

Basi di Dati: Strutture ed Algoritmi

Appelli del 2002

Appello del 10.1.2002

1. Si consideri la relazione:

Impiegati(Codice, Nome, AnnoNascita, Dipartimento)

Per le seguenti interrogazioni si dia un piano di accesso e si stimi il costo considerando due casi: il primo con l'interrogazione nel modo in cui è scritta e il secondo riscrivendo l'interrogazione in modo da trovare un piano migliore:

- a) Esiste un indice a B^+ -albero su AnnoNascita:

```
SELECT    Dipartimento
FROM      Impiegati
WHERE     AnnoNascita = 1960 OR AnnoNascita = 1970 ;
```

- b) Esiste un indice a B^+ -albero su AnnoNascita:

```
SELECT    Dipartimento
FROM      Impiegati
WHERE     AnnoNascita < 1970 AND AnnoNascita > 1968 ;
```

- c) Esiste un indice a B^+ -albero su AnnoNascita:

```
SELECT    Dipartimento
FROM      Impiegati
WHERE     2*AnnoNascita = 3920;
```

- d) Non esistono indici:

```
SELECT    Dipartimento, AVG(AnnoNascita)
FROM      Impiegati
GROUP BY  Dipartimento
HAVING    Dipartimento = 20;
```

- e) Si consideri anche la tabella

Dipartimenti(Numero, NomeD, Indirizzo, Telefono, Direttore)

in cui Direttore è una chiave esterna per Impiegati, con chiave primaria Codice, e possono esistere dipartimenti privi di direttori:

```
SELECT    Codice
FROM      Impiegati, Dipartimenti
WHERE     Codice = Direttore;
```

2. Si considerino le transazioni:

$$T_1 = \{r_1[A], w_1[A], r_1[B], w_1[B]\},$$

$$T_2 = \{r_2[B], r_2[A], w_2[A], w_2[B]\}$$

Si supponga che il serializzatore aggiunga le richieste di blocco esclusivo come segue:

$$T_1 = \{wl_1(A), r_1[A], w_1[A], wl_1(B), r_1[B], w_1[B], commit, wu_1(A), wu_1(B)\}$$

$$T_2 = \{wl_2(B), r_2[B], wl_2(A), r_2[A], w_2[A], w_2[B], commit, wu_2(A), wu_2(B)\}$$

Supponendo che le transazioni vengano eseguite concorrentemente, si risponda alle seguenti domande (giustificando le risposte):

- La storia prodotta è c-serializzabile?
- Può verificarsi uno stallo? Supponendo che parta prima T_1 , quale transazione riparte se si adotta il protocollo di prevenzione dello stallo *wait-die*?
- Può verificarsi una cascata di abort?

Si supponga che il serializzatore aggiunga le richieste di blocco esclusivo in questo altro modo:

$$T_1 = \{wl_1(A), r_1[A], w_1[A], wl_1(B), wu_1(A), r_1[B], w_1[B], wu_1(B)\}$$

$$T_2 = \{wl_2(A), wl_2(B), r_2[B], r_2[A], w_2[A], wu_2(A), w_2[B], wu_2[B]\}$$

- La storia prodotta è c-serializzabile?
- Può verificarsi uno stallo? Supponendo che parta prima T_1 , quale transazione riparte se si adotta il protocollo di prevenzione dello stallo *wound-wait*?
- Può verificarsi una cascata di abort?

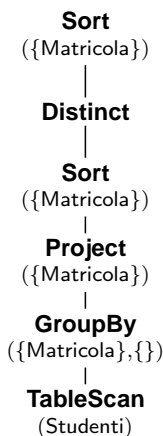
3. Spiegare come viene eseguito dal gestore del buffer il comando di lettura di una pagina. Quando il gestore del buffer scrive una pagina nella memoria permanente?

