

Politecnico di Torino

Sistemi di gestione di basi di dati

6 Febbraio 2012

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

```
TECNICO(TId, Cognome, DataNascita, Qualifica, Laboratorio)
PAZIENTE(PIId, DataNascita, Malattia, OId)
ANALISI(Data, PIId, TId, Tipo, Risultati)
OSPEDALE(OId, Nome, Indirizzo, Città)
```

Si ipotizzino le seguenti cardinalità per le tabelle:

- $\text{card}(\text{TECNICO}) = 10^4$ tuple,
 $\text{MIN}(\text{DataNascita}) = 1-1-1960$, $\text{MAX}(\text{DataNascita}) = 31-12-1989$,
valori distinti di Qualifica $\simeq 10$,
- $\text{card}(\text{PAZIENTE}) = 10^6$ tuple,
 $\text{MIN}(\text{DataNascita}) = 1-1-1912$, $\text{MAX}(\text{DataNascita}) = 31-12-2011$,
valori distinti di Malattia $\simeq 100$
- $\text{card}(\text{ANALISI}) = 10^7$ tuple,
 $\text{MIN}(\text{Data}) = 1-1-2011$, $\text{MAX}(\text{Data}) = 31-12-2011$
valori distinti di Tipo $\simeq 10$
- $\text{card}(\text{OSPEDALE}) = 10^3$ tuple,
valori distinti di Città $\simeq 10$

Inoltre si ipotizzi il seguente fattore di riduzione per la condizione di group by:

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

Si consideri la seguente query SQL:

```
select Città, Count(*)
from OSPEDALE O, PAZIENTE P
where O.PId=P.PId and
      and PId in (select PId from PAZIENTE P1, ANALISI A, TECNICO T
                 where P1.PId=A.PId and A.TId=T.TId
                 and P1.DataNascita = 1991 and Qualifica <>'Infermiere'
                 and A.Data  $\geq$  1/12/2011 and A.Data  $\leq$  31/12/2011
                 and Tipo <> 'Esame del sangue'
                 group by PId
                 having count (*) > 10)
group by Città;
```

Per l'interrogazione SQL

- 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.
- 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 *).

PRODOTTO (CodP, NomeP, Prezzo, PuntiProdotto)
CARTA_FEDELTÀ (CodC, NomeCliente, PuntiTotali)
ACQUISTO (CodA, CodP, Data, CodC*, NumeroPezzi)
PREMIO (CodPremio, DescrizPremio, PuntiNecessari)
RICHIESTA_NOTIFICA(CodN, CodC, CodPremio, DescrizPremio)

Si vogliono gestire alcune attività relative alle carte fedeltà di un supermercato. La tabella `CARTA_FEDELTÀ` riporta, per ogni carta fedeltà, i punti complessivamente acquisiti da un cliente (attributo `PuntiTotali`). La tabella `PREMIO` descrive i premi disponibili. Per ciascun premio, il valore in punti è definito dall'attributo `PuntiNecessari`. Si scrivano i trigger per gestire le seguenti attività.

(1) *Attribuzione dei punti per un acquisto ed eventuale selezione del premio.* Si scriva il trigger per aggiornare la situazione della carta fedeltà del cliente che ha eseguito l'acquisto. Quando viene eseguito un nuovo acquisto (inserimento nella tabella `ACQUISTO`), devono essere aggiornati i punti totali conseguiti dal cliente. I punti indicati nell'attributo `PuntiProdotto` sono relativi all'acquisto di un solo pezzo. Per calcolare i punti totali conseguiti nell'acquisto occorre considerare il numero totale di pezzi acquistati (attributo `NumeroPezzi`). Quando il prodotto non è associato all'acquisizione di punti, l'attributo `PuntiProdotto` vale zero. Si noti che gli acquisti non sono necessariamente associati a una carta fedeltà. Quando l'acquisto non è associato a una carta fedeltà, il valore dell'attributo `CodC` è `NULL` e non deve essere eseguito alcun aggiornamento della tabella `CARTA_FEDELTÀ`.

