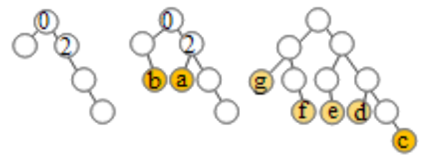
The height of a height-balanced binary tree

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 is O(size-of-tree).

The first tree on the right above is not balanced because the root's left subtree has height -1 and its right subtree has height 1. The second tree is 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. We add as few nodes as possible to make it balanced, and as far to the right as possible: node a and then node b. The tree is balanced. The right subtree of the root has twice as many nodes as the left subtree, but the tree is height-balanced.

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

Nevertheless, we can still prove that the height of a balanced tree of 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 prove the theorem based on two lemmas, which are themselves proved after the theorem.

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

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

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

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

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

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

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

That was the easy part of the proof! We now turn to the hard part, proving a lower bound on the number of nodes in a tree based on its height.

Lemma 1. 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: $2^{1/2 - 1} < 2$, 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: $2^{2/2 - 1} = 2^0 = 1 < 4$, and the theorem holds for $h = 2$.

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

One of the subtrees 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. From this, we see that:

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

We calculate:

$\quad min(h)$
$= \quad$ < the above formula>
$\quad 1 + min(h-1) + min(h-2)$
$> \quad$ <arithmetic>
$\quad min(h-1) + min(h-2)$
$> \quad$ <arithmetic, since $min(h-1) > min(h-2)$>
$\quad 2\ min(h-2)$
$> \quad$ <inductive hypothesis>
$\quad 2 * 2^{(h-2)/2 - 1}$
$> \quad$ <arithmetic>
$\quad 2 * 2^{h/2 - 2}$
$> \quad$ <arithmetic>
$\quad 2^{h/2 - 1}$

Q.E.D.

The following Lemma 2 uses Lemma 1 to give the lower bound on the number of nodes in a balanced tree given its height.

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

Proof. From lemma 1, we have:

$\quad min(h) > 2^{h/2 - 1}$
$= \quad$ <take lg of both sides>
$\quad \lg min(h) > h/2 - 1$
$= \quad$ <arithmetic>
$\quad h/2 < \lg min(h) + 1$
$= \quad$ <arithmetic>
$\quad h < 2 \lg min(h) + 2$