

UNIVERSITÀ DEGLI STUDI DELL'INSUBRIA DIPARTIMENTO DI SCIENZE TEORICHE E APPLICATE

Basi di Dati II

28 Giugno 2023 -- Pag 1/2 (Tempo a disposizione totale: 2 ore)

Quiz 1 (pt 1) (in giallo è evidenziata la risposta corretta)

Si consideri il seguente schema relazionale:

Studenti(matricola, nomeS, cittaS, annoNascita)

Esami(codiceE, nomeE, CorsoLaurea)

Voti(esame^{esame}, studente^{studenti}, voto, data)

Indicare quale di queste espressioni restituisce le informazioni sugli esami del corso di laurea di informatica sostenuti con voto superiore a 25 da almeno uno studente nato dopo il 2000

[1.a] $(\sigma_{CorsoLaurea="Informatica"}(Esami)) \cap ((\sigma_{voto>25}(Voti)) \bowtie_{studente=matricola}(\sigma_{annoNascita>2000}(Studenti)))$ [1.b] $\pi_{nomeE,codiceE}((\sigma_{CorsoLaurea="Informatica"}(Esami)) \bowtie_{CodiceE=Esame}((\sigma_{voto>25}(Voti)) \bowtie_{studente=matricola}(\sigma_{voto>25}(Voti)))$

annoNascita>2000 (Studenti))))

[1.c] $\pi_{\text{nomeE}}(\text{Voti} \bowtie_{\text{studente=matricola}}(\sigma_{\text{annoNascita>2000}}(\text{Studenti}))) \cap \pi_{\text{nomeE}}(((\sigma_{\text{voto>25}}(\text{Voti})) \bowtie_{\text{esame=codiceE}}(\sigma_{\text{corsoLaurea="Informatica"}}(\text{Esami})))$

[1.d] $\pi_{\text{matricola}}(\sigma_{\text{annoNascita}}) \cap \pi_{\text{studente}}((\sigma_{\text{voto}})) \bowtie_{\text{esame}}(\sigma_{\text{corsoLaurea}}))$

Quiz 2 (pt 1.5) (in giallo è evidenziata la risposta corretta)

Si consideri il seguente file di log (visualizzato in due colonne per ridurre lo spazio)

- 1. B(T5)
- 2. B(T6)
- 3. U(T6, O1, B1, A1)
- 4. I(T5, O2, A2)
- 5. B(T7)
- 6. C(T5)
- 7. B(T8)
- 8. U(T6, O2, B3, A3)
- 9. U(T7, O2, B4, A4)

- 10. U(T8, O3, B5, A5)
- 11. C(T8)
- 12. CK(T6,T7)
- 13. U(T7, O5, B6, A6)
- 14. U(T6, O5, B7, A7)
- 15. A(T7)
- 16. B(T9)
- 17. U(T9, O2, B8, A8)
- 18. C(T6)

Ipotizzando un failure all'instante 19, dopo una ripresa a caldo che valori hanno gli oggetti O2 e O5

[2.a] O2= B8, O5= B6

[2.b] O2= B4, O5= B6

[2.c] O2= A3, O5= A7

[2.d] O2= B3, O5= B7

Quiz 3 (pt 1.5) (in giallo è evidenziata la risposta corretta)

Si supponga di costruire un B⁺-tree su un campo di ricerca V che occupa 10 byte, usando blocchi di dimensione 1000 byte, con puntatore al blocco che occupa 5 byte e puntatori ai record che occupano 4 byte. Qual è l'ordine massimo di un nodo interno del B⁺-tree?

[3.a] 66

[3.b] 45

[3.c] 71

[3.d] 70

[3.e] nessuno dei valori indicati

ESERCIZIO 1 (pt. 7)

Si consideri il seguente schema di base di dati per la gestione di conferenze CONFERENZE(<u>titolo</u>, Anno, Città, Nazione)

PERSONE(<u>CF</u>, Nome, Cognome, email, AnnoNascita)

ORGANIZZA(titoloConf^{conferenze}, persona^{persone}, funzione)

- 1.1 Scrivere un'espressione algebrica che restituisce il nome e cognome delle persone nate prima del 2000 e che hanno partecipato all'organizzazione di conferenze nel 2020 con la funzione di "Program chair"
- 1.2 Per l'espressione ottenuta al punto 1, disegnare il query tree (albero dell'interrogazione) ottimizzato
- 1.3 Indicare quante operazioni I/O richiede l'esecuzione ottimizzata di un join tra ORGANIZZA e PERSONE con l'algoritmo merge sort ipotizzando che:
 - i due file non sono ordinati fisicamente
 - Il buffer mette a disposizione 10 blocchi
 - la tabella PERSONE occupa 1000 blocchi
 - la tabella ORGANIZZA occupa 100 blocchi

