

Basi di dati II — Prova parziale — 28 marzo 2014 — Compito A

Tempo a disposizione: un'ora.

Cognome _____ Nome _____ Matricola _____

Domanda 1 (20%) Si consideri un disco con una singola faccia che contenga $T=20.000$ tracce e con una velocità di rotazione di $v=10.000$ giri al minuto. Ogni traccia contiene $N=240$ blocchi e ogni blocco contiene $L=4KB$. Rispondere alle seguenti domande mostrando formula e valore numerico numero (N.B. non servono calcolatrici, i risultati sono semplici, approssimare $1GB=1000MB=1.000.000KB$).

Qual è la capacità del disco (in GB)?

Qual è la massima velocità di trasferimento (in MB al secondo)?

Supponendo che il tempo di seek t_S sia di circa 4msec, qual è il tempo medio necessario per leggere un'intera traccia?

Domanda 2 (20%) Si consideri una relazione $R(CodiceImpiegato, Cognome, Nome, Mansione)$ con $L = 1.000.000$ ennuple, con fattore di blocco $f_R = 100$. Con riferimento alla ricerca di tutti gli impiegati con una certa mansione, indicare il costo dell'accesso sequenziale e di quello diretto con indice su *Mansione* (con profondità $p = 4$) nei due casi seguenti (mostrare formule e valori numerici):

1. campo selettivo ($m_1 = 100.000$ valori diversi per *Mansione*)

costo accesso sequenziale:

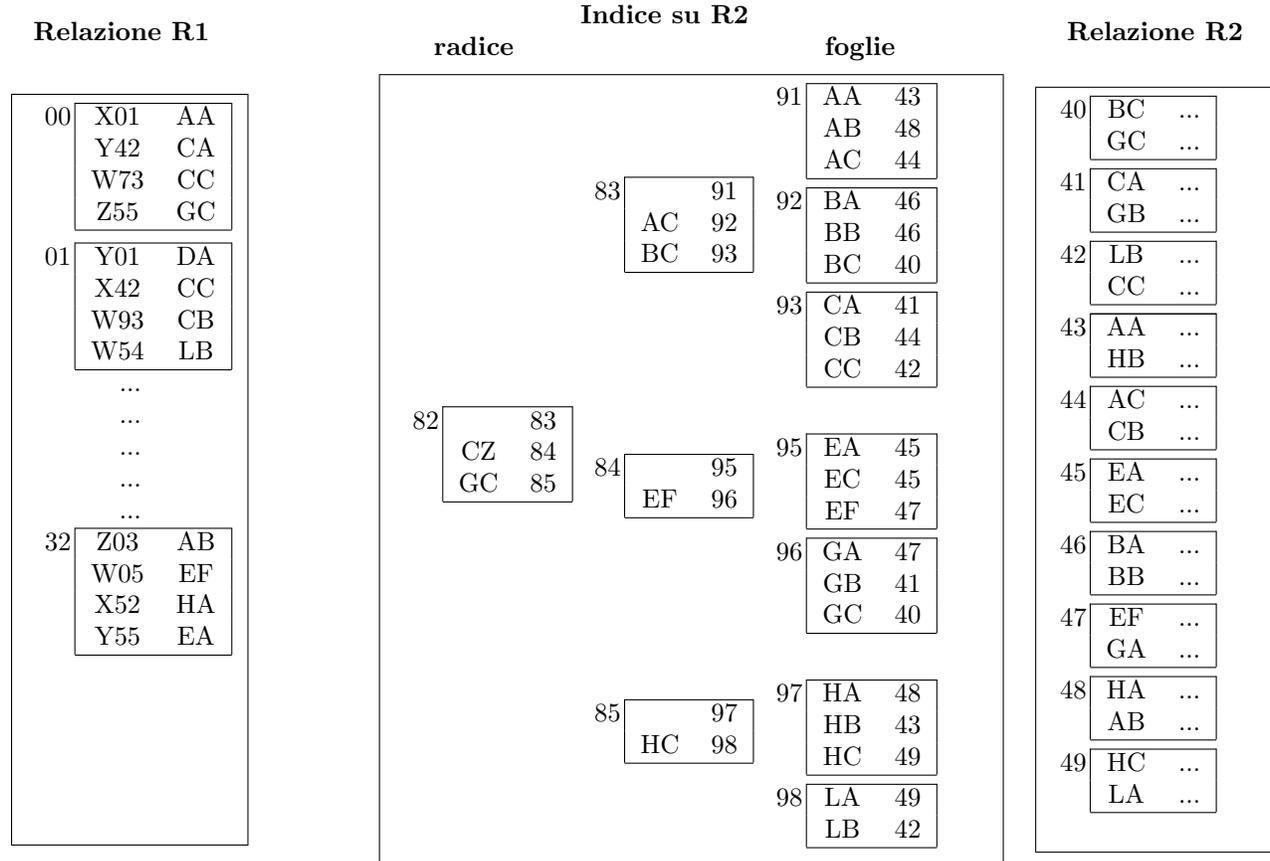
costo accesso diretto:

2. campo poco selettivo ($m_2 = 20$ valori diversi per *Mansione*)

costo accesso sequenziale:

costo accesso diretto:

Domanda 3 (40%) Considerare le relazioni R1 ed R2 e l'indice I2 su R2 schematizzati sotto. I riquadri interni indicano i blocchi e il numero a fianco a ciascun riquadro indica l'indirizzo del blocco. Nell'indice, i valori numerici sono riferimenti ai blocchi (blocchi dell'indice, per la radice e il livello intermedio, e blocchi di R2 per le foglie).



Si supponga di disporre di un buffer di otto pagine, con una politica di rimpiazzo di tipo LRU (*least recently used*, che, come noto, è di solito la migliore).

Considerare l'esecuzione del join di R1 ed R2, sulla base dei valori del secondo attributo di R1 e del primo di R2, con il metodo nested loop con accesso diretto tramite l'indice di R2.

Indicare gli indirizzi dei blocchi su cui si eseguono operazioni di pin (o fix) per produrre le prime due ennuple del risultato.

Indicare gli indirizzi dei blocchi effettivamente letti da memoria secondaria e caricati nel buffer (nell'ordine) per produrre le prime due ennuple del risultato.

Indicare gli indirizzi dei blocchi che si può presumere si trovino nel buffer al termine dell'esecuzione del join.

Basi di dati II — 28 marzo 2014 — Compito A

Domanda 4 (20%)

Si consideri un B+-tree con nodi intermedi che contengono tre chiavi e quattro puntatori e foglie con tre chiavi, in cui vengano inserite chiavi (a partire dall'albero vuoto) nel seguente ordine: 13, 17, 32, 41, 57, 61, 71, 89, 21, 22. Mostrare l'albero dopo l'inserimento di cinque chiavi, di otto chiavi e alla fine. (Si ricorda che nel B+-tree ogni chiave compare in una foglia)



Mostrare poi l'albero dopo l'eliminazione della chiave 57 dall'ultimo albero ottenuto in risposta alla domanda precedente.



Basi di dati II — Prova parziale — 28 marzo 2014 — Compito A

Cenni sulle soluzioni

(solo Compito A, le varianti del testo sono in rosso)

Tempo a disposizione: un'ora.

Cognome _____ Nome _____ Matricola _____

Domanda 1 (20%) Si consideri un disco con una singola faccia che contenga $T=20.000$ tracce e con una velocità di rotazione di $v=10.000$ giri al minuto. Ogni traccia contiene $N=240$ blocchi e ogni blocco contiene $L=4\text{KB}$. Rispondere alle seguenti domande mostrando formula e valore numerico numero (N.B. non servono calcolatrici, i risultati sono semplici, approssimare $1\text{GB}=1000\text{MB}=1.000.000\text{KB}$).

Qual è la capacità del disco (in GB)?

$$T \times N \times L = 20.000 \times 240 \times 4\text{KB} = 19.200.000\text{KB} = 19,2\text{GB}$$

Qual è la massima velocità di trasferimento (in MB al secondo)?

$$v/60 \times N \times L = 10.000/60 \times 240 \times 4\text{KB} = 160.000\text{KB}/\text{sec} = 160\text{MB}/\text{sec}$$

Supponendo che il tempo di seek t_S sia di circa 4msec, qual è il tempo medio necessario per leggere un'intera traccia?

$$t_S + 1/v = 4 + 60/10.000 \times 10^3 = 10 \text{ msec}$$

Domanda 2 (20%) Si consideri una relazione $R(\text{CodiceImpiegato}, \text{Cognome}, \text{Nome}, \text{Mansione})$ con $L = 1.000.000$ ennuple, con fattore di blocco $f_R = 100$. Con riferimento alla ricerca di tutti gli impiegati con una certa mansione, indicare il costo dell'accesso sequenziale e di quello diretto con indice su *Mansione* (con profondità $p = 4$) nei due casi seguenti (mostrare formule e valori numerici):

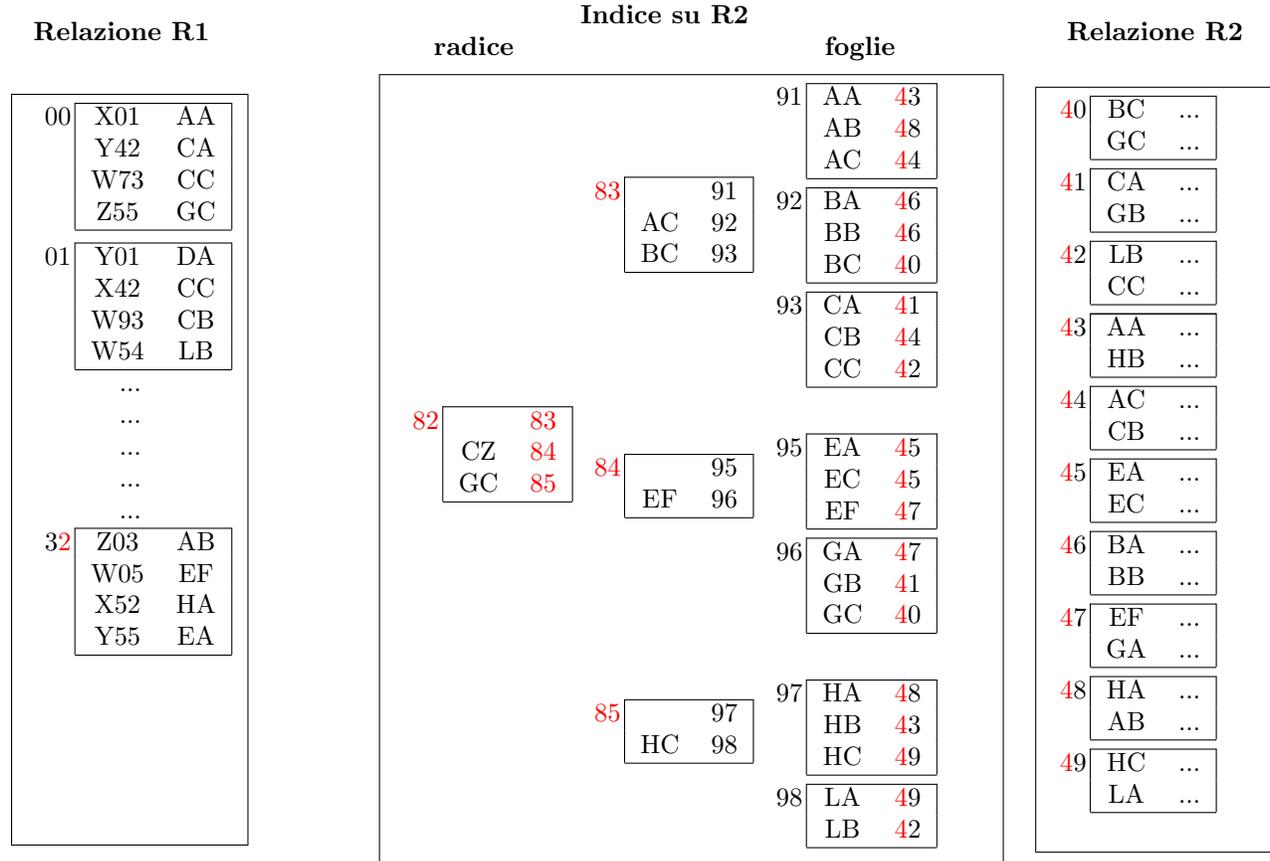
1. campo selettivo ($m_1 = 100.000$ valori diversi per *Mansione*)

costo accesso sequenziale: $L/f_R = 10.000$	costo accesso diretto: $p + L/v_1 \approx 14$
---	---

2. campo poco selettivo ($m_2 = 20$ valori diversi per *Mansione*)

costo accesso sequenziale: $L/f_R = 10.000$	costo accesso diretto: $p + L/v_2 + \dots \approx 50.000$
---	---

Domanda 3 (40%) Considerare le relazioni R1 ed R2 e l'indice I2 su R2 schematizzati sotto. I riquadri interni indicano i blocchi e il numero a fianco a ciascun riquadro indica l'indirizzo del blocco. Nell'indice, i valori numerici sono riferimenti ai blocchi (blocchi dell'indice, per la radice e il livello intermedio, e blocchi di R2 per le foglie).



Si supponga di disporre di un buffer di otto pagine, con una politica di rimpiazzo di tipo LRU (*least recently used*, che, come noto, è di solito la migliore).

Considerare l'esecuzione del join di R1 ed R2, sulla base dei valori del secondo attributo di R1 e del primo di R2, con il metodo nested loop con accesso diretto tramite l'indice di R2.

Indicare gli indirizzi dei blocchi su cui si eseguono operazioni di pin (o fix) per produrre le prime due ennuple del risultato.

00, 82, 83, 91, 43, 82, 83, 93, 41

Indicare gli indirizzi dei blocchi effettivamente letti da memoria secondaria e caricati nel buffer (nell'ordine) per produrre le prime due ennuple del risultato.

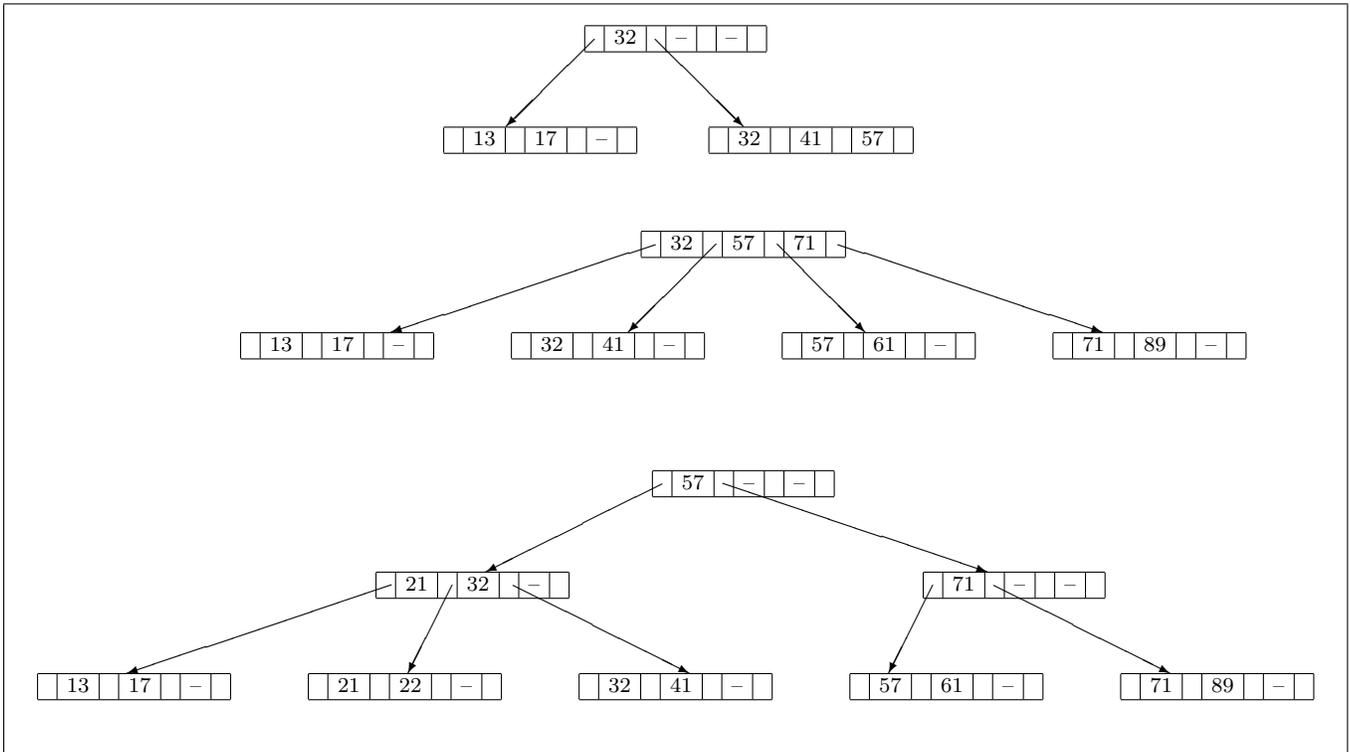
00, 82, 83, 91, 43, 93, 41

Indicare gli indirizzi dei blocchi che si può presumere si trovino nel buffer al termine dell'esecuzione del join.

85, 97, 48, 32, 82, 84, 95, 45

Domanda 4 (20%)

Si consideri un B+-tree con nodi intermedi che contengono tre chiavi e quattro puntatori e foglie con tre chiavi, in cui vengano inserite chiavi (a partire dall'albero vuoto) nel seguente ordine: 13, 17, 32, 41, 57, 61, 71, 89, 21, 22. Mostrare l'albero dopo l'inserimento di cinque chiavi, di otto chiavi e alla fine. (Si ricorda che nel B+-tree ogni chiave compare in una foglia)



Mostrare poi l'albero dopo l'eliminazione della chiave 57 dall'ultimo albero ottenuto in risposta alla domanda precedente.

