

# Basi di dati, primo modulo

## 19 aprile 2001

Tempo a disposizione: due ore. Libri chiusi.

**Domanda 1** (15%) Indicare (con una brevissima giustificazione) quali delle seguenti affermazioni sono vere e quali false.

1. il controllo di concorrenza ha l'obiettivo di garantire l'atomicità delle transazioni
2. il controllo di concorrenza ha l'obiettivo di garantire l'isolamento delle transazioni
3. se le transazioni fossero eseguite serialmente il controllo di concorrenza non sarebbe necessario
4. il controllo di concorrenza ha l'obiettivo di garantire la persistenza delle transazioni

**Domanda 2** (15%) Indicare (con una brevissima giustificazione) quali delle seguenti affermazioni sono vere e quali false.

1. il controllo di affidabilità ha l'obiettivo di garantire la persistenza delle transazioni
2. il controllo di affidabilità ha l'obiettivo di garantire l'isolamento delle transazioni
3. il controllo di affidabilità ha l'obiettivo di garantire l'atomicità delle transazioni
4. se le transazioni fossero eseguite serialmente il controllo di affidabilità non sarebbe necessario

**Domanda 3** (30%) Considerare i seguenti schedule:

1.  $r_1(x)w_1(x)r_2(x)w_2(x)r_0(y)w_1(y)$
2.  $w_0(z)r_1(z)r_2(z)w_2(z)w_2(y)$
3.  $r_1(x)r_2(z)w_2(z)w_1(x)r_2(x)w_2(x)$
4.  $r_2(x)w_2(x)r_1(x)w_1(x)$

Specificare, con una breve giustificazione, a quali delle seguenti classi ciascuno di essi appartiene: S (seriale), VSR (view-serializzabile), CSR (conflict-serializzabile), 2PL (generabile da uno scheduler basato sul lock a due fasi) and TS (generabile da uno scheduler che utilizzi il metodo dei timestamp; si assuma che gli identificatori delle transazioni corrispondano ai timestamp).

**Domanda 4** (40%) Considerare un file corrispondente ad una relazione sugli attributi **Matricola**, **Cognome** e **Dipartimento**, sotto le seguenti ipotesi:

- il file contiene  $N = 1.000.000$  record, che hanno lunghezza fissa  $l = 40$  byte; i campi **Matricola**, **Cognome** e **Dipartimento** occupano rispettivamente  $l_m = 15$ ,  $l_c = 20$  e  $l_d = 5$  byte;
- ci sono  $v_c = 100.000$  cognomi diversi e  $v_d = 5.000$  dipartimenti
- il file system utilizzato prevede blocchi che rendono disponibili alla struttura fisica  $B = 500$  byte; i puntatori (sia i puntatori a blocchi sia quelli a singoli record) occupano  $l_p = 5$  byte.

Le operazioni principali su tale file sono le seguenti:

1. ricerca approssimata (sottostringa iniziale) sul **Cognome** con frequenza  $q_1$ ; mediamente si ottengono  $r_1 = 100$  risposte;
2. ricerca puntuale sul **Dipartimento** (che però, si ricordi, non è chiave), con frequenza  $q_2$

Si supponga che il sistema a disposizione permetta organizzazioni disordinate e ordinate (con indice primario) e indici secondari (B+-tree). Di conseguenza, il file dovrà avere una struttura primaria basata su un ordinamento su **Cognome** oppure su **Dipartimento**, con un indice primario su tale campo e un indice secondario sull'altro.

1. Calcolare il numero di accessi a blocchi nell'unità di tempo in funzione di  $q_1$  e  $q_2$  a seconda della scelta della struttura primaria;
2. Indicare per quali valori del rapporto fra  $q_1$  e  $q_2$  risulti conveniente ciascuna delle due scelte.