

COMPITO DI SISTEMI INFORMATIVI

21 giugno 2012 (Tot. 16) Tempo: 2h

Esercizio 1 (punti 3)

Si consideri il seguente log:

- | | |
|----------------------|----------------------|
| 1. I(T1,O1,A1) | 13. D(T2,O6,B13) |
| 2. B(T2) | 14. I(T4,O7,A14) |
| 3. B(T3) | 15. C(T2) |
| 4. U(T2,O1,B4,A4) | 16. B(T5) |
| 5. D(T3,O2,B5) | 17. U(T5,O8,B17,A17) |
| 6. U(T1,O3,B6,A6) | 18. I(T5,O9,A18) |
| 7. C(T1) | 19. B(T6) |
| 8. U(T2,O4,B8,A8) | 20. U(T6,O9,B20,A20) |
| 9. B(T4) | 21. B(T7) |
| 10. CK(T2,T3,T4) | 22. U(T7,O9,B22,A22) |
| 11. I(T4,O5,A11) | 23. C(T5) |
| 12. U(T3,O5,B12,A12) | |

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

Esercizio 2 (punti 4)

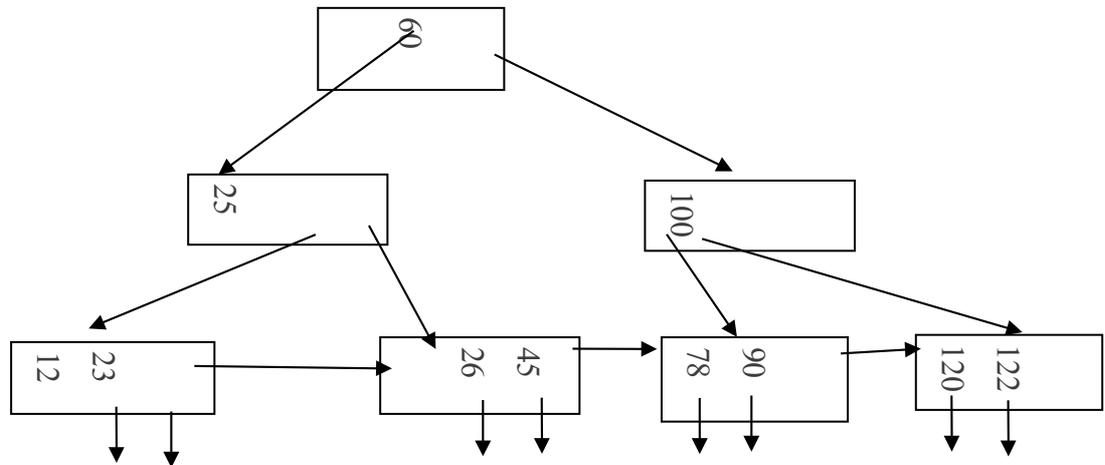
Dato il seguente schedule:

$r1(x), r2(y), r1(y), r3(x), w3(x), w1(z), w3(y), w2(y), r2(u)$

si indichi se è conflict serializzabile.

Esercizio 3 (punti 3)

Sia dato il seguente B+ tree di ordine 3.



Si mostri l'albero risultante dalla cancellazione della chiave 90

Esercizio 4 (punti 6)

Sia data la seguente relazione

Agenzia(CodAg, Nome, Via, Città, CAP, Telefono, Fax)

Cliente(CodiceCli, Nome, Cognome, Via, Città, CAP, Telefono, CodAg, Età)

la query

```
SELECT A.*, C.*  
FROM Agenzia AS A, Cliente AS C  
WHERE A.Città = C.Città AND C.Età >=30
```

e i parametri:

buffer di memoria centrale disponibili: $M=10000$

dimensione del buffer: $B=4000$ bytes

numero di tuple: $T(\text{Cliente})=200000$, $T(\text{Agenzia})=50000$,

dimensione delle tuple: $S(\text{Cliente})=2000$, $S(\text{Agenzia})=2000$, $S(\text{Città})=200$

numero di valori $V(\text{Cliente}, \text{Città})=20000$, $V(\text{Agenzia}, \text{Città})=1000$

