

Politecnico di Torino

Sistemi di gestione di basi di dati

21 Settembre 2011

1. (6 punti) Sono date le relazioni seguenti (le chiavi primarie sono sottolineate):

```
ATTORE(AId, Nome, Cognome, Cittadinanza, DataNascita)
COMMEDIA(ComId, Titolo, Regista, NumScene, Anno)
ATTORE-IN-COMMEDIA(ComId, AId, Ruolo)
PROGRAMMAZIONE(ComId, Teatro, Data, OraInizio, Durata)
```

Si ipotizzino le seguenti cardinalità per le tabelle:

- $\text{card}(\text{ATTORE}) = 10^4$ tuple,
 $\text{MIN}(\text{DataNascita}) = 1-1-1960$, $\text{MAX}(\text{DataNascita}) = 31-12-1999$,
- $\text{card}(\text{COMMEDIA}) = 10^3$ tuple,
valori distinti di NumScene $\simeq 15$,
- $\text{card}(\text{ATTORE-IN-COMMEDIA}) = 10^6$ tuple,
valori distinti di Ruolo $\simeq 30$,
- $\text{card}(\text{PROGRAMMAZIONE}) = 10^8$ tuple,
 $\text{MIN}(\text{Data}) = 1-1-2010$, $\text{MAX}(\text{Data}) = 31-12-2010$,
 $\text{MIN}(\text{Durata}) = 81$, $\text{MAX}(\text{Durata}) = 180$,

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

- $\text{having count}(\text{DISTINCT Teatro}) \geq 50 \simeq \frac{1}{10}$.

Si consideri la seguente query SQL:

```
select Titolo, Regista
from COMMEDIA C, PROGRAMMAZIONE P, ATTORE-IN-COMMEDIA AC
where P.ComId=C.ComId and AC.ComId=C.ComId
      and Durata=180 and NumScene>12
      and AC.Aid in (select Aid from ATTORE
                    where DataNascita  $\geq$  1996)
group by ComId, Titolo, Regista
having count(DISTINCT Teatro)  $\geq$  50;
```

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. (7 Punti) Sono date le relazioni seguenti (le chiavi primarie sono sottolineate, gli attributi opzionali hanno l'asterisco):

```
DIPENDENTE(CodD, NomeD, Qualifica)
COSTO_ORARIO_DIPENDENTE(TipoAttività, Qualifica, CostoOrario)
TipoAttività={Ordinaria, Straordinaria}
RIEPILOGO_GIORNALIERO(CodD, Data, NumeroOreSvolte)
BUSTA_PAGA(CodD, Mese, ImportoComplessivo)
RICHIESTA_BUSTA_PAGA(CodR, CodD, Mese)
```

Si vuole gestire il calcolo delle buste paga mensili per i dipendenti di un'azienda.

Per i dipendenti dell'azienda, la retribuzione mensile è calcolata in base al costo orario del dipendente e al numero di ore svolte dal dipendente nel mese. Il costo orario può variare in base alla qualifica del dipendente e al tipo di attività svolta (ordinaria o straordinaria). Ogni mese, le ore svolte dal dipendente sono pagate come attività 'ordinaria' fino ad un massimo di 160 ore. Le ore svolte dal dipendente nel mese oltre le 160 ore, vengono pagate come attività 'straordinaria'.

La tabella DIPENDENTE riporta la qualifica dei dipendenti. La tabella COSTO_ORARIO_DIPENDENTE riporta il costo orario distinto per tipo di attività (ordinaria e straordinaria, attributo TipoAttività) e qualifica dei dipendenti. La tabella RIEPILOGO_GIORNALIERO riporta, per ogni dipendente, il numero di ore svolte in ciascuna data.

Si scrivano i trigger per gestire le seguenti attività.

(1) *Calcolo di una nuova busta paga.* Viene richiesto il calcolo di una nuova busta paga (un nuovo record è inserito nella tabella RICHIESTA_BUSTA_PAGA) per un dipendente in un dato mese. Per il calcolo della busta paga, devono essere conteggiate tutte le ore di attività ordinaria e straordinaria svolte dal dipendente nel mese considerato. Nella tabella BUSTA_PAGA deve essere inserito un nuovo record con le informazioni sulla busta paga calcolata. Il mese può essere ricavato dall'attributo Data nella tabella RIEPILOGO_GIORNALIERO utilizzando la funzione TOMONTH.

