

Gestione dei dati – appello del 16/7/2012

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 Data la seguente DTD:

```
<!DOCTYPE z [  
<!ELEMENT z ((x,y)+)>  
<!ELEMENT x (y,(q|s|t),x*)>  
<!ELEMENT y (q,v)>  
<!ELEMENT q (#PCDATA)>  
<!ELEMENT s (x*,t,s*,p)>  
<!ELEMENT t (x|y|q|s|t)+>  
<!ELEMENT p (#PCDATA)>  
<!ELEMENT v (#PCDATA)>  
<!ATTLIST s val CDATA #REQUIRED>  
<!ATTLIST v attr 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 2 Si consideri una base di dati contenente la relazione **Libro** con attributi **titolo**, **anno**, **codiceLibro**, **genere**, **numeroPagine**, **codiceEditore** e la relazione **Editore** con attributi **codiceEditore**, **nome**, **nazione**, **regione**, **indirizzo**, **annoDiFondazione**. La relazione **Libro** contiene 20.000 record, e ogni record di tale relazione ha dimensione $N/50$. La relazione **Editore** contiene 600 record, e ogni record di tale relazione ha dimensione $N/15$, dove N è la dimensione di una pagina di memoria.

Si consideri la seguente query Q :

```
SELECT L.titolo, L.genere, E.nome  
FROM Libro L, Editore E  
WHERE L.codiceEditore=E.codiceEditore AND E.nazione="Italia" AND E.regione="Lazio"
```

Si assuma che: (i) il sistema possa eseguire i join mediante l'algoritmo Index Nested Loop; (ii) il buffer abbia a disposizione 12 slot liberi per eseguire la query; (iii) la relazione **Editore** abbia 60 record con valore dell'attributo **nazione** pari a "Italia" e valore dell'attributo **regione** pari a "Lazio". Si richiede di: (i) definire l'organizzazione fisica delle relazioni **Libro** e **Editore** e il query plan per la query Q tali che risulti minimo il numero di accessi alla memoria di massa nell'esecuzione della query; (ii) calcolare il costo dell'esecuzione del query plan definito al punto precedente, espresso come numero di accessi alla memoria di massa. Si ricorda che i query plan da considerare devono necessariamente prevedere relazioni primitive come relazioni interne delle operazioni di join.

Esercizio 3 Dato lo schedule $S = r_2(W) w_4(X) r_5(V) w_1(Y) r_3(Z) w_2(T) r_2(X) w_4(T) r_5(X) w_1(W) r_3(Y) r_2(Z) w_1(T) r_3(W)$

1. dire se S è view-serializzabile, motivando la risposta;
2. dire se S è conflict-serializzabile, motivando la risposta. In caso positivo, scrivere uno schedule seriale equivalente a S ;
3. sia S' lo schedule ottenuto da S eliminando le istruzioni della transazione T_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;
4. dire se è possibile inserire nello schedule S' (definito al punto precedente) le istruzioni di commit delle transazioni T_1, T_2, T_3, T_5 in modo tale che lo schedule risultante sia stretto e, in caso positivo, scrivere tale schedule.

Esercizio 4 Si consideri una base di dati contenente la relazione **Studente** con attributi **matricola**, **cognome**, **nome**, **annoNascita**, **numeroEsamiSostenuti**, **annoImmatricolazione**. La relazione **Studente** contiene 10^6 record, e ogni record di tale relazione ha dimensione $N/200$, 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 e nome di tutti gli studenti nati nell'anno a e che hanno sostenuto esattamente n esami". 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 4 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 l'anno di nascita di tutti gli studenti aventi anno di immatricolazione i , matricola m e numero di esami sostenuti compreso tra n_1 e n_2 ". 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 4 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.