

## Basi di dati II, primo modulo Prova parziale — 22 marzo 2010 — Compito A

### Cenni sulle soluzioni

Rispondere su questo foglio. Usare il foglio protocollo come brutta copia.  
Tempo a disposizione: un'ora.

Cognome \_\_\_\_\_ Nome \_\_\_\_\_ Matricola \_\_\_\_\_ Ordin. \_\_\_\_\_

**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** Si consideri una relazione  $R(CodiceCliente, Cognome, Nome, Categoria)$  con  $N = 1.000.000$  ennuple. Con riferimento alla ricerca di tutti i clienti di una certa categoria, indicare il costo dell'accesso sequenziale e di quello diretto con indice su *Categoria* nei due casi seguenti (mostrare formule e valori numerici):

1. campo selettivo ( $v_1 = 100.000$  valori diversi per *Categoria*)

$$\text{costo accesso sequenziale: } N/f_R = 20.000 \quad \text{costo accesso diretto: } p + N/v_1 \approx 14$$

2. campo poco selettivo ( $v_2 = 20$  valori diversi per *Categoria*)

$$\text{costo accesso sequenziale: } N/f_R = 20.000 \quad \text{costo accesso diretto: } p + N/v_2 + \dots \approx 50.000$$

Supporre che l'indice abbia profondità  $p = 4$  e che i fattori di blocco del file e dell'indice siano rispettivamente  $f_R = 50$  e  $f_C = 200$ .

**Domanda 3** Si consideri una relazione  $R(\underline{A} \ B \ C \ D \ E)$ , in cui gli attributi hanno tutti la stessa dimensione  $a$  (ad esempio, ma è irrilevante, 4Byte), molto più piccola della dimensione del blocco pari a  $P$ . Si supponga che la relazione sia molto grande ( $N$  ennuple) e che le operazioni più frequenti su di essa siano le seguenti:

$o_1$  `SELECT * FROM R ORDER BY A`, con frequenza  $f_1$

$o_2$  `SELECT A, B, C FROM R ORDER BY A`, con frequenza  $f_2 = 10 \times f_1$

Valutare le due seguenti alternative di memorizzazione, calcolando il costo complessivo (riportare la formula che indica il numero di accessi nell'unità di tempo, in base alle variabili sopra citate):

- (i) memorizzazione della relazione  $R(\underline{A} \ B \ C \ D \ E)$  ordinata su  $A$

$$\text{costo unitario di } o_1: \quad c_1 = N/(P/(5a)) = 5aN/P \quad \text{costo unitario di } o_2: \quad c_2 = N/(P/(5a)) = 5aN/P$$

$$\text{costo complessivo: } c_1 f_1 + c_2 f_2 = N/(P/(5a)) = 55f_1 aN/P$$

- (ii) memorizzazione delle proiezioni  $R1(\underline{A} \ B \ C)$  e  $R2(\underline{A} \ D \ E)$ , entrambe ordinate su  $A$

$$\text{costo unitario di } o_1: \quad c_1 = 2N/(P/(3a)) = 6aN/P \quad \text{costo unitario di } o_2: \quad c_2 = N/(P/(3a)) = 3aN/P$$

$$\text{costo complessivo: } c_1 f_1 + c_2 f_2 = 36f_1 aN/P$$

Supporre che il join possa essere eseguito con il metodo merge-join (e che il costo del join stesso sia trascurabile rispetto alle due scansioni).

## Compito A

**Domanda 4** Si consideri un B-tree con nodi intermedi che contengono due chiavi e tre puntatori e foglie con due chiavi, in cui vengano inserite chiavi (a partire dall'albero vuoto) nel seguente ordine: 12, 22, 32, 42, 52, 13, 14, 15, 16, 17, 18. Mostrare l'albero dopo l'inserimento di tre, cinque, sette chiavi e alla fine.

provare con l'applet suggerita a lezione: <http://slady.net/java/bt/view.php?w=700>

**Domanda 5** Si consideri una base di dati sulle seguenti relazioni, ognuna delle quali ha una struttura heap e un indice secondario sulla chiave (si suppongano le relazioni tutte molto grandi, con  $S_1$  dieci volte più piccola delle altre due)

- $S_1(\underline{ABC})$  con vincolo di integrità referenziale fra  $C$  e la chiave  $D$  di  $S_2$
- $S_2(\underline{DEF})$
- $S_3(\underline{GHL})$

e con una vista definita come segue:

- `CREATE VIEW V AS SELECT * FROM (S1 JOIN S2 ON B=D) LEFT JOIN S3 ON C=G`

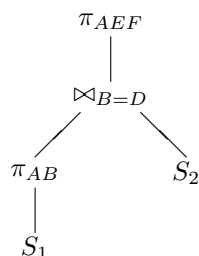
In tale contesto, supponendo che il sistema esegua join (senza trarre profitto significativo dal buffer) solo con nested loop, utilizzando gli indici ove definiti, mostrare un possibile piano di esecuzione (in termini di operatori dell'algebra relazionale e loro realizzazioni) per ciascuna delle seguenti interrogazioni

1. `SELECT A, B, C FROM V`
2. `SELECT A, E, F FROM V`
3. `SELECT D, E, F FROM V`

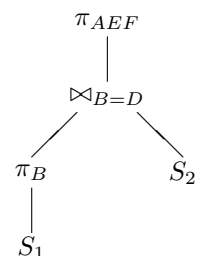
1.

$S_1$   
scansione

2.



3.



Nei join è indicata a sinistra la tabella esterna e viene eseguito un accesso diretto su quella interna.