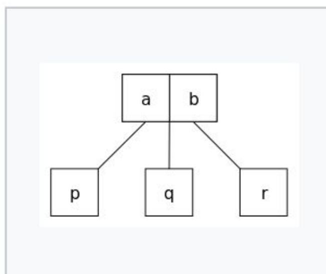
We'll focus on 2-3-4 trees, which are one flavor of B-tree.

Definition:

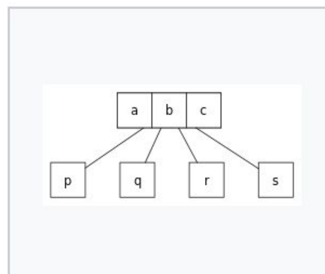
> In computer science, a **2–3–4 tree** (also called a 2–4 tree) is a self-balancing data structure...where every node with children (internal node) has either two, three, or four child nodes:
>
> - a 2-node has one data element, and if internal has two child nodes
> - a 3-node has two data elements, and if internal has three child nodes
> - a 4-node has three data elements, and if internal has four child nodes



2-node



3-node



4-node

So, the biggest node has 3 values and 4 children. Bigger than that, and the node splits.

Properties:

- ➢ Every node (leaf or internal) is a 2-node, 3-node or a 4-node, and holds one, two, or three data elements, respectively.
- ➢ All leaves are at the same depth (the bottom level).
- ➢ All data is kept in sorted order.

Since our tree is always balanced, then search, insert, and delete are all **O(log n)**!

2–3–4 trees are an isometry of red–black trees, meaning that they are equivalent data structures.

Sedgewick covers 2-3 trees. (shrug) Can you figure out the difference between these and 2-3-4 trees?

Insertion pseudocode

```
insert( key)
    add new key to an existing leaf node, ala BST
    L1: if adding to 4-node (too large)
        split and push value up to parent
        if parent is null, then make node root
        goto L1 for parent
```