

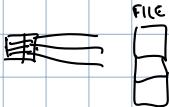
Esercizio 4:

Si consideri un file contenente 100.000.000 record di dimensione prefissata pari a 500 byte, memorizzati in blocchi di dimensione pari a 4096 byte in modo unspanned. La dimensione del campo chiave primaria V sia 14 byte; la dimensione del puntatore a blocco P sia 6 byte. Si chiede di confrontare fra loro le seguenti soluzioni, in termini di numero medio di accessi a blocco e di dimensione dell'indice.

- Ricerca basata su un indice (primario) costruito sul campo chiave primaria V .
- Ricerca basata su un indice multilivello statico ottenuto a partire dall'indice primario.
- Ricerca basata su un B^+ -albero, con campo di ricerca il campo chiave primaria V , puntatore ai dati di dimensione pari a 7 byte e puntatore ausiliario di dimensione pari a 6 byte, assumendo che ciascun nodo del B^+ -albero sia pieno al 70%.

$$M = 100.000.000 \quad \text{dim rec} = 500 \quad B = 4096$$

a. $bfr = \left\lceil \frac{4096}{500} \right\rceil = 8 \text{ record per blocco}$



bfr $\rightarrow L$ ce ne stanno meno

mb $\rightarrow \lceil \frac{M}{bfr} \rceil$ dato che è unspanned

ricerca $\rightarrow \lceil \frac{M}{bfr} \rceil$

$$mb = \left\lceil \frac{1000000000}{4096} \right\rceil = 1250000 \text{ blocchi per il file}$$

$$bfr = \left\lceil \frac{4096}{500} \right\rceil = 204 \text{ record per blocco}$$

indice PRIMARIO $\rightarrow 1 \text{ entry per OGNI BLOCCO del FILE}$ $m_{bi} = \frac{mb}{bfr}$

indice SECONDARIO $\rightarrow 1 \text{ entry per OGNI RECORD del FILE}$ $m_{bi} = \frac{mr}{bfr}$

in ogni index entry ho un blocco $\rightarrow 204$ blocchi del file in un blocco del file indice

$$m_{bi} = \left\lceil \frac{1250000}{204} \right\rceil = 61275 \text{ blocchi per l'indice}$$

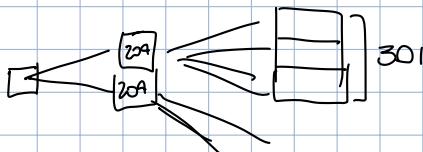
204 \rightarrow con ricerca DIC mom/2 con RIC UN faccio/2

una ricerca richiede (IN MEDIA) $\lceil \log_2(61275) \rceil + 1 = 17$ accessi

- b. trasformare indice primario in multilivello statico.

$$f_0 = bfr = 204$$

$$\text{indice di 2 livello PRIM: } m_{b2} = \left\lceil \frac{61275}{204} \right\rceil = 301 \text{ blocchi}$$



$$\text{indice di 3 livello PRIM: } m_{b3} = \left\lceil \frac{301}{204} \right\rceil = 2 \text{ blocchi}$$

$$\text{indice di 4 livello PRIM: } m_{b4} = \left\lceil \frac{2}{204} \right\rceil = 1 \text{ blocco}$$

una ricerca in media richiede $4 + 1 = 5$ accessi

perché a differenza del caso con un solo indice (dove va cercato nell'indice) nel multilivello io mi "guido" dentro l'albero e ad ogni livello cerco solo in un blocco (e al suo interno faccio scansione lineare dei record)

Esercizio 4:

Si consideri un file contenente 100.000.000 record di dimensione prefissata pari a 500 byte, memorizzati in blocchi di dimensione pari a 4096 byte in modo unspanned. La dimensione del campo chiave primaria V sia 14 byte; la dimensione del puntatore a blocco P sia 6 byte. Si chiede di confrontare fra loro le seguenti soluzioni, in termini di numero medio di accessi a blocco e di dimensione dell'indice.

- Ricerca basata su un indice (primario) costruito sul campo chiave primaria V .
- Ricerca basata su un indice multilivello statico ottenuto a partire dall'indice primario.
- Ricerca basata su un B^+ -albero, con campo di ricerca il campo chiave primaria V , puntatore ai dati di dimensione pari a 7 byte e puntatore ausiliario di dimensione pari a 6 byte, assumendo che ciascun nodo del B^+ -albero sia pieno al 70%.

