

Corso di Basi di dati

Prova scritta parziale — 31 marzo 2001 — Compito A

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

Nota bene: rispondere con riferimento alla numerazione del testo, in modo ordinato e il più possibile sintetico (nella misura in cui ciò non pregiudichi la comprensibilità).

Domanda 1 (30%) Si consideri il protocollo di commit a due fasi (2PC).

1. Spiegare perché
 - (a) un guasto del coordinatore (TM) può avere conseguenze molto pesanti anche sulle prestazioni dei partecipanti (RM);
 - (b) un guasto di un partecipante non ha conseguenze particolari sulle prestazioni degli altri partecipanti (a parte l'eventuale abort di transazioni).

Per ovviare alle conseguenze negative di un guasto del coordinatore, alcune implementazioni del 2PC prevedono la possibilità di comunicazione fra i partecipanti (mentre la versione base prevede solo comunicazione fra il coordinatore e ciascuno dei partecipanti). In particolare, un partecipante che abbia una transazione in stato di "ready" può chiedere agli altri partecipanti informazioni sullo stato di tale transazione (che può essere "prima-del-ready", "ready", "commit" o "abort") presso di loro.

2. In tale contesto indicare
 - (a) come (e in quali casi) il partecipante può trarre profitto dalle risposte ottenute;
 - (b) quali insiemi di risposte il partecipante può ricevere e quali invece no.

Domanda 2 (25%) Considerare un file corrispondente ad una relazione su vari attributi fra cui CodiceFiscale e Cognome, sotto le seguenti ipotesi:

- il file contiene 200.000 record, che hanno lunghezza fissa pari a 110 byte; i campi CodiceFiscale e Cognome occupano rispettivamente 6 e 20 byte;
- il file system utilizzato prevede blocchi che rendono disponibili alla struttura fisica 1000 byte; i puntatori (sia i puntatori a blocchi sia quelli a singoli record) occupano 5 byte.

Indicare il numero di blocchi approssimativamente necessari per ciascuna delle seguenti strutture:

1. il file;
2. un indice primario statico su Cognome (su cui quindi il file è ordinato);
3. un indice secondario (B-tree) su CodiceFiscale.

Domanda 3 (25%) Con riferimento alla domanda precedente, calcolare (approssimativamente) il numero di accessi a blocchi, nell'unità di tempo (ad esempio l'ora), per le seguenti operazioni:

1. ricerca di un record sulla base del CodiceFiscale (eseguita 1000 volte l'ora);
2. ricerca dei record con una certa sottostringa iniziale di Cognome, che restituisca mediamente 25 record (200 volte l'ora).

Inoltre, indicare quali vantaggi o svantaggi si avrebbero, sempre con riferimento alle operazioni citate, in ciascuno dei seguenti casi (da considerare separatamente):

3. l'indice primario viene sostituito da una struttura hash;
4. nell'indice secondario il B-tree viene sostituito da una struttura hash (sempre secondaria).

Domanda 4 (20%) Si consideri una base di dati relativa alla biglietteria di un circo che prevede l'assegnazione dei posti (cioè ogni biglietto contiene l'indicazione del relativo posto). In particolare, si tiene traccia di quali siano i posti disponibili e quali quelli venduti per ciascuno spettacolo. Le vendite si aprono al mattino e i dati vengono eliminati la sera; ci sono più spettacoli in una giornata, ad esempio mattutino, pomeridiano, serale. I posti sono suddivisi per categoria. In tale contesto, si vuole definire un datawarehouse per studiare l'andamento delle presenze nel tempo sui posti delle varie categorie. Mostrare

1. per la base di dati, lo schema concettuale;
2. per il datawarehouse:
 - (a) lo schema concettuale;
 - (b) i fatti, le misure e le dimensioni;
 - (c) lo schema dimensionale a stella.