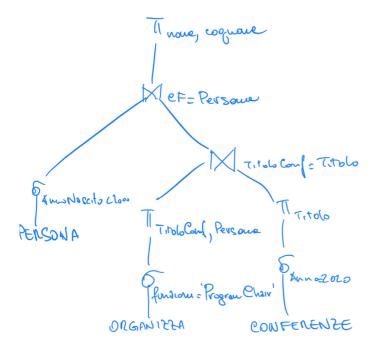
SOLUZIONE ESERCIZIO 1

1.1 Una possibile espressione algebrica è la seguente:

 $\pi_{\text{nome,cognome}}(\sigma_{\text{Anno-Nascita} < 2000} \text{ (Persone)} \bowtie_{\text{CF=persona}} (\sigma_{\text{Anno-2020 AND funzione='Program Chair'}}(\text{ORGANIZZA} \bowtie_{\text{titoloConf=titolo}} \text{CONFERENZE})))$

1.2 Il query tree associato all'espressione definita nel punto 1.1 è il seguente

La definire la versione ottimizzata si sfrutta la proprietà commutativa del σ , si sposta ogni operatore σ quanto più possibile verso il basso del query tree. Sfruttando la proprietà commutativa del π , si sposta l'operatore π quanto più possibile verso il basso dell'albero di interrogazione, mantenendo solo gli attributi necessari per eseguire le operazioni successive. Il query ottimizzato è il seguente:



1.3 I due file non sono ordinati, per poter applicare il join merge sort è necessario prima ordinarli entrambi.

Il costo dell'ordinamento con l'algoritmo di ordinamento esterno merge sort $2N * (1 + \lceil \log_{B-1} \lceil N/B \rceil))$

Nel caso dell'ordinamento del file che contiene la tabella PERSONA questo è

$$2*1000*(1+[log_{9}[1000/10]]) = 2000*(1+[log_{9}100]) = 2000*(1+3) = 8000$$

Nel caso dell'ordinamento del file che contiene la tabella ORGANIZZA questo è $2*100*(1+\lceil\log_9\lceil 100/10\rceil])=200*(1+\lceil\log_910\rceil)=200*(1+2)=600$

Il costo del merge sort esterno è dato dal costo di ordinamento dei file, più i costi di lettura di tutti i blocchi dei due file, quindi:

8000 + 600 + 1000 + 1000 = 9700 I/O

Esercizio 2 (pt 6.5)

Considerare un file che occupa 2.500.000 blocchi nella memoria secondaria. Si ipotizzi l'esecuzione di un algoritmo di merge sort esterno generico con 12 blocchi di buffer disponibili. Rispondere alle seguenti domande:

- 2.1) Quanti passaggi saranno necessari per ordinare completamente il file?
- 2.2) Qual è il costo totale di I/O dell'ordinamento del file?
- 2.3) Di quanti blocchi di buffer si ha bisogno per ordinare completamente il file in due soli passaggi?

SOLUZIONE ESERCIZIO 2

2.1

L'algoritmo di merge sort esterno richiede un passo iniziale (Passo 0) per il sort iniziale. In questo passo si ottengono [2.500.000/12] = 208.334 run.

Seguono un numero di passi di merge sui run ottenuti al passo 0. Il numero di merge con 12 buffer è pari a $\lceil \log_{11} \lceil 208.334 \rceil \rceil = 6$.

Il numero di passi totale è quindi 1+6=7

2.2

Il costo totale è 2N* numero passi, dove N è il numero di blocchi da ordinare. Quindi 2.500.000 *7= 17.500.000

2.3

L'ordinamento in solo due passi implica che viene eseguita un solo passo di merge. Sia X il numero di blocchi di buffer, per poter eseguire un solo merge su 2.500.000

```
Log <sub>x-1</sub> (2.500.000/x)=1
(x-1)<sup>1</sup>= (2.500.000/x)
\sim x<sup>2</sup>= 2.500.000 x=1582
```

Esercizio 3 (pt. 7)

Si consideri la relazione R(A,B,C....) con la seguente configurazione

- T(R)= 100.000.000 tuple, ognuna di 40 byte
- A occupa 4 byte
- C occupa 8 byte ed è unique
- B=4000, il blocco B ha dimensione 4000 byte
- P, puntatore al blocco, occupa 6 byte
- Pr, puntatore al record, occupa 7 byte
- 3.1) Assumendo che su A è costruito un indice B⁺-tree con fattore di riempimento al 70%, indicare la stima di costo della seguente operazione: SELECT * FROM R WHERE A = 1000
- 3.2) Assumendo che su C è costruito un indice secondario, indicare la stima di costo della seguente operazione: E = 1000
- 3.3) Assumendo che su C è costruito un indice multilivello, indicare la stima di costo della seguente operazione: SELECT * FROM R WHERE C = 1000

SOLUZIONE ESERCIZIO 3

3.1

La query SELECT * FROM R WHERE A = 1000 richiede di attraversare il B+-tree costruito su A per trovare il puntatore al record avente A=1000, più un accesso al blocco che contiene il record.

