

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

Esercizio I

(10 punti)

Si supponga di utilizzare un'unità disco con i seguenti parametri: tempo di posizionamento $s = 12$ ms, ritardo di rotazione $rd = 7$ ms, tempo di trasferimento di blocco $btt = 1$ ms, dimensione del blocco $B = 2048$ byte. Un puntatore ha dimensione $P = 8$ byte. Si consideri un file TELEFONATE che abbia $r = 256000$ record di lunghezza fissa con i seguenti campi: *NUM_CHIAMANTE* (16 byte), *NUM_CHIAMATO* (16 byte), *DATA_CHIAMATA* (10 byte), *ORA_INIZIO* (8 byte), *ORA_FINE* (8 byte), *PIANO_TARIFFARIO* (3 byte), *OPZIONE* (2 byte). Un byte aggiuntivo viene utilizzato per indicare la cancellazione di un record.

- (a) Calcolare la dimensione R di un record in byte, il fattore di blocco bfr e il numero di blocchi b necessari a memorizzare la relazione assumendo un'organizzazione senza spanning e un fattore di riempimento $f = 100\%$.
- (b) Si supponga che il file sia ordinato cronologicamente (secondo i campi *DATA_CHIAMATA* e *ORA_INIZIO*) e che si voglia costruire un indice su *NUM_CHIAMANTE* per cercare di ottimizzare le ricerche delle telefonate effettuate. Si assuma inoltre che il numero medio di telefonate effettuate da ciascun *NUM_CHIAMANTE* sia pari a $n = 250$, che il fattore di riempimento per l'indice sia pari a $f' = 75\%$ e che il fattore di riempimento per il livello dei puntatori ai record del file sia pari a $f'' = 100\%$.
Si calcoli il fattore di blocco $bfri$ dell'indice (che sarà anche il suo fan-out), il numero di voci ri dell'indice al primo livello ed il numero di blocchi bi al primo livello; si determini inoltre il numero di blocchi necessari nel livello intermedio per memorizzare tutti i puntatori ai record; si calcoli poi il numero di livelli necessari se si volesse costruire un indice multilivello e il numero totale di blocchi per l'indice multilivello.
- (c) Si determini il numero di accessi minimo $nmin$ e massimo $nmax$ ai blocchi e il tempo minimo $tmin$ e massimo $tmax$ necessario per trovare tutte le chiamate effettuate da un singolo *NUM_CHIAMANTE* utilizzando l'indice realizzato al punto (b).
- (d) Ripetere i calcoli del punto (c), i.e. $nmin'$, $nmax'$, $tmin'$ e $tmax'$, ipotizzando di non utilizzare l'indice e di accedere direttamente al file dei dati. Quale deve essere la dimensione del buffer affinché sia possibile ottenere effettivamente $tmin'$?

Esercizio 2

(6 punti)

Si consideri lo schema relazionale composto dalle seguenti relazioni:

R (A, B, C, D)S (D, E, F)in cui $PK(R) = A$, $PK(S) = D$ e $FK(R \rightarrow S) = D$ e si consideri l'interrogazione:

```
SELECT A, B, F
FROM R, S
WHERE R.D = S.D AND C = 7;
```

- Si costruisca l'albero canonico (o iniziale) per l'esecuzione dell'interrogazione;
- si ottimizzi l'albero canonico per minimizzare i costi di esecuzione;
- quale algoritmo di esecuzione dell'operazione di JOIN è più conveniente utilizzare per questa interrogazione supponendo che le tuple di R siano $nR = 1000$, le tuple di S siano $nS = 50$ e la selettività dell'operazione di selezione su C sia $sC = 10$? Giustificare la risposta.

Esercizio 3

(6 punti)

Sia dato il seguente *schedule*:S: $r1(X), r1(Y), r2(X), w2(X), w1(X), w1(Y)$

- (a) Verificare la conflict-serializzabilità e/o la view-serializzabilità di *S*. In caso affermativo indicare tutti gli *schedule* seriali ad esso equivalente.
- (b) Indicare se lo *schedule S* presenta delle anomalie (perdita di aggiornamento, inserimento fantasma, etc...). In caso affermativo indicare quali operazioni elementari determinano l'anomalia.
- (c) Ipotizzando di utilizzare il protocollo di lock a 2 fasi (2PL), valutare se lo *schedule S* presenta situazioni di deadlock. In caso affermativo indicare le situazioni di attesa ed una possibile soluzione.

Esercizio 4

(8 punti)

Si consideri il problema di gestione delle telefonate implementato nell'esercizio I con la relazione TELEFONATE.

- (a) Esempificare la struttura di un documento XML che utilizzi lo stesso approccio;
- (b) scrivere il DTD che renda valido il documento XML realizzato;
- (c) Indicare la dimensione del documento XML considerando il numero di telefonate ($r = 256000$) e le dimensioni medie degli elementi fornite nei dati dell'esercizio I;
- (d) Scrivere un ulteriore documento XML (ed il relativo DTD) che riduca le dimensioni del documento stesso e che faciliti la ricerca delle telefonate effettuate da un *NUM_CHIAMANTE* specifico.