

COMPITO DI SISTEMI INFORMATIVI/BASI DI DATI II

12 luglio 2007 (Tot. 16) Tempo: 2h

Esercizio 1 (punti 3)

Si consideri il seguente log:

- | | |
|----------------------|----------------------|
| 1. I(T1,O1,A1) | 15. CK(T3,T4,T5) |
| 2. U(T2,O1,B2,A2) | 16. C(T3) |
| 3. B(T3) | 17. U(T4,O4,B17,A17) |
| 4. U(T3,O1,B4,A4) | 18. U(T5,O5,B18,A18) |
| 5. I(T2,O2,A5) | 19. B(T6) |
| 6. C(T2) | 20. D(T6,O5,B20) |
| 7. C(T1) | 21. B(T7) |
| 8. B(T4) | 22. I(T7,O6,A22) |
| 9. D(T4,O1) | 23. U(T5,O6,B23,A23) |
| 10. I(T4,O3,A10) | 24. I(T6,O7,A24) |
| 11. U(T3,O3,B11,A11) | 25. D(T7,O4,B25) |
| 12. U(T4,O2,B12,A12) | 26. C(T6) |
| 13. B(T5) | 27. U(T7,O3,B27,A27) |
| 14. I(T5,O4,A14) | |

si mostrino le operazioni di recovery da effettuare supponendo che il guasto avvenga subito dopo l'ultimo record del log.

Esercizio 2 (punti 4)

Dato il seguente schedule:

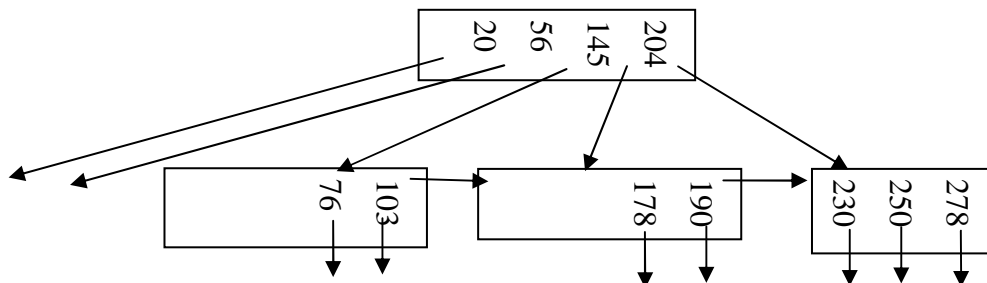
r1(x) r2(x) w1(x) w1(w) w3(y) r2(x) w2(x) r3(w) w2(z) r2(w) r3(x)

si indichi se è view-equivalente allo schedule seriale

r1(x) w1(x) w1(w) r2(x) r2(x) w2(x) w2(z) r2(w) w3(y) r3(w) r3(x)

Esercizio 3 (punti 3)

Sia dato il seguente B+ tree di ordine 4.



Si mostri come si modifica l'albero nel caso di cancellazione della chiave 178

Esercizio 4 (punti 6)

Siano date le relazioni

CD(CodiceCD, CodiceMus, Titolo, Anno, Durata)

Musicista(CodiceMus, CasaDiscografica, Nazionalità, Età)

la query

```
SELECT C.*, M.*
```

```
FROM CD AS C, Musicista AS M
```

```
WHERE C.CodiceMus=M.CodiceMus AND C.Durata>30 AND C.Durata<70
```

e i parametri:

buffer di memoria centrale disponibili: $M=20.000$

dimensione del buffer: $B=4.000$ bytes

numero di tuple: $T(\text{CD})=81967$, $T(\text{Musicista})=40000$,

dimensione delle tuple: $S(\text{CD})=2000$ bytes, $S(\text{Musicista})=3000$,

$S(\text{CodMusicista})=200$

numero di valori: $V(\text{CD}, \text{CodiceMus})=40000$,

$\text{Min}(\text{CD}, \text{Durata})=25$ $\text{Max}(\text{CD}, \text{Durata})=80$

Si calcoli il costo minimo di ciascuna delle seguenti sequenze:

$(\sigma_{\text{Durata}>30 \text{ AND } \text{Durata}<70} \text{CD}) \triangleright \triangleleft \text{Musicista}$

$\sigma_{\text{Durata}>30 \text{ AND } \text{Durata}<70} (\text{CD} \triangleright \triangleleft \text{Musicista})$

Si supponga che i record siano impaccati nei buffer e che tutto lo spazio nei buffer sia occupato dai record. Si supponga che CD e Musicista siano ordinate sulla base del campo CodiceMus. Se ne tenga conto nel calcolo del costo del join.