Per avere una stima sul costo di ricerca su B⁺-tree devo calcolare il numero di livelli. Calcolo prima l'ordina massimo per i nodi interni e nodi foglia.

Ordine massimo nei nodi interni

Pp + A(p-1) <= 4000 6p+4(p-1) <= 4000 6p+4p-1<= 4000 10p<= 4001

p_{interno}=400

Ordine massimo nei nodi foglia p*A + pPr + P <=4000 4p+ 7p+ 6<=4000 11p +6<=4000 11p <= 3994

p_{foglia}=363

Con fattore di riempimento al 70%, l'ordine massimo diventa

p_{interno}= 400 *0.70= 280 p_{foglia}= 363*0.70= 254

Il numero di livelli per ospitare 100.000.000 tuple è calcolato

Livello	#nodi	#valori di chiavi	#puntatori record
Root	1	(280-1)=279	
1 liv.	280	280 x 279= 78.120	
2 liv.	280 ²	$280^2 \times 279 = 21.873.600$	
3 liv./foglie	2803	$280^3 \times 254 = 5.575.808.000$	5.575.808.000

Quindi ricerca un record in questo B+-tree richiede 4 accessi I/O. Il costo della query è quindi 4+1= 5 I/O

3.2

La query SELECT * FROM R WHERE C = 1000 richiede di attraversare l'indice secondario costruito su C per trovare il puntatore al record avente C=1000, più un accesso al blocco che contiene il record.

Per avere una stima sul costo di ricerca sull'indice secondario, devo calcolare il numero di blocchi contenenti le voci dell'indice su cui fare la ricerca binaria.

Il file non è ordinato sul C, quindi l'indice è denso. E' necessaria una voce per ogni record, quindi 100.000.000 voci.

Ogni voce richiede 14 byte, 8 per il campo C e 6 per il puntatore al record (N.b. si poteva anche usare il puntatore al record)

In un blocco da 4000 byte si possono memorizzare [4000/14]=285 voci

L'indice secondario occupa, quindi, [100.000.000/285]= 350.878 blocchi

La ricerca sull'indice secondario è una ricerca binaria su 350.878 blocchi, richiede \log_2 (350.878)= 19 operazioni I/O

Il costo della query è quindi 19 + 1 = 20 operazioni I/O.

3.3

Per stimare il costo della query devo calcolare di quanti livelli è costituito l'indice multilivello su C. Il primo livello è l'indice secondario calcolato nel punto 3.2.

Livello 2 - Il secondo livello è un indice sparso sull'indice primo livello, con una voce per blocco dell'indice del primo livello. 350.878 voci richiedono un numero di blocchi pari a [350.878/285] = 1232 Livello 3 – come per il precedente. 1232 voci richiedono un numero di blocchi pari a [1232/285] = 5 Livello 4 – 5 voci richiede un solo blocco.

Per ricerca il record con C=1000 nell'indice multilivello è necessario leggere 4 blocchi (uno per livello). Il costo della query è quindi 4 (ricerca sull'indice) + 1 (lettura blocco dati), ovvero 5 operazioni I/O.

Esercizio 4 (pt. 6.5)

Indicare quale dei seguenti schedule è CSR, VSR o non serializzabili

- 4.1) r3(z) w1(y) w3(z) r2(z) r2(y) w3(y) w2(y)
- 4.2) r1(a) w2(a) r3(a) r1(b) r4(c) w3(d) r4(d) w4(b) w5(b) w5(c) w2(b) r1(d)
- 4.3) r1(x) w2(x) r3(x) r1(y) w3(v) w2(y) r1(v)

SOLUZIONE ESERCIZIO 4.1

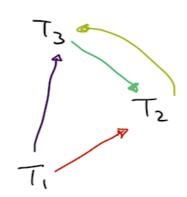
1 CONSIDERS I CONFLIT, SULLE RISONSE

8= 13 w3 12 4= w, 12 w3 w2

CONFUITA:

ω₃ Γ₂
ω, Γ₂
ω, ω₃
ω, ω₂
Γ₂
ω₃
ω, ω₂

(2) COSTRUISCO IL GRAFO DEI CONFLITA



E' CICLICO, QUINDI SINON E' CBR

E' USR?

- (W₂(2), 1₂(2))
 (W₁(4), 1₂(M))
- Per y: W2 (y)

 Per Z: W3 (2)
- (3) CONSIDERS I POSSIBILI SCHEDUE SERVALLY $T_1 = W_1(y)$ $T_2 = \Gamma_2(z) \Gamma_2(y) \ \omega_2(y)$ $T_3 = \Gamma_3(z) \ \omega_3(z) \ \omega_3(y)$

PERL ESSERE USR CON SI, LO SCHEDIE DEUE AJERE LE STESSE RELATIONI LEGGI-DA C SUNTIURE FINALI

LE SCRITURE FINAL' DI SI ESCUPDONO
GLI SCHEDUE SERVALI DOUE TO E ESEGUL
TA RER ULTIMA, PERCHÉ PONTEREBRE AD
BANA SCRITURA FINALE DI M, W, (4), DUERSA
DA QUELLA DI SI

LO STESSO RAGIONAMENTO ESCLUDE

TO COME ULTIMA THANSARIONE NEL SCHEDNE

SERIALE, PERCHE *UREMMO ULTIMA SCRITUM

SU y DATA DA ULZ(2), MENTIRE IN SI

E' Di Tz (Wz(2))

GIR SCHEDULE SERIALL' CHE POTREBBERD ESSERE USR equivalent' A SA DEJOHO AVERE TO FINALE

SIZZ TZ TZ TZ Z FINTRAMBI HANNO LE STESSE ULTIME SCRITTURE
Mi SZ

CONTROLLO LEGGI DA

I'M NON E' USIL con 51.

S12=

$$(\omega_3(z), r_2(z))$$
 $(\omega_1(y), r_2(y))$
REVATION: LEGGI-DA

SIZE' USAL EQUIVALENTE A S2

SI & USR

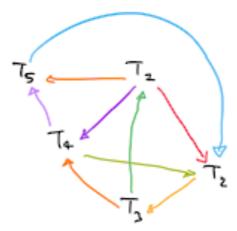
SOLUZIONE ESERCIZIO 4.2

52=1, (a) w2 (a) 3 (a) 12 (b) 14 (c) w3 (d) 4 (d) w4 (b) w5 (c) w2 (b) 12 (d)

E' CSR?

CONFLITTI





So Now E' CSR.

E' USE!

1 RELATION! LEGGI-DA

1 ULTIME SOUTHURE

(3) PER TROVARE UND SCHEDULE SERIALE
USIL E QUI VALENTE A S, (4.2) DENO
COMSIDERANE GLI SCHEDULE SERIALI
CHE SI POSSONO COMPORRE DA QUESTE
5 TRANSATIONI:

Ti = r2 (x) r2 (b) r2 (d)

Tzz Wz(a) w2 (b)

T3= 13 (x) w3 (d)

T4= r4 (e) r4 (d) w4 (b)

T5= w5 (b) w5(e)

PRIMA DI CONSIDERADE TUTI I CASI
POSSIBILI POSSO ESCLUDENE DELLE
COMBINAZIONI CHE DANNO SCHEDVIE
SERIALI MON VIEW EQUIVAIGNTI,
DOUGNO CON RELAZIONI LECGI-DA
DIVERSE E(O SCRITIVRE FINALI DIUERSÈ.

DSSERVAZIONE 1

CONSIDERO LA SCRITTURA FINALE SUL...

POSSO DEBURNE CHE TO DEUE ESSENE

ESEGUITO DOPO TITE, INFATTI ENTRANGI

DUESTE DUE TRANSATION' SCRIUDIND SUL
L'OGGETTO D.

SE ANCHE SOLO UNA DI LORD USIN'SSE

ESEGUITA DOPO TO ADREMINO UND SCHEDILE

SERVALE CON UNA SCRITTURA DUENDA
SU D, DUINDI NON UTEUL EDUINHUENTE

A Se
QUINDI DEUO CONSIDEMAR SCRIEDILE

SERVALL' DOJE

TA cTs PNECEDONO TZ

OSSERNATIONE 2

CONSIDERO BUESTA RELAZIONE LEGGI-DA

(W2(a), (3(a))

PER AVENE LA STESSA RECARIONE DEVE AVENE TZ PRIMA DI T3.

QUINDI

To ets PRECEDONS TZ PRECEDE T3.

