

## COMPITO DI SISTEMI INFORMATIVI/BASI DI DATI II

13 giugno 2007 (Tot. 16) Tempo: 2h

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

Si consideri il seguente log:

- |                      |                      |
|----------------------|----------------------|
| 1. U(T1,O1,B1,A1)    | 16. D(T3,O3,B16)     |
| 2. I(T2,O2,A2)       | 17. C(T4)            |
| 3. C(T1)             | 18. I(T3,O5,A18)     |
| 4. B(T3)             | 19. U(T5,O5,B19,A19) |
| 5. D(T3,O1,B5)       | 20. B(T6)            |
| 6. B(T4)             | 21. U(T6,O4,B21,A21) |
| 7. U(T4,O2,B7,A7)    | 22. I(T5,O6,A22)     |
| 8. C(T2)             | 23. C(T5)            |
| 9. I(T3,O3,A9)       | 24. B(T7)            |
| 10. U(T4,O3,B10,A10) | 25. I(T7,O7,A25)     |
| 11. B(T5)            | 26. U(T6,O7,B26,A26) |
| 12. I(T5,O4,A12)     | 27. C(T3)            |
| 13. U(T3,O4,B13,A13) | 28. U(T7,O6,B28,A28) |
| 14. CK(T3,T4,T5)     |                      |
| 15. U(T5,O3,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)**

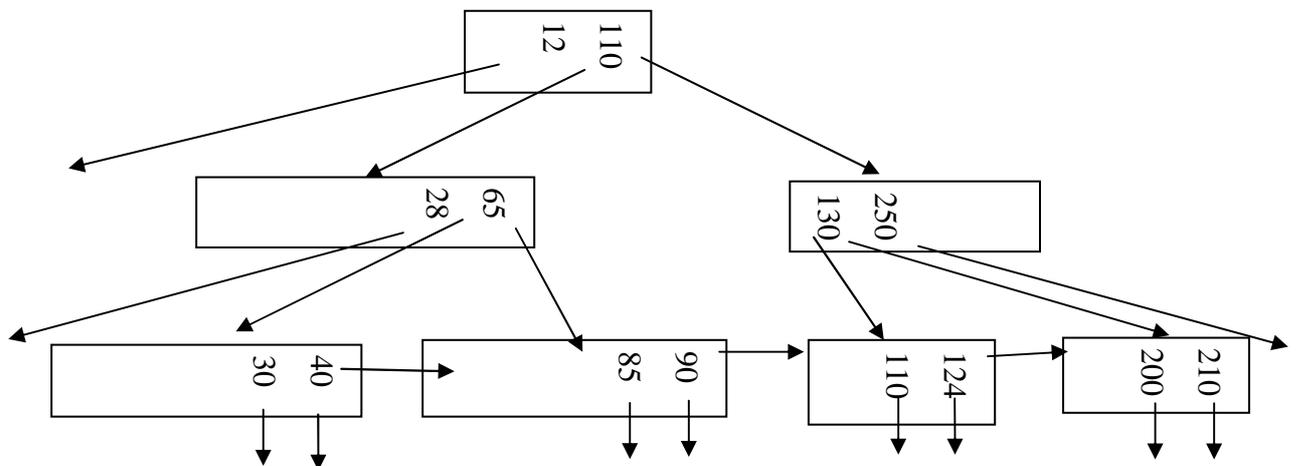
Dato il seguente schedule:

r2(x) w3(y) w1(x) r5(p) w5(x) w2(z) r4(w) r3(z) r5(y) w4(p)

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 cancellazione della chiave 85

### Esercizio 4 (punti 6)

Siano date le relazioni

Automobile(Targa,Proprietario,DataImmatricolazione,Città, Cilindrata)

Conducente(Codice,CittàResidenza,Età)

la query

SELECT V.\*, A.\*

FROM Automobile AS A, Conducente AS C

WHERE A.Città=C.Città AND A.Cilindrata>1251 AND A.Cilindrata<1500

e i parametri:

buffer di memoria centrale disponibili:  $M=20.000$

dimensione del buffer:  $B=8.000$  bytes

numero di tuple:  $T(\text{Automobile})=480000$ ,  $T(\text{Conducente})=160000$ ,

dimensione delle tuple:  $S(\text{Automobile})=2000$  bytes,  $S(\text{Conducente})=1500$ ,

