

## Basi di dati II

### Esercizi di autovalutazione — 26 aprile 2012

### Possibili soluzioni

**Domanda 1** Considerare uno schema dimensionale relativo agli esami, che utilizzi, come tabella dei fatti e come una delle dimