

COMPITO DI SISTEMI INFORMATIVI/BASI DI DATI II

7 luglio 2008 – Compito (Tot. 16) Tempo: 2h

Esercizio 1 (punti 3)

Si consideri il seguente log:

- | | |
|----------------------|----------------------|
| 1. I(T1,O1,A1) | 16. B(T5) |
| 2. B(T2) | 17. U(T5,O3,B17,A17) |
| 3. I(T2,O2,A3) | 18. U(T3,O1,B18,A18) |
| 4. U(T1,O2,B4,A4) | 19. B(T6) |
| 5. B(T3) | 20. I(T6,O6,A20) |
| 6. U(T3,O3,B6,A6) | 21. C(T3) |
| 7. D(T2,O2,B7) | 22. U(T2,O6,B22,A22) |
| 8. B(T4) | 23. I(T2,O7,A23) |
| 9. U(T4,O1,B9,A9) | 24. U(T6,O7,B24,A24) |
| 10. U(T1,O1,B10,A10) | 25. C(T2) |
| 11. C(T1) | 26. U(T4,O6,B26,A26) |
| 12. U(T2,O4,B12,A12) | 27. D(T6,O6,B27) |
| 13. CK(T2,T3,T4) | |
| 14. D(T3,O4,B14) | |
| 15. U(T2,O5,B15,A15) | |

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

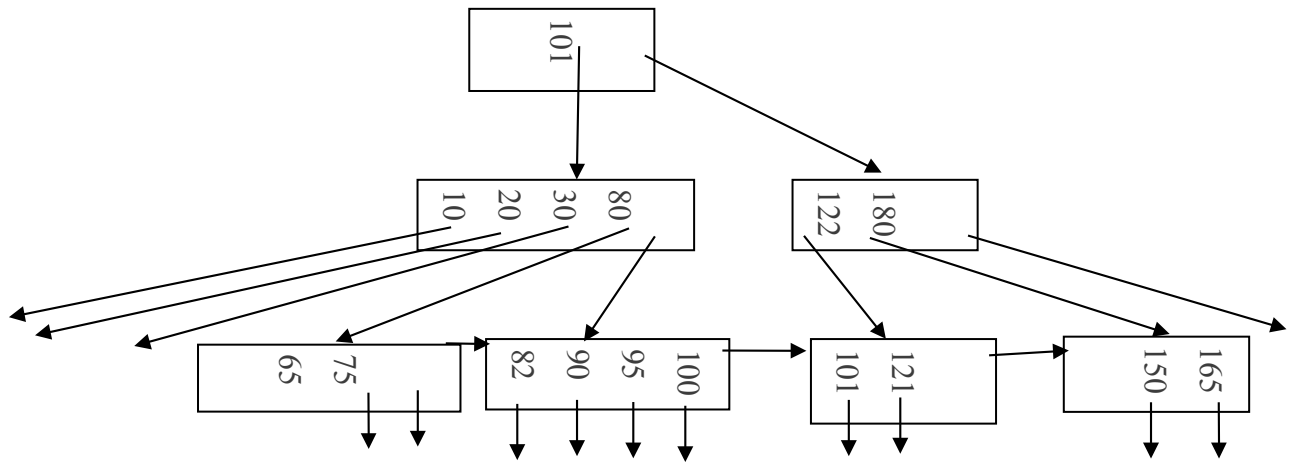
Esercizio 2 (punti 4)

r1(x) r3(w) r4(x) r2(x) w2(y) w4(z) w1(y) w4(x) w3(y)

si indichi se rispetta il two-phase locking. Nel caso lo rispetti, si indichi una sequenza di acquisizione e rilascio di lock compatibile con lo schedule e che rispetti il two-phase locking. Nel caso non lo rispetti, si indichi almeno uno dei conflitti.

Esercizio 3 (punti 3)

Sia dato il seguente B+ tree di ordine 4.



Si mostri come si modifica l'albero nel caso di inserimento della chiave 91.