C.

$$B = 4096 \text{ byte}$$

$$V = 14 \text{ byte}$$

$$P_R = 7 \text{ byte}$$

$$P = 6 \text{ byte}$$

$$P \cdot 0,7$$

$$P_{LEAF} \cdot 0,7$$

$$B \leq P \cdot p + (p-1) \cdot V$$

$$B \leq P_{LEAF} (V + P_R) + P$$

devo calcolare p, P_{LEAF} e poi vedo quanti record ho per blocco

$$4096 \leq 6p + (p-1) \cdot 14 \Rightarrow 4096 \leq 6p + 14p - 14 \Rightarrow 4110 \leq 20p \Rightarrow p = 205$$

$$4096 \leq 21p_{LEAF} + 6 \Rightarrow 4090 \leq 21p_{LEAF} \Rightarrow p_{LEAF} = 194$$

$$\text{Se ciascun nodo è occupato al 70\%: } P_{LEAF} \cdot 0,7 \approx 136$$

$$P \cdot 0,7 \approx 144$$

quindi ogni nodo foglia avrà mediamente 136 chiavi e 136 puntatori
interno avrà mediamente 143 chiavi e 144 puntatori

Quanto scendiamo? fino a gestire 100.000.000 record

quindi costruisca la tabella:

	modi	valori	puntatori	la radice è un nodo interno
RADICE	1	143	144 <small>144 \cdot 144</small>	
livello 1	144	20592	20593 <small>20593 \cdot 20736</small>	
livello 2	20736	2965248	2985384	
livello 3	2985384	426335712 <small>426335712 \cdot 136</small>	406093824 <small>406093824 \cdot 136 (è foglia)</small>	

con 3 livelli ci arriviamo e il numero di accessi è $3+1+1 = 5$ accessi

Esercizio 1.4 Sia dato un file contenente $r = 200000$ record, memorizzati su un disco con dimensione del blocco $B = 1024$ byte. I record abbiamo lunghezza fissa $R = 100$ byte e siano di tipo unspanned. La dimensione del campo chiave sia $V1 = 10$ byte; la dimensione del puntatore ad un blocco sia $P = 5$ byte.

- Si assuma che il file non sia ordinato rispetto ad alcun campo. Determinare il numero medio di accessi a blocco richiesto da una ricerca basata sulla chiave primaria.
- Si assuma che il file sia ordinato rispetto alla chiave primaria. Determinare il numero medio di accessi a blocco richiesto da una ricerca basata sulla chiave primaria.
- Si assuma che il file sia ordinato rispetto alla chiave primaria. Determinare il numero medio di accessi a blocco richiesto da una ricerca con indice primario.
- Si assuma che il file sia ordinato rispetto alla chiave primaria. Determinare il numero medio di accessi a blocco richiesto da una ricerca con indice secondario costruito su una chiave candidata (diversa dalla chiave primaria) di dimensione $V2 = 10$ byte.

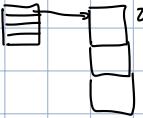
a. $bfr = \lfloor \frac{1024}{100} \rfloor = 10$ record

$$mb = \lceil \frac{200000}{10} \rceil = 20000 \text{ blocchi}$$

di media, facendo ricerca lineare, accedo a $\frac{20000}{2} = 10000$ blocchi

b. di media, facendo ricerca binaria, accedo a $\lceil \log_2(20000) \rceil = 15$ blocchi

c. $bfr_i = \lfloor \frac{1024}{15} \rfloor = 68$ record



$$mbi = \lceil \frac{20000}{68} \rceil = 295 \text{ blocchi}$$

di media, la ricerca richiede $\lceil \log_2(295) \rceil + 1 = 10$ accessi

d. $bfr_i = \lfloor \frac{1024}{15} \rfloor = 68$ record

$$\#index entry = 200000$$

$$mbi = \lceil \frac{200000}{68} \rceil = 2942 \text{ blocchi}$$

di media il #di accessi è $\lceil \log_2(2942) \rceil + 1 = 13$ accessi

Esercizio 1.5 Sia dato un file ordinato con $r = 200000$ record memorizzati su un disco con dimensioni del blocco $B = 2056$ byte. I record abbiano lunghezza fissa $R = 400$ byte e siano di tipo unspanned. La dimensione del campo chiave primaria sia $V1 = 20$ byte, mentre la dimensione del campo chiave candidata (diversa dalla chiave primaria) sia $V2 = 15$ byte. La dimensione del puntatore a blocco sia $P = 5$ byte.

