

# Gestione dei dati – appello del 16/7/2010

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 **Studente** con attributi **cognome**, **nome**, **matricola**, **annoNascita**, **nomeComune**, **nomeProvincia**. La relazione **Studente** contiene 500000 record, e ogni record di tale relazione ha dimensione  $N/100$ , dove  $N$  la dimensione di una pagina di memoria. Inoltre **matricola** è chiave per la relazione **Studente**. La relazione **Studente** è soggetta ad aggiornamenti frequenti.

1. Si consideri la seguente query Q1:

```
select S.cognome, S.nome, S.matricola from Studente S
where S.nomeComune = "Roma" and S.matricola >= 10000 and S.matricola <= 50000
```

Scegliere l'organizzazione fisica per la relazione **Studente** che minimizza il costo dell'esecuzione di tale query;

2. assumendo un tempo medio di accesso ad una pagina in memoria di massa di 5 millisecondi, e trascurando i tempi di elaborazione in memoria centrale, calcolare il tempo necessario all'esecuzione della query Q1 in base all'organizzazione fisica per la relazione **Studente** scelta al punto 1;
3. si consideri ora la seguente query Q2:

```
select S.cognome, S.nome, S.matricola from Studente S
where S.nomeComune >= "B" and S.nomeComune <= "C" and S.matricola = 30000
```

Scegliere l'organizzazione fisica per la relazione **Studente** che minimizza il costo dell'esecuzione di tale query. Si assuma che per ogni nome di comune  $n$  siano presenti in media 10000 record il cui valore di **nomeComune** è uguale a  $n$ ;

4. assumendo un tempo medio di accesso ad una pagina in memoria di massa di 5 millisecondi, e trascurando i tempi di elaborazione in memoria centrale, calcolare il tempo necessario all'esecuzione della query Q2 in base all'organizzazione fisica per la relazione **Studente** scelta al punto 3.

**Esercizio 2** Si consideri una base di dati contenente la relazione **Studente** con attributi **cognome**, **nome**, **matricola**, **dataNascita**, **nomeComune**, **nomeProvincia** e la relazione **Comune** con attributi **nomeComune**, **numeroAbitanti**, **nomeProvincia**, **superficie**.

La relazione **Studente** contiene 500000 record, e ogni record di tale relazione ha dimensione  $N/100$ , dove  $N$  la dimensione di una pagina di memoria. Si assuma inoltre che la relazione **Comune** contenga 10000 record, e che ogni record di tale relazione abbia dimensione  $N/50$ . Si supponga inoltre che **matricola** sia chiave per la relazione **Comune** e che (**nomeComune**,**nomeProvincia**) sia chiave per la relazione **Comune**. Le suddette relazioni non sono soggette ad aggiornamenti.

Si consideri la seguente query Q:

```
select S.cognome, S.nome, S.matricola, C.numeroAbitanti from Studente S, Comune C
where S.nomeComune = C.nomeComune
```

(1) Con l'obiettivo di minimizzare il costo dell'esecuzione di Q, e assumendo che il buffer abbia a disposizione 12 slot liberi per eseguire Q, scegliere sia un algoritmo per l'esecuzione del join che una organizzazione fisica per le relazioni **Studente** e **Comune**; (2) sulla base di tali scelte, calcolare il costo dell'esecuzione di Q in termini di numero di accessi a pagine in memoria di massa.

**Esercizio 3** Dato il seguente schedule  $S = w_1(X) r_2(Y) w_3(T) w_3(V) c_3 w_4(V) w_1(Y) r_1(Z) c_1 r_4(T) r_2(X) w_2(Z) r_2(T) c_2 w_4(X) c_4$

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

**Esercizio 4** Data la seguente DTD:

```
<!DOCTYPE corsodilaurea [
  <!ELEMENT corsodilaurea (studente+) >
  <!ELEMENT studente (cognome, nome+, matricola, annoNascita, comune) >
  <!ELEMENT cognome (#PCDATA)>
  <!ELEMENT nome (#PCDATA)>
  <!ELEMENT matricola (#PCDATA)>
  <!ELEMENT annoNascita (#PCDATA)>
  <!ELEMENT comune (nome, provincia, persona*)>
  <!ELEMENT provincia (#PCDATA)>
  <!ELEMENT persona (studente|(cognome,nome))* >
  <!ATTLIST comune regione CDATA #REQUIRED>
  <!ATTLIST annoNascita bisestile (si|no) "no">
]>
```

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.