

Sistemi per la Gestione delle Basi di Dati

27 gennaio 2016

1. (7 Punti) Sono date le seguenti relazioni (le chiavi primarie sono sottolineate):

FARMACO (CodFarmaco, Nome_Farmaco, Categoria, Casa_Farmaceutica, Salvavita)
PRINCIPIO_ATTIVO_IN_FARMACO (CodFarmaco, CodPriAt, Max_milligrammi_per_giorno)
FARMACIA (CodFarmacia, Nome_Farmacia, Stato, Indirizzo, Numero_Telefono)
CONSEGNA(CodFarmaco, CodFarmacia, Data, Quantitá)

Si ipotizzino le seguenti cardinalità:

- $\text{card}(\text{FARMACO}) = 10^5$ tuple,
valori distinti di categoria $\simeq 10$,
valori distinti di salvavita: {Si, No},
- $\text{card}(\text{PRINCIPIO_ATTIVO_IN_FARMACO}) = 10^6$ tuple,
 $\text{MIN}(\text{Max_milligrammi_per_giorno}) = 1$, $\text{MAX}(\text{Max_milligrammi_per_giorno}) = 500$,
- $\text{card}(\text{FARMACIA}) = 10^5$ tuples,
valori distinti di stato $\simeq 10$.
- $\text{card}(\text{CONSEGNA}) = 10^{10}$ tuples,
 $\text{MIN}(\text{Data}) = 1-1-1990$, $\text{MAX}(\text{Data}) = 31-12-2009$,

Inoltre, ipotizzare il seguente fattore di riduzione per la condizione group by:

- $\text{having count}(\ast) \leq 100 \simeq \frac{1}{10}$.

Si consideri la seguente query SQL:

```
select F.CodFarmaco, Nome_Farmaco
from FARMACO F, PRINCIPIO_ATTIVO_IN_FARMACO PAF
where F.CodFarmaco = PAF.CodFarmaco and F.Salvavita= 'No'
and F.Categoria = 'Antibiotico' and PAF.Max_milligrammi_per_giorno >= 450
and F.CodFarmaco IN (select C.CodFarmaco
                      from CONSEGNA C, FARMACIA FA
                      where C.CodFarmacia = FA.CodFarmacia and
                           FA.Stato <> 'Italia'
                      and C.Data >= 31-12-1999
                      group by C.CodFarmaco
                      having count(*) <= 100)
```

Per l'interrogazione SQL

- (a) Si scriva l'espressione algebrica corrispondente, indicando le operazioni svolte, la cardinalità e la selettività di ogni operazione. Dove necessario, si ipotizzi la distribuzione dei dati. Discutere la possibilità di anticipare l'operatore GROUP BY.
- (b) Si scelgano le strutture fisiche accessorie per migliorare le prestazioni dell'interrogazione. Si motivi la scelta e si definisca il piano di esecuzione (ordine e tipo dei join, accesso alle tabelle e/o indici, etc.).

2. (8 Punti) Sono date le relazioni seguenti (le chiavi primarie sono sottolineate, gli attributi opzionali sono indicati con *).

```
AUTOVETTURA(Targa, StatoCorrente)
CLIENTE(CodCliente, NomeCliente)
NOLEGGIO(Targa, TimeStampInizioNoleggio, TimeStampFineNoleggio*, CodCliente)
TARIFFE_NOLEGGIO(TipoAttività, CostoAlMinuto)
ATTIVITÀ_NOLEGGIO(Targa, TimeStampInizio, TipoAttività, TimeStampFine*)
RICEVUTA_NOLEGGIO(CodRicevuta, CodCliente, Targa, TimeStampInizioNoleggio,
                  TimeStampFineNoleggio, Costo)
```

Si vuole gestire in modo automatico il processo di noleggio di autovetture in modalità car sharing. La tabella AUTOVETTURA memorizza le informazioni sullo stato delle autovetture di proprietà della società che gestisce il noleggio. L'attributo StatoCorrente può assumere i valori libera se l'autovettura è a disposizione per il noleggio e occupata se l'autovettura è attualmente noleggiata. La tabella CLIENTE descrive i clienti registrati per il noleggio di autovetture. La tabella NOLEGGIO contiene le informazioni generali sul noleggio, quali i timestamp di inizio e, al termine, di fine. Mentre il noleggio è in corso, il timestamp di fine assume il valore NULL. La tabella TARIFFE_NOLEGGIO contiene il costo di ogni tipologia di attività svolta durante il noleggio: uso dell'autovettura (con motore acceso) e sosta dell'autovettura (con motore spento). Tutte le tariffe sono applicate a minuti, a partire dal primo minuto di noleggio. Durante un noleggio, il cliente può alternare attività di uso e di sosta. La tabella ATTIVITÀ_NOLEGGIO elenca tutti le attività relative all'ultimo noleggio in corso di ogni autovettura. In particolare, indica tutti i periodi temporali in cui l'autovettura è stata utilizzata (uso) ed, eventualmente, lasciata in sosta (sosta). L'ultima attività in corso per l'autovettura è caratterizzato da valore NULL per l'attributo TimeStampFine. La tabella RICEVUTA_NOLEGGIO contiene le ricevute emesse alla fine di un periodo di noleggio. Contiene le informazioni sul momento di inizio e fine del periodo complessivo di noleggio e sul costo complessivo del noleggio.

Si scrivano i trigger per gestire le seguenti attività.

