

Cena pretrage

Pretraga heap i uređenih fajlova

Heap i uređeni fajlovi

Heap File



Sorted File



For illustration, records are just integers

Cilj

- Odrediti troškove operacija (u zavisnosti od organizacije fajla)
 - Skeniranja celog fajla
 - Pronalaženja jednog sloga
 - Pronalaženja slogova iz opsega vrednosti
 - Dodavanja
 - Brisanja
-

Model troškova (cost model)

- Odrediti
 - Koliko I/O operacija nad blokovima se obavlja.
 - Prosečno vreme za obavljanje operacija
 - U određivanje troškova ulaze
 - B – broj strana
 - R – broj slogova u bloku
 - D – prosečno vreme I/O operacije nad jednim blokom (čitanje/pisanje sa/na disk)
 - Zanimaju se
 - Organizacija blokova na disku (da li su susedni ili ne), uzima se prosečno vreme I/O
 - Pre-fetch
 - Vreme potrebno za izvršavanje operacija kada su podaci u memoriji, ...
-

Model troškova (cost model)

- Pretpostavke
 - Organizacija blokova na disku se ne razmatra (da li su susedni ili ne), uzima se prosečno vreme I/O
 - Zanimarujemo postojanje direktorijum strana ili header strane heap fajla
 - Ne uzima se u obzir pre-fetch
 - Zanimaruje se vreme potrebno za izvršavanje operacija kada su podaci u memoriji
 - Za pretragu jedne vrednosti se podrazumeva da postoji samo jedan odgovarajući slog
 - Podrazumeva se da se radi dodavanje/brisanje jednog sloga
 - Dodavanje u heap fajl podrazumeva da se slog dodaje na kraj
 - Sortirani fajl ima pakovanu organizaciju – sortirani je po ključu pretrage, brisanje podrazumeva reorganizaciju prostora na strani/ama
-

Skeniranje fajla

Troškovi operacija

File scan

Skeniranje celog fajla

- Heap fajl
 - $B * D$
- Uređeni fajl
 - $B * D$

B – broj strana

R – broj slogova u bloku

D – prosečno I/O vreme

Heap File



Sorted File



For illustration, records are just integers

$$B = 5, R = 2, D = 5 \text{ ms}$$

Pretraga jednog sloga

Složenost pretrage

Equality Search

B – broj strana

R – broj slogova u bloku

D – prosečno I/O vreme

Heap fajl

- $P(i)$ – verovatnoća da se traženi slog nađe na i -toj strani
 - $P(i) = 1 / B$
- $T(i)$ – broj učitavanih strana ako je traženi slog na strani i
 - $T(i) = i$
- Procenjen broj strana

$$\sum_{i=1}^B T(i)P(i) = \sum_{i=1}^B i \frac{1}{B} = \frac{1}{B} \sum_{i=1}^B i = \frac{1}{B} \frac{B(B+1)}{2} \approx \frac{B}{2}$$

- Procenjeno vreme

$$\frac{B}{2} D$$

Uređeni fajl

- Binarna pretraga. U najgorem slučaju će biti učitano $\log_2 B$
- $P(i) = 2^{i-1} / B$ – ako je učitano i strana, verovatnoća da se slog nalazi u i -toj
- $T(i) = i$
- Procenjen broj strana

$$\sum_{i=1}^{\log_2 B} T(i)P(i) = \sum_{i=1}^{\log_2 B} i \frac{2^{i-1}}{B} = \frac{1}{B} \sum_{i=1}^{\log_2 B} i 2^{i-1} = \log_2 B - \frac{B-1}{B}$$

