

Corso di Basi di dati
Prova scritta parziale — 1 marzo 2001
Compito A

Tempo a disposizione: un'ora e trenta minuti. Libri e appunti chiusi.

Nota bene: rispondere in modo ordinato e il più possibile sintetico (nella misura in cui ciò non pregiudichi la comprensibilità).

Domanda 1 (35%)

Un'applicazione di una compagnia aerea prevede molte transazioni come la seguente, in cui x denota il numero di posti disponibili su un certo volo e y il numero di posti che vengono prenotati, su tale volo, con la transazione stessa:

$$\begin{aligned} & read(x) \\ & x := x - y \\ & write(x) \end{aligned}$$

Con riferimento a tale contesto:

1. Indicare quale tipo di anomalia potrebbe nascere, in assenza di controllo di concorrenza; mostrare uno schedule con tale anomalia.
2. Mostrare come il controllo di concorrenza basato su timestamp impedisca l'anomalia.
3. Mostrare come il 2PL impedisca l'anomalia, commentando anche i vantaggi e svantaggi che si avrebbero nei due casi seguenti
 - (a) vengono utilizzati inizialmente lock condivisi, che vengono poi "promossi" a lock esclusivi;
 - (b) vengono utilizzati sin dall'inizio lock esclusivi.

Domanda 2 (20%)

Illustrare brevemente una procedura di recovery (semplificata rispetto a quella usuale) utilizzabile nell'ipotesi che il protocollo di scrittura preveda che le scritture nella base di dati avvengano tutte dopo il commit. Spiegare perché questa opportunità, pur permettendo una procedura di recovery più semplice ed efficiente, non viene molto utilizzata in pratica.

Domanda 3 (35%)

Considerare un file corrispondente ad una relazione su vari attributi fra cui Matricola e Cognome, sotto le seguenti ipotesi:

- i record hanno lunghezza fissa pari a 110 byte e i campi Matricola e Cognome ne occupano rispettivamente 4 e 20
- il file contiene 10.000 record
- il file system utilizzato prevede blocchi che rendono disponibili alla struttura fisica 500 byte (si possono cioè ignorare i byte gestiti dal sistema)
- nel file system utilizzato, i puntatori occupano 4 byte (e supponiamo per semplicità che ciò sia vero sia per puntatori a blocchi sia per puntatori a singoli record)

Schematizzare, anche per mezzo di un disegno, una struttura fisica che preveda:

1. un indice primario statico su Cognome (e quindi una struttura ordinata su di esso)
2. un indice secondario (B-tree) su Matricola

Indicare il numero di blocchi approssimativamente necessari per ciascuna struttura (cioè per il file e per ciascuno degli indici).

Domanda 4 (10%)

Con riferimento alla domanda precedente, indicare come potrebbe essere realizzato un indice secondario hash su Matricola (utilizzando una funzione hash semplice, anche se non realistica, quale il resto della divisione per un numero facile da gestire, quale 50 o 100).

Corso di Basi di dati
Prova scritta parziale — 1 marzo 2001
Compito B

Tempo a disposizione: un'ora e trenta minuti. Libri e appunti chiusi.

Nota bene: rispondere in modo ordinato e il più possibile sintetico (nella misura in cui ciò non pregiudichi la comprensibilità).

Domanda 1 (35%)

Un'applicazione di una compagnia aerea prevede molte transazioni come la seguente, in cui y denota il numero di posti disponibili su un certo volo e z il numero di posti che vengono prenotati, su tale volo, con la transazione stessa:

$$\begin{aligned} & read(y) \\ & y := y - z \\ & write(y) \end{aligned}$$

Con riferimento a tale contesto:

1. Indicare quale tipo di anomalia potrebbe nascere, in assenza di controllo di concorrenza; mostrare uno schedule con tale anomalia.
2. Mostrare come il controllo di concorrenza basato su timestamp impedisca l'anomalia.
3. Mostrare come il 2PL impedisca l'anomalia, commentando anche i vantaggi e svantaggi che si avrebbero nei due casi seguenti
 - (a) vengono utilizzati inizialmente lock condivisi, che vengono poi "promossi" a lock esclusivi;
 - (b) vengono utilizzati sin dall'inizio lock esclusivi.

Domanda 2 (20%)

Illustrare brevemente una procedura di recovery (semplificata rispetto a quella usuale) utilizzabile nell'ipotesi che il protocollo di scrittura preveda che le scritture nella base di dati avvengano tutte prima del commit. Spiegare perché questa opportunità, pur permettendo una procedura di recovery più semplice ed efficiente, non viene molto utilizzata in pratica.

Domanda 3 (35%)

Considerare un file corrispondente ad una relazione su vari attributi fra cui *Matricola* e *CodiceFiscale*, sotto le seguenti ipotesi:

- i record hanno lunghezza fissa pari a 110 byte e i campi *Matricola* e *CodiceFiscale* ne occupano rispettivamente 4 e 16
- il file contiene 20.000 record
- il file system utilizzato prevede blocchi che rendono disponibili alla struttura fisica 500 byte (si possono cioè ignorare i byte gestiti dal sistema)
- nel file system utilizzato, i puntatori occupano 4 byte (e supponiamo per semplicità che ciò sia vero sia per puntatori a blocchi sia per puntatori a singoli record)

Schematizzare, anche per mezzo di un disegno, una struttura fisica che preveda:

1. un indice primario statico su *CodiceFiscale* (e quindi una struttura ordinata su di esso)
2. un indice secondario (B-tree) su *Matricola*

Indicare il numero di blocchi approssimativamente necessari per ciascuna struttura (cioè per il file e per ciascuno degli indici).

Domanda 4 (10%)

Con riferimento alla domanda precedente, indicare come potrebbe essere realizzato un indice secondario hash su *Matricola* (utilizzando una funzione hash semplice, anche se non realistica, quale il resto della divisione per un numero facile da gestire, quale 50 o 100).