# B-trees

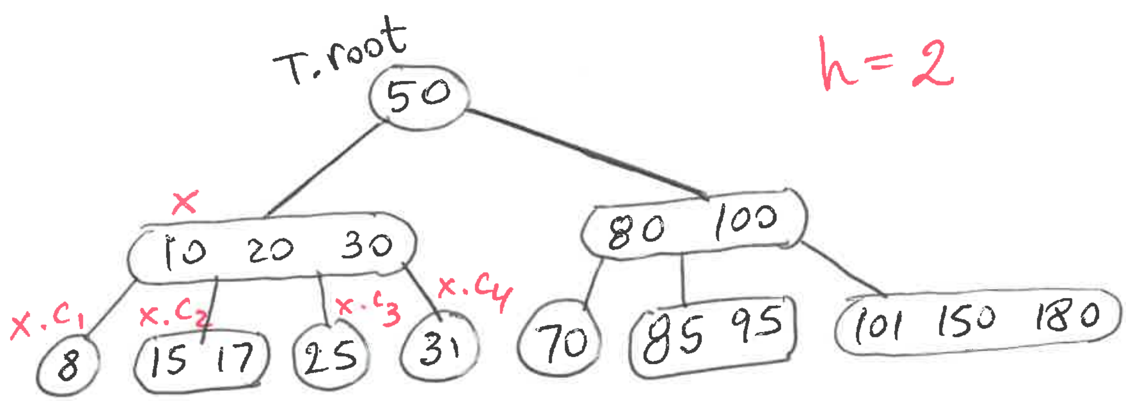
$t$ - minimum degree

- if $x$ is not the root, then
  - the number of keys $x.n$, $t-1 \le x.n \le 2t-1$
    a node $x$ is full if it has $2t-1$ keys
  - the number of children is $x.n+1$, therefore
    $t \le$ no. of children $\le 2 \cdot t$
- if $x$ is the root, then
  - the number of keys $x.n$ $\quad 1 \le x.n \le 2t-1$
    the node is full if it has $2t-1$ keys
  - the number of children is $x.n+1$, therefore
    $2 \le$ no. of children $\le 2t$

example

$t = 2$

# Keys : 1..3




$x.n = 3$

$x.key_1 = 10$

$x.key_2 = 20$

$x.key_3 = 30$

$x.leaf$ = FALSE

## B-tree definition

A B-tree T is a rooted tree (where T.root is the root) with the following properties:

1. every node $x$ has the following attributes
   a. $x.n$ – the number of keys currently stored in x
   b. the keys $x.key_1$, $x.key_2$, …, $x.key_{x.n}$ so that

      $x.key_1 \leq x.key_2 \leq \ldots \leq x.key_{x.n}$

   c. $x.leaf$ – a boolean value which is TRUE if $x$ is a leaf and FALSE if $x$ is an internal node
2. each internal node $x$ has $x.n+1$ pointers $x.c_1$, $x.c_2$, …, $x.c_{x.n+1}$ to its children; if $x$ is a leaf then its $c_i$ attributes are undefined
3. if $k_i$ is any key stored in the subtree with root $x.c_i$ then:

   $k_1 \leq x.key_1 \leq k_2 \leq x.key_2 \leq \ldots \leq x.key_{x.n} \leq k_{x.n+1}$

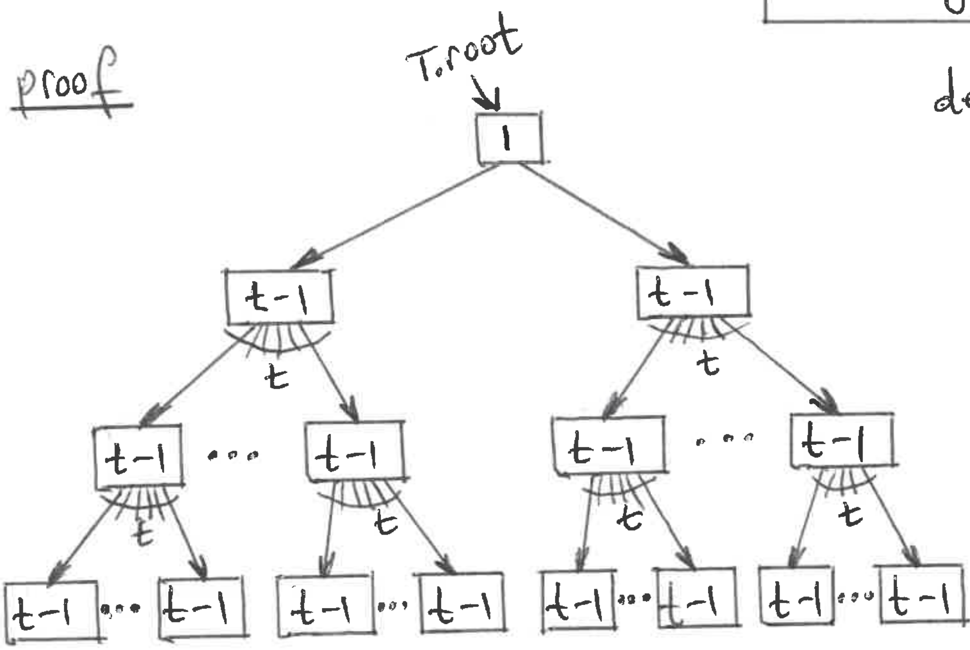
4. all the leaves have the same depth, which is the tree high $h$.
5. the B-tree has a **minimum degree** $t$ ( $t$ is an integer $t \geq 2$):
   - every node other than the root must have $\geq$ t – 1 keys and $\geq$ t children; if B-tree is nonempty, then the root has at least one key
   - every node has $\leq$ 2t – 1 keys and $\leq 2t$ children

A node is full is it has 2t – 1 keys.

# The height of a B-tree

**Theorem:** if $n \geq 1$, then for any $n$-Key B-tree $T$ of height $h$ and minimum degree $t$, $\boxed{h \leq \log_t \frac{n+1}{2}}$

proof




| depth | number of nodes | number of Keys |
|---|---|---|
| 0 | 1 | 1 |
| 1 | 2 | $2(t-1)$ |
| 2 | $2t$ | $2t(t-1)$ |
| 3 | $2t^2$ | $2t^2(t-1)$ |
| ... | | |
| $h$ | $2t^{h-1}$ | $2t^{h-1}(t-1)$ |

$$n \geq 1 + 2(t-1) + 2t(t-1) + 2t^2(t-1) + \dots + 2 \cdot t^{h-1}(t-1) =$$

$$= 1 + 2(t-1)\underbrace{(1 + t + t^2 + t^3 + \dots + t^{h-1})}_{\text{geometric series}} = 1 + 2 \cdot \cancel{(t-1)} \cdot \frac{t^h - 1}{\cancel{t-1}} =$$

$$= 1 + 2 \cdot t^h - 2 = 2 \cdot t^h - 1$$

$$n \geq 2 \cdot t^h - 1$$

$$t^h \leq \frac{n+1}{2}$$

$$\log_t (t^h) \leq \log_t \left(\frac{n+1}{2}\right)$$

$$h \leq \log_t \left(\frac{n+1}{2}\right)$$

$$h = O\left(\log_t \frac{n+1}{2}\right)$$

$$\boxed{h = O(\log_t n)}$$

$$\boxed{h = O(\lg n)}$$