

Tecnologia delle basi di dati (ex Basi di dati, primo modulo)

20 giugno 2007 — Compito A

Tempo a disposizione: 2 ore e 15 minuti. **Nota:** è richiesta una “bella copia” comprensibile e ordinata.

Domanda 1 (25%) Si consideri la seguente base di dati relazionale, relativa alle ricette acquisite da un insieme di farmacie:

- Ricette(Numero, CodFarmacia, CFPaziente, Data)
- Farmacie(CodFarmacia, Nome, Via, NumeroCivico, Città)
- ElementiRicetta(NumeroRicetta, CodFarmaco, Quantità)
- 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);
- prescrizione di farmaci nella stessa ricetta
- ASL di residenza e fascia d'età (ad esempio, 0-3,4-17, 18-30, ...; ma potrebbero variare) dei pazienti;
- ASL della farmacia

Supporre che, per ovvie ragioni di privacy, non possano essere riportati dati che permettano di risalire alle identità dei pazienti (CF, cognome, nome, data di nascita e indirizzo)

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 (20%) Illustrare, brevemente ma con ordine, vantaggi e svantaggi (relativamente a vari tipi di operazioni) delle seguenti strutture fisiche, definite con riferimento ad uno stesso attributo A non chiave:

1. indice primario (sparso) multilivello statico;
2. indice secondario multilivello statico;
3. B+-tree secondario;
4. hash.

Inoltre, indicare, in ciascuno dei quattro casi (sia in forma simbolica sia in forma numerica), il costo di operazioni di ricerca basate su A (“trovare i record con un certo valore per l'attributo A ”), con riferimento ad una relazione R contenente $L = 100.000.000$ ennuple di $R = 25$ byte ciascuna, di cui $a = 12$ per l'attributo A ; si abbiano mediamente $m = 3$ record con lo stesso valore su A ; supporre che i blocchi abbiano dimensione $B = 2\text{KB}$, approssimabile come $B = 2.000$ e che i puntatori ai blocchi abbiano lunghezza $p = 4$.

Domanda 3 (20%) Si consideri una forma di equivalenza fra schedule denominata *final-state-equivalenza*, secondo la quale due schedule S_1 e S_2 sono equivalenti se, per ogni istanza i della base di dati, essi la trasformano nello stesso modo (e quindi $S_1(i) = S_2(i)$, se con $S(i)$ indichiamo l'istanza della base di dati ottenuta applicando a i le operazioni dello schedule S).

1. Formulare una definizione per questa proprietà, variante della definizione di view-equivalenza.
2. Spiegare, almeno intuitivamente, il rapporto che esiste fra final-state-equivalenza e view-equivalenza (sono equivalenti, incomparabili, oppure una implica l'altra?).

Domanda 4 (15%) Illustrare il protocollo del commit a due fasi utilizzando un esempio della vita quotidiana e in particolare l'organizzazione di una partita a tennis la quale è prevista la partecipazione di quattro specifiche persone insieme alla disponibilità di un campo presso un circolo.

Domanda 5 (20%) Specificare in quali casi (e perché) un indice può risultare utile per eseguire una proiezione con eliminazione di duplicati.

Tecnologia delle basi di dati (ex Basi di dati, primo modulo)

20 giugno 2007 — Compito B

Tempo a disposizione: 2 ore e 15 minuti. **Nota:** è richiesta una “bella copia” comprensibile e ordinata.

Domanda 1 (25%) Si consideri la seguente base di dati relazionale, relativa alle ricette acquisite da un insieme di farmacie:

- Ricette(Numero, CodFarmacia, CFPaziente, Data)
- Farmacie(CodFarmacia, Nome, Via, NumeroCivico, Città)
- ElementiRicetta(NumeroRicetta, CodFarmaco, Quantità)
- 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);
- prescrizione di farmaci nella stessa ricetta
- ASL di residenza e fascia d'età (ad esempio, 0-3,4-17, 18-30, ...; ma potrebbero variare) dei pazienti;
- ASL della farmacia

Supporre che, per ovvie ragioni di privacy, non possano essere riportati dati che permettano di risalire alle identità dei pazienti (CF, cognome, nome, data di nascita e indirizzo)

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 (20%) Illustrare, brevemente ma con ordine, vantaggi e svantaggi (relativamente a vari tipi di operazioni) delle seguenti strutture fisiche, definite con riferimento ad uno stesso attributo A non chiave:

1. indice primario (sparso) multilivello statico;
2. indice secondario multilivello statico;
3. B+-tree secondario;
4. hash.

Inoltre, indicare, in ciascuno dei quattro casi (sia in forma simbolica sia in forma numerica), il costo di operazioni di ricerca basate su A (“trovare i record con un certo valore per l'attributo A ”), con riferimento ad una relazione R contenente $N = 35.000.000$ ennuple di $L = 20$ byte ciascuna, di cui $k = 6$ per l'attributo A ; si abbiano mediamente $m = 3$ record con lo stesso valore su A ; supporre che i blocchi abbiano dimensione $B = 1\text{KB}$, approssimabile come $B = 1.000$ e che i puntatori ai blocchi abbiano lunghezza $p = 4$.

Domanda 3 (20%) Si consideri una forma di equivalenza fra schedule denominata *final-state-equivalenza*, secondo la quale due schedule S_1 e S_2 sono equivalenti se, per ogni istanza d della base di dati, essi la trasformano nello stesso modo (e quindi $S_1(d) = S_2(d)$, se con $S(d)$ indichiamo l'istanza della base di dati ottenuta applicando a d le operazioni dello schedule S).

1. Formulare una definizione per questa proprietà, variante della definizione di view-equivalenza.
2. Spiegare, almeno intuitivamente, il rapporto che esiste fra final-state-equivalenza e view-equivalenza (sono equivalenti, incomparabili, oppure una implica l'altra?).

Domanda 4 (15%) Illustrare il protocollo del commit a due fasi utilizzando un esempio della vita quotidiana e in particolare l'organizzazione di una riunione per la quale è prevista la partecipazione di tre specifiche persone insieme alla disponibilità di un tavolo presso un certo ristorante.

Domanda 5 (20%) Specificare in quali casi (e perché) un indice può risultare utile per eseguire una unione in SQL (con eliminazione di duplicati).