

Gestione dei dati – appello del 11/2/2011

COGNOME:
NOME:
MATRICOLA:

Autorizzo la pubblicazione del mio voto di questo esame sul sito web http://www.dis.uniroma1.it/~rosati/gd , secondo quanto prevede il decreto legislativo 196/2003 (codice in materia di protezione dei dati personali) che dichiaro di conoscere. In fede,

Esercizio 1 Si consideri una base di dati contenente la relazione **Impiegato** con attributi **cognome**, **nome**, **matricola**, **dataNascita**, **nazione**, **stipendio** e la relazione **Nazione** con attributi **nome**, **numeroAbitanti**, **capitale**, **continente**, **superficie**.

La relazione **Impiegato** contiene 50000 record, e ogni record di tale relazione ha dimensione $N/50$, dove N la dimensione di una pagina di memoria. La relazione **Nazione** contiene 10000 record, e ogni record di tale relazione ha dimensione $N/50$. Inoltre, **matricola**, **nazione** è chiave per la relazione **Impiegato** e **nome** è chiave per la relazione **Nazione**. Le suddette relazioni sono soggette ad aggiornamenti. Inoltre sulla relazione **Impiegato** è dichiarato un indice B⁺-tree clusterizzato con chiave di ricerca **nazione** e fan-out 200, mentre sulla relazione **Nazione** è dichiarato un indice B⁺-tree clusterizzato con chiave di ricerca **nome** e fan-out 100.

Si consideri la seguente query Q:

```
select I.cognome, I.nome, I.matricola, N.numeroAbitanti
from Impiegato I, Nazione N
where I.nazione = N.nome
```

Calcolare il costo (espresso in millisecondi) dell'esecuzione della precedente query, assumendo che: (i) il buffer abbia a disposizione 7 slot liberi per eseguire Q; (ii) il sistema esegua l'algoritmo Block Nested Loop; (iii) il tempo medio di accesso ad una pagina in memoria di massa sia di 5 millisecondi; (iv) i tempi di elaborazione in memoria centrale siano trascurabili.

Esercizio 2 Dato il seguente schedule $S = w_4(W) r_3(Z) r_2(X) r_1(Y) w_3(W) c_3 w_2(Z) w_1(X) c_1 w_4(Y) c_4 c_2$

1. dire se S è recuperabile, motivando la risposta;
2. dire se S è conflict-serializzabile, motivando la risposta;
3. dire se S è view-serializzabile, motivando la risposta;
4. dire se S è eseguibile da uno scheduler che segue il protocollo 2PL con lock esclusivi e condivisi. In caso positivo, completare lo schedule S con le opportune istruzioni di lock e unlock. In caso negativo, motivare la risposta.

Esercizio 3 Data la seguente DTD:

```
<!DOCTYPE root [
  <!ELEMENT root (x+) >
  <!ELEMENT x (s*, w, z, y+, (t|x)*) >
  <!ELEMENT s (v*) >
  <!ELEMENT v (#PCDATA) >
  <!ELEMENT w (#PCDATA) >
  <!ELEMENT z (#PCDATA) >
  <!ELEMENT y (v,s) >
  <!ELEMENT t (w,(t|x)*) >
  <!ATTLIST x a CDATA #REQUIRED >
  <!ATTLIST z b CDATA #REQUIRED
            c CDATA #REQUIRED >
]>
```

1. Scrivere lo schema relazionale ottenuto applicando l'algoritmo schema-driven di XML shredding alla DTD precedente.
2. Scrivere un documento XML valido rispetto alla DTD precedente e scrivere l'istanza della base di dati ottenuta traducendo tale documento nello schema relazionale ottenuto al punto 1.

Esercizio 4 Si consideri una base di dati contenente la relazione **Studente** con attributi **cognome**, **nome**, **matricola**, **indirizzo**, **comune**, **annoDiNascita**. La relazione **Studente** contiene 10^5 record, e ogni record di tale relazione ha dimensione $N/250$, dove N è la dimensione di una pagina di memoria. Inoltre **matricola** è chiave per la relazione **Studente**. La relazione **Studente** è soggetta ad aggiornamenti.

1. Si considerino query del tipo: "selezionare cognome, nome e matricola di tutti gli studenti aventi una certa matricola m e nati in un certo anno a ". Scegliere l'organizzazione fisica per la relazione **Studente** che minimizza il costo dell'esecuzione di tali query;
2. assumendo un tempo medio di accesso ad una pagina in memoria di massa di 3 millisecondi, e trascurando i tempi di elaborazione in memoria centrale, calcolare il tempo (espresso in millisecondi) necessario all'esecuzione delle query di cui al punto 1 in base all'organizzazione fisica per la relazione **Studente** scelta al punto 1;
3. si considerino ora query del tipo: "selezionare cognome, nome e matricola di tutti gli studenti nati in un certo anno a e la cui matricola m è compresa tra 100000 e 120000". Scegliere l'organizzazione fisica per la relazione **Studente** che minimizza il costo dell'esecuzione di tali query;
4. assumendo un tempo medio di accesso ad una pagina in memoria di massa di 3 millisecondi, e trascurando i tempi di elaborazione in memoria centrale, calcolare il tempo (espresso in millisecondi) necessario all'esecuzione delle query di cui al punto 3 in base all'organizzazione fisica per la relazione **Studente** scelta al punto 3.