

## COMPITO DI SISTEMI INFORMATIVI/BASI DI DATI II

13 febbraio 2008 (Tot. 16) Tempo: 2h

### **Esercizio 1 (punti 3)**

Si consideri il seguente log:

- |                      |                      |
|----------------------|----------------------|
| 1. B(T1)             | 16. I(T3,O8,A16)     |
| 2. I(T1,O1,A2)       | 17. U(T1,O8,B17,A17) |
| 3. U(T2,O2,B3,A3)    | 18. C(T1)            |
| 4. B(T3)             | 19. U(T4,O8,B19,A19) |
| 5. D(T3,O5,B5)       | 20. D(T3,O8,B20)     |
| 6. I(T3,O6,A6)       | 21. B(T5)            |
| 7. B(T4)             | 22. U(T5,O7,B22,A22) |
| 8. U(T4,O2,B8,A8)    | 23. I(T3,O9,A23)     |
| 9. D(T2,O2,B9)       | 24. B(T6)            |
| 10. C(T2)            | 25. U(T6,O9,B25,A25) |
| 11. U(T1,O5,B11,A11) | 26. C(T4)            |
| 12. D(T1,O5,B12)     | 27. U(T5,O9,B27,A27) |
| 13. I(T4,O6,A13)     | 28. C(T5)            |
| 14. CK(T1,T3,T4)     |                      |
| 15. U(T4,O7,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)**

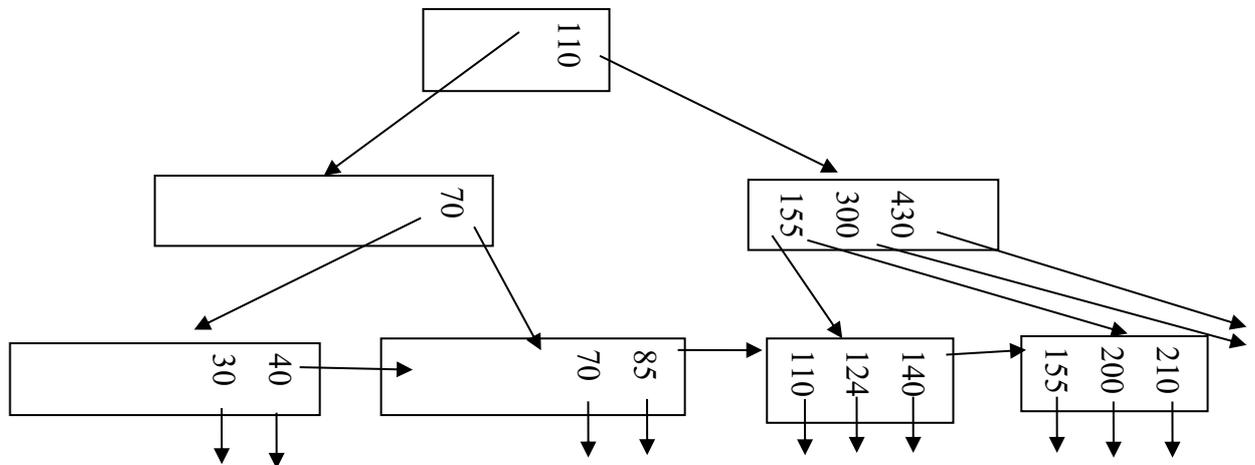
Dato il seguente schedule:

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

si indichi se 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 3.



Si mostri come si modifica l'albero nel caso di cancellazione della chiave 70

### Esercizio 4 (punti 6)

Siano date le relazioni

Paziente(CodicePaz, Nome, Cognome, Età, CodiceMed)

Medico(CodiceMed, Nome, Cognome, Indirizzo)

la query

```
SELECT P.*, C.*
```

```
FROM Paziente AS P, Medico AS M
```

```
WHERE P.CodiceMed=M.CodiceMed AND P.Età>10 AND P.Età<20
```

e i parametri:

buffer di memoria centrale disponibili:  $M=1000$

dimensione del buffer:  $B=1.000$  bytes

numero di tuple:  $T(\text{Paziente})= 80000$ ,  $T(\text{Medico})= 10000$ ,

dimensione delle tuple:  $S(\text{Paziente})=500$  bytes,  $S(\text{Medico})=400$ ,

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

Si calcoli il costo minimo di ciascuna delle seguenti sequenze:

$\sigma_{\text{Età}>10 \text{ AND } \text{Età}<20}(\text{Paziente} \bowtie \text{Medico})$

$(\sigma_{\text{Età}>10 \text{ AND } \text{Età}<20} \text{Paziente}) \bowtie \text{Medico}$

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 primario su Medico.CodiceMed e un indice secondario su Paziente.Età. Per il calcolo del costo e della cardinalità della selezione su Età si utilizzi la stima dei valori nel range, sapendo che il range di Età è  $[1,105]$ .

