

COMPITO DI SISTEMI INFORMATIVI/BASI DI DATI II

17 giugno 2008 – Compito A (Tot. 16) Tempo: 2h

Esercizio 1 (punti 3)

Si consideri il seguente log:

- | | |
|----------------------|----------------------|
| 1. U(T1,O1,B1,A1) | 16. C(T3) |
| 2. B(T2) | 17. I(T2,O6,A17) |
| 3. B(T3) | 18. U(T4,O3,B18,A18) |
| 4. I(T2,O2,A4) | 19. D(T2,O3,B19) |
| 5. D(T3,O1,B5) | 20. B(T6) |
| 6. U(T3,O2,B6,A6) | 21. U(T6,O2,B21,A21) |
| 7. C(T1) | 22. U(T6,O7,B22,A22) |
| 8. I(T3,O3,A8) | 23. D(T4,O5,B23) |
| 9. B(T4) | 24. C(T4) |
| 10. U(T4,O3,B10,A10) | 25. I(T6,O8,A25) |
| 11. CK(T2,T3,T4) | 26. U(T5,O8,B26,A26) |
| 12. U(T2,O4,B12,A12) | 27. D(T2,O8,B27) |
| 13. I(T3,O5,A13) | 28. C(T2) |
| 14. B(T5) | |
| 15. U(T5,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)

Dato il seguente schedule:

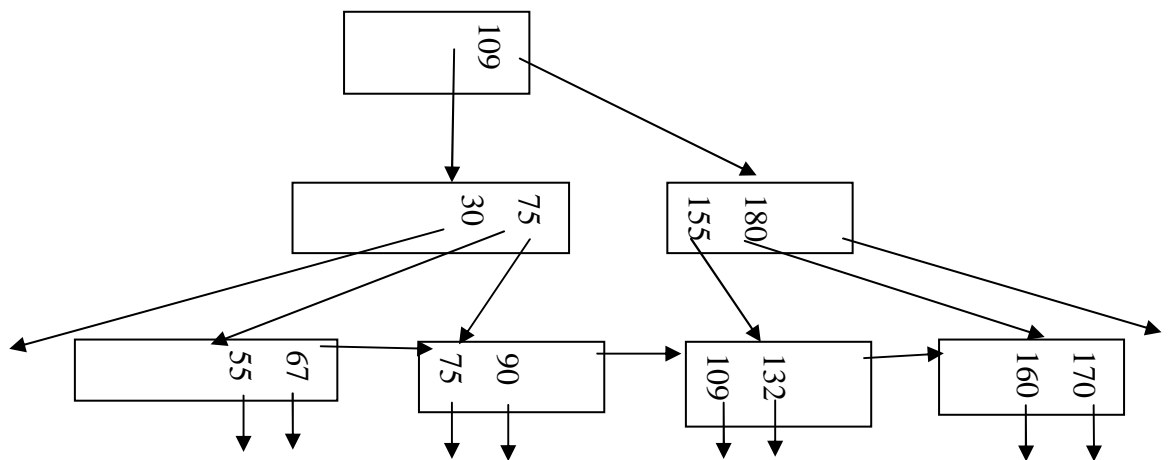
w1(x) r2(x) r3(y) r1(y) r2(w) r3(z) w1(w) w3(y) r1(x) w2(x) w1(w) r2(y)

si indichi se è view equivalente al seguente schedale seriale

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

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 75.

Esercizio 4 (punti 6)

Siano date le relazioni

Attore(CodiceAtt, Nome, Cognome, Età, Città, Studio)

Film(CodiceFilm, Titolo, Durata, Studio)

la query

SELECT A.*, F.*

FROM Attore AS A, Film AS F

WHERE A.Studio=F.Studio AND A.Città='Los Angeles'

e i parametri:

buffer di memoria centrale disponibili: $M=250.000$

dimensione del buffer: $B=4.000$ bytes

numero di tuple: $T(\text{Attore})=60.000.000$, $T(\text{Film})=500.000$,

dimensione delle tuple: $S(\text{Attore})=2.000$ bytes, $S(\text{Film})=4.000$,

$S(\text{Studio})=100$

numero di valori $V(\text{Attore}, \text{Città})=510$, $V(\text{Attore}, \text{Studio})=2.000$ $V(\text{Film}, \text{Studio})=2.000$

