

Tecnologie delle Basi di Dati M

Appello del 7/6/2013

Esercizio 1 (3 punti)

Si considerino le seguenti transazioni:

```
T1: read(T1, PA), write(T1, PA)
T2: read(T2, PA), write(T2, PA)
```

ed il seguente schedule:

```
read(T1, PA), read(T2, PA), write(T2, PA), abort(T2),
write(T1, PA), commit(T1)
```

si risponda alle seguenti domande, giustificando le risposte:

1. Lo schedule è serializzabile?
2. Lo schedule potrebbe essere prodotto da un protocollo Strict-2PL?
3. Quale schedule verrebbe generato da un protocollo Conservative-2PL che ricevesse i comandi nell'ordine indicato in precedenza (si supponga che ogni comando di lettura/scrittura sia preceduto dalla corrispondente richiesta di lock)?
4. A quale schedule seriale è equivalente lo schedule generato al passo precedente?

Esercizio 2 (4 punti)

Data la relazione con schema:

```
Personale(matricola, nome, data, luogo, stipendio, responsabile)
```

si ottimizzi l'esecuzione della seguente interrogazione SQL:

```
SELECT P.matricola, R.matricola
FROM Personale P, Personale R
WHERE P.responsabile=R.matricola
      AND R.stipendio > 9000
      AND P.luogo IN ('Milano', 'Bologna')
```

tenendo conto che dai cataloghi della base di dati risulta:

- Numero di tuple `Personale` = 80K
- Numero di pagine `Personale` = 5K
- Indice unclustered (TID ordinate) su `luogo`: numero foglie = 200, numero chiavi = 100
- Indice clustered su `stipendio`: numero foglie = 1K, valore minimo = 1000, valore massimo = 11000.
- Indice unclustered (TID ordinate) su `responsabile`: numero foglie = 8K, numero di responsabili = 200.
- Indice unclustered su `matricola`: numero foglie = 3000

Si disegni infine l'albero corrispondente al piano di accesso di costo minimo e stimi il numero di risultati dell'interrogazione.

Suggerimento: per la formula di Cardenas si utilizzino i seguenti valori, validi per $P = 2000$:

R	$\Phi(R, 5000)$
100	97.56
200	190.37
300	278.65
400	362.62
500	442.50
600	518.47
700	590.75
800	659.49
900	724.89
1000	787.09

R	$\Phi(R, 5000)$
1100	846.26
1200	902.54
1300	956.08
1400	1007.00
1500	1055.44
1600	1101.52
1700	1145.35
1800	1187.04
1900	1226.70
2000	1264.43

R	$\Phi(R, 5000)$
2100	1300.31
2200	1334.44
2300	1366.91
2400	1397.79
2500	1427.17
2600	1455.11
2700	1481.69
2800	1506.98
2900	1531.03
3000	1553.91

Esercizio 3 (5 punti)

Si illustrino le varie possibilità di gestione degli overflow per le strutture di static hashing.

Esercizio 4 (3 punti)

Si illustri quale sia l'effetto di incrementare/decrementare la dimensione della pagina dati su una struttura ad indice multi-dimensionale paginata quale R-tree, discutendone in particolare l'impatto sui costi di inserimento e ricerca.

Soluzione Esercizio 1

1. Sì, lo schedule è equivalente allo schedule seriale (T2, T1).
2. No, in quanto T2 non è in grado di ottenere un lock esclusivo su PA fino al termine di T1.
3. `read(T1, PA), write(T1, PA), commit(T1), read(T2, PA), write(T2, PA), abort(T2)`
4. Lo schedule precedente è equivalente allo schedule seriale (T1, T2).

Soluzione Esercizio 2

Selettività dei predicati:

Predicato R.stipendio > 6000 = $(11000-9000)/(11000-1000) = 0.2$

Predicato P.luogo IN ('Milano', 'Bologna') = $1/100 = 0.01$ per ogni valore di luogo

Accesso a P:

Costo scan sequenziale = **5000**

Costo indice su luogo: $2 \times (NL \times 0.01 + \Phi(NT \times 0.01, NP)) = 2 \times (200 \times 0.01 + \Phi(80K \times 0.01, 5K)) = 2 \times (2 + \Phi(800, 5000)) = 2 \times (2 + 660) = 1324$

Costo indice su responsabile: $NL \times 0.005 + \Phi(NT \times 0.005, NP) = 8K \times 0.005 + \Phi(80K \times 0.005, 5K) = 40 + \Phi(400, 5000) = 40 + 363 = 403$

Numero tuple residue = $NT \times 2 \times 0.01 = 1600$

Accesso a R:

Costo scan sequenziale = **5000**

Costo indice su stipendio: $NL \times 0.2 + NP \times 0.2 = 1000 \times 0.2 + 5000 \times 0.2 = 200 + 1000 = 1200$

Costo indice su matricola: $1 + 1 = 2$

Numero tuple residue = $NT \times 0.2 = 16000$

Costi di join:

P esterna: costo = costo indice su luogo + $800 \times$ costo indice matricola = $1324 + 1600 \times 2 = 4524$

R esterna: costo = costo indice su stipendio + $1000 \times$ costo indice su responsabile = $1200 + 16000 \times 403 = 6449200$

Il numero di risultati dell'interrogazione è $80K \times 0.2 \times 2 \times 0.01 = 320$