## SOLUZIONE

### Esercizio 1

1. B(T1)
2. I(T1,O1,A2)
3. U(T2,O2,B3,A3)
4. B(T3)
5. D(T3,O5,B5)
6. I(T3,O6,A6)
7. B(T4)
8. U(T4,O2,B8,A8)
9. D(T2,O2,B9)
10. C(T2)
11. U(T1,O5,B11,A11)
12. D(T1,O5,B12)
13. I(T4,O6,A13)
14. CK(T1,T3,T4)
15. U(T4,O7,B15,A15)
16. I(T3,O8,A16)
17. U(T1,O8,B17,A17)
18. C(T1)
19. U(T4,O8,B19,A19)
20. D(T3,O8,B20)
21. B(T5)
22. U(T5,O7,B22,A22)
23. I(T3,O9,A23)
24. B(T6)
25. U(T6,O9,B25,A25)
26. C(T4)
27. U(T5,O9,B27,A27)
28. C(T5)

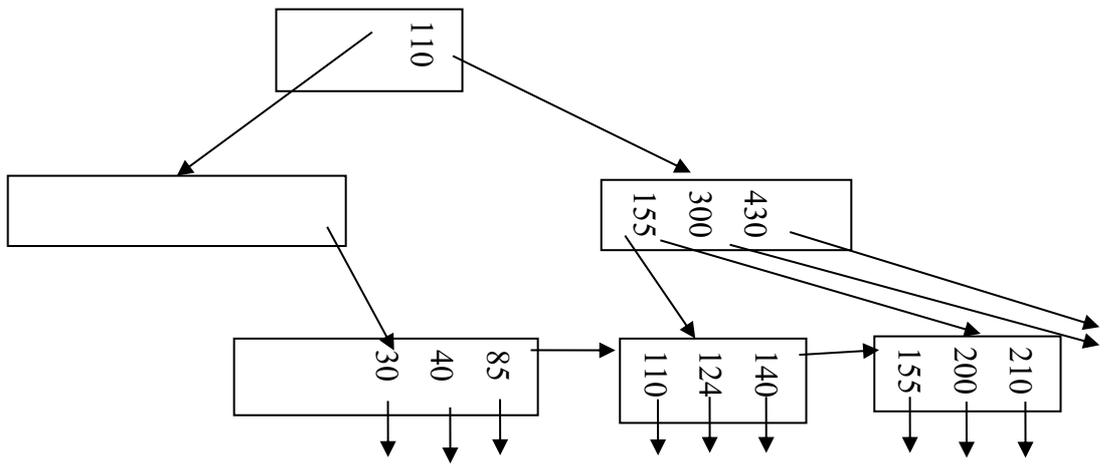
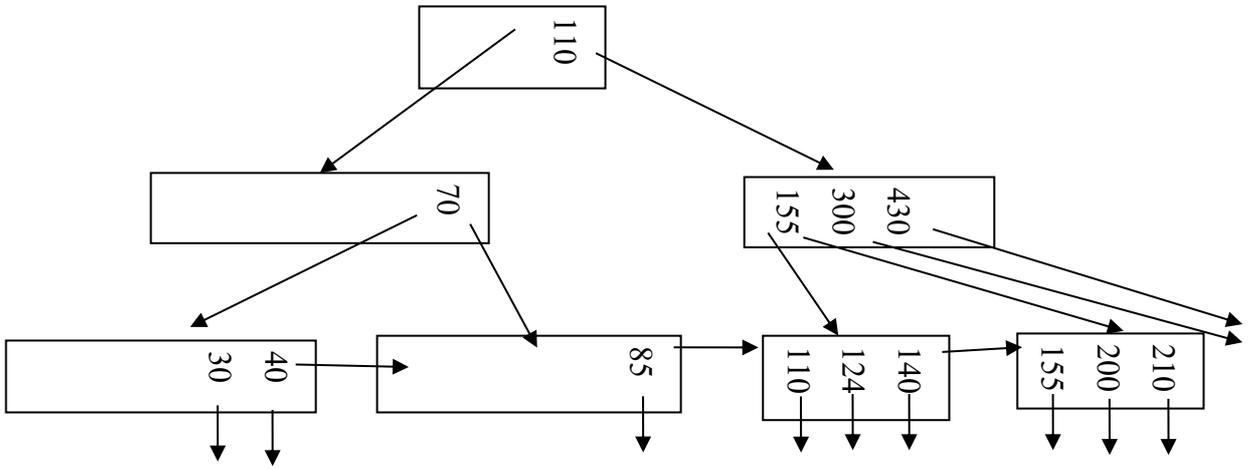
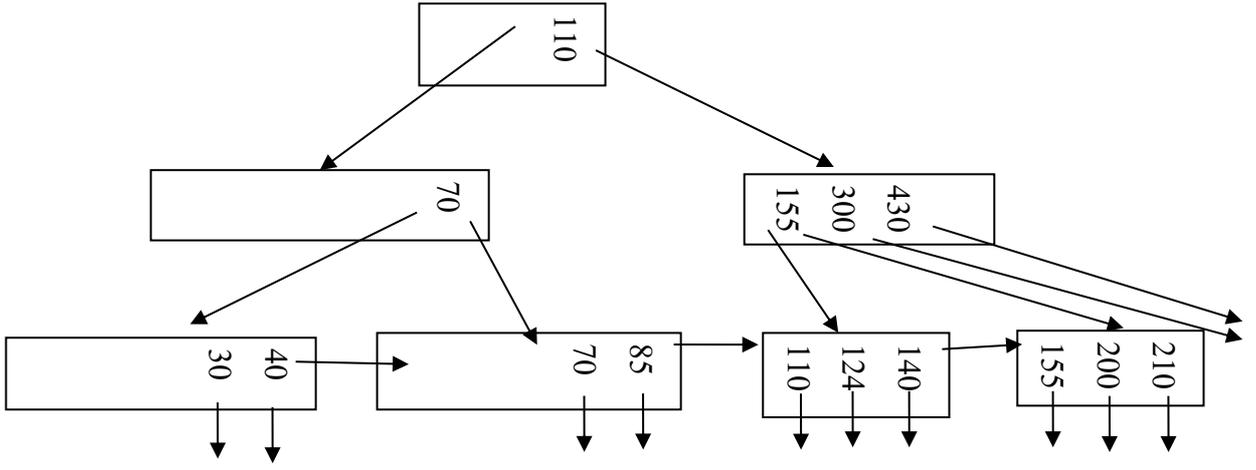
- 14 UNDO={T1,T3,T4 } REDO={}
- 18 UNDO={T3,T4} REDO={T1}
- 21 UNDO={T3,T4,T5} REDO={T1}
- 24 UNDO={T3,T4,T5,T6 } REDO={T1}
- 26 UNDO={T3, T5,T6 } REDO={T1,T4}
- 28 UNDO={T3, T6 } REDO={T1,T4,T5}
- UNDO
- 25 O9=B25
- 23 D(O9)
- 20 I(O8,B20)
- 16 D(O8)
- 6 D(O6)
- 5 I(O5,B5)
- REDO
- 2 I(O1,A2)
- 8 O2=A8
- 11 O5=A11
- 12 D(O5)
- 13 I(O6,A13)
- 15 O7=A15
- 17 O8=A17
- 19 O8=A19
- 22 O7=A22
- 27 O9=A27