$\text{Min}(\text{Cliente}, \text{Età})=20$ $\text{Max}(\text{Cliente}, \text{Età})=60$

Si calcoli il costo minimo di ciascuna delle seguenti sequenze:

$(\sigma_{\text{Età} \geq 30} \text{ Cliente}) \triangleright \triangleleft \text{ Agenzia}$

$\sigma_{\text{Età} \geq 30} (\text{ Cliente } \triangleright \triangleleft \text{ Agenzia})$

Si indichi quale delle sequenze ha costo inferiore.

Si supponga che i record siano impaccati nei buffer e che tutto lo spazio nei buffer sia occupato dai record.

Si supponga di avere un indice secondario su Cliente.Età.

SOLUZIONE

Esercizio 1

1. I(T1,O1,A1)
2. B(T2)
3. B(T3)
4. U(T2,O1,B4,A4)
5. D(T3,O2,B5)
6. U(T1,O3,B6,A6)
7. C(T1)
8. U(T2,O4,B8,A8)
9. B(T4)
10. CK(T2,T3,T4)
11. I(T4,O5,A11)
12. U(T3,O5,B12,A12)
13. D(T2,O6,B13)
14. I(T4,O7,A14)
15. C(T2)
16. B(T5)
17. U(T5,O8,B17,A17)
18. I(T5,O9,A18)
19. B(T6)
20. U(T6,O9,B20,A20)
21. B(T7)
22. U(T7,O9,B22,A22)
23. C(T5)

- 10 UNDO={T2,T3,T4} REDO={}
15 UNDO={T3,T4} REDO={T2}
16 UNDO={T3,T4,T5} REDO={T2}
19 UNDO={ T3,T4,T5,T6} REDO={T2}
21 UNDO={T3, T4,T5,T6,T7} REDO={ T2}
23 UNDO={ T3, T4,T6,T7} REDO={ T2, T5}

UNDO

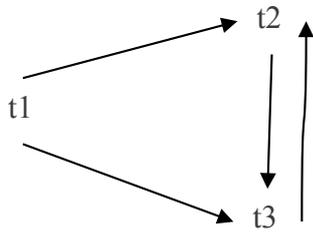
- 22 O9=B22
20 O9=B20
14 D(O7)
12 O5=B12
11 D(O5)
5 I(O2,B5)

REDO

- 4 O1=A4
8 O4=A8
13 D(O6)
17 O8=A17
18 I(O9,A18)

