

COMPITO DI SISTEMI INFORMATIVI/BASI DI DATI II

15 settembre 2006 (Tot. 16) Tempo: 2h

Esercizio 1 (punti 3)

Si consideri il seguente log:

- | | |
|----------------------|----------------------|
| 1. I(T1,O1,A1) | 14. I(T3,O6,A14) |
| 2. U(T2,O2,B2,A2) | 15. U(T5,O6,B15,A15) |
| 3. B(T3) | 16. D(T4,O6,B16) |
| 4. D(T3,O2,B4) | 17. B(T6) |
| 5. C(T1) | 18. C(T5) |
| 6. C(T2) | 19. D(T6,O5,B19) |
| 7. B(T4) | 20. B(T7) |
| 8. U(T4,O3,B8,A8) | 21. I(T7,O7,A21) |
| 9. I(T3,O4,A9) | 22. U(T4,O7,B22,A22) |
| 10. B(T5) | 23. C(T4) |
| 11. U(T5,O4,B11,A11) | 24. U(T7,O4,B24,A24) |
| 12. I(T4,O5,A12) | 25. I(T6,O8,A25) |
| 13. CK(T3,T4,T5) | 26. U(T3,O8,B26,A26) |

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

Esercizio 2 (punti 4)

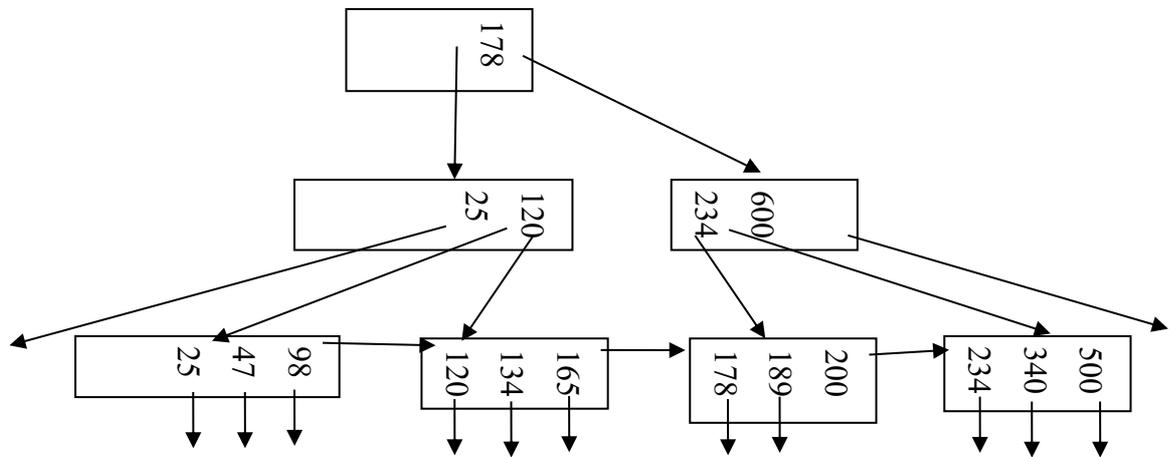
Dato il seguente schedule:

w1(x) r4(y) r2(x) w2(y) w3(x) r3(z) w2(w) r5(w) r5(y) w5(z)

si indichi se e' conflict serializzabile.

Esercizio 3 (punti 3)

Sia dato il seguente B+ tree di ordine 5.



Si mostri come si modifica l'albero nel caso di cancellazione della chiave 47.

Esercizio 4 (punti 6)

Siano date le relazioni

Lavoratore(CodLav, Nome, Cognome, DataDiNascita, CittaDiNascita, Stipendio)

UfficioPostale(CodUff, Nome, Citta, Indirizzo)

la query

SELECT L.*, U.*

FROM Lavoratore AS L, UfficioPostale AS U

WHERE L.CittaDiNascita=U.Citta AND L..Stipendio>20000 AND L.Stipendio<21000

e i parametri:

buffer di memoria centrale disponibili: $M=1.000$

dimensione del buffer: $B=1.000$ bytes

numero di tuple: $T(\text{Lavoratore})=1.000.000$, $T(\text{UfficioPostale})=15.000$,

dimensione delle tuple: $S(\text{Lavoratore})=250$ bytes, $S(\text{UfficioPostale})=200$,

$S(\text{Citta})=S(\text{CittaDiNascita})=50$

numero di valori: $V(\text{Lavoratore}, \text{CittaDiNascita})=12.000$, $V(\text{UfficioPostale}, \text{Citta})=10.000$,

$\text{Min}(\text{Lavoratore}, \text{Stipendio})=10.000$ $\text{Max}(\text{Lavoratore}, \text{Stipendio})=30.000$

Si stabilisca qual'è l'ordine migliore con cui eseguire le operazioni supponendo di utilizzare l'Hash Join Ibrido. In particolare, occorrerà calcolare il costo delle due sequenze:

$(\sigma_{20000 < \text{Lavoratore.Stipendio} < 21000} \text{Lavoratore}) \bowtie \text{UfficioPostale}$ e

$\sigma_{20000 < \text{Lavoratore.Stipendio} < 21000} (\text{Lavoratore} \bowtie \text{UfficioPostale})$.

