

Gestione dei dati – appello del 24/9/2012

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 a [  
  <!ELEMENT a ((b,c)+)>  
  <!ELEMENT b (b*,(e|d|h),c)>  
  <!ELEMENT c (e,g)>  
  <!ELEMENT d (h,b*,f,d*)>  
  <!ELEMENT e (#PCDATA)>  
  <!ELEMENT f (#PCDATA)>  
  <!ELEMENT g (#PCDATA)>  
  <!ELEMENT h (b|c|e|d|h)+>  
  <!ATTLIST b val CDATA #REQUIRED>  
  <!ATTLIST c 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 **Film** con attributi **titolo**, **anno**, **codiceRegista**, **genere**, **durata** e la relazione **Regista** con attributi **codiceRegista**, **cognome**, **nome**, **nazionalita**, **annoDiNascita**. La relazione **Film** contiene 20.000 record, e ogni record di tale relazione ha dimensione $N/50$. La relazione **Regista** contiene 3000 record, e ogni record di tale relazione ha dimensione $N/30$, dove N è la dimensione di una pagina di memoria.

Si consideri la seguente query Q :

```
SELECT F.titolo, R.cognome, R.nome  
FROM Film F, Regista R  
WHERE F.codiceRegista=R.codiceRegista AND R.nazionalita="italiana" AND R.annoDiNascita="1950"
```

Si assuma che: (i) il sistema possa eseguire i join mediante l'algoritmo Block Nested Loop; (ii) il buffer abbia a disposizione 22 slot liberi per eseguire la query; (iii) la relazione **Regista** abbia 90 record con valore dell'attributo **nazionalita** pari a "italiana" e valore dell'attributo **annoDiNascita** pari a "1950". Si richiede di: (i) definire l'organizzazione fisica delle relazioni **Film** e **Regista** 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 $S = w_5(C) r_3(E) w_4(F) r_2(A) w_3(D) r_1(B) r_3(C) w_5(D) r_2(C) w_4(E) r_3(B) r_1(F) r_1(E) w_4(D)$

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_3 . 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_4, 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 **Film** con attributi **titolo**, **anno**, **regista**, **genere**, **durata**. La relazione **Film** contiene 10^5 record, e ogni record di tale relazione ha dimensione $N/80$, dove N è la dimensione di una pagina di memoria. Inoltre $\langle \text{titolo}, \text{regista} \rangle$ è chiave per la relazione **Film**. La relazione **Film** è soggetta ad aggiornamenti.

1. Si considerino query del tipo: "selezionare titolo e regista di tutti i film di un certo anno a e di un certo genere g ". Scegliere l'organizzazione fisica per la relazione **Film** 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 **Film** scelta al punto 1;
3. si considerino ora query del tipo: "selezionare l'anno di tutti i film aventi titolo t , regista r e durata compresa tra d_1 e d_2 ". Scegliere l'organizzazione fisica per la relazione **Film** 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 **Film** scelta al punto 3.