Esercizio 4 (punti 6)

Siano date le relazioni

Articolo(CodiceArt, Titolo, Rivista, Volume, Numero, AutorePrincipale)

Autore(CodiceAut, Nome, Cognome, DataDiNascita, Istituzione)

la query

SELECT P.*, C.*

FROM Articolo, Autore

WHERE Articolo.AutorePrincipale=Autore.CodiceAut AND Articolo.Volume \geq 2 AND
Articolo.Volume $<$ 5

e i parametri:

buffer di memoria centrale disponibili: M=400.000

dimensione del buffer: B=1.000 bytes

numero di tuple: T(Articolo)= 100.000, T(Autore)= 100.000,

dimensione delle tuple: S(Articolo)=8.000 bytes, S(Autore)=5.000 bytes, S(CodiceAut)=100 bytes

numero di valori V(Articolo, AutorePrincipale)=50.000, Min(Articolo, Volume)=1

Max(Articolo, Volume)=12

Si calcoli il costo minimo di ciascuna delle seguenti sequenze:

$(\sigma_{\text{Volume} \geq 2 \text{ AND Volume} < 5} \text{ Articolo}) \triangleright \triangleleft \text{Autore}$

$\sigma_{\text{Volume} \geq 2 \text{ AND Volume} < 5} (\text{Articolo} \triangleright \triangleleft \text{Autore})$

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

Si supponga di avere indici primari su Articolo.Volume e su Autore.CodiceAut. Per il calcolo del costo e della cardinalità della selezione su Volume si utilizzi la stima dei valori nel range, con i valori massimo e minimo sopra indicati.