Esercizio 2

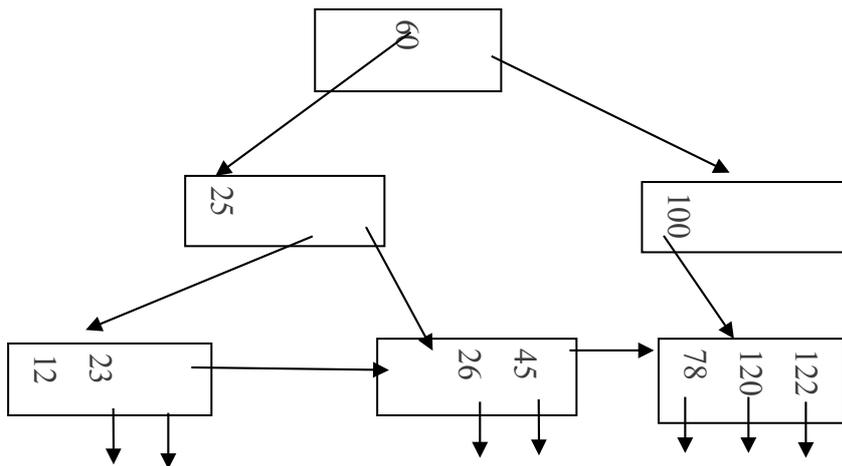
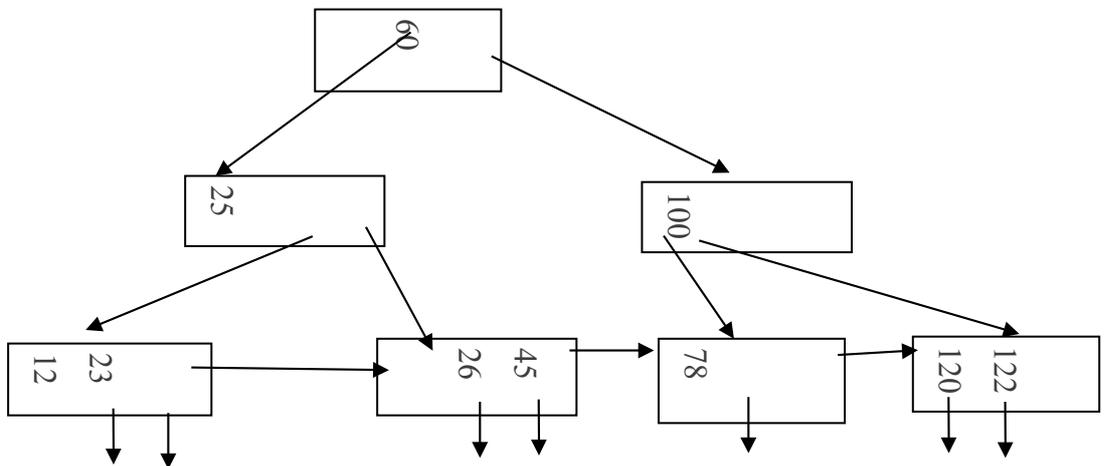
$r1(x), r2(y), r1(y), r3(x), w3(x), w1(z), w3(y), w2(y), r2(u)$

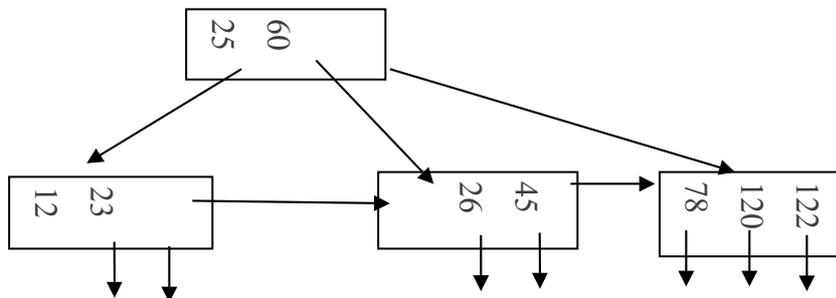
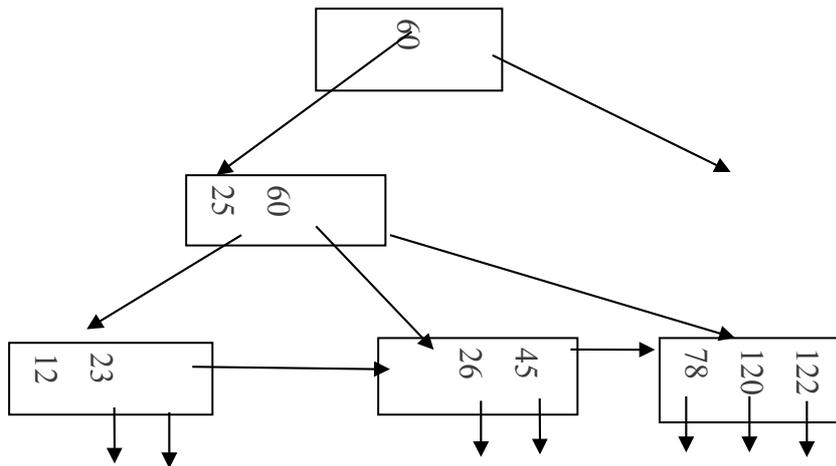


Lo schedule non è conflict serializzabile.

