Corso di Basi di Dati 2

Primo Compitino

9 Aprile 1999

1. (Indicativamente punti 10) Dato il seguente insieme di chiavi:

$$10, 5, 7, 9, 14, 32, 24, 18, 42, 8, 2, 4, 37, 39, 16, 12, 15$$

   (a) mostrare il B-albero di ordine 5 ottenuto inserendo un elemento dopo l'altro nell'ordine dato;

   (b) mostrare il B-albero di ordine 5 che si sarebbe invece ottenuto applicando il bilanciamento in fase di inserimento dei dati;

   (c) mostrare il B-albero ottenuto dalla cancellazione dell'elemento 9 e poi dell'elemento 2 dal B-albero in (a);

   (d) mostrare il B^+-albero di ordine 4 ottenuto inserendo un elemento dopo l'altro nell'ordine dato;

   (e) mostrare il B^+-albero ottenuto dalla cancellazione dell'elemento 7 e poi dell'elemento 8 dal B^+-albero in (c).

2. (Indicativamente punti 10)

   (a) Dato il seguente insieme di chiavi:

$$24, 18, 39, 15, 7, 13, 25, 5, 69, 3, 12, 8, 10, 23, 71$$

      i. mostrare la struttura di hash virtuale, con funzioni $H_i(k) = k \; mod \; 7 * 2^i$ e capacità di ogni pagina 2, ottenuta inserendo gli elementi nell'ordine dato;

      ii. mostrare le operazioni effettuate per ricercare in tale struttura le chiavi 7 e 36.

   (b) Mostrare la struttura di hash estensibile, con capacità di ogni pagina 2, supponendo di inserire nell'ordine i record le cui pseudochiavi hanno i seguenti valori:

$$h(r_1) = 1001000 \quad h(r_2) = 1011000 \quad h(r_3) = 1110100 \quad h(r_4) = 1000000 \quad h(r_5) = 0100000 \quad h(r_6) = 1111000$$

3. (Indicativamente, punti 6) Sia dato il seguente schema relazionale: $R(A, B, C, D, E)$ dove B, D, E sono attributi di tipo intero. Sia data inoltre una relazione r di schema R contenente 5000 tuple memorizzate in 500 pagine. Si supponga che il range dei valori assunti dall'attributo B nelle tuple di r sia [0...5000] e che i range dei valori assunti dagli attributi D ed E nelle tuple di r siano entrambi [0...100]. Si supponga inoltre di avere i seguenti indici sulla relazione r:

   - un indice I_1 clusterizzato sull'attributo B contenente 150 entrate e con 10 foglie;
   - un indice I_2 non clusterizzato sull'attributo A contenente 100 entrate e con 20 foglie;
   - un indice I_3 non clusterizzato sull'attributo C contenente 80 entrate e con 10 foglie.

   Data l'interrogazione

```
SELECT *
FROM R
WHERE A IN (k1, k2, k3) AND B >= 1000 AND
      (C = k OR D < 50) AND E BETWEEN (25,75)
```

   si valutino le possibili strategie per l'esecuzione dell'interrogazione (indicandone i relativi costi) e si determini la strategia ottima. Si fornisca inoltre una stima del numero di tuple restituite dall'interrogazione.

4. (Indicativamente punti 4) Siano date le relazioni R ed S, entrambe contenenti un attibuto A, chiave per S. La relazione R contenga 5000 tuple memorizzate in 500 pagine, e la relazione S contenga 300 tuple memorizzate in 25 pagine. Sia inoltre allocato un indice sull'attributo A della relazione R, con 300 entrate e 10 foglie. Determinare le possibili strategie per eseguire $R \bowtie S$ ed i relativi costi.