

Basi di dati II, primo modulo

Terzo homework — 3 aprile 2009

Domanda 1 Si consideri la seguente base di dati relazionale, relativa alle prescrizioni di farmaci acquisite da un insieme di farmacie:

- Ricette(Numero, CodFarmacia, CFPaziente, Data)
- Farmacia(CodFarmacia, Nome)
- ElementiRicetta(NumeroRicetta, NumeroLinea, CodFarmaco)
- Farmaci(Codice, Descrizione, CodMolecola, CodCasa, Prezzo, Fascia)
- Molecole(CodMolecola, Descrizione)
- Pazienti(CF, Cognome, Nome, DataNascita, Via, NumeroCivico, Città)
- CaseFarmaceutiche(CodCasa, Nome)
- ASL(Codice, Nome)
- Territorio(Via, Città, NumeroCivico, ASL)

Si noti che ci sono dati che cambiano nel tempo fra cui prezzi e fasce ('A', 'B' o 'C') dei farmaci e indirizzi dei pazienti.

Costruire, in tale contesto, uno schema a stella che permetta di analizzare le prescrizioni (quantità e prezzi complessivi) rispetto a

- data (dimensione standard i cui dettagli possono essere omessi);
- farmaci, con le loro proprietà (molecola e casa farmaceutica);
- ASL di residenza e fascia d'età (ad esempio, 0-3,4-17, 18-30, ...; ma potrebbero variare) dei pazienti;
- farmacia.

Supporre che, per ovvie ragioni di privacy, non possano essere riportati dati che permettano di risalire alle identità dei pazienti.

Specificare la grana dei fatti e indicare sinteticamente come si ottiene la relazione dei fatti da quelle della base di dati (supponendo disponibili, ove necessario, opportune tabelle per la conversione delle chiavi).

Domanda 2 Considerare la seguente interrogazione in SQL:

```
SELECT A, D, H FROM R, S, T WHERE E = B AND C = G
```

definita con riferimento a tre relazioni, definite e frammentate come segue (per essere poi distribuite):

- $R(A, C, E)$ frammentata orizzontalmente: $R_a = \sigma_{C > 100}(R)$; $R_b = \sigma_{C \leq 100}(R)$
- $S(B, D, F)$ frammentata verticalmente: $S_a = \pi_{B,D}(S)$; $S_b = \pi_{B,F}(S)$
- $T(G, H)$ frammentata orizzontalmente: $T_a = \sigma_{G > 200}(T)$; $T_b = \sigma_{G \leq 200}(T)$

Mostrare (ad esempio sotto forma di albero) l'espressione dell'algebra relazionale definita sui frammenti che realizza in modo efficiente tale interrogazione.

Domanda 3 Si consideri il protocollo di commit a 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) 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.