(1) *Termine del noleggio di un'autovettura.* Il termine del noleggio di un'autovettura, è definito dall'aggiornamento del valore dell'attributo TimeStampFine della tabella NOLEGGIO con il timestamp corrispondente al momento di termine del noleggio. Quando termina un noleggio, occorre modificare lo stato corrente dell'autovettura, che deve essere posto a libera e deve essere aggiornato il timestamp di fine per l'ultima attività svolta durante il noleggio (tabella ATTIVITÀ_NOLEGGIO).

Si deve inoltre calcolare il costo complessivo del noleggio, tenendo conto di tutte le attività svolte durante il periodo di noleggio. Per entrambi i tipi di attività, occorre calcolare la durata complessiva ed il costo corrispondente. Il costo complessivo è calcolato tenendo conto di tutti i periodi (dati dalla differenza di timestamp di inizio e di fine) di uso, e di eventuali periodi di sosta. Si utilizzi la funzione TO_MINUTES() per convertire le durate dei periodi in minuti. Deve poi essere inserita la ricevuta corrispondente all'importo complessivo del noleggio nella tabella RICEVUTA_NOLEGGIO. Si consideri che il codice della ricevuta (attributo CodRicevuta) è un contatore che deve essere incrementato per ogni nuova ricevuta.

Infine, tutte le informazioni sulle attività svolte durante l'ultimo noleggio dell'autovettura devono essere eliminate dalla tabella ATTIVITÀ_NOLEGGIO.

(2) *Vincolo di integrità sulle attività associate ad un'autovettura.* Quando si inserisce una nuova attività nella tabella ATTIVITÀ_NOLEGGIO, il valore di TimeStampFine deve sempre essere NULL. Ogni azione sulla tabella ATTIVITÀ_NOLEGGIO che causi la violazione del vincolo non deve essere eseguita.

3. Progettazione Data Warehouse

L'Unione Europea nell'ambito della valutazione dei risultati della ricerca scientifica finanziata con fondi pubblici ha deciso di analizzare le pubblicazioni prodotte dai ricercatori delle università europee.

I ricercatori e i professori delle università europee presentano i risultati delle proprie attività di ricerca scrivendo degli articoli scientifici, denominati genericamente pubblicazioni. Ogni pubblicazione ricade in una tipologia specifica (es. articolo a conferenza, articolo di giornale, capitolo di libro, ecc.) ed è caratterizzata da una data specifica di pubblicazione, da uno o più autori e da un luogo di pubblicazione (es. conferenza, workshop, rivista, libro, ecc.), il quale ne determina la tipologia specifica. Il luogo di pubblicazione (rivista, conferenza, ecc.) è tipicamente caratterizzato da una o più edizioni, delle quali si è interessati all'anno (es. conferenza internazionale delle Basi di Dati edizione 2015, conferenza internazionale delle Basi di Dati edizione 2016, ecc.) e da un editore (es. Elsevier, ecc.).

Tra gli autori della pubblicazione è identificato un autore principale, il quale afferisce ad uno specifico dipartimento universitario. Ogni dipartimento appartiene ad un campus universitario e ciascuna università può essere formata da uno o più campus. Inoltre, le università sono suddivise per dimensione (piccole, medie, grandi, in base al numero di ricercatori che vi lavorano). Ogni singolo dipartimento è infine caratterizzato da un settore scientifico-disciplinare di afferenza. Per esempio, il Dipartimento di Automatica e Informatica afferisce al settore scientifico-disciplinare ING-INF/05, fa parte del campus Cittadella Politecnica, a sua volta appartenente all'università Politecnico di Torino.

L'Unione Europea è interessata ad analizzare il numero di pubblicazioni al variare di:

- mese, bimestre, trimestre, semestre, anno della pubblicazione;
- anno accademico (da settembre ad agosto);
- mese di sospensione delle attività didattiche (luglio e agosto di tutti gli anni);
- mese dell'anno;
- dipartimento dell'autore principale, relativo campus e università;
- dimensione dell'università, stato europeo dell'università, settore scientifico-disciplinare di afferenza del dipartimento;
- numero di autori (da 1 a 10 con incrementi singoli, e una categoria per più di 10 autori), luogo in cui è avvenuta la pubblicazione (conferenza, rivista, ecc.), edizione (l'anno in cui si è tenuta la conferenza o è stata pubblicata la rivista), editore, e tipo della pubblicazione (articolo a conferenza, articolo di giornale, capitolo di libro, ecc.).

Progettazione

- (a) (6 Punti) Progettare il data warehouse (modello concettuale e tabelle dei fatti e delle dimensioni) in modo da soddisfare le richieste descritte nelle specifiche del problema. Il data warehouse progettato deve inoltre permettere di rispondere in modo efficiente alle seguenti interrogazioni.
- (b) (8 Punti) Esprimere le due interrogazioni seguenti utilizzando il linguaggio SQL esteso. Indicare la risposta nello spazio sottostante ciascuna domanda.

(a) Per ogni università europea, per ogni tipo di pubblicazione, e per ogni anno, calcolare il numero medio mensile di pubblicazioni, e il totale cumulativo annuale delle pubblicazioni. Si considerino solo i mesi in cui c'è stata almeno una pubblicazione.

(b) Per ogni anno, calcolare la percentuale di pubblicazioni del dipartimento rispetto al totale della sua università. Assegnare inoltre un rank globale ai dipartimenti di tutte le università per numero di pubblicazioni totali (il primo dipartimento è quello con il maggior numero di pubblicazioni).