

Tecnologia delle basi di dati

Basi di dati, primo modulo

22 luglio 2004

Tempo a disposizione: un'ora e quarantacinque minuti. Libri chiusi.

Domanda 1

Per ragionare su alcuni concetti e tecniche di controllo di concorrenza, è utile includere negli schedule anche le operazioni di commit: ad esempio c_i potrebbe indicare il commit della transazione i . Un esempio di schedule potrebbe quindi essere: $w_1(x)r_2(x)c_2w_3(y)c_3w_1(y)c_1$. In tale contesto, una classe di schedule interessante è la seguente:

Uno schedule s è *commit order-preserving conflict serializable (COCSR)* quando, per ogni coppia di transazioni t_i e t_j che vanno in commit in s , se due operazioni $o_i(x)$ di t_i e $o_j(x)$ di t_j sono in conflitto e $o_i(x)$ precede $o_j(x)$ in s , allora c_i precede c_j in s .

Informalmente, per tutte le transazioni che vanno in commit, l'ordine delle operazioni in conflitto è coerente con l'ordine dei commit.

Dimostrare che la classe di schedule COCSR è propriamente contenuta nella classe CSR (facendo riferimento, per l'ipotesi di commit-proiezione, alle sole transazioni che vanno in commit); mostrare cioè che COCSR è contenuta in CSR e che non è vero il viceversa.

Domanda 2

Si consideri un B-tree con nodi intermedi che contengono due chiavi e tre puntatori e foglie con due chiavi. Mostrare un possibile contenuto della struttura a seguito di inserimenti delle chiavi nel seguente ordine (a partire dall'albero vuoto): 19, 16, 18, 5, 1, 8, 15, 20, 17.

Domanda 3

Ripetere l'esercizio precedente con riferimento ad un B+-tree.

Domanda 4

Considerare la seguente interrogazione in SQL:

```
SELECT DISTINCT R1.B, R3.L
FROM R1, R2, R3
WHERE R1.C = R2.E AND R2.D=R3.H AND R1.B=R3.G AND R1.C=10
```

con riferimento ad una base di dati su cui sono note le seguenti informazioni:

- la relazione $R_1(\underline{A}BC)$ ha 100.000 ennuple, una struttura heap e un indice secondario sulla chiave A ; l'attributo C ha 100 valori diversi, distribuiti in modo uniforme; l'attributo B ha un vincolo di riferimento verso la chiave G di R_3
- la relazione $R_2(\underline{D}EF)$ ha 100.000 ennuple, una struttura heap e un indice secondario sulla chiave D ;
- la relazione $R_3(\underline{G}HL)$ ha 200.000 ennuple, una struttura hash sulla chiave G ; l'attributo H ha un vincolo di riferimento verso la chiave D di R_2

In tale contesto:

1. Mostrare un possibile piano di esecuzione (in termini di operatori dell'algebra relazionale e loro realizzazioni; giustificare brevemente le scelte più significative.
2. Stimare, nella misura in cui ciò sia possibile, le dimensioni dei risultati intermedi e finale.
3. Commentare sulle esigenze implementative imposte, in questo caso, dalla parola chiave DISTINCT.

Domanda 5

Si supponga di avere un DBMS con due livelli di isolamento, READ COMMITTED e SERIALIZABLE. Si supponga di avere una relazione:

Prenotazioni(Volo, Data, Posto, Libero)

e una transazione che (i) cerca i posti liberi su un volo:

```
select Posto from Prenotazioni where Libero = 'True' and Volo = ?1 and Data = ?2
```

(ii) li mostra all'utente chiedendo di scegliere (iii) acquisisce la scelta dell'utente, fissando il posto scelto come occupato (e stampando una ricevuta):

```
update Prenotazioni set Libero = 'False' where Volo = ?1 and Data = ?2 and Posto = ?3.
```

Spiegare di quali problemi soffre la transazione, nei due casi della scelta sul livello di isolamento. Individuare una possibile soluzione pragmatica, indicando anche le possibili conseguenze.