Si supponga che i record siano impaccati nei buffer e che tutto lo spazio nei buffer sia occupato dai record. Si supponga inoltre che ci sia un indice secondario su Lavoratore.Stipendio. Si supponga che non vengano costruiti indici su Lavoratore > < UfficioPostale. Per il calcolo del costo e della cardinalità della selezione si utilizzi la stima dei valori nel range, sapendo che il range di stipendio è [10000,30000].

SOLUZIONE

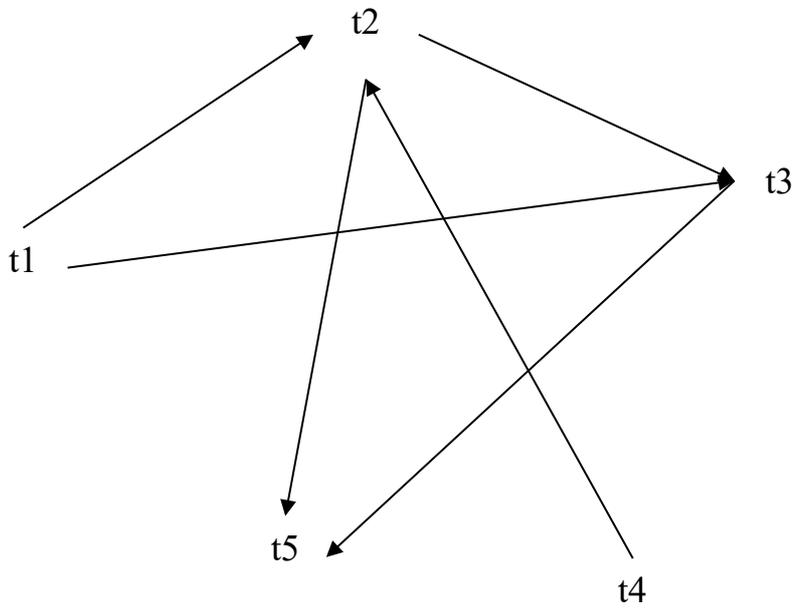
Esercizio 1

1. I(T1,O1,A1)
2. U(T2,O2,B2,A2)
3. B(T3)
4. D(T3,O2,B4)
5. C(T1)
6. C(T2)
7. B(T4)
8. U(T4,O3,B8,A8)
9. I(T3,O4,A9)
10. B(T5)
11. U(T5,O4,B11,A11)
12. I(T4,O5,A12)
13. CK(T3,T4,T5)
14. I(T3,O6,A14)
15. U(T5,O6,B15,A15)
16. D(T4,O6,B16)
17. B(T6)
18. C(T5)
19. D(T6,O5,B19)
20. B(T7)
21. I(T7,O7,A21)
22. U(T4,O7,B22,A22)
23. C(T4)
24. U(T7,O4,B24,A24)
25. I(T6,O8,A25)
26. U(T3,O8,B26,A26)