- Si assuma che il file sia ordinato rispetto al campo chiave primaria. Determinare il numero di accessi a blocco richiesti rispettivamente da una ricerca con indice primario e da una ricerca con indice secondario sulla chiave candidata (diversa dalla chiave primaria).
- Si assuma che il file non sia ordinato rispetto ad alcun campo. Determinare il numero di accessi a blocco richiesti rispettivamente da una ricerca con indice secondario sulla chiave primaria e da una ricerca con indice secondario sulla chiave candidata (diversa dalla chiave primaria).
- Confrontare tra loro le soluzioni proposte ai punti (a) e (b) e discuterne vantaggi e svantaggi.

a. $bfr_2 = \left\lfloor \frac{2056}{400} \right\rfloor = 5$ record

$$mb = \left\lceil \frac{200000}{5} \right\rceil = 40000 \text{ blocchi}$$

con indice prim: $bfr_1 = \left\lfloor \frac{2056}{25} \right\rfloor = 82$ record

$$mb_1 = \left\lceil \frac{40000}{82} \right\rceil = 488 \text{ blocchi} \rightarrow \text{accedo a } \lceil \log_2(488) \rceil + 1 = 10 \text{ accessi}$$

con indice sec: $bfr_2 = \left\lfloor \frac{2056}{20} \right\rfloor = 102$ record

$$mb_1 = \left\lceil \frac{200000}{102} \right\rceil = 1961 \text{ blocchi} \rightarrow \text{accedo a } \lceil \log_2(1961) \rceil + 1 = 12 \text{ accessi}$$

- b. Se il file è disordinato e il file indice sec usa chiave secondaria $\rightarrow 12$ accessi (non cambia che sia ordinato o no il file dei dati)

$$bfr_1 = \left\lfloor \frac{2056}{25} \right\rfloor = 82$$

$$mb_1 = \left\lceil \frac{200000}{82} \right\rceil = 2440 \text{ blocchi} \rightarrow \text{accedo a } \lceil \log_2(2440) \rceil + 1 = 13 \text{ accessi}$$

a liv. di blocco

(non cambia che sia ordinato o no il file dei dati)

↳ se è ordinato faccio

la ricerca dicotomica DENTRO al blocco

se non è ordinato faccio la

RICERCA LINEARE DENTRO al blocco

Esercizio 1.7 Sia dato un file ordinato contenente $r = 100000$ record memorizzati su un disco con dimensioni del blocco $B = 1024$ byte. I record abbiano lunghezza fissa $R = 200$ byte e siano di tipo unspanned. La dimensione del campo chiave sia $V1 = 10$ byte. Il puntatore ad un blocco sia lungo $P = 5$ byte.

- (a)] Si assuma che il file sia ordinato rispetto al campo chiave. Determinare il numero di accessi a blocco richiesti rispettivamente da una ricerca binaria e da una ricerca con indice primario.
- (b) Si assuma che il file sia ordinato rispetto ad un campo non chiave di dimensione $V2 = 20$ byte. Determinare il numero di accessi a blocco richiesti rispettivamente da una ricerca lineare e da una ricerca con indice secondario denso costruito su un campo chiave di dimensione $V3 = 15$ byte.
- (c) Si consideri l'indice secondario denso definito al punto precedente e lo si converta in un indice multilivello. Calcolare il numero di livelli dell'indice, il numero di blocchi per ogni livello ed il numero di accessi necessario per recuperare un blocco del file di dati.

$$bfr = \left\lfloor \frac{1024}{200} \right\rfloor = 5 \text{ record} \quad mb = \left\lceil \frac{100000}{5} \right\rceil = 20000 \text{ blocchi}$$

a. ricerca binaria $\rightarrow \lceil \log_2(20000) \rceil = 15$ accessi in media

$$bfr_i = \left\lfloor \frac{1024}{15} \right\rfloor = 68 \text{ record} \quad mbi = \left\lceil \frac{20000}{68} \right\rceil = 295 \text{ blocchi}$$

com file indice $\rightarrow \lceil \log_2(295) \rceil + 1 = 10$ accessi in media

b. ricerca lineare in media: $\frac{20000}{2} = 10000$ accessi

$$bfr_2 = \left\lfloor \frac{1024}{20} \right\rfloor = 51 \text{ record a blocco}$$

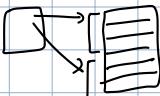
$$mb_2 = \left\lceil \frac{100000}{51} \right\rceil = 1961 \text{ blocchi}$$

la ricerca binaria com indice secondario $\rightarrow \lceil \log_2(1961) \rceil + 1 = 12$ accessi

c. $f_0 = bfr_2 = 51$

ogni record punta a 1 blocco

un blocco dell'indice di 2 liv. punta a 51 blocchi



$$mb_2 = \left\lceil \frac{1961}{51} \right\rceil = 39 \text{ blocchi}$$

$$mb_3 = \left\lceil \frac{39}{51} \right\rceil = 1 \text{ blocco}$$

ci vogliono 3 livelli \rightarrow per una ricerca ci vogliono 4 accessi

Esercizio 5:

Si consideri un file contenente 100.000.000 record di dimensione prefissata pari a 500 byte, memorizzati in blocchi di dimensione pari a 4096 byte in modo unspanned. La dimensione del campo chiave primaria V sia 14 byte; la dimensione del puntatore a blocco P sia 6 byte. Si chiede di confrontare fra loro le seguenti soluzioni, in termini di numero medio di accessi a blocco e di dimensione dell'indice.