(2) *Gestione del vincolo di integrità sul giorno lavorativo.* Il vincolo di integrità richiede che, quando viene inserito un nuovo record nella tabella RIEPILOGO_GIORNALIERO, questo contenga il numero di ore di lavoro svolte dal dipendente nella *data corrente*. Non è possibile inserire un nuovo record contenente il numero di ore svolte in date antecedenti o successive a quella corrente. Se viene inserito un record relativo ad una data diversa da quella corrente, l'operazione di inserimento deve essere annullata. La data corrente può essere ricavata dalla data di sistema (funzione `sysdate`) utilizzando la funzione TODAY. Si scriva il trigger per la gestione del vincolo di integrità.

3. Progettazione Data Warehouse

Un'azienda di videosorveglianza gestisce sistemi di telecontrollo installati presso le numerose sedi di clienti in tutto il mondo. L'azienda intende progettare un datawarehouse in grado di raccogliere i dati provenienti dalle diverse sorgenti per rispondere in modo efficiente alle analisi da effettuare. Ogni sito da sorvegliare prevede l'installazione di alcune telecamere. Il sito può essere pubblico (es. parco) o privato (es. uffici) ed è di proprietà di un cliente dell'azienda.

Le telecamere registrano un video e lo elaborano per riconoscere le sagome delle persone, e sono in grado di associare ad ogni sagoma una delle seguenti classi: maschi adulti, femmine adulte, bambini. Ogni associazione ha un livello di confidenza relativo al riconoscimento (basso, medio, alto) che indica quanto il classificatore è sicuro della corretta identificazione della classe della sagoma. Nell'attuale base di dati dell'azienda, tali informazioni sono memorizzate per contare i passaggi delle persone (maschi, femmine, bambini) nei diversi punti in cui sono posizionate le telecamere. La posizione di ogni telecamera può essere distinta in: interna, esterna, all'ingresso, all'uscita. Infine, di ogni telecamera è noto il tipo di tecnologia che adotta (es. infrarossi, alta-risoluzione, ecc., un solo tipo per ogni telecamera) e per ogni riconoscimento è noto anche se è avvenuto con illuminazione da luce solare o con luce artificiale.

L'azienda di videosorveglianza è interessata ad analizzare, per ogni telecamera, il numero di passaggi di persone, di maschi, di femmine e di bambini in funzione di:

- posizionamento e tecnologia della telecamera.
- sito in cui si trova la telecamera, cliente a cui appartiene e tipologia di sito (pubblico o privato)
- città, regione, stato in cui si trova il sito
- livello di riconoscimento delle sagome (alto, basso, medio) e tipo di illuminazione (solare, artificiale)
- ora del giorno e fascia oraria (0-8, 8-12, 12-18 e 18-24)
- data, giorno del mese, giorno della settimana, mese, anno
- giorno festivo o feriale.

Il data warehouse realizzato deve contenere le informazioni relative agli anni 2006-2010. 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):

- l'azienda ha mediamente 10 telecamere per ogni sito
- sono presenti circa 10 siti in ogni città
- l'azienda ha installazioni presenti in circa 1000 città per ognuna delle 10 nazioni in cui è presente
- l'azienda ha 10 mila clienti
- le telecamere hanno 4 posizionamenti diversi e sono costruite con 4 tecnologie diverse

Sono riportate di seguito alcune delle interrogazioni frequenti di interesse per l'azienda, alle quali il datawarehouse deve poter rispondere in modo efficiente:

- (a) Calcolare la percentuale di maschi rispetto al totale delle persone riconosciute (inclusi bambini) e la percentuale di femmine rispetto al totale delle persone riconosciute (inclusi bambini) separatamente per ogni giorno della settimana e per ogni cliente. Considerare solo i dati dell'anno 2010 con un livello di riconoscimento "alto".
- (b) Calcolare la percentuale di persone nei giorni festivi rispetto al totale, separatamente per ogni sito e considerando solo il livello di riconoscimento "alto".
- (c) Per ogni mese, calcolare il totale cumulativo di persone da inizio anno, il totale mensile di persone e la percentuale di persone riconosciute nel mese rispetto al totale annuo, separatamente per ogni sito.

Progettazione

- (a) (7 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) (4 Punti) Esprimere l'interrogazione frequente (b) utilizzando il linguaggio SQL esteso.
- (c) (*Opzionale*: 5 Punti) Esprimere l'interrogazione frequente (c) utilizzando il linguaggio SQL esteso.