Si calcoli il costo minimo di ciascuna delle seguenti sequenze:

$\sigma_{\text{Città}='Los Angeles'}(\text{Attore} \bowtie \text{Film})$

$(\sigma_{\text{Città}='Los Angeles'} \text{Attore}) \bowtie \text{Film}$

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 Attore.Città e un indice secondario su Film.Studio.

SOLUZIONE

Esercizio 1

1. U(T1,O1,B1,A1)
2. B(T2)
3. B(T3)
4. I(T2,O2,A4)
5. D(T3,O1,B5)
6. U(T3,O2,B6,A6)
7. C(T1)
8. I(T3,O3,A8)
9. B(T4)
10. U(T4,O3,B10,A10)
11. CK(T2,T3,T4)
12. U(T2,O4,B12,A12)
13. I(T3,O5,A13)
14. B(T5)
15. U(T5,O5,B15,A15)
16. C(T3)
17. I(T2,O6,A17)
18. U(T4,O3,B18,A18)
19. D(T2,O3,B19)
20. B(T6)
21. U(T6,O2,B21,A21)
22. U(T6,O7,B22,A22)
23. D(T4,O5,B23)
24. C(T4)
25. I(T6,O8,A25)
26. U(T5,O8,B26,A26)
27. D(T2,O8,B27)
28. C(T2)

- 11 UNDO={T2,T3,T4 } REDO={ }
- 14 UNDO={T2,T3,T4,T5} REDO={ }
- 16 UNDO={T2,T4,T5} REDO={T3}
- 20 UNDO={T2,T4,T5,T6} REDO={T3}
- 24 UNDO={ T2, T5,T6} REDO={T3,T4}
- 28 UNDO={ T5,T6} REDO={T2,T3,T4}
- UNDO
- 26 O8=B26
- 25 D(O8)
- 22 O7=B22
- 21 O2=B21
- 15 O5=B15
- REDO
- 4 I(O2,A4)
- 5 D(O1)
- 6 O2=A6
- 8 I(O3,A8)
- 10 O3=A10
- 12 O4=A12
- 13 I(O5,A13)
- 17 I(O6,A17)
- 18 O3=A18
- 19 D(O3)
- 23 D(O5)
- 27 D(O8)

Esercizio 2

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

"legge-da"={ (r2(x),w1(x)),(r2(y),w3(y)) }

scritture finali={ (t2,x),(t3,y),(t1,w) }

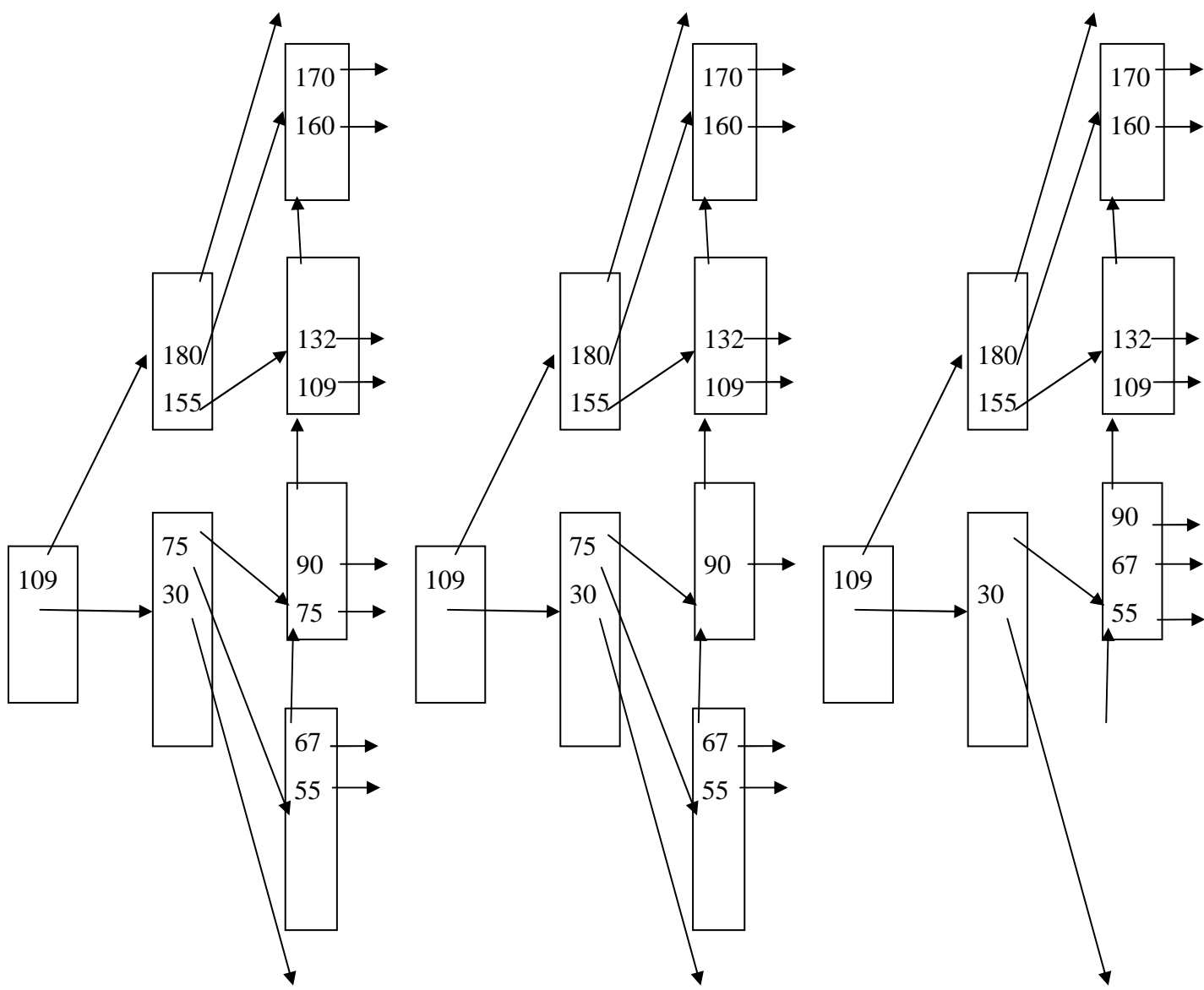
w1(x) r1(y) w1(w) r1(x) w1(w) r2(x) r2(w) w2(x) r2(y) r3(y) r3(z) w3(y)

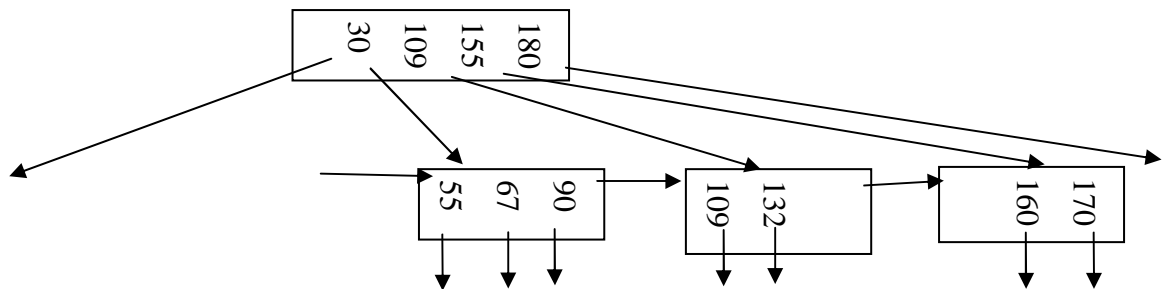
"legge-da"={ (r2(x),w1(x)),(r2(w),r1(w)) }

