

Basi di Dati II
Corso di Laurea in Informatica
Prova Scritta
30 marzo 2011

Esercizio I

(9 punti)

Si consideri 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.

Esercizio 2

(9 punti)

Si consideri la seguente relazione (la chiave primaria è sottolineata):

MULTA(NumSerie, Data, Ora, TipoInfrazione, CodFiscale)

Si ipotizzino le seguenti cardinalità:

 $\text{card}(\text{MULTA}) \approx 10^5$ tuple, $\text{card}(\sigma_{\text{Data} \geq 1/1/2005 \text{ and } \text{Data} \leq 31/12/2005} \text{MULTA}) \approx 10^5$ tuple, $\text{card}(\sigma_{\text{TipoInfrazione}='Tipo10'} \text{MULTA}) \approx 10^3$ tuple,

Per la seguente interrogazione SQL:

```
SELECT Data, COUNT(*)  
FROM MULTA  
WHERE Data ≥ 1/10/2005 and Data ≤ 1/12/2005 and TipoInfrazione='Tipo10'  
GROUP BY Data;
```

- Si scriva l'espressione algebrica corrispondente.
- Si scriva il piano di esecuzione che potrebbe essere scelto dall'ottimizzatore in assenza di strutture fisiche accessorie.
- Si scelgano una o più strutture fisiche accessorie per migliorare le prestazioni dell'interrogazione. Si motivi la scelta e si definisca il nuovo piano di esecuzione ipotizzato.

Esercizio 3

(9 punti)

Dati i seguenti schedule:

- (i) $r_2(x), r_1(x), w_1(x), w_2(x)$;
- (ii) $r_2(x), r_1(x), w_1(x), r_2(x)$;
- (iii) $r_2(y), r_2(z), r_1(x), r_1(z), w_1(z), w_1(x), r_2(x)$;
- (iv) $w_1(y), r_2(x), w_1(x), r_2(x), r_3(y), r_0(x), w_0(x), w_2(y)$;

- (a) indicare quali sono view-serializzabili e/o conflict-serializzabili, giustificando la risposta;
- (b) per ogni schedule serializzabile, determinare gli schedule seriali equivalenti.

Esercizio 4

(6 punti)

Si progetti il DTD per la gestione di una lista di indirizzi postali e si scriva un file XML valido per il DTD progettato, tenendo in considerazione i seguenti vincoli:

- Una lista di indirizzi contiene almeno un'informazione
- Ogni informazione include il nome, zero o più indirizzi, zero o più numeri di telefono, zero o più indirizzi email, eventualmente la nazionalità, zero o più note.
- Il nome contiene il nome proprio, zero o più secondi nomi, ed il cognome.
- Un indirizzo è caratterizzato da almeno una via, un indirizzo postale, eventualmente la provincia, e sicuramente il paese.

Matricola: _____

Nome e Cognome: _____