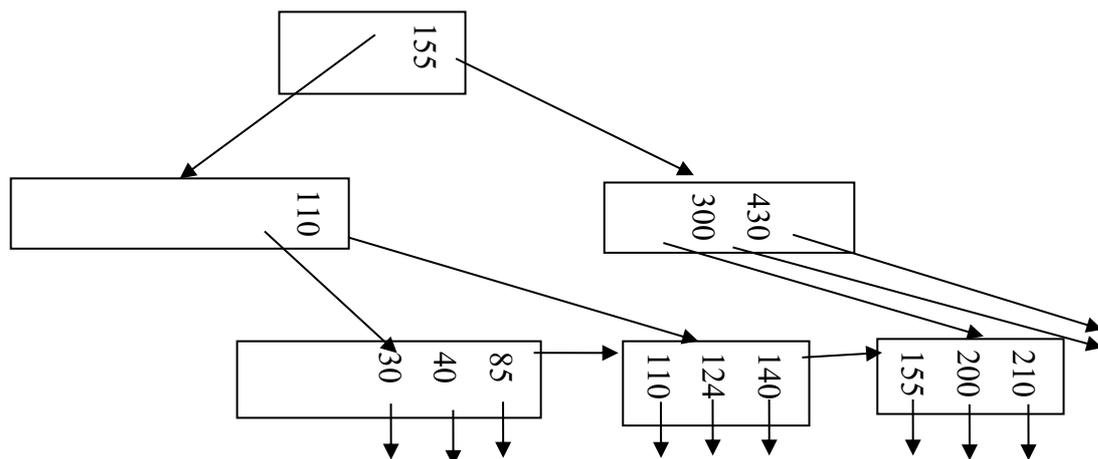
**Esercizio 2**

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

Lo schedale non rispetta il two-phase locking in quanto la transazione t2 deve rilasciare il lock su x dopo r2(x) perche' t1 deve scrivere su x ma t2 deve successivamente scrivere su x

