B-strom (B-Tree)

B-strom je zobecněním binárního vyhledávacího stromu. Na rozdíl od něj však může mít každý uzel více než dvě děti (potomky) a obsahovat více než jeden klíč. Písmeno „B“ v názvu nebylo nikdy autory (Rudolf Bayer a Edward M. McCreight) přesně vysvětleno, ale nejčastěji se interpretuje jako **Balanced** (vyvážený), protože strom automaticky udržuje všechny své listy ve stejné hloubce.

Hlavní výhodou B-stromu je minimalizace počtu operací čtení z disku, což je u velkých databází nejpomalejší část procesu.

Klíčové vlastnosti

- **Vysoký stupeň větvení:** Uzly mohou mít stovky i tisíce potomků. To znamená, že strom je velmi „nízký“ (mělký). I v miliardě záznamů lze najít konkrétní prvek na 3 až 4 průchody uzly.
- **Vždy vyvážený:** Algoritmus vkládání a mazání zajišťuje, že vzdálenost od kořene k jakémukoliv listu je vždy stejná.
- **Efektivita disku:** Velikost jednoho uzlu v B-stromu se obvykle nastavuje tak, aby odpovídala velikosti bloku na disku (např. 4 KB nebo 8 KB). Při jednom čtení z disku se tak načte celý uzel s mnoha klíči.

Jak probíhá vyhledávání

Vyhledávání v B-stromu je podobné listování v telefonním seznamu:

```
1. Začnete v **kořenovém uzlu**.
2. Porovnáte hledaný klíč s klíči v uzlu.
3. Najdete interval, do kterého váš klíč patří (např. klíč je mezi 10 a 20).
4. Sledujete ukazatel na odpovídajícího potomka (podstrom).
5. Opakujete, dokud nenajdete shodu nebo nedosáhnete listu.
```

Díky tomu, že jsou klíče v každém uzlu seřazené, lze uvnitř uzlu použít binární vyhledávání, což proces dále urychluje.

Vkládání a štěpení uzlů

Když se do B-stromu přidávají nová data, strom roste „odspodu nahoru“:

```
1. Data se vloží do příslušného listu.
2. Pokud je list plný, dojde k jeho **štěpení** (split). Prostřední prvek se
přesune o úroveň výš do rodičovského uzlu.
3. Pokud se tím zaplní i rodičovský uzel, štěpí se i ten. Takto se může
zaplnit až kořen, což je jediný okamžik, kdy se výška stromu zvýší o 1
úroveň.
```

Varianty B-stromu

V praxi se nejčastěji setkáme s variantami, které základní model vylepšují:

Varianta	Rozdíl oproti základnímu B-stromu	Použití
B+ strom	Skutečná data jsou pouze v listech. Listy jsou navzájem propojeny lineárně.	Nejpoužívanější v MySQL, PostgreSQL, NTFS.
B* strom	Vyžaduje, aby uzly byly zaplněny alespoň ze 2/3 (místo 1/2), což šetří místo.	Souborové systémy (HFS+).

Proč je to důležité?

Bez B-stromů by operace typu „najdi všechny objednávky od 1. ledna do 15. ledna“ vyžadovaly prohledání celého pevného disku. B+ strom díky propojení listů umožní najít začátek intervalu a pak prostě „přečíst“ sousední listy, což je bleskové.

Související pojmy: Indexování, RDBMS, Databáze, Latence, RAM, SQL, Binární strom.

From:
https://serviceit.cz/ - **IT ENCYKLOPEDIE**

Permanent link:
https://serviceit.cz/doku.php?id=b-tree

Last update: **2025/12/31 19:43**