$S(\text{Città})=200$

numero di valori:  $V(\text{Automobile},\text{Città})=10000$ ,  $V(\text{Conducente},\text{Città})=20000$

$\text{Min}(\text{Automobile},\text{Cilindrata})=501$   $\text{Max}(\text{Automobile},\text{Cilindrata})=2500$

Si calcoli il costo minimo di ciascuna delle seguenti sequenze:

$(\sigma_{\text{Cilindrata}>1251 \text{ AND } \text{Cilindrata}<1500} \text{ Automobile}) \triangleright \triangleleft \text{Conducente}$

$\sigma_{\text{Cilindrata}>1251 \text{ AND } \text{Cilindrata}<1500} (\text{Automobile} \triangleright \triangleleft \text{Conducente})$

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 Automobile.Cilindrata. Per il calcolo del costo e della cardinalità della selezione su cilindrata si utilizzi la stima dei valori nel range, sapendo che il range di Cilindrata è  $[501,2500]$ .

## SOLUZIONE

### Esercizio 1

1. U(T1,O1,B1,A1)
2. I(T2,O2,A2)
3. C(T1)
4. B(T3)
5. D(T3,O1,B5)
6. B(T4)
7. U(T4,O2,B7,A7)
8. C(T2)
9. I(T3,O3,A9)
10. U(T4,O3,B10,A10)
11. B(T5)
12. I(T5,O4,A12)
13. U(T3,O4,B13,A13)
14. CK(T3,T4,T5)
15. U(T5,O3,B15,A15)
16. D(T3,O3,B16)
17. C(T4)
18. I(T3,O5,A18)
19. U(T5,O5,B19,A19)
20. B(T6)
21. U(T6,O4,B21,A21)
22. I(T5,O6,A22)
23. C(T5)
24. B(T7)
25. I(T7,O7,A25)
26. U(T6,O7,B26,A26)
27. C(T3)
28. U(T7,O6,B28,A28)

14 UNDO={T3,T4,T5} REDO={}  
17 UNDO={T3,T5} REDO={T4}  
20 UNDO={T3,T5,T6} REDO={T4}  
23 UNDO={T3,T6} REDO={T4,T5}  
24 UNDO={T3,T6,T7} REDO={T4,T5}  
27 UNDO={T6,T7} REDO={T3,T4,T5}

UNDO

28 O6=B28  
26 O7=B26  
25 D(O7)  
21 O4=B19

REDO

5 D(O1)  
7 O2=A7  
9 I(O3,A9)  
10 O3=A10  
12 I(O4,A12)  
13 O4=A13  
15 O3=A15  
16 D(O3)  
18 I(O5,A18)  
19 O5=A19  
22 I(O6,A22)