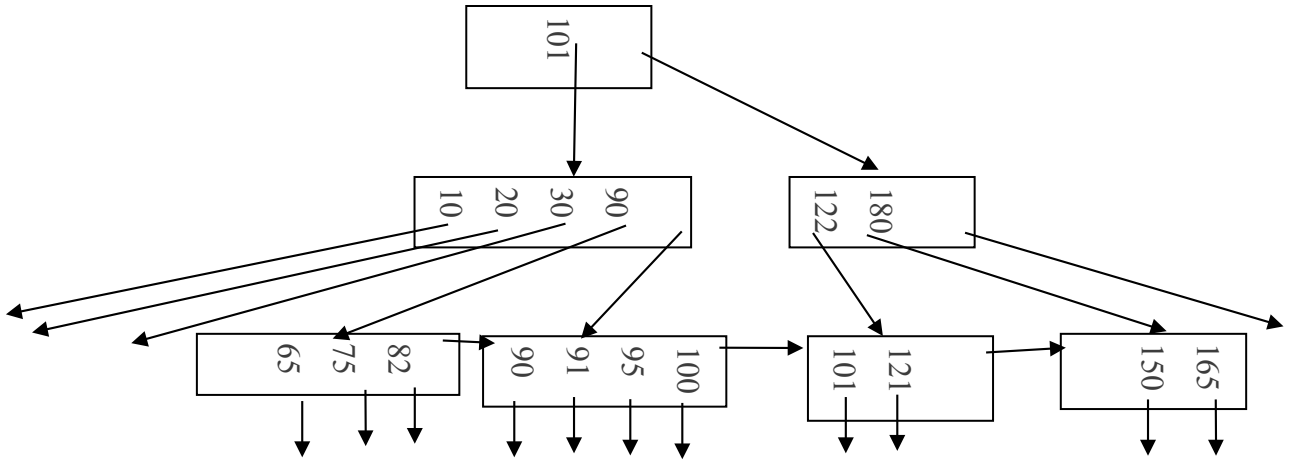
SOLUZIONE

Esercizio 1

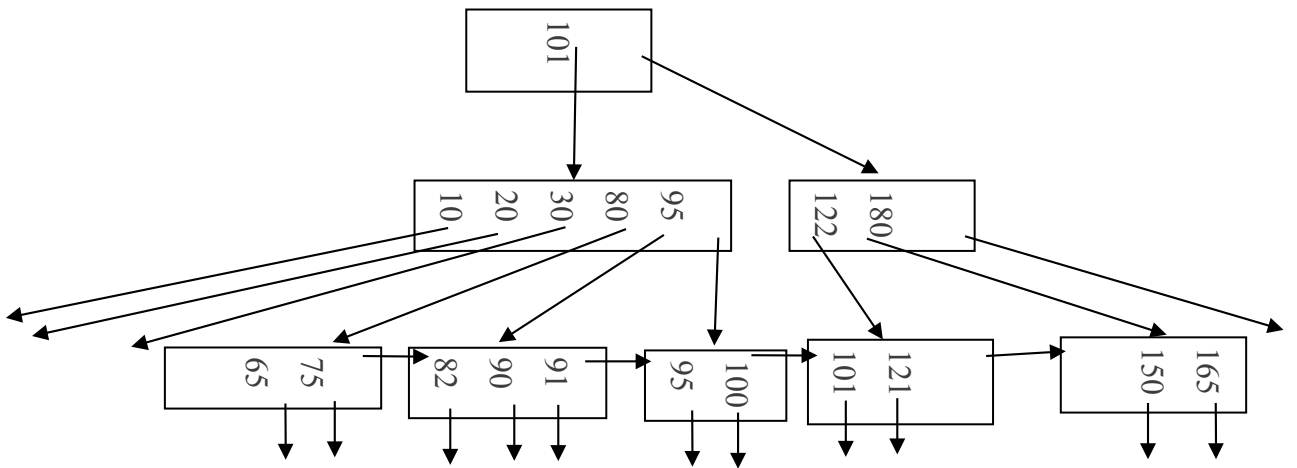
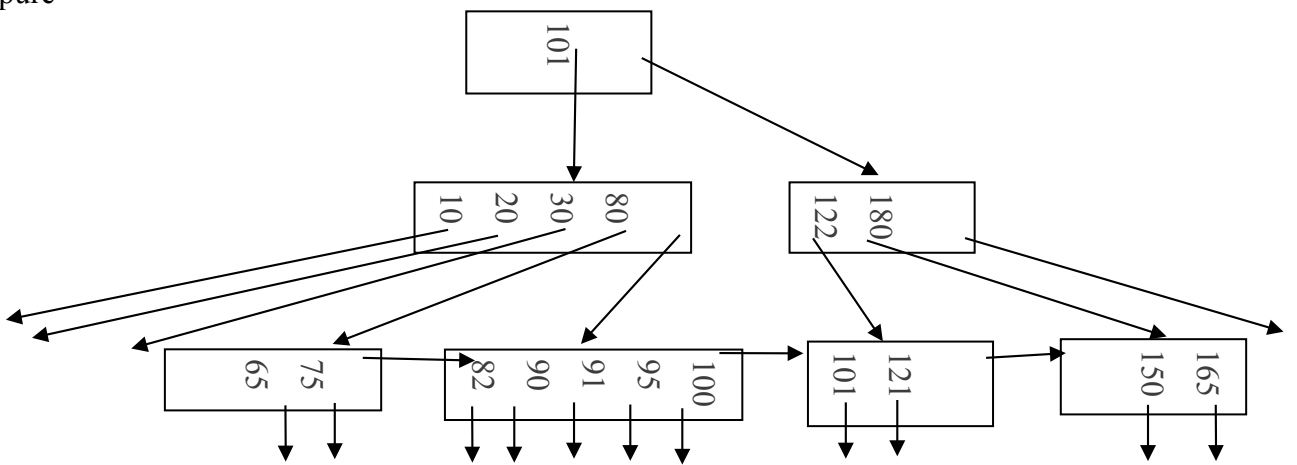
1. I(T1,O1,A1)
2. B(T2)
3. I(T2,O2,A3)
4. U(T1,O2,B4,A4)
5. B(T3)
6. U(T3,O3,B6,A6)
7. D(T2,O2,B7)
8. B(T4)
9. U(T4,O1,B9,A9)
10. U(T1,O1,B10,A10)
11. C(T1)
12. U(T2,O4,B12,A12)
13. CK(T2,T3,T4)
14. D(T3,O4,B14)
15. U(T2,O5,B15,A15)
16. B(T5)
17. U(T5,O3,B17,A17)
18. U(T3,O1,B18,A18)
19. B(T6)
20. I(T6,O6,A20)
21. C(T3)
22. U(T2,O6,B22,A22)
23. I(T2,O7,A23)
24. U(T6,O7,B24,A24)
25. C(T2)
26. U(T4,O6,B26,A26)
27. D(T6,O6,B27)

