

# Politecnico di Torino

## Sistemi per la gestione di basi di dati

21 giugno 2012

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

```
CONFERENZA(IDC, Titolo, Città, Stato, Sponsor, Data_Inizio, Data_Fine)
ARTICOLO(ID_AR, Titolo, #Pagine, Tipo, IDA)
AUTORE(IDA, Nome, Data_Nascita, NomeUniversità, Nazione, Email)
REGISTRAZIONE(IDC, ID_AR, IDA, Data, Costo)
```

Si ipotizzino le seguenti cardinalità:

- $\text{card}(\text{CONFERENZA}) = 10^4$  tuple,  
valori distinti di Città  $\simeq 10^2$ ,  
valori distinti di Stato  $\simeq 10$ ,  
valori distinti di Sponsor  $\simeq 10$ ,  
 $\text{MIN}(\text{Data\_Inizio}) = 1/1/2010$ ,  $\text{MAX}(\text{Data\_Fine}) = 31/12/2011$ ,
- $\text{card}(\text{ARTICOLO}) = 5 \cdot 10^5$  tuple,  
 $\text{MIN}(\#Pagine) = 2$ ,  $\text{MAX}(\#Pagine) = 12$ ,
- $\text{card}(\text{AUTORE}) = 10^6$  tuple,  
 $\text{MIN}(\text{Data\_Nascita}) = 1/1/1960$ ,  $\text{MAX}(\text{Data\_Nascita}) = 31/12/1989$ .  
valori distinti di NomeUniversità  $\simeq 10^3$ ,  
valori distinti di Nazione  $\simeq 10^2$ ,
- $\text{card}(\text{REGISTRAZIONE}) = 5 \cdot 10^9$  tuple,  
 $\text{MIN}(\text{Data}) = 1/1/2010$ ,  $\text{MAX}(\text{Data}) = 31/12/2011$ ,  
 $\text{MIN}(\text{Costo}) = 100$  \$,  $\text{MAX}(\text{Costo}) = 600$  \$.

Inoltre si ipotizzi il seguente fattore di riduzione per le condizioni di group by:

- $\text{having count}(\ast) > 1 \simeq \frac{1}{4}$ ,

Si consideri la seguente query SQL:

```
select Titolo, Tipo
from ARTICOLO
where #Pagine  $\geq$  8 and
```

```
    IDA not in (select IDA from AUTORE A, CONFERENZA C, REGISTRAZIONE R
                where R.IDC=C.IDC and R.IDA=A.IDA
                and Data  $\geq$  1/11/2011
                and C.Stato = 'USA' and A.Nazione  $\neq$  'USA'
                and Data_Nascita  $\geq$  1/1/1987
                group by IDA
                having count(\ast) > 1)
```

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 \*):

FARMACO (CodF, NomeF, CasaFarmaceutica, Prezzo, PercentualeSconto\*)  
CARRELLO (CodC, IDCliente, Data, Stato)  
CONTENUTO\_CARRELLO (CodC, CodF, NumeroPezzi, Importo, StatoDisponibilità)  
LIVELLO\_MAGAZZINO (CodF, NumConfezioniDisponibili)  
RICHIESTA\_SPEDIZIONE (CodC, CodF, NumeroPezzi)  
SELEZIONE\_FARMACO(CodS, CodC, CodF, NumeroPezzi)

Si vogliono gestire alcune attività relative a un sito per la vendita on-line di farmaci. La tabella CONTENUTO\_CARRELLO elenca i farmaci contenuti nel carrello elettronico descritto dalla tabella CARRELLO. Si scrivano i trigger per gestire le seguenti attività.

(1) *Selezione di un farmaco per l'acquisto.* Si scriva il trigger per aggiornare il contenuto del carrello elettronico quando è selezionato un nuovo prodotto da inserire nel carrello elettronico (inserimento nella tabella SELEZIONE\_FARMACO). Occorre verificare la disponibilità del farmaco nel magazzino (tabella LIVELLO\_MAGAZZINO). Si ipotizzi che il farmaco sia sempre presente nel magazzino, eventualmente con zero confezioni disponibili. Se il numero di confezioni disponibili è superiore al numero di confezioni richieste, deve essere inserito il farmaco richiesto nel carrello elettronico (tabella CONTENUTO\_CARRELLO), con stato disponibilità pari a 'D' (cioè disponibile). Occorre anche calcolare l'importo, in base al prezzo del farmaco e all'eventuale sconto praticato (attributo PercentualeSconto, valore espresso in percentuale). Infine, deve essere aggiornata la disponibilità a magazzino. Se il numero di confezioni disponibili è inferiore al numero di confezioni richieste, il farmaco deve essere inserito nel carrello elettronico per la quantità disponibile, con stato disponibilità pari a 'P' (cioè parziale). La quantità disponibile in magazzino deve essere aggiornata di conseguenza. Infine, se non vi è disponibilità, il farmaco non deve essere inserito nel carrello elettronico.

(2) *Vincolo di integrità sul numero massimo di farmaci venduti con sconto elevato.* Nello stesso carrello elettronico (identificato dall'attributo CodC) non possono essere presenti complessivamente più di 20 pezzi venduti con uno sconto superiore al 40%. Si scriva il trigger che gestisce il vincolo d'integrità.

### 3. Progettazione Data Warehouse

Il mercato della musica digitale è in forte espansione e le piattaforme Internet che offrono la possibilità di scaricare musica legale cresce di giorno in giorno. Ogni piattaforma memorizza localmente tutte le informazioni sui download dei brani musicali richiesti dai propri clienti. Per poter valutare l'efficienza di questo nuovo mercato e capire in quale direzione può essere ampliato, vi è la necessità di progettare un data warehouse che offra una visione d'insieme su questo nuovo mercato. In particolare, il data warehouse deve essere progettato per analizzare:

- il tempo medio di download per brano musicale,
- il numero medio di download giornaliero,
- il ricavo giornaliero medio per brano musicale.

L'analisi deve essere effettuata in funzione di:

- data, mese, trimestre, quadrimestre, semestre e anno in cui è stato effettuato il download,
- fascia oraria (8:00-12:00, 12:00-16:00, 16:00-20:00, etc) e periodo del giorno (a.m., p.m.) in cui è stato effettuato il download,
- brano musicale, casa discografica che ha curato la registrazione del brano, album in cui il brano è stato pubblicato, anno di incisione dell'album,
- autore del brano musicale,
- genere musicale del brano,
- cantante/band e rispettiva nazionalità,
- città, provincia, regione e stato dell'utente che ha effettuato il download,
- metodo di pagamento (carta di credito, carta prepagata, etc.).

Il data warehouse conterrà le informazioni relative agli anni 2007, 2008, 2009, 2010, 2011. Sono riportate di seguito alcune delle interrogazioni frequenti di interesse, alle quali il data warehouse deve poter rispondere in modo efficiente:

- (a) Considerando solo i download effettuati nel secondo trimestre dell'anno 2010 da utenti situati in Abruzzo, visualizzare, separatamente per ogni album, il numero totale di download e il tempo medio di download.
- (b) Per ogni autore di un brano musicale e per ogni fascia oraria, visualizzare il numero totale di download separatamente per ogni casa discografica. Ordinare e assegnare un rank in ordine crescente di numero totale di download.
- (c) Per ogni brano musicale di cantanti italiani e considerando solo i download effettuati nel 2011, visualizzare per ogni mese il ricavo complessivo, il ricavo giornaliero medio, il numero medio giornaliero di download, il numero totale di download cumulativo dall'inizio dell'anno, separatamente per ogni genere musicale.

#### **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.