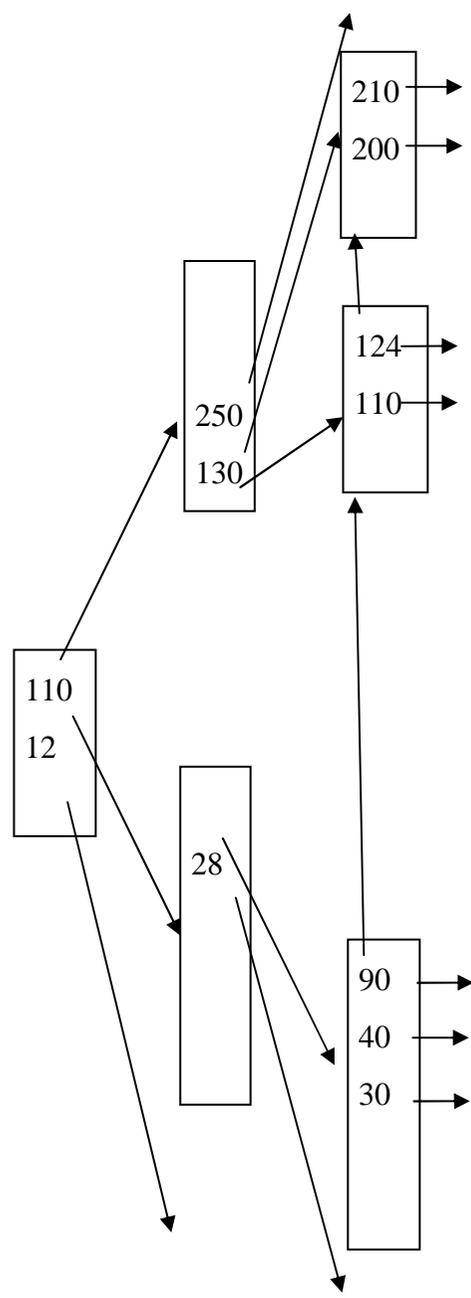
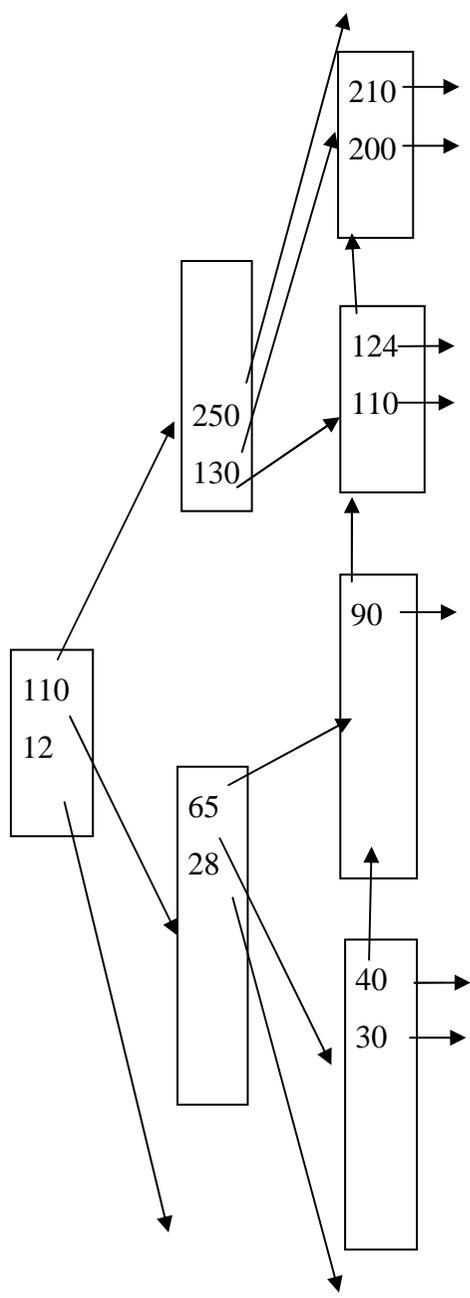
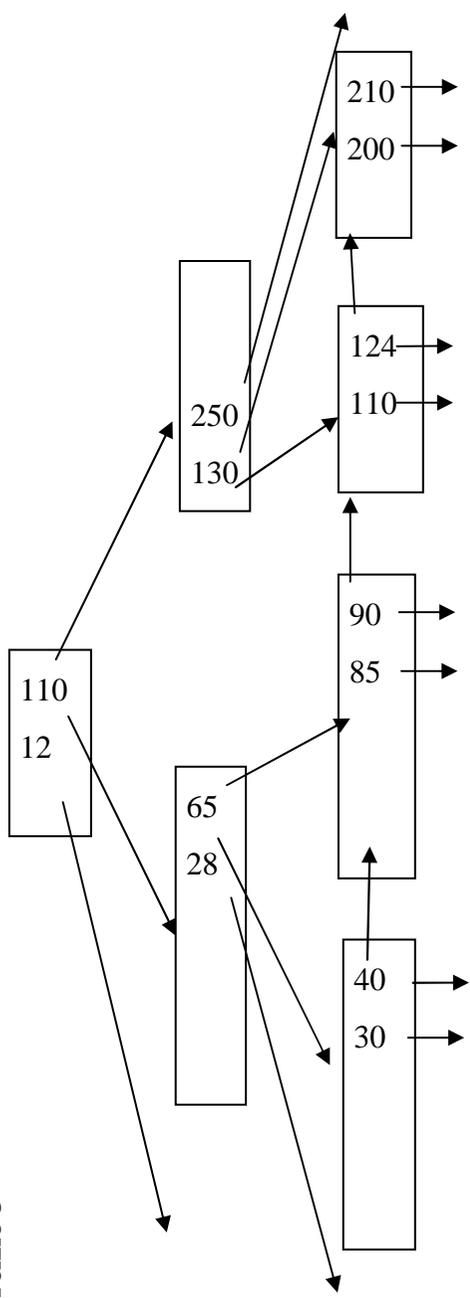
**Esercizio 2**