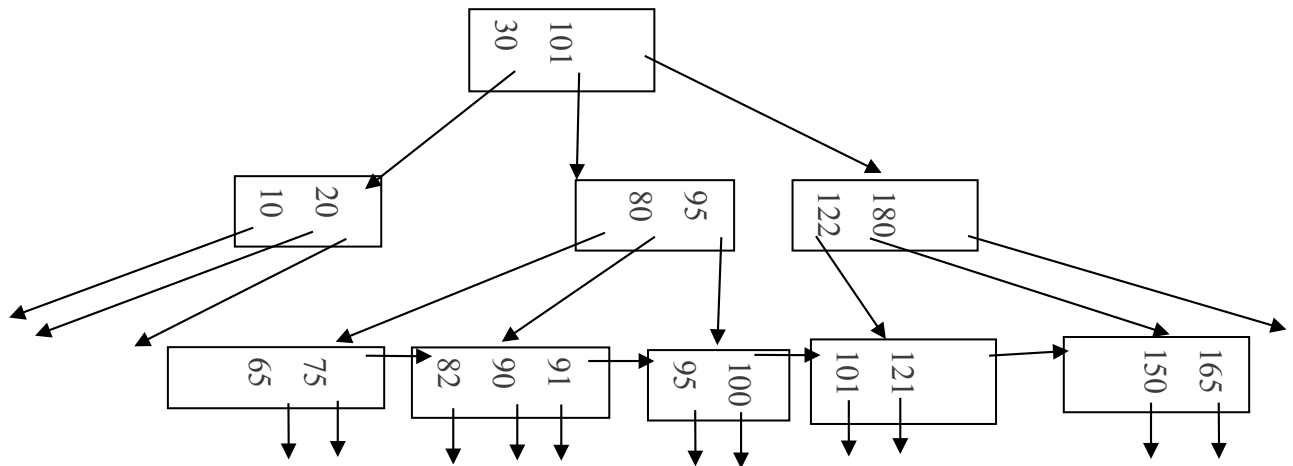
- 13 UNDO={T2,T3,T4} REDO={}
- 16 UNDO={T2,T3,T4,T5} REDO={ }
- 19 UNDO={T2,T3,T4,T5,T6} REDO={ }
- 21 UNDO={ T2, T4,T5,T6} REDO={T3}
- 25 UNDO={ T4,T5,T6} REDO={T2,T3}
- UNDO
- 27 I(O6,B27)
- 26 O6=B26
- 24 O7=B24
- 20 D(O6)
- 17 O3=B17
- 9 O1=B9
- REDO
- 3 I(O2,A3)
- 6 O3=A6
- 7 D(O2)
- 12 O4=A12
- 14 D(O4)
- 15 O5=A15
- 18 O1=A18
- 22 O6=A22
- 23 I(O7,A23)

Esercizio 3



Oppure





Esercizio 4

buffer di memoria centrale disponibili: $M=400.000$

dimensione del buffer: $B=1.000$ bytes

numero di tuple: $T(\text{Articolo})=100.000$, $T(\text{Autore})=100.000$,

dimensione delle tuple: $S(\text{Articolo})=8.000$ bytes, $S(\text{Autore})=5000$,
 $S(\text{CodiceAut})=100$

numero di valori $V(\text{Articolo}, \text{AutorePrincipale})=50.000$

$\text{Min}(\text{Articolo}, \text{Volume})=1$

$\text{Max}(\text{Articolo}, \text{Volume})=12$

Si calcoli il costo minimo di ciascuna delle seguenti sequenze:

$(\sigma_{\text{Volume}>2 \text{ AND } \text{Volume}<5} \text{ Articolo}) \triangleright \triangleleft \text{Autore}$

$\sigma_{\text{Volume}>2 \text{ AND } \text{Volume}<5} (\text{Articolo} \triangleright \triangleleft \text{Autore})$

Calcoliamo innanzitutto il numero di blocchi occupati da ciascuna relazione:

$B(\text{Articolo})=T(\text{Articolo}) \cdot S(\text{Articolo})/B = 100000 \cdot 8000/1000=800.000$

$B(\text{Autore}) = 100000 \cdot 5000/1000=500.000$

Prima sequenza: $(\sigma_{\text{Volume}>2 \text{ AND } \text{Volume}<5} \text{ Articolo}) \triangleright \triangleleft \text{Autore}$

Sia $X = \sigma_{\text{Volume}>2 \text{ AND } \text{Volume}<5} \text{ Articolo}$

C'è un indice primario su Volume

$F=(5-2)/(12-1+1)=0.25$

$\text{Costo}(X) = 3 + \lceil B(\text{Articolo}) \cdot f \rceil = 3 + \lceil 800000 \cdot 0.25 \rceil = 200003$

$T(X) = 100000 \cdot 0.25 = 25000$

$B(X) = 25000 \cdot 8000/1000 = 200000$

X sta in memoria centrale quindi posso usare il join ad un passo

$\text{Costo}(X \triangleright \triangleleft \text{Autore}) = B(X) + B(\text{Autore}) = 200000 + 500000 = 700.000$

Costo totale = $200003 + 700000 = 900.003$

Seconda sequenza: $\sigma_{\text{Volume} \geq 2 \text{ AND } \text{Volume} < 5}$ (Articolo \triangleright \triangleleft Autore)

Sia $Z = \text{Articolo} \triangleright \triangleleft \text{Autore}$

Né Articolo né Autore stanno in memoria centrale, quindi si prova con il join nested-loop basato sui blocchi, con l'hash join ibrido e con il join con indice visto che c'è un indice primario su Autore.CodiceAut.

CostoJNL(Z)= $B(\text{Autore})+B(\text{Autore}) * B(\text{Articolo}) / M = 500000 + 500000 * 800000 / 400000 = 1.500.000$

Hash join ibrido: si può fare perché $M > \sqrt{B(\text{Autore})} \Rightarrow 400000 > \sqrt{500000} \Rightarrow 400000 > 707$

CostoHJI(Z)= $(3 - 2M / B(\text{Autore})) (B(\text{Autore}) + B(\text{Articolo})) = (3 - 2 * 400000 / 500000) * (500000 + 800000) = 1.820.000$

CostoIJ(Z)= $B(\text{Articolo}) + T(\text{Articolo}) \lceil B(\text{Autore}) / V(\text{Autore}, \text{CodiceAut}) \rceil = 800000 + 100000 * \lceil 500000 / 100000 \rceil = 1.300.000$

$T(Z) = T(\text{Articolo}) * T(\text{Autore}) / \max \{ V(\text{Articolo}, \text{AutorePrincipale}), V(\text{Autore}, \text{CodiceAut}) \} = 100000 * 100000 / 100000 = 100000$

$S(Z) = S(\text{Articolo}) + S(\text{Autore}) - S(\text{CodiceAut}) = 8000 + 5000 - 100 = 12900$

$B(Z) = 100000 * 12900 / 1000 = 1.290.000$

Dato che non ci sono indici su Z :

Costo: ($\sigma_{\text{Volume} > 2 \text{ AND } \text{Volume} < 5} Z$)= $B(Z) = 1.290.000$

Costo totale= $1290000 + 1300000 = 2.590.000$