B-Trees, Part II

Review of Properties

Let $t$ be the order of the B-tree (parameter given to constructor)

**BALANCED** - all leaf nodes are at same height. So an internal node with $x$ elements has $x+1$ non-empty children

**NODE UTILIZATION** - With the exception of root all nodes have $\geq t$ children ($\Rightarrow \geq t-1$ elements). The root has $\geq 2$ children. All nodes have $\leq 2t$ children ($\Rightarrow \leq 2t-1$ elements)
Relation between B-tree and red-black tree

For $t=2$, a B-tree is called a 2-3-4 tree.

Red-black tree is a representation of a 2-3-4 tree as a binary search tree.
\[ x \times x \times y \times \rightarrow \text{split} \rightarrow x \times y \times z \]

- You're part of your parent.
Split (+ Merge)

Insertion

Split

Merge

delete

t/\t-1
elements elements

roam for growth
Top-Down Insertion

Follow path to leaf where you'd insert (with natural extension of binary search tree insertion)

Whenever a full node (2t+1 children) encountered, split it and then continue

Goal: minimize possibility of a page fault occurring twice on same page
Bottom-up Insertion

Do standard insertion in leaf if there is room.

Otherwise split leaf (which could propagate to the root). Stop as soon as parent is not full.

Reduces unnecessary splits.
Analysis of height

Suppose the B-tree has height $h$

What is the minimum number of elements it could hold?

$n \geq f(h)$

Then solve for $h$ as a function of $n + t$

$h \leq \underline{\phantom{0}}$
B-tree with minimum utilization

\[
\begin{align*}
\text{level} & \quad \text{# nodes} & \quad \text{# elements} \\
0 & \quad 1 & \quad 1 \\
1 & \quad 2 \times (t-1) & \\
2 & \quad 2t \times (t-1) & \\
3 & \quad 2t^2 \times (t-1) & \\
\vdots & \vdots & \\
\end{align*}
\]

\( h = \text{the # of page faults} \)

\( h \geq 1 + 2(t-1) + 2t(t-1) + \cdots + 2t^{h-1}(t-1) \)

\( = 1 + 2(t-1) \left( 1 + t + t^2 + \cdots + t^{h-1} \right) \)

\( = 1 + 2(t-1) \left( \frac{t^{h-1} - 1}{t-1} \right) = 1 + 2(t^{h-1}) \)
\begin{align*}
    n & \geq 1 + 2(t^h - 1) \\
    n-1 & \geq 2(t^h - 1) \\
    \frac{n-1}{2} & \geq t^h - 1 \\
    t^h & \leq \frac{n+1}{2} \\
    h & \leq \log_t \left(\frac{n+1}{2}\right) = \log_t (n+1) - \log_t 2
\end{align*}
If $t=1000$, $h=5$

If no more than 5 page faults

\[ n \geq 1 + 2(t^h - 1) \]

\[ = 1 + 2(1000^5 - 1) \approx 2 \times 10^{15} \]
B+ tree

$t = 2$ For letters “abstraction”