

Basi di dati II, primo modulo

Esercizi di autovalutazione — 4 marzo 2011

Cenni sulle soluzioni

Note

- Gli esercizi sono facoltativi ma è estremamente importante svolgerli (e anche discuterli ad esempio sul forum), perché le domande proposte nei compiti di esame possono essere molto simili.

Domanda 1 Si consideri un disco con una singola faccia che contenga $T=20.000$ tracce e con una velocità di rotazione di $v=10.000$ giri al minuto. Ogni traccia contiene $N=240$ blocchi e ogni blocco contiene $L=2KB$. Rispondere alle seguenti domande mostrando formula e valore numerico numero (N.B. non servono calcolatrici, i risultati sono semplici, approssimare $1GB=1000MB=1.000.000KB$).

1. Qual è la capacità del disco (in GB)?

$$T \times N \times L = 20.000 \times 240 \times 2KB = 9.600.000KB = 9,6GB$$

2. Qual è la massima velocità di trasferimento (in MB al secondo)?

$$v/60 \times N \times L = 10.000/60 \times 240 \times 2KB = 80.000KB/sec = 80MB/sec$$

Domanda 2 Considerare un disco con una velocità di rotazione di $v = 10.000$ giri al minuto e un tempo medio di posizionamento della testina (tempo di seek) $t_S = 5$ msec. Ogni traccia contiene $N = 240$ blocchi e ogni blocco contiene $B = 4$ KB. Rispondere alle seguenti domande mostrando formula e valore numerico numero (N.B. non servono calcolatrici, i risultati sono semplici, approssimare $1 MB = 1000 KB$).

1. Qual è tempo medio di latenza (attesa dovuta alla rotazione) t_L ?

$$\text{è il tempo necessario per mezzo giro: } 1/2 \times 1/v \times 60 = 1/2 \times 1/10.000 \times 60 = 0,003 \text{ sec} = 3 \text{ msec}$$

2. Qual è la massima velocità di trasferimento (in MB al secondo)?

$$v/60 \times N \times B = 10.000/60 \times 240 \times 4 \text{ KB} = 160.000 \text{ KB/sec} = 160 \text{ MB/sec}$$

3. Qual è il tempo minimo t_B necessario per leggere un blocco?

$$1/v \times 60 \times 1/N = 1/10.000 \times 60 \times 1/240 = 1/10.000 \times 1/240 \times 60 \text{ sec} = 25 \mu\text{sec}$$

4. Qual è il tempo medio necessario per leggere un blocco?

$$t_S + t_L + t_B = 5 + 3 + 0,025 \text{ msec} = \text{ca } 8 \text{ msec}$$

5. Qual è il tempo medio necessario per leggere $n = 40$ blocchi contigui?

$$t_S + t_L + n \times t_B = 5 + 3 + 40 \times 0,025 \text{ msec} = \text{ca } 9 \text{ msec}$$

Domanda 3 Installare un DBMS commerciale o di pubblico dominio, ad esempio uno fra DB2, Oracle, SQLServer, Postgres, Derby, MySQL (sono più o meno tutti scaricabili in rete).

Domanda 4 Installare il DBMS didattico SimpleDB (<http://www.cs.bc.edu/sciore/simplydb/>).

Domanda 5 Si supponga di avere un file heap con numero di record $N=1.000.000$ di $L=100$ byte ciascuno con un sistema che utilizzi blocchi di dimensione $D=4K$ byte. Quanti accessi a memoria secondaria sono necessari per trovare un record con un certo valore su un campo?

Nel caso peggiore, un numero di accessi pari al numero di blocchi del file. Se assumiamo che il record sia presente, possiamo pensare che ne bastino mediamente la metà (o anche meno se ci sono ripetizioni). Il numero di blocchi è pari a $N/(D/L) \approx 25.000$.

Domanda 6 Si supponga di avere un file definito su vari campi fra cui un identificatore che viene costruito concatenando un numero d'ordine relativo alla giornata con la data corrente (ad esempio, 1223022011 è il dodicesimo identificatore generato il 23 febbraio 2011). Spiegare perché, su questo campo, una funzione hash basata sul resto della divisione sarebbe estremamente inefficace.

In effetti, l'affermazione è vera solo rispetto ad una divisione per una potenza di 10 o un suo multiplo (o, se si lavora sulla rappresentazione binaria della stringa, per una potenza di 2). In tal caso, in presenza di una probabile ripetizione del valore dell'anno, gli identificatori avrebbero spesso le stesse ultime cifre e quindi i valori della funzione hash sarebbero tutt'altro che uniformemente distribuiti.