Esercizio 3

Cancellazione della chiave 90





Esercizio 4

buffer di memoria centrale disponibili: $M=10000$
 dimensione del buffer: $B=4000$ bytes
 numero di tuple: $T(\text{Cliente})=200000$, $T(\text{Agenzia})=50000$,
 dimensione delle tuple: $S(\text{Cliente})=2000$, $S(\text{Agenzia})=2000$, $S(\text{Città})=200$
 numero di valori $V(\text{Cliente}, \text{Città})=20000$, $V(\text{Agenzia}, \text{Città})=1000$
 $\text{Min}(\text{Cliente}, \text{Età})=20$ $\text{Max}(\text{Cliente}, \text{Età})=60$

Calcoliamo innanzitutto il numero di blocchi occupati da ciascuna relazione:
 $B(\text{Cliente})=T(\text{Cliente}) \cdot S(\text{Cliente})/B = 200000 \cdot 2000/4000=100000$
 $B(\text{Agenzia}) = 50000 \cdot 2000/4000=25000$

$(\sigma_{\text{Età} \geq 30} \text{ Cliente}) \triangleright \triangleleft \text{ Agenzia}$

$\sigma_{\text{Età} \geq 30} (\text{Cliente} \triangleright \triangleleft \text{ Agenzia})$

Prima sequenza: $(\sigma_{\text{Età} \geq 30} \text{ Cliente}) \triangleright \triangleleft \text{ Agenzia}$

Sia $X = (\sigma_{\text{Età} \geq 30} \text{ Cliente})$

C'è un indice secondario su Cliente.Età

$f=(60-30+1)/(60-20+1)=0.756$

$\text{Costo}(X) = 3 + f \cdot T(\text{Cliente}) = 3 + 0.756 \cdot 200000 = 151203$

$$B(X) = 0.756 * 100000 = 75600$$

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

Sia $Z = X \triangleright \triangleleft$ Agenzia

$$\text{CostoJNL}(Z) = B(\text{Agenzia}) + B(X) * B(\text{Agenzia}) / M = 25000 + 75600 * 25000 / 10000 = 214000$$

$$\text{CostoHJI}(Z) = (3 - 2M / B(\text{Agenzia})) (B(X) + B(\text{Agenzia})) = (3 - 2 * 10000 / 25000) * (75600 + 25000) = 221320$$

$$\text{Costo totale} = 151203 + 214000 = 365203$$

Seconda sequenza: $\sigma_{\text{Età} >= 30}$ (Cliente $\triangleright \triangleleft$ Agenzia)

Sia $Z = \text{Cliente} \triangleright \triangleleft$ Agenzia

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

$$\text{CostoJNL}(Z) = B(\text{Agenzia}) + B(\text{Cliente}) * B(\text{Agenzia}) / M = 25000 + 100000 * 25000 / 10000 = 275000$$

$$\text{CostoHJI}(Z) = (3 - 2M / B(\text{Agenzia})) (B(\text{Cliente}) + B(\text{Agenzia})) = (3 - 2 * 10000 / 25000) * (100000 + 25000) = 275000$$

$$T(Z) = T(\text{Agenzia}) * T(\text{Cliente}) / \max\{V(\text{Cliente}, \text{Città}), V(\text{Agenzia}, \text{Città})\} = 50000 * 200000 / 20000 = 500000$$

$$S(Z) = S(\text{Cliente}) + S(\text{Agenzia}) - S(\text{Città}) = 2000 + 2000 - 200 = 3800$$

$$B(Z) = 500000 * 3800 / 4000 = 475000$$

Dato che non ci sono indici su Z:

$$\text{Costo } \sigma_{\text{Età} >= 30}(Z) = B(Z) = 475000$$

$$\text{Costo totale} = 275000 + 475000 = 750000$$

La sequenza di costo minore è la prima