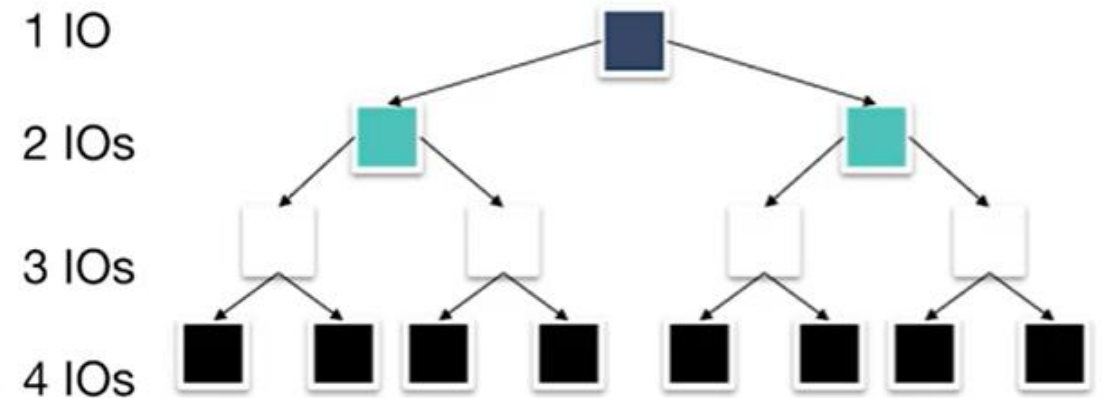
- Procenjeno vreme

$$\log_2 B * D$$

B – broj strana

R – broj slogova u bloku

D – prosečno I/O vreme



$$\sum_{i=1}^n i 2^{i-1} = \frac{1 - (n+1)r^n + nr^{n+1}}{(1-r)^2}$$

Pretraga u opsegu

Složenost pretrage

Range Search

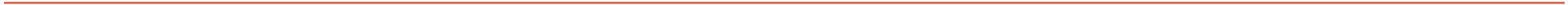
Heap fajl

- Moraju se pretražiti sve strane
- Procenjen broj strana B
- Procenjeno vreme BD

B – broj strana

R – broj slogova u bloku

D – prosečno I/O vreme



B – broj strana

R – broj slogova u bloku

D – prosečno I/O vreme

Uređeni fajl

- Binarna pretraga dok se ne pronađe prvi slog u opsegu. Nakon se učitava onoliko strana koliko je potrebno da se obuhvate sve vrednosti koje pripadaju opsegu.
- Procenjen broj strana

$$\log_2 B - \frac{B - 1}{B} + \#strana_u_opsegu$$

- Procenjeno vreme

$$(\log_2 B + \#strana_u_opsegu) * D$$

Dodavanje i brisanje

Složenost pretrage

Insert, delete

Dodavanje

- Heap fajl
 - Pretpostavka – dodavanje na kraj
 - Procenjeno vreme $2D$
 - Učitavanje jedne strane i pisanje izmenjene na disk
 - Uređeni fajl
 - $\log_2 B * D$ - za pronalaženje pozicije
 - Plus izmena svih strana 'desno' od dodatog sloga $D * 2 * \frac{B}{2} = BD$
 - Ukupno $(\log_2 B + B) * D$
-

Brisanje

- Heap fajl
 - Pretpostavka – dodavanje na kraj
 - Procenjeno vreme $(\frac{B}{2} + 1)D$
 - Pronalaženje, brisanje, pisanje izmenjene strane na disk
 - Uređeni fajl
 - $\log_2 B * D$ - za pronalaženje pozicije
 - Plus izmena svih strana 'desno' od dodatog sloga $D * 2 * \frac{B}{2} = BD$
 - Ukupno $(\log_2 B + B) * D$
-

Sumarna tabela

Operacija	Prosečno vreme					
	Heap	Sortiran	B = 5 D = 5 ms		B = 5000 D = 5 ms	
Skeniranje	$B * D$	$B * D$	25ms	25ms		
Jedna vrednost	$B * D / 2$	$\log_2 B * D$	12,5ms	11,6ms	12500ms	61.45ms
Opseg vrednosti	$B * D$	$(\log_2 B + \text{strane u opsegu}) * D$	25ms	11,6ms*	25000ms	61.45ms*
Dodavanje	$2 * D$	$(\log_2 B + B) * D$	10ms	36.6ms		
Brisanje	$(B/2+1)^* D$	$(\log_2 B + B) * D$	17,5ms	36.6ms		