Si supponga di avere un indice secondario su CD.Durata. Per il calcolo del costo e della cardinalità della selezione su Durata si utilizzi la stima dei valori nel range. Si noti che CodiceMus è chiave primaria di Musicista.

SOLUZIONE

Esercizio 1

1. I(T1,O1,A1)
2. U(T2,O1,B2,A2)
3. B(T3)
4. U(T3,O1,B4,A4)
5. I(T2,O2,A5)
6. C(T2)
7. C(T1)
8. B(T4)
9. D(T4,O1,B9)
10. I(T4,O3,A10)
11. U(T3,O3,B11,A11)
12. U(T4,O2,B12,A12)
13. B(T5)
14. I(T5,O4,A14)
15. CK(T3,T4,T5)
16. C(T3)
17. U(T4,O4,B17,A17)
18. U(T5,O5,B18,A18)
19. B(T6)
20. D(T6,O5,B20)
21. B(T7)
22. I(T7,O6,A22)
23. U(T5,O6,B23,A23)
24. I(T6,O7,A24)
25. D(T7,O4,B25)
26. C(T6)
27. U(T7,O3,B27,A27)

15 UNDO={T3,T4,T5} REDO={}
16 UNDO={T4,T5} REDO={T3}
19 UNDO={T4,T5,T6} REDO={T3}
21 UNDO={T4,T5,T6,T7} REDO={T3}
26 UNDO={ T4,T5,T7} REDO={T3,T6}

UNDO

27 O3=B27
25 I(O4,B25)
23 O6=B23
22 D(O6)
18 O5=B18
17 O4=B17
14 D(O4)
12 O2=B12
10 D(O3)
9 I(O1,B9)

REDO

4 O1=A4
11 O3=A11
20 D(O5)
24 I(O7,A24)

Esercizio 2

r1(x) r2(x) w1(x) w1(w) w3(y) r2(x) w2(x) r3(w) w2(z) r2(w) r3(x)

"legge-da"={(t2,t1),(t3,t1),(t3,t2)}

scritture finali={(t2,x),(t3,y),(t2,z),(t1,w)}

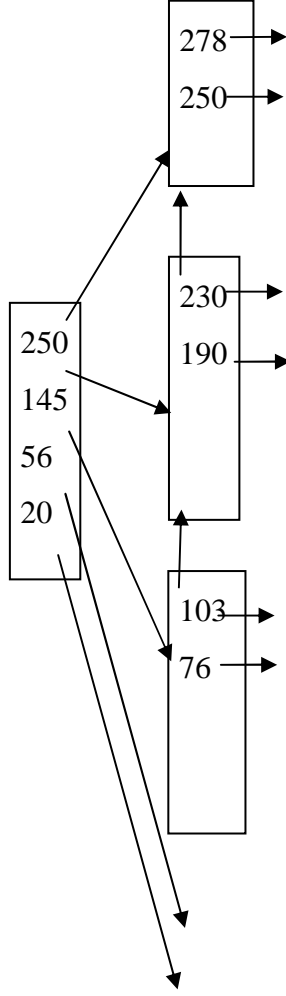
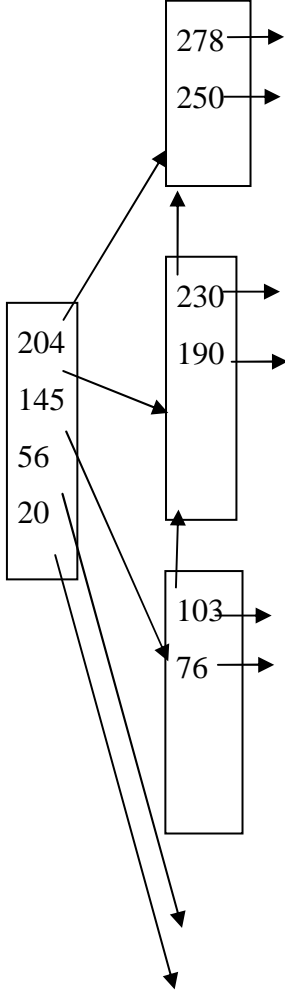
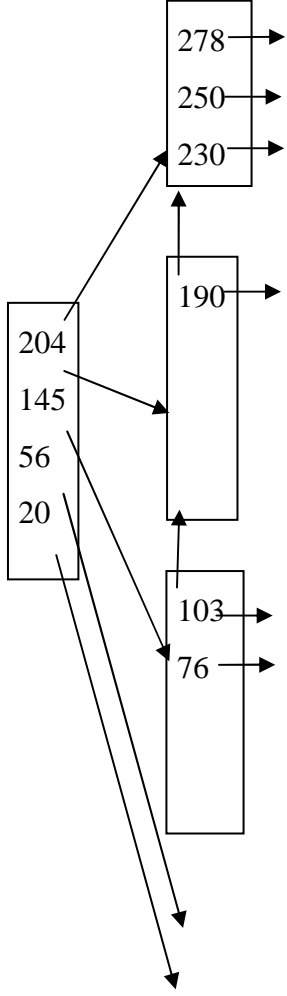
r1(x) w1(x) w1(w) r2(x) r2(x) w2(x) w2(z) r2(w) w3(y) r3(w) r3(x)