scritture finali={ (t2,x),(t3,y), (t1,w) }

I due schedule non sono view-equivalenti

Esercizio 3





Esercizio 4

buffer di memoria centrale disponibili: $M=250.000$

dimensione del buffer: $B=4.000$ bytes

numero di tuple: $T(\text{Attore})= 60.000.000$, $T(\text{Film})= 500.000$,

dimensione delle tuple: $S(\text{Attore})=2000$ bytes, $S(\text{Film})=4000$,

$S(\text{Studio})=100$

numero di valori $V(\text{Attore}, \text{Città})=510$, $V(\text{Attore}, \text{Studio})=2000$ $V(\text{Film}, \text{Studio})=2000$

Si calcoli il costo minimo di ciascuna delle seguenti sequenze:

$\sigma_{\text{Città}='Los Angeles'}(\text{Attore} \bowtie \text{Film})$

$(\sigma_{\text{Città}='Los Angeles'} \text{Attore}) \bowtie \text{Film}$

Calcoliamo innanzitutto il numero di blocchi occupati da ciascuna relazione:

$B(\text{Attore})=T(\text{Attore}) \cdot S(\text{Attore})/B = 60000000 \cdot 2000/4000 = 30.000.000$

$B(\text{Film}) = 500000 \cdot 4000/4000 = 500.000$

Prima sequenza: $\sigma_{\text{Città}='Los Angeles'}(\text{Attore} \bowtie \text{Film})$

Sia $Z = \text{Attore} \bowtie \text{Film}$

Né Attore né Film 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 secondario su Film. Studio.

$\text{CostoJNL}(Z)=B(\text{Film})+B(\text{Attore}) \cdot B(\text{Film})/M=500000+30000000 \cdot 500000/250000= 60.500.000$

Hash join ibrido: si può fare perché $M > \sqrt{B(\text{Film})} \Rightarrow 250000 > \sqrt{500000} \Rightarrow 250000 > 707.1$

$\text{CostoHJI}(Z)=(3-2M/B(\text{Film}))(B(\text{Film})+B(\text{Attore})) = (3-2 \cdot 250000/500000) \cdot (500000+30000000) = 61.000.000$

$\text{CostoIJ}(Z)=B(\text{Attore})+T(\text{Attore}) \lceil T(\text{Film})/V(\text{Film}, \text{Studio}) \rceil = 30000000+60000000 \cdot 500000/2000=15.030.000.000$

$T(Z)=T(\text{Attore}) \cdot T(\text{Film})/\max\{V(\text{Attore}, \text{Studio}), V(\text{Film}, \text{Studio})\} =$

$60000000 \cdot 500000/2000=15.000.000.000$

$S(Z)=S(\text{Attore})+S(\text{Film})-S(\text{Studio})= 2000+4000-100=5900$

$B(Z) = 15000000000 \cdot 5900/4000=22.125.000.000$

Dato che non ci sono indici su Z:

Costo: $(\sigma_{\text{Città}='Los Angeles'}Z)=B(Z)=22.125.000.000$

Costo totale= $60500000+22125000000=22.185.500.000$

Seconda sequenza: $(\sigma_{\text{Città}='Los Angeles'}\text{Attore}) \triangleright \triangleleft \text{Film}$

Sia $X = \sigma_{\text{Città}='Los Angeles'}\text{Attore}$

C'è un indice primario su Attore.Città

Costo(X) = $3 + \lceil B(\text{Attore})/V(\text{Attore},\text{Città}) \rceil = 3 + \lceil 30000000/510 \rceil = 3 + \lceil 58823,53 \rceil = 58827$

$T(X) = 60000000/510 = 117647.1 \approx 117647$

$B(X) = T(X)*S(X)/B = 117647*2000/4000 = 58823,5 \approx 58.824$

X sta in memoria centrale, quindi si utilizza il join a un passo

Costo($X \triangleright \triangleleft \text{Film}$) = $B(\text{Film}) + B(X) = 500.000 + 58.824 = 558.824$

Costo totale = $58.827 + 558.824 = 617.651$