Successivamente, si deve verificare se i punti accumulati consentono di ottenere un premio (cioè se esiste almeno un premio di valore (attributo `PuntiNecessari`) inferiore o uguale ai punti totali accumulati). In caso positivo, deve essere scelto il premio di valore massimo che può essere ricevuto con i punti accumulati. Si supponga che ci sia al più un premio che soddisfa questa condizione. Si deve infine richiedere la notifica (inserimento nella tabella `RICHIESTA_NOTIFICA`) della possibilità di ricevere il premio selezionato al possessore della carta fedeltà. La chiave primaria `CodN` è un contatore che deve essere incrementato ogni volta che è inserita una nuova notifica (si tenga conto che le notifiche completamente evase potrebbero essere eliminate dalla base di dati).

(2) *Vincolo di integrità sul valore massimo in punti di un prodotto.* Per ogni prodotto, il valore in punti (attributo `PuntiProdotto` nella tabella `PRODOTTO`) non può essere superiore a $1/10$ del prezzo di vendita (attributo `Prezzo`). Si scriva il trigger che gestisce il vincolo d'integrità, assegnando il valore massimo, pari a $1/10$ del prezzo, quando il valore massimo consentito viene superato.

3. Progettazione Data Warehouse

Una società di servizi per la promozione del commercio internazionale vuole valutare la quantità e la qualità dei servizi erogati dalle proprie filiali sparse in tutto il mondo analizzando il volume d'affari di ogni filiale.

La società è interessata a progettare e realizzare un data warehouse che offra una visione d'insieme della propria attività economica, analizzando il volume d'affari medio giornaliero e il numero di prestazioni erogate delle proprie filiali in funzione di:

- mese, quadrimestre e anno in cui è stato realizzato il volume d'affari,
- data, giorno della settimana, giorno del mese (1-31),
- filiale,
- servizio e tipologia di servizio (finanziario, consulenza commerciale, assicurativo, ecc.)
- modalità di erogazione del servizio, per esempio se il servizio è disponibile on-line oppure al telefono
- tipologia di cliente (privato, professionista, PMI, ecc.),
- modalità di pagamento
- provincia, regione, stato e continente in cui è sita la filiale.

Il data warehouse realizzato deve contenere le informazioni relative agli anni 2004-2011. Al fine di una corretta realizzazione del data warehouse sono state fornite le seguenti informazioni (le informazioni ritenute necessarie ma non presenti in questo elenco possono essere ipotizzate e stimate dal candidato):

- numero di filiali $\simeq 1500$
- numero di tipologia di cliente $\simeq 5$
- numero di servizi $\simeq 300$

Sono riportate di seguito alcune delle interrogazioni frequenti di interesse per la società, alle quali il data warehouse deve poter rispondere in modo efficiente:

- (a) Per ogni tipologia di servizio e considerando solo il volume d'affari del 2009, visualizzare per ogni filiale e quadrimestre, il volume d'affari totale, il numero totale di prestazioni erogate, il volume d'affari medio giornaliero e il numero medio giornaliero di prestazioni erogate.
- (b) Per ogni coppia (città della filiale, anno) visualizzare il volume d'affari totale e il numero totale di prestazioni erogate.
- (c) Per ogni filiale sita in Francia, visualizzare il volume d'affari quadrimestrale, il volume d'affari medio giornaliero per ogni quadrimestre e il volume d'affari totale quadrimestrale cumulativo dall'inizio dell'anno, separatamente per ogni tipologia di servizio.

Progettazione

- (a) (6 Punti) Progettare il data warehouse in modo da soddisfare le richieste descritte nelle specifiche del problema. Il data warehouse progettato deve inoltre permettere di rispondere in modo efficiente a tutte le interrogazioni frequenti indicate.
- (b) (8 Punti) Esprimere le interrogazioni frequenti (a) e (c) utilizzando il linguaggio SQL esteso.