"legge-da"={(t2,t1),(t3,t1),(t3,t2)}

scritture finali={(t2,x),(t3,y),(t2,z),(t1,w)}

I due schedule sono view-equivalenti

Esercizio 3



Esercizio 4

buffer di memoria centrale disponibili: $M=20.000$
dimensione del buffer: $B=4.000$ bytes
numero di tuple: $T(\text{CD})=81967$, $T(\text{Musicista})=40000$,
dimensione delle tuple: $S(\text{CD})=2000$ bytes, $S(\text{Musicista})=3000$,
 $S(\text{CodMusicista})=200$
numero di valori: $V(\text{CD}, \text{CodiceMus})=40000$,
 $\text{Min}(\text{CD}, \text{Durata})=25$ $\text{Max}(\text{CD}, \text{Durata})=80$

Calcoliamo innanzitutto il numero di blocchi occupati da ciascuna relazione:

$$B(\text{CD})=T(\text{CD}) * S(\text{CD}) / B = 81967 * 2000 / 4000 = 40984$$

$$B(\text{Musicista}) = 40000 * 3000 / 4000 = 30000$$

Prima sequenza: $(\sigma_{\text{Durata} > 30 \text{ AND } \text{Durata} < 70} \text{CD}) \triangleright \triangleleft \text{Musicista}$

Sia $X = \sigma_{\text{Durata} > 30 \text{ AND } \text{Durata} < 70} \text{CD}$

Utilizzando l'indice secondario su Durata

$$f = (70 - 30 + 1) / (80 - 25 + 1) = 0.732$$

$$\text{Costo}(X) = 3 + f * T(\text{CD}) = 3 + 0.732 * 81967 = 60003$$

$$T(X) = f * T(\text{CD}) = 0.732 * 81967 = 60000$$

$$B(X) = 60000 * 2000 / 4000 = 30000$$

Né X né Musicista stanno in memoria centrale, quindi si prova con il join nested-loop basato sui blocchi e con l'hash join ibrido.

$$\text{CostoJNL}(X \triangleright \triangleleft \text{Musicista}) = B(\text{Musicista}) + B(X) * B(\text{Musicista}) / M = 30000 + 30000 * 30000 / 20000 = 75000$$

$$\text{CostoHJI}(X \triangleright \triangleleft \text{Musicista}) = (3 - 2M / B(\text{Musicista})) * (B(X) + B(\text{Musicista})) = (3 - 2 * 20000 / 30000) * (30000 + 30000) = 100000$$

$$\text{Costo totale} = 60003 + 75000 = 135003$$

Seconda sequenza: $\sigma_{\text{Durata} > 30 \text{ AND } \text{Durata} < 70} (\text{CD} \triangleright \triangleleft \text{Musicista})$

Sia $Z = \text{CD} \triangleright \triangleleft \text{Musicista}$

CD e Musicista sono ordinati quindi posso usare il sort join senza il passo preliminare di ordinamento

$$\text{CostoSJ}(Z) = B(\text{CD}) + B(\text{Musicista}) = 40984 + 30000 = 70984$$

$$T(Z) = T(\text{CD}) * T(\text{Musicista}) / \max\{V(\text{CD}, \text{CodiceMus}), V(\text{Musicista}, \text{CodiceMus})\} = 81967 * 40000 / \max\{40000, 40000\} = 81967$$

$$S(Z) = S(\text{CD}) + S(\text{Musicista}) - S(\text{CodiceMus}) = 2000 + 3000 - 200 = 4800$$

$$B(Z) = 81967 * 4800 / 4000 = 98361$$

Dato che non ci sono indici su Z :

$$\text{Costo}(\sigma_{\text{Durata} > 30 \text{ AND } \text{Durata} < 70} (\text{CD} \triangleright \triangleleft \text{Musicista})) = B(Z) = 98361$$

$$\text{Costo totale} = 70984 + 98361 = 169345$$