- Ricerca basata su un indice multilivello statico ottenuto a partire da un indice secondario costruito sul campo chiave primaria V .
- Ricerca basata su un B -albero, con campo di ricerca il campo chiave primaria V , puntatore ai dati di dimensione pari a 7 byte e puntatore ausiliario di dimensione pari a 6 byte, assumendo che ciascun nodo del B -albero sia pieno al 70%.
- Ricerca basata su un B^+ -albero, con campo di ricerca il campo chiave primaria V , puntatore ai dati di dimensione pari a 7 byte e puntatore ausiliario di dimensione pari a 6 byte, assumendo che ciascun nodo del B^+ -albero sia pieno al 70%.

a) indice secondario \rightarrow demo

$$bf_{ri} = \left\lceil \frac{4096}{20} \right\rceil = 204$$



$$m_{blocchi} = \left\lceil \frac{100.000.000}{204} \right\rceil = 490197 \text{ blocchi}$$

ora devi mappare 204 blocchi in un blocco

$$f_o = bf_{ri}$$

$$m_{blocchi_2} = \left\lceil \frac{490197}{204} \right\rceil = 2403 \text{ blocchi}$$

\Rightarrow una ricerca richiede $1 + 4 = 5$ accessi
numero di blocchi dell'indice è = 492613

$$m_{blocchi_3} = \left\lceil \frac{2403}{204} \right\rceil = 12 \text{ blocchi}$$

$$m_{blocchi_4} = \left\lceil \frac{12}{204} \right\rceil = 1 \text{ blocco}$$

b) $V = 14 \text{ byte}$
 $P_r = 7 \text{ byte}$
 $P = 6 \text{ byte}$
 $p \cdot 0,7$

$$\text{in un } B\text{-albero: } p \cdot P + (p-1)(P+P_r) \leq B$$

$$6p + 21p - 21 \leq 4096$$

$$p \leq 152 \rightarrow p = 152 \cdot 0,7 = 106$$

in ogni modo ho 106 puntatori e 105 chiavi \rightarrow con quanti livelli copro 100.000.000 record?

livello	modi	puntatori	chiavi
1	1	106	105
2	106	11236	11130
3	11236	1191016	1179780
4	1191016	126247696	125056680

e gli accessi sono: $4+1=5$

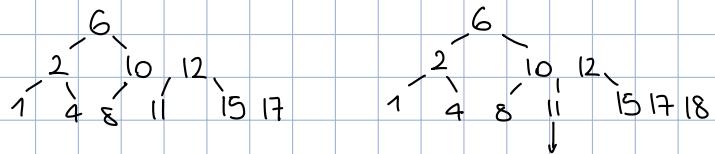
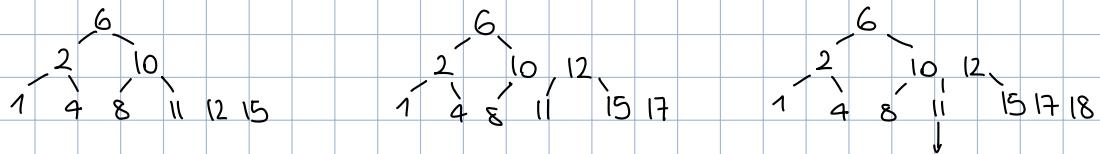
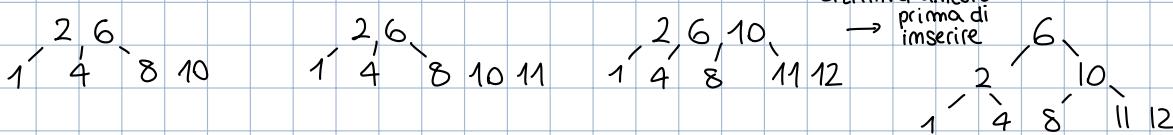
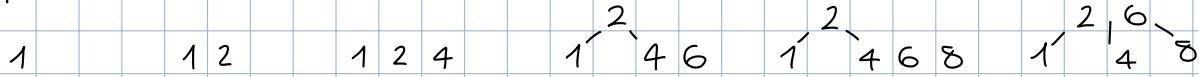
Esercizio 5:

Dato il seguente insieme di chiavi:

1, 2, 4, 6, 8, 10, 11, 12, 15, 17, 18,

mostrare il *B*-albero di ordine 4 ottenuto inserendo un elemento dopo l'altro nell'ordine dato (riportando la sequenza di alberi generata dal processo di inserimento).

$$p=4$$



es. point query per trovare 11: 3 accessi ai blocchi + 1 accesso ai dati

per trovare 16: 3 accessi e esito negativo

per le range query: quanti blocchi per es. 8-17? → con B-ALBERI: 5 (devo fare su e giù)

con B₊ALBERI: mi sposto tra le foglie