Corso di Basi di dati

Prova scritta parziale — 31 marzo 2001 — Compito B

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

Nota bene: rispondere con riferimento alla numerazione del testo, in modo ordinato e il più possibile sintetico (nella misura in cui ciò non pregiudichi la comprensibilità).

Domanda 1 (30%) Si consideri il protocollo di commit a a due fasi (2PC).

1. Spiegare perché
 - (a) un guasto di un partecipante non ha conseguenze particolari sulle prestazioni degli altri partecipanti (a parte l'eventuale abort di transazioni);
 - (b) un guasto del coordinatore (TM) può avere conseguenze molto pesanti anche sulle prestazioni dei partecipanti (RM).

Per ovviare alle conseguenze negative di un guasto del coordinatore, alcune implementazioni del 2PC prevedono la possibilità di comunicazione fra i partecipanti (mentre la versione base prevede solo comunicazione fra il coordinatore e ciascuno dei partecipanti). In particolare, un partecipante che abbia una transazione in stato di “ready” può chiedere agli altri partecipanti informazioni sullo stato di tale transazione (che può essere “prima-del-ready”, “ready”, “commit” o “abort”) presso di loro.

2. In tale contesto indicare
 - (a) quali insiemi di risposte il partecipante può ricevere e quali invece no;
 - (b) come (e in quali casi) il partecipante può trarre profitto dalle risposte ottenute.

Domanda 2 (25%) Considerare un file corrispondente ad una relazione su vari attributi fra cui CodiceFiscale e Via, sotto le seguenti ipotesi:

- il file contiene 100.000 record, che hanno lunghezza fissa pari a 110 byte; i campi CodiceFiscale e Via occupano rispettivamente 6 e 16 byte;
- il file system utilizzato prevede blocchi che rendono disponibili alla struttura fisica 1000 byte; i puntatori (sia i puntatori a blocchi sia quelli a singoli record) occupano 5 byte.

Indicare il numero di blocchi approssimativamente necessari per ciascuna delle seguenti strutture:

1. il file;
2. un indice primario statico su Via (su cui quindi il file è ordinato);
3. un indice secondario (B-tree) su CodiceFiscale.

Domanda 3 (25%) Con riferimento alla domanda precedente, calcolare (approssimativamente) il numero di accessi a blocchi, nell'unità di tempo (ad esempio l'ora), per le seguenti operazioni:

1. ricerca di un record sulla base del CodiceFiscale (eseguita 1000 volte l'ora);
2. ricerca dei record con una certa sottostringa iniziale di Via, che restituisca mediamente 25 record (200 volte l'ora).

Inoltre, indicare quali vantaggi o svantaggi si avrebbero, sempre con riferimento alle operazioni citate, in ciascuno dei seguenti casi (da considerare separatamente):

3. l'indice primario viene sostituito da una struttura hash;
4. nell'indice secondario il B-tree viene sostituito da una struttura hash (sempre secondaria).

Domanda 4 (20%) Si consideri una base di dati relativa alle prenotazioni di un ristorante che prevede l'assegnazione dei tavoli (cioè quando il cliente prenota gli viene comunicato il numero del tavolo assegnato). In particolare, interessano solo le prenotazioni per la cena, per la quale sono previsti più turni (comunque prefissati, ad esempio, primo e secondo); quindi per ogni tavolo e per ogni turno, in ogni momento si può dire se è prenotato o libero. Le prenotazioni si aprono al mattino e non si accettano clienti che non abbiano prenotato; i dati vengono eliminati la sera. I tavoli sono suddivisi per categoria (ad esempio: due posti, quattro posti, sei posti). In tale contesto, si vuole definire un datawarehouse per studiare l'andamento nel tempo del numero di persone che sono state servite nei vari turni. Mostrare

1. per la base di dati, lo schema concettuale;
2. per il datawarehouse:
 - (a) lo schema concettuale;
 - (b) i fatti, le misure e le dimensioni;
 - (c) lo schema dimensionale a stella.