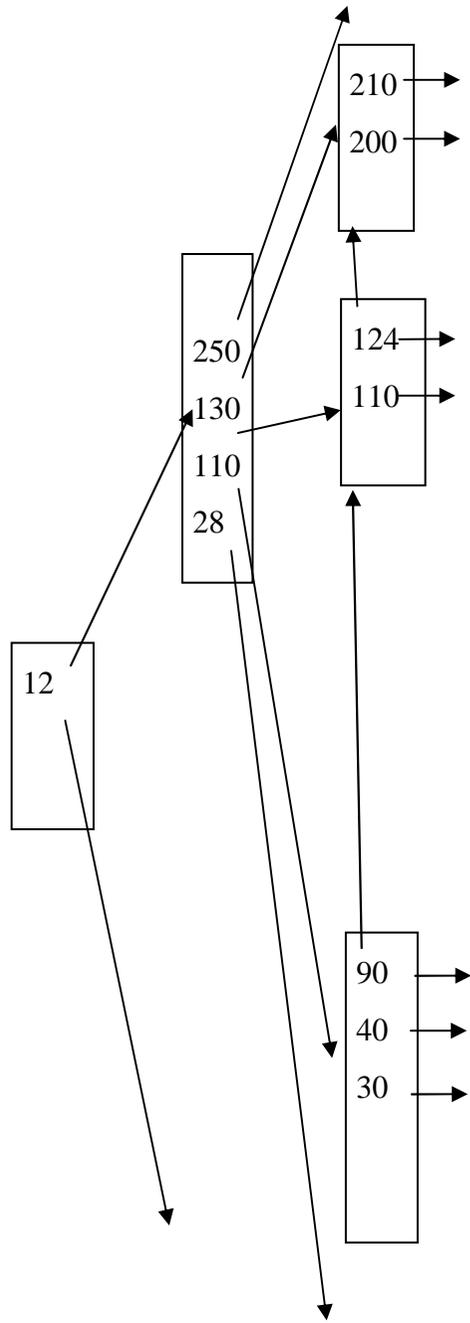
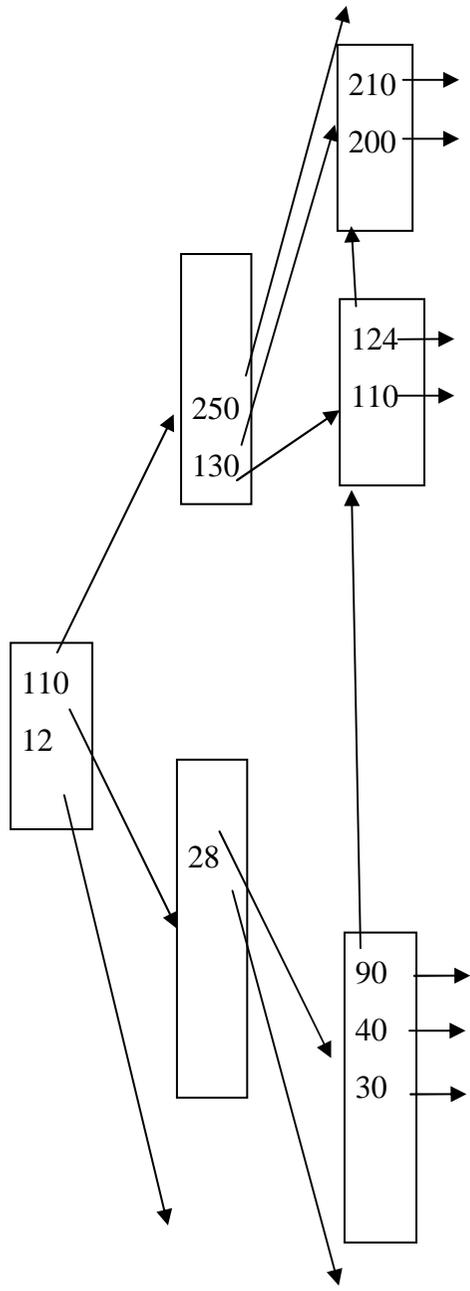
r2(x) w3(y) w1(x) r5(p) w5(x) w2(z) r4(w) r3(z) r5(y) w4(p)

Lo schedule rispetta il two-phase locking. Una sequenze di acquisizione e rilascio dei lock compatibile con lo schedule e che rispetti il two-phase locking è la seguente:

Transazione 1	Transazione 2	Transazione 3	Transazione 4	Transazione 5
	r_lock(x)			
		w_lock(y)		
	w_lock(z)			
	unlock(x)			
w_lock(x)				
unlock(x)				
				r_lock(p)
				w_lock(x)
	unlock(z)			
			r_lock(w)	
		r_lock(z)		
		unlock(y)		
				r_lock(y)
				unlock(p)
			w_lock(p)	
			unlock(p)	
		unlock(z)		
			unlock(w)	
				unlock(x)
				unlock(y)

**Esercizio 3**





#### Esercizio 4

buffer di memoria centrale disponibili:  $M=20.000$

dimensione del buffer:  $B=8.000$  bytes

numero di tuple:  $T(\text{Automobile})=480000$ ,  $T(\text{Conducente})=160000$ ,

dimensione delle tuple:  $S(\text{Automobile})=2000$  bytes,  $S(\text{Conducente})=1500$ ,  
 $S(\text{Città})=200$

numero di valori:  $V(\text{Automobile}, \text{Città})=10000$ ,  $V(\text{Conducente}, \text{Città})=20000$

$\text{Min}(\text{Automobile}, \text{Cilindrata})=501$   $\text{Max}(\text{Automobile}, \text{Cilindrata})=2500$

Calcoliamo innanzitutto il numero di blocchi occupati da ciascuna relazione:

$B(\text{Automobile})=T(\text{Automobile}) * S(\text{Automobile}) / B = 480000 * 2000 / 8000 = 120000$

$B(\text{Conducente}) = 160000 * 1500 / 8000 = 30000$

Prima sequenza:  $(\sigma_{\text{Cilindrata} > 1251 \text{ AND } \text{Cilindrata} < 1500} \text{ Automobile}) \triangleright \triangleleft \text{Conducente}$

Sia  $X = \sigma_{\text{Cilindrata} > 1251 \text{ AND } \text{Cilindrata} < 1500} \text{ Automobile}$

Utilizzando l'indice secondario su Cilindrata

$f = (1500 - 1251 + 1) / (2500 - 501 + 1) = 0.125$

$\text{Costo}(X) = 3 + f * T(\text{Automobile}) = 3 + 0.125 * 480000 = 60003$

$T(X) = f * T(\text{Automobile}) = 0.125 * 480000 = 60000$

$B(X) = 60000 * 2000 / 8000 = 15000$

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

$\text{Costo}_{\text{JIP}}(X \triangleright \triangleleft \text{Conducente}) = B(X) + B(Y) = 15000 + 30000 = 45000$

$\text{Costo totale} = 60003 + 45000 = 105003$

Seconda sequenza:  $\sigma_{\text{Cilindrata} > 1251 \text{ AND } \text{Cilindrata} < 1500} (\text{Automobile} \triangleright \triangleleft \text{Conducente})$

Sia  $Z = \text{Automobile} \triangleright \triangleleft \text{Conducente}$

Né Automobile né Conducente stanno in memoria centrale, quindi provo con il join nested-loop basato sui blocchi e con l'hash join ibrido

Join nested-loop basato sui blocchi:

$\text{Costo}_{\text{NLJ}}(Z) = B(\text{Conducente}) + B(\text{Automobile})B(\text{Conducente}) / M =$   
 $= 30000 + 120000 * 30000 / 20000 = 210000$

Hash join ibrido: si può fare perchè  $20000 > \sqrt{30000} = 173$

$\text{Costo}_{\text{HJ}}(Z) = (3 - 2 * 20000 / 30000) * (120000 + 30000) = 250000$

Il migliore e' il join nested-loop basato sui blocchi

$T(Z) = T(\text{Automobile}) * T(\text{Conducente}) / \max\{V(\text{Automobile}, \text{Città}), V(\text{Conducente}, \text{Città})\} =$   
 $480000 * 160000 / \max\{10000, 20000\} = 3840000$

$S(Z) = S(\text{Automobile}) + S(\text{Conducente}) - S(\text{Città}) = 2000 + 1500 - 200 = 3300$

$B(Z) = 3840000 * 3300 / 8000 = 1584000$

Dato che non ci sono indici su Z:

$\text{Costo}(\sigma_{\text{Cilindrata} > 1251 \text{ AND } \text{Cilindrata} < 1500} (\text{Automobile} \triangleright \triangleleft \text{Conducente})) = B(Z) = 1584000$

$\text{Costo totale} = 1584000 + 210000 = 1794000$