

Basi di dati, primo modulo — 5 febbraio 2004

Tempo a disposizione: due ore. Libri chiusi.

Domanda 1 (20%) Considerare i seguenti schedule:

1. $r_1(x)r_2(z)w_2(z)w_1(x)r_2(x)w_2(x)$
2. $r_2(x)w_2(x)r_3(x)w_3(x)r_1(y)w_2(y)$
3. $r_1(y)r_4(z)r_4(y)w_4(y)w_4(z)r_1(z)r_2(z)r_3(z)w_3(z)$
4. $r_4(z)r_4(y)w_4(y)w_4(z)r_2(x)w_2(x)r_1(x)w_1(x)$

Specificare, con una breve giustificazione, a quali delle seguenti classi ciascuno di essi appartiene: S (seriale), VSR, CSR, 2PL e TS; si assuma che l'ordinamento degli identificatori delle transazioni corrisponda a quello dei timestamp).

Domanda 2 (20%) Si supponga di dover eseguire una interrogazione che calcola statistiche su una base di dati (ad esempio: trovare per ogni corso la media dei voti assegnati). Indicare (con un breve commento, non più di tre righe) quale livello di isolamento (READ UNCOMMITTED, READ COMMITTED, REPEATABLE READ o SERIALIZABLE) si potrebbe scegliere in ciascuno dei seguenti casi

1. per tutti i corsi sono già presenti molti esami e l'interrogazione è eseguita mentre vengono inseriti alcuni esami, comunque pochi per corso (rispetto a quelli già presenti); sono accettabili risultati "approssimati"
2. per tutti i corsi sono già presenti molti esami e l'interrogazione è eseguita mentre vengono inseriti alcuni esami, comunque pochi per corso (rispetto a quelli già presenti); *non* sono accettabili risultati "approssimati"
3. per tutti i corsi sono già presenti molti esami e l'interrogazione è eseguita mentre vengono corretti (cioè modificati) i voti di alcuni esami, comunque pochi per corso (rispetto a quelli già presenti) e senza inserirne di nuovi; *non* sono accettabili risultati "approssimati"
4. per tutti i corsi sono già presenti molti esami e l'interrogazione è eseguita mentre vengono corretti (cioè modificati) i voti di alcuni esami, comunque pochi per corso (rispetto a quelli già presenti) e senza inserirne di nuovi; sono accettabili risultati "approssimati"

Domanda 3 (20%)

Spiegare perché il locking a due fasi stretto, con un commit globale, garantisce la serializzabilità in contesto distribuito mentre il il locking a due fasi locale non la garantisce.

Domanda 4 (25%)

Alcuni DBMS permettono una tecnica di memorizzazione chiamata "co-clustering" o "clustering eterogeneo," in cui un file contiene record di due o più relazioni e tali record sono allocati (ad esempio ordinati) secondo i valori di opportuni campi dell'una e dell'altra relazione. Ad esempio, date due relazioni

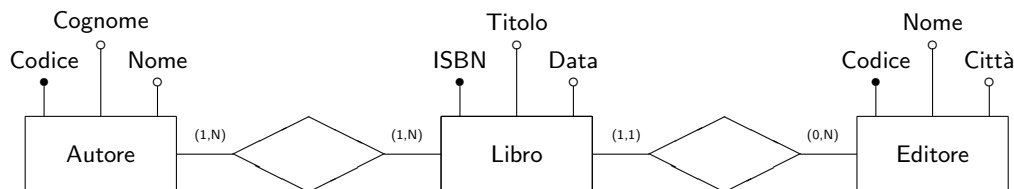
- *Ordini*(*CodiceOrdine*, *Cliente*, *Data*, *Totale*)
- *LineeOrdine*(*CodiceOrdine*, *Linea*, *Prodotto*, *Importo*)

questa tecnica (con riferimento agli attributi *CodiceOrdine* delle due relazioni) permetterebbe una memorizzazione contigua di ciascun ordine con le rispettive "linee d'ordine," cioè dei prodotti ordinati (ciascun ordine fa riferimento a più prodotti, ognuno su una "linea").

Con riferimento all'esempio, indicare quali delle seguenti operazioni possono trarre vantaggio dall'uso di questa opportunità e quali ne possono essere penalizzate (spiegare la risposta possibilmente anche in termini quantitativi, attraverso l'uso di esempi):

1. stampa dei dettagli (cioè delle linee d'ordine) di tutti gli ordini (ordinati per codice)
2. stampa dei dettagli di un ordine
3. stampa delle informazioni sintetiche (codice, cliente, data, totale) di tutti gli ordini

Domanda 5 (15%) Considerare i dati rappresentati dal seguente schema E-R



Indicare (con una breve motivazione) in quale (o quali) dei seguenti contesti

1. in una base di dati orientata agli oggetti
2. in una base di dati relazionale a oggetti
3. in un sito Web

può avere senso prevedere una struttura complessa (classe nel caso 1, relazione nidificata nel caso 2, schema di pagina nel caso 3) Autore, ogni istanza della quale contenga informazioni aggregate, e precisamente i dati personali di un autore e la lista dei libri che ha scritto, con tutti i dati, incluso il relativo editore (con tutti i dati).