**Esercizio 3**





#### Esercizio 4

buffer di memoria centrale disponibili:  $M=1000$

dimensione del buffer:  $B=1.000$  bytes

numero di tuple:  $T(\text{Paziente})= 80000$ ,  $T(\text{Medico})= 10000$ ,

dimensione delle tuple:  $S(\text{Paziente})=500$  bytes,  $S(\text{Medico})=400$ ,  
 $S(\text{CodiceMed})=200$

Si calcoli il costo minimo di ciascuna delle seguenti sequenze:

$\sigma_{\text{Età}>10 \text{ AND } \text{Età}<20} (\text{Paziente} \triangleright \triangleleft \text{Medico})$

$(\sigma_{\text{Età}>10 \text{ AND } \text{Età}<20} \text{Paziente}) \triangleright \triangleleft \text{Medico}$

Calcoliamo innanzitutto il numero di blocchi occupati da ciascuna relazione:

$B(\text{Paziente})=T(\text{Paziente}) \cdot S(\text{Paziente})/B = 80000 \cdot 500/1000=40000$

$B(\text{Medico}) = 10000 \cdot 400/1000=4000$

Prima sequenza:  $(\sigma_{\text{Età}>10 \text{ AND } \text{Età}<20} \text{Paziente}) \triangleright \triangleleft \text{Medico}$

Sia  $X = \sigma_{\text{Età}>10 \text{ AND } \text{Età}<20} \text{Paziente}$

C'è un indice secondario su Paziente

$F=(20-10-1)/(105-1+1)=9/105=0.086$

Costo(X) =  $3 + f \cdot T(\text{Paziente})=3 + 0.086 \cdot 80000=6883$

$T(X) = 0.086 \cdot 80000=6880$

$B(X) = 0.086 \cdot 40000=3440$

Né X né Medico 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 su CodiceMed.

CostoJNL( $X \triangleright \triangleleft \text{Medico}$ )= $B(X)+B(X) \cdot B(\text{Medico})/M = 3440+3440 \cdot 4000/1000=17200$

CostoHJI( $X \triangleright \triangleleft \text{Medico}$ )= $(3-2M/B(\text{Medico}))(B(X)+B(\text{Medico})) = (3-2 \cdot 1000/3440) \cdot (3440+4200) = 18478$

CostoIJ( $X \triangleright \triangleleft \text{Medico}$ )= $B(X)+T(X) \cdot \lceil B(\text{Medico})/V(\text{Medico}, \text{CodiceMed}) \rceil = 3440+6880 \cdot 1=11840$

Costo totale =  $6883+11840=18723$

Seconda sequenza:  $\sigma_{\text{Età}>10 \text{ AND } \text{Età}<20} (\text{Paziente} \triangleright \triangleleft \text{Medico})$

Sia  $Z = \text{Paziente} \triangleright \triangleleft \text{Medico}$

Né Paziente né Medico 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 su CodiceMed.

CostoJNL(Z)=B(Medico)+B(Paziente)\*B(Medico)/M=4000+4000\*40000/1000=164000

CostoHJI(Z)=(3-2M/B(Medico))(B(Medico)+B(Paziente)) = (3-2\*1000/4000)\*(4000+40000) = 110000

CostoIJ(Z)=B(Paziente)+T(Paziente)  $\lceil$  B(Medico)/V(Medico,CodiceMed)  $\rceil$  = 40000+80000\*1=120000

T(Z)=T(Medico)\*T(Paziente)/max{V(Medico,CodiceMed), V(Paziente, CodiceMed)} = 80000\*10000/10000=80000

S(Z)=S(Medico)+S(Paziente)-S(CodiceMed)= 400+500-200=700

B(Z) = 80000\*700/1000=56000

Dato che non ci sono indici su Z:

Costo:  $(\sigma_{\text{Età}>10 \text{ AND } \text{Età}<20} Z)=B(Z)= 56000$

Costo totale=56000+110000=166000