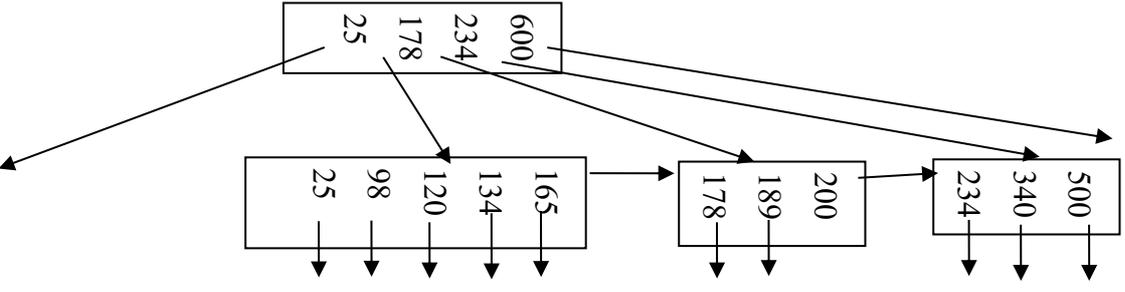
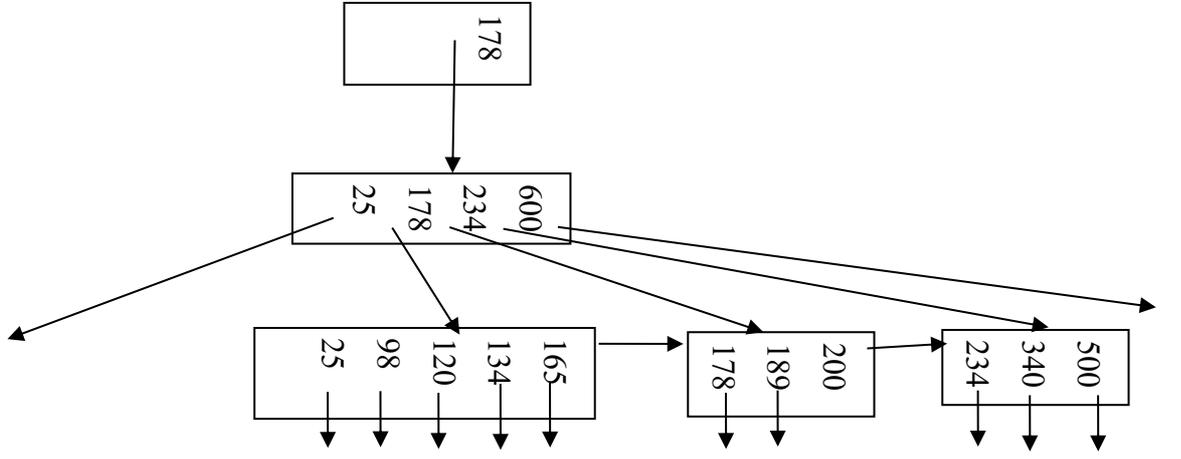
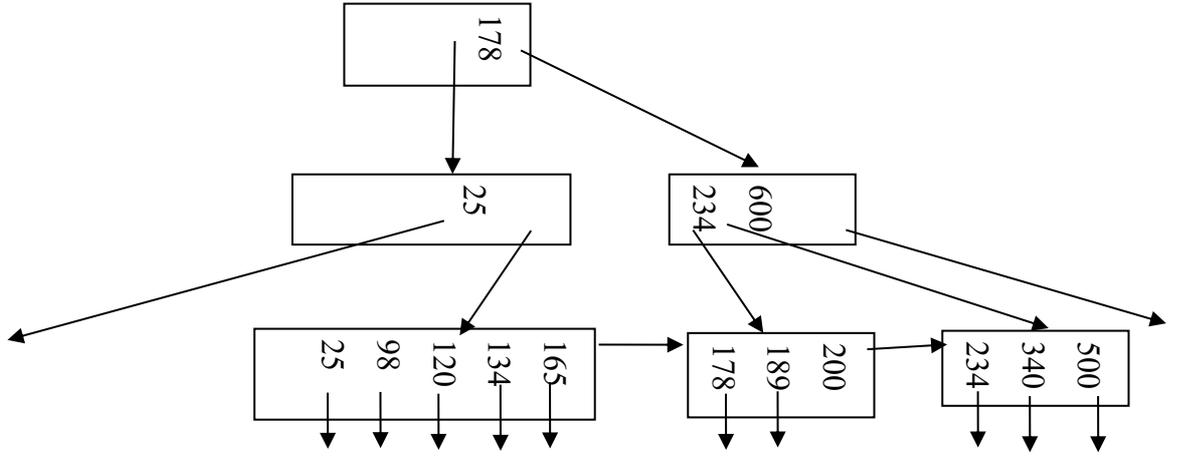
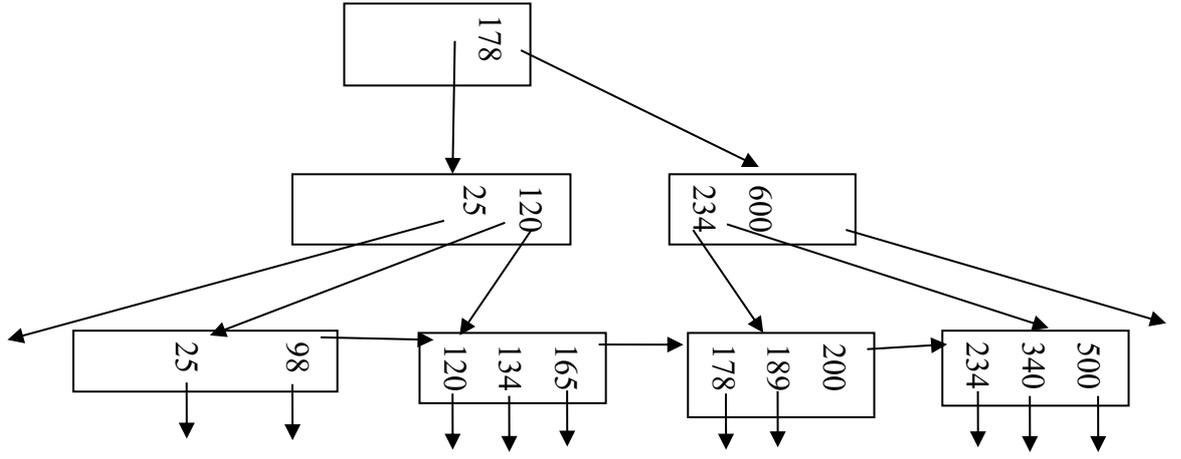
- 13 UNDO={T3,T4,T5} REDO={}
- 17 UNDO={T3,T4,T5,T6} REDO={}
- 18 UNDO={T3,T4,T6} REDO={T5}
- 20 UNDO={T3,T4,T6,T7 } REDO={T5}
- 23 UNDO={T3,T6,T7} REDO={T4,T5}
- UNDO
- 26 O8=B26
- 25 D(O8)
- 24 O4=B24
- 21 D(O7)
- 19 I(O5,B19)
- 14 D(O6)
- 9 D(O4)
- 4 I(O2,B4)
- REDO
- 8 O3=A8
- 11 O4=A11
- 12 I(O5,A12)
- 15 O6=A15
- 16 D(O6)
- 22 O7=A22