Appello del 14.2.2002

1. Si consideri la relazione `Studenti(Matricola, Nome, AnnoNascita)`, ordinata sulla chiave primaria `Matricola`, e l'interrogazione:

```
SELECT    DISTINCT Matricola, COUNT(*)
FROM      Studenti
WHERE     AnnoNascita = 1974
GROUP BY  Matricola
ORDER BY  Matricola;
```

Si dica se il seguente piano d'accesso produce il risultato cercato. Se non va bene, lo si modifichi aggiungendo le parti mancanti:



Si semplifichi il piano di accesso eliminando gli operatori inutili e si dia una stima del suo costo. Come cambia il piano di accesso e il suo costo se la relazione non è ordinata su `Matricola`?

2. Siano S_1 e S_2 due storie sullo stesso insieme di transazioni. Si supponga che S_1 sia c-serializzabile e che i grafi di serializzabilità delle due storie $GS(S_1)$ e $GS(S_2)$ siano uguali. Si considerino le seguenti affermazioni su S_2 :

- (a) S_2 è seriale
- (b) S_2 è c-equivalente a S_1
- (c) S_2 è c-serializzabile.

Giustificare quali delle seguenti conclusioni sono vere:

- (a) sono vere 1, 2 e 3; (b) solo 2 e 3 sono vere; (c) solo 2; (d) solo 3; (e) nessuna.

3. Come viene garantita l'atomicità e la persistenza delle transazioni dal gestore dell'affidabilità ?

Appello del 27.5.2002

1. Si consideri lo schema:

Fornitori(Fid, FNome, Città)
Parti(Pid, PNome, Colore, Peso)
Ordini(Fid*, Pid*, PGid*, Qta)

e l'interrogazione:

```
SELECT    DISTINCT FNome, sum(Qta)
FROM      Ordini O, Fornitori F
WHERE     O.Fid = F.Fid AND O.Pid < 'P5' AND F.Città = 'Pisa'
GROUP BY  FNome
ORDER BY  FNome
```

Si fornisca un piano di accesso usando indici a vostra scelta e si dia una stima del suo costo.

2. Si definiscano le nozioni di grafo di serializzabilità e di storia c-serializzabile. Si considerino le storie:

$$S_1 = \{r_1[A], w_1[B], r_2[B], w_2[C], r_3[C], w_3[A]\}$$

$$S_2 = \{r_1[A], r_2[A], w_1[B], w_2[B], r_1[B], r_2[B], w_2[C], w_1[D]\}$$

Si dica se sono c-serializzabili. Se lo sono, si dica quali sono le storie seriali equivalenti.

3. Spiegare come viene eseguito dal gestore del buffer il comando di lettura di una pagina. Quando il gestore del buffer scrive una pagina nella memoria permanente?

Appello del 18.6.2002

1. Si consideri lo schema $R(\underline{K}, A)$, $S(\underline{KE}, B)$, e l'interrogazione:

```
SELECT *
FROM R, S
WHERE K = KE;
```

Si dia il piano di accesso per eseguire l'interrogazione con il metodo MergeJoin senza uso di indici, oppure usando gli indici su K e KE , e si dia una stima del costo dei due piani. Spiegare in quali dei due casi seguenti non conviene usare gli indici:

- (a) R e S hanno un record per pagina,
- (b) R e S hanno molti record per pagina.

2. In figura sono mostrate le operazioni di una transazione con la notazione:

- START inizio transazione
- END fine transazione
- READ(X, t) copia il valore dell'elemento X della BD nella variabile t
- WRITE(X, t) copia il valore della variabile t nell'elemento X della BD che si trova nel buffer
- FLUSH(X) trasferisce X dal buffer al disco
- X Buf, valore di X nel buffer
- X disco, valore di X nel disco

Si completi la seguente tabella con le operazioni di FLUSH(X) e i valori delle colonne supponendo che la transazione sia gestita con l'algoritmo *Disfare-NonRifare*:

Azione	t	A Buf	B Buf	A disco	B disco	Log
START	–	–	–	8	8	(Begin T)
READ(A,t)	8	8	–	8	8	
$t = t*2$	16	8	–	8	8	
WRITE(A,t)						
READ(B,t)						
$t = t*2$						
WRITE(B,t)						
END						

3. Si definiscano le nozioni di grafo di serializzabilità e di grafo di attesa. Si dica come vengono usati nella realizzazione di un DBMS.

