

Tecnologie delle Basi di Dati M

Appello del 19/9/2013

Esercizio 1 (3 punti)

Si consideri un sistema che gestisca le transazioni secondo la procedura *Undo-Redo*. Si supponga che il log del sistema abbia il seguente contenuto al verificarsi di un *system crash*:

1. BEGIN(T1)
2. UPDATE(T1, PA, VA1, VA2)
3. BEGIN(T2)
4. UPDATE(T2, PA, VA2, VA3)
5. COMMIT(T1)
6. UPDATE(T2, PB, VB1, VB2)
7. END(T1)
8. CHECKPOINT
9. UPDATE(T2, PC, VC1, VC2)
10. BEGIN(T3)
11. UPDATE(T3, PD, VD1, VD2)
12. COMMIT(T2)

si risponda alle seguenti domande, giustificando le risposte:

1. Il log può essere generato da un sistema che utilizzi il protocollo 2PL?
 - a. Se sì, mostrare una possibile sequenza di richieste da parte delle transazioni.
 - b. Se no, si indichi quali record andrebbero spostati (e in che posizione) per ottenere un log “corretto”.
2. Quali pagine verrebbero incluse nella tabella delle pagine sporche dall’algoritmo ARIES? Quale valore avrebbe ciascuna di tali pagine all’inizio della procedura di Redo?

Esercizio 2 (4 punti)

Date le relazioni con schema:

Esami(matricola, codice, data, voto)

Studenti(matricola, nome, cognome, data, luogo, cdl)

si ottimizzi l’esecuzione della seguente interrogazione SQL:

```
SELECT S.nome, S.cognome, E.voto
FROM Studenti S, Esami E
WHERE S.matricola=E.matricola
      AND E.voto in ('30','30L')
      AND S.cognome LIKE P%
```

tenendo conto che dai cataloghi della base di dati risulta:

- Numero di tuple Esami = 80K
- Numero di pagine Esami = 2K
- Numero di tuple Studenti = 20K
- Numero di pagine Studenti = 2K
- Indice unclustered (TID ordinate) su voto: numero foglie = 240, numero voti = 16
- Indice clustered su E.matricola: numero foglie = 200
- Indice unclustered su S.matricola: numero foglie = 100

- Indice clustered su cognome: numero foglie = 100 i cognomi iniziano con una lettera dell’alfabeto italiano (ma nessun cognome inizia per H)

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	$\Phi(R, 2000)$
200	190.37
400	362.62
600	518.47
800	659.49
1000	787.09
1200	902.54
1400	1007.00
1600	1101.52
1800	1187.04
2000	1264.43

R	$\Phi(R, 2000)$
2200	1334.44
2400	1397.79
2600	1455.11
2800	1506.98
3000	1553.91
3200	1596.37
3400	1634.79
3600	1669.55
3800	1701.00
4000	1729.46

R	$\Phi(R, 2000)$
4200	1755.22
4400	1778.52
4600	1799.60
4800	1818.67
5000	1835.93
5200	1851.55
5400	1865.68
5600	1878.47
5800	1890.03
6000	1900.50

Esercizio 3 (5 punti)

Si illustrino i principali problemi dovuti alla concorrenza di transazioni e si mostri come il protocollo Strict 2PL ne impedisca il verificarsi.

Esercizio 4 (3 punti)

Descrivere la realizzazione della scrittura di una pagina, il commit e l’abort in un sistema che adotti una politica *noUndo-Redo*.

Soluzione Esercizio 1

1. Sì: l'ordine delle richieste potrebbe, ad esempio, essere:
`read(T1, PA), write(T1, PA), read(T2, PA), write(T2, PA),`
`read(T2, PB), write(T2, PB), commit(T1), read(T2, PC), write(T2,`
`PC), read(T3, PD), write(T3, PD), commit(T2).`
2. La tabella delle pagine sporche comprenderebbe PC (con valore VC1) e PD (con valore VD1).

Soluzione Esercizio 2

Selettività dei predicati:

Predicato E.voto in ('30', '30L') = $1/16 = 0.0625$ per ogni valore di voto

Predicato S.cognome LIKE P% = $1/20 = 0.05$

Predicato di join = $1/20K$ (chiave esterna)

Accesso a S:

Costo scan sequenziale = **2000**

Costo indice su cognome: $NL / 20 + NP / 20 = 100 \times 0.05 + 2000 \times 0.05 = 5 + 100 = 105$

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

Numero tuple residue = $NT / 20 = 1000$

Accesso a E:

Costo scan sequenziale = **1000**

Costo indice su voto: $2 \times (NL \times 0.0625 + \Phi(NT \times 0.0625, NP)) = 2 \times (240/16 + \Phi(80K/16, 1K)) =$

$2 \times (15 + 1836) = 3702$

Costo indice su matricola: $NL/20K + NP/20K = 1 + 1 = 2$

Numero tuple residue = $NT / 8 = 10000$

Costi di join:

S esterna: costo = costo indice su cognome + $1000 \times$ costo indice matricola = $105 + 1000 \times 2 = 2105$

E esterna: costo = costo scan sequenziale + $10000 \times$ costo indice su matricola = $2000 + 10000 \times 2 = 22000$

Il numero di risultati dell'interrogazione è $80K \times 2 \times 0.0625 \times 0.05 = 500$