Esercizio 2

w1(x) r4(y) r2(x) w2(y) w3(x) r3(z) w2(w) r5(w) r5(y) w5(z)
Lo schedule e' conflict serializzabile



Esercizio 3



Esercizio 4

buffer di memoria centrale disponibili: $M=1.000$

dimensione del buffer: $B=1.000$ bytes

numero di tuple: $T(\text{Lavoratore})=1.000.000$, $T(\text{UfficioPostale})=15.000$,

dimensione delle tuple: $S(\text{Lavoratore})=250$ bytes, $S(\text{UfficioPostale})=200$,

$S(\text{Citta})=S(\text{CittaDiNascita})=50$

numero di valori: $V(\text{Lavoratore}, \text{CittaDiNascita})=12.000$, $V(\text{UfficioPostale}, \text{Citta})=10.000$,

$\text{Min}(\text{Lavoratore}, \text{Stipendio})=10.000$ $\text{Max}(\text{Lavoratore}, \text{Stipendio})=30.000$

Calcoliamo innanzitutto il numero di blocchi occupati da ciascuna relazione:

$B(\text{Lavoratore})=T(\text{Lavoratore}) * S(\text{Lavoratore}) / B = 1000000 * 250 / 1000 = 250000$

$B(\text{UfficioPostale}) = 15000 * 200 / 1000 = 3000$

Prima sequenza: $(\sigma_{20000 < \text{Lavoratore.Stipendio} < 21000} \text{Lavoratore}) \triangleright \triangleleft \text{UfficioPostale}$

Sia $X = \sigma_{20000 < \text{Lavoratore.Stipendio} < 21000} \text{Lavoratore}$

$f = (21000 - 20000 - 1) / (30000 - 10000 + 1) = 0,0499$

$\text{Costo}(X) = 3 + f * T(\text{Lavoratore}) = 3 + 0,0499 * 1000000 = 49903$

$T(X) = f * T(\text{Lavoratore}) = 49900$

$B(X) = 49900 * 250 / 1000 = 12475$

$\text{Costo}(X \triangleright \triangleleft \text{UfficioPostale}) = (3 - 2 * 1000 / 3000) * (3000 + 12475) = 36108$

$\text{Costo}((\sigma_{20000 < \text{Lavoratore.Stipendio} < 21000} \text{Lavoratore}) \triangleright \triangleleft \text{UfficioPostale}) = 49903 + 36108 = 86011$

Seconda sequenza: $\sigma_{20000 < \text{Lavoratore.Stipendio} < 21000}(\text{Lavoratore} \triangleright \triangleleft \text{UfficioPostale})$

Sia $Y = \text{Lavoratore} \triangleright \triangleleft \text{UfficioPostale}$

$\text{Costo}(Y) = (3 - 2 * 1000 / 3000) * (3000 + 250000) = 590333$

$S(Y) = S(\text{Lavoratore}) + S(\text{UfficioPostale}) - S(\text{Citta}) = 250 + 200 - 50 = 400$

$T(Y) = T(\text{Lavoratore}) * T(\text{UfficioPostale}) / \max(V(\text{Lavoratore}, \text{CittaDiNascita}), V(\text{UfficioPostale}, \text{Citta})) = 1000000 * 15000 / 12000 = 1250000$

$B(Y) = 1250000 * 400 / 1000 = 500000$

$\text{Costo}(\sigma_{20000 < \text{Lavoratore.Stipendio} < 21000} Y) = B(Y) = 500000$

$\text{Costo}(\sigma_{20000 < \text{Lavoratore.Stipendio} < 21000}(\text{Lavoratore} \triangleright \triangleleft \text{UfficioPostale})) = 590333 + 500000 = 1090333$

La sequenza migliore per realizzare l'operazione è la prima.