Appello del 15.7.2002

1. Si consideri la relazione Impiegati(Codice, Nome, AnnoNascita). Per le seguenti interrogazioni si dia un piano di accesso (supponendo di eliminare i duplicati e di eseguire il GroupBy con l'algoritmo di ordinamento) e si giustifichi in quali casi non occorre l'operatore fisico Sort. Si stimi il costo del piano di accesso che ritenete più costoso.

```
SELECT    DISTINCT AnnoNascita
FROM      Impiegati
WHERE     AnnoNascita > 1950;
```

```
SELECT    DISTINCT AnnoNascita
FROM      Impiegati
WHERE     AnnoNascita = 1950;
```

```
SELECT    AnnoNascita
FROM      Impiegati
WHERE     AnnoNascita = 1950
ORDER BY  AnnoNascita;
```

```
SELECT    COUNT(*)
FROM      Impiegati
WHERE     AnnoNascita = 1950
GROUP BY  AnnoNascita;
```

2. Si consideri la storia:

$$S = \{r_1[A], w_1[B], r_2[A], w_2[B], r_3[A], w_3[B]\}$$

Quante storie si possono definire con le stesse operazioni e c-equivalenti a S ?

3. Descrivere il protocollo 2PL stretto e la tecnica di gestione dei blocchi.

Appello dell'11.9.2002

1. Si definiscano le nozioni di grafo di serializzabilità e di grafo di attesa. Si dica come vengono usati nella realizzazione di un DBMS.
2. Si consideri la base di dati

Impiegati(CodiceI, Nome, Stipendio, Età, Dipartimento*)
Dipartimenti(CodiceD, Budget, Università, Direttore*)

Stipendio ha valori fra 10M e 100M, Età ha valori fra 20 e 80, Budget ha valori fra 10 e 100M, ogni dipartimento ha in media 30 impiegati e ci sono 30 Università. Si supponga una distribuzione uniforme dei valori.

Per ognuna delle seguenti interrogazioni, quali degli indici proposti definireste per agevolare l'esecuzione dell'interrogazione? Dare il piano di accesso e una stima del suo costo. Se il DBMS non considera la possibilità di usare solo indici per rispondere all'interrogazione (ovvero, proceda sempre con accessi agli indici e poi ai dati, pur avendo tutte le informazioni necessarie per la risposta nell'indice), come cambierebbe la vostra risposta?

- a) Trovare Nome, Età e Stipendio di tutti gli impiegati
 - a1) nessun indice
 - a2) indice a B^+ -albero su (Nome, Età, Stipendio)
 - a3) indice a B^+ -albero su (Nome, Età)
 - b) Trovare i codici dei dipartimenti dell'Università di Pisa con un budget inferiore a 30
 - b1) nessun indice
 - b2) indice di ordinamento a B^+ -albero su (Università, Budget)
 - b3) indice a B^+ -albero su (Università, Budget)
 - b4) indice di ordinamento a B^+ -albero su Budget
 - c) Trovare i nomi degli impiegati che dirigono dipartimenti ed hanno uno stipendio maggiore di 80M
 - c1) nessun indice
 - c2) indice a B^+ -albero su CodiceD
 - c3) indice a B^+ -albero su CodiceI
 - c4) indice di ordinamento a B^+ -albero su Stipendio
3. Dire quali delle seguenti affermazioni è vera o falsa e giustificate la risposta:
 - (a) Il costo di una ricerca per intervallo della chiave K con l'uso di un indice su K è sempre inferiore alla scansione del file.

- (b) Con il metodo rifare, tutte le modifiche di una T devono essere riportate nella BD prima che il record di commit sia scritto nel giornale.
- (c) Tutte le storie prodotte dal 2PL-stretto sono seriali.