E GHCIZALUBEZO

DA QUESTA REUDIONE LEGGI- DA

(w2 (d), (4 (d))

TO DECE ESSERE PRIMAR TA

QUIND' ABBIAND CHE

OSSERVAZIONI 1.2 IMPLICAND

To ets precessons Tz, precess Tz.

OSSERVAZIONE 3 IMPLICA

T3 PRECEDE TO

NOW PUD ESISTERE UND SCHEDULE

CHE BODDISTA LA CONDIZIONE DATA

DALLE DESERVATIONI 1.2 E CONTEMPORA

NEMENTE LA CONDIZIONE UNPLICATA

DALL' OSSERVATIONE 3

QUININI POSSIAND CONCLUIDATE CHE

NON ESISTE UND SCHEDULE SERLALE

CHE HA LE STESSE LELAZIONI LEGGI-DA

DI SZ & LE STESSE SCRITURE

FINANI
SZ NDN E' USR-

SOLUZIONE ESERCIZIO 4.3

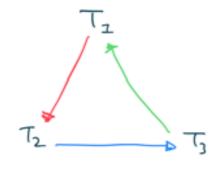
E' CSR?

1 CONSIDERO I CONFLITTI SULLE RISONSE

N= 12 W2 N= 12 W2

CONFLITTI

ω₂ ι₂ ω₂ ι₃ υ, ω₂



IL GRATED HA CICLI, S3 NON E' CBR.

E' USR?

- ($\omega_2(\kappa)$, $f_3(\kappa)$)
- Per y: Wz(y)

Per V: W, (v)

(3) PER TROUBLE OND SCHEDULE SERIALE USR EQUIVACENTE & S3 CONSIDERD LE POSSIBILI COMBINATION' OI QUESTE TRANSA7: ON1:

$$T_1 = r_2(x) r_1(y) r_1(y)$$
 $T_2 = w_2(x) w_2(y)$
 $T_3 = r_3(x) w_3(x)$

· T2 T2 T3

$$= \Gamma_{2}(x) \Gamma_{2}(y) \Gamma_{2}(y) \omega_{2}(x) \omega_{2}(y) \Gamma_{3}(x) \omega_{3}(y)$$

ULTIME SCRITTIONS

Per
$$Y = W_2(Y)$$
 = ALLE OLTIME
Per $Y = W_3(Y)$ SCRUTTURE DI

scrance ois

LEGGI- DA

· ななを

= (x) (x) (1 (y) (1 (x) (3 (x) W2 (x) W2 (x)

ULTIME SOUTHINE

Per
$$y = \omega_2(y) = ALLE OLTIME

Per $y = \omega_2(y) = ALLE OLTIME

SCRUTTURE Disg$$$

LEGGI-BA

€ T3 T1 T2

ULTIME SOUTHINE

€ T3 T2 T1

= (3 (x) W3 (v) W2 (x) W2 (y) (2 (x) 12 (y) (1 (v)

ULTIME SOUTHINE

Per
$$y = \omega_2(y) = ALLE OLTIME

Per $y = \omega_2(y) = ALLE OLTIME

SCRUTTURE Disg$$$

UEGG+DA .

$$(\omega_{3}(v), r_{2}(v))$$

 $(\omega_{2}(y), r_{1}(y)) \neq \text{DAUE RELAZION!}$
 $(\omega_{2}(y), r_{2}(y)) \neq \text{LEGGI DA DI S_{3}}$
 $(\omega_{2}(x), r_{3}(x))$

· T2 T1 73

ULTIME SOUTHINE

Per
$$y = \omega_2(y) = ALLE ULTIME
Per $y = \omega_2(y) = SCRUTURE Disgreen$$$

LEGGI-DA .

$$(\omega_{2}(x), r_{3}(x))$$
 $(\omega_{2}(y), r_{2}(y)) \neq \text{DAUE RELAZION!}$
 $(\omega_{2}(y), r_{2}(y)) \neq \text{LEGGI DA DI S_{3}}$

· 72 73 71

= W2 (x) W2 (y) 13 (x) W3 (x) 12 (x) 12 (y) 12 (x)

ULTIME SOUTHURE

Per
$$y = \omega_2(y) = ALLE OLTIME

Per $y = \omega_2(y) = SCRTTURE Dis$$$

UBGGI-DA:

$$(\omega_3(v), c_1(v))$$

 $(\omega_2(y), c_1(y)) \neq 0$
 $(\omega_2(x), c_1(x)) \neq 0$
 $(\omega_2(x), c_1(x)) \neq 0$
 $(\omega_2(x), c_1(x)) \neq 0$
 $(\omega_2(x), c_2(x))$

S3 MON E' USIL, PENCHE' MON ESISTOND SCHEDULE SELLALL VIEW EQUIVALENTI A S1.