Hash files

Assume we have a file with **250,000 records**, each of them of size **300** bytes, **75** of which for the key field. Each block is of size **1024** bytes. A block pointer is **4 bytes**.
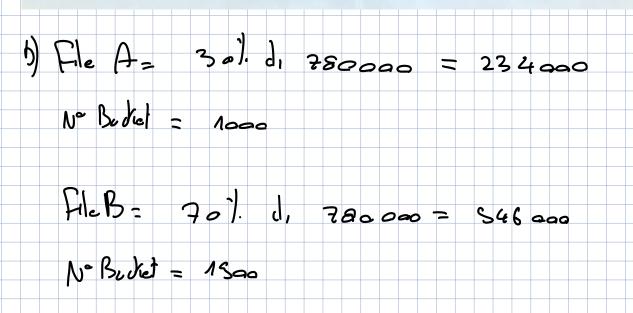
a) If we use a hash organization with **1200** buckets, how many blocks do we need for the bucket directory?

b) How many blocks do we need for the buckets, assuming a uniform distribution of records into buckets?

c) Still assuming a uniform distribution of records, what is the average number of accesses needed to find a record in the file?

d) How many buckets should we create to have instead an average number of accesses less or equal 10, still assuming a uniform distribution of records in the buckets?

---

a) 1200 puichè corrispondere bioni voci

$$\text{Qnti Pointer} \times \text{Blocco (PBlck)} = \left\lfloor \frac{\text{Block Size}}{\text{Pointer size}} \right\rfloor = \frac{1024}{4} = 256$$

↳ App. per blf.

$$\text{n° Blocchi} \times \text{Buddet dr} = \left\lceil \frac{\text{N° Buodet}}{\text{PBlck}} \right\rceil = \frac{1200}{256} = 5$$

↳ App per ecc.

$$\text{Record} \times \text{Blocco} = \frac{\text{Block Size} - \text{pointer Size}}{\text{Record Size}} = \frac{1024 - 4}{300} = 3$$

↳ per blf

$$\text{Record} \times \text{Buodet} = \frac{\text{N Record}}{\text{N° BKT}} = \frac{250\,000}{1200} = 209$$

↳ per eccesso

$$\text{Blocchi per Buodet} = \frac{\text{Rec} \times \text{Buodet}}{\text{Record} \times \text{Blocco}} = \frac{209}{3} = 70$$

$$\text{Tot Blocchi} = \text{Blocchi per Budget} \cdot \text{N Blocchi} = 70 \cdot 1200 = 84000$$

c) n° MEDIO BLOCCHI DA CONTROLLARE PER TROVARE RECORD

$$\text{AvgTime} = \frac{\text{Blocchi per Budget}}{2} = 35$$

↳ per eccesso

d) Voglio che AvgTime passi da 35 a ≤ 10

L'unico parametro modificabile è n° budget (1200) aumentandolo

devo rifare quindi tutti i conti all'indietro:

$$\frac{\text{Blck} \cdot \text{Budget}}{2} \le 10 \implies \text{Blck} \cdot \text{Budget} \le 20$$

$$\implies \frac{\text{Record} \times \text{Budget}}{\underbrace{\text{Record} \times \text{Blocco}}_{\to \text{NON MODIFICABILE}}} \le 20 \qquad \implies$$

$$\frac{\text{Record} \times \text{Budget}}{3} \le 20 \implies \text{Record} \times \text{Budget} \le 60$$

↳ NON MODIFICABILE

$$\underbrace{\frac{\text{N° record}}{\text{N° Budget}}}_{} \le 60$$

$$\text{N° Budget} \cdot 60 \ge \text{N° record}$$

$$\text{N° Budget} \ge \frac{\text{N° record}}{60}$$

$$\text{N° Budget} \ge \frac{2500000}{60} =$$
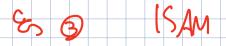
$$\text{N° Budget} \ge 4167$$

Assume we have a file with **780.000** records, each of them of size **250** bytes. Each block is of size **1024** bytes. A block pointer is **4** bytes. We are using a hash organization with **2500** buckets.

a) How many blocks do we need to use for the bucket directory and for the buckets, assuming a uniform distribution of records?

b) How many blocks do we need to use, assuming that 30% of records are uniformly distributed in **1000** buckets, and the remaining 70% is uniformly distributed in the remaining **1500** buckets?

b) File A = 30% di 780000 = 234000

N° Buckt = 1000

File B = 70% di 780000 = 546000

N° Buckt = 1500

Rifaccio gli stessi calcoli dovo calcolare Tot Block

Assume we have a file of **150.000** records, each of them of size **250** bytes, **50** of which are used for the key field. Each block is of size **1024** bytes. A block pointer is of size **4** bytes.

   a) If we use an ISAM organization, we assume that blocks are filled at most at **70%** of their capacity, how many blocks do we need for the index file?
   b) If we use binary search, what is the maximum number of accesses to blocks to find a record in the file for case a)?

a)

$$\text{Block Size} = 1024 \qquad\qquad\qquad \text{Per difette}$$

$$\text{Spazio Usabile (Non +70\%.)} = \frac{1024 \cdot 70}{100} \overset{\uparrow}{\simeq} 716$$

$$\text{Record} \times \text{Block} = \frac{\text{Block Size}}{\text{Record Size}} = \frac{716}{250} = 2$$

$$\text{Tot Block} = \frac{\text{N Record}}{\text{Record} \times \text{Block}} = \frac{150000}{2} = 75000$$

$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{per dif}$$

$$\text{Record per Blocco} = \frac{\text{Block Size}}{\text{key Size} + \text{Pointer Size}} = \frac{716}{54} \overset{!}{\simeq} 13$$

$$\text{N° Blocco Indice} = \frac{\text{Tot Block}}{\text{Record} \times \text{Blocco}} = \frac{75000}{13} = 5770$$

b) Search Time $= \lceil \log_2 (N^{\circ} Block) \rceil + 1 = 13 + 1 = 14$