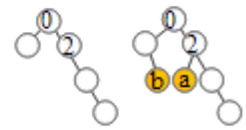
The height of a height-balanced binary tree

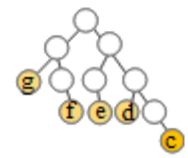
A binary tree is *height-balanced* if, for every node, the heights of its two children differ by at most one. We will prove that the height of a height-balanced tree with n nodes is $O(n)$.

The first tree on the right above is not height-balanced because the root's empty left subtree has height -1 (By convention, the empty tree has height -1) and its right subtree has height 1. The second tree is height-balanced.

A height-balanced tree can be heavily weighted to one side. Look at the first tree to the right. It is not balanced: Node 2's subtrees have heights -1 and 1, and node 0's subtrees have heights 0 and 2. In the second tree, we add as few nodes as possible, and as far right as possible, to make it balanced: node a and then b. The right subtree of the root (of this height-balanced tree) has twice as many nodes as the left subtree, but the tree is height-balanced.



Suppose we add one more node, c, as to the right; then, to balance the tree, we add nodes d, e, f, and g in that order, as far right as possible. The resulting tree is height-balanced, but it's heavily "weighted" to the right: The root's right subtree has almost twice as many nodes as its left subtree.



Nevertheless, we can prove that the height of a balanced tree with n nodes is $O(n)$.

Below, we use simply *tree* for *binary tree* and *balanced* for *height-balanced.* Also, we use lg n for the base-2 logarithm of n. Finally, we define:

> $\mathtt{min}(h)$ is the minimum number of nodes in a balanced tree of height h.

We start by proving the important Lemma 1. It seems backward: Instead of giving an upper bound on the height for a given tree size, it gives a lower bound on the tree size given the height. But it works!

Lemma 1. For a balanced tree of height $h > 0$, $\mathtt{min}(h) > 2^{h/2 - 1}$.

Proof. The proof is by induction on h.

Case $h = 1$. A balanced tree of height 1 has at least 2 nodes. We have: $2 > 2^{1/2 - 1}$, so the theorem holds for $h = 1$.

Case $h = 2$. A balanced tree of height 2 has at least 4 nodes —see the second tree at the top of this page. We have: $4 > 2^{2/2 - 1} = 2^0 = 1$, so the theorem holds for $h = 2$.

Case $h \geq 3$. Assume the theorem holds for values in the range 1..h-1. We prove it holds for h. A balanced tree of height h with a minimum number of nodes has a root and two subtrees. Since the tree has a minimum number of nodes, so do it subtrees.

Since the tree has height h, one subtree has height h-1. Since the subtrees have a minimum number of nodes, the other subtree has height h-2 (see the tree to the right). It can't have a smaller height because the tree is balanced —for any node the heights of its subtrees differ by at most 1. It should also be clear that the shorter subtree has fewer nodes. From this, we see that:

$$\mathtt{min}(h) = 1 + \mathtt{min}(h\text{-}1) + \mathtt{min}(h\text{-}2)$$

We calculate:

$\mathtt{min}(h)$

= < the above formula>

$1 + \mathtt{min}(h\text{-}1) + \mathtt{min}(h\text{-}2)$

> <arithmetic —delete the 1 and use fact that $\mathtt{min}(h\text{-}1) > \mathtt{min}(h\text{-}2)$>

$2\ \mathtt{min}(h\text{-}2)$

> <inductive hypothesis>

$2 * 2^{(h-2)/2 - 1}$

> <arithmetic>

$2^{h/2 - 1}$

Q.E.D. (Quit.End.Done.)

The lemma proves that $\mathtt{min}(\mathtt{h}) > 2^{h/2 - 1}$. The next lemma shows what happens if we take the lg of both sides of this formula and isolate h on one side of the relation.

Lemma 2. $\mathtt{h} < 2 \lg \mathtt{min}(\mathtt{h}) + 2$.

Proof. From Lemma 1, we have:

$\mathtt{min}(\mathtt{h}) > 2^{h/2 - 1}$

= <take lg of both sides>

$\lg \mathtt{min}(\mathtt{h}) > \mathtt{h}/2 - 1$

= <arithmetic>

$\mathtt{h}/2 < \lg \mathtt{min}(\mathtt{h}) + 1$

= <arithmetic>

$\mathtt{h} < 2 \lg \mathtt{min}(\mathtt{h}) + 2$

Q.E.D.

Theorem: The height of a balanced tree of n nodes is $O(\lg \mathtt{n})$.

Proof. Let h be the height of a balanced tree. Lemma 2 proves that

$$\mathtt{h} < 2 (\lg \mathtt{min}(\mathtt{h})) + 2$$

Since $\mathtt{min}(\mathtt{h})$ is the minimum number of nodes in a balanced tree of height h, we have

$\mathtt{h} < 2 (\lg \mathtt{n}) + 2$ for *any* balanced tree of height h with $\mathtt{n} > 0$ nodes.

Therefore, h is $O(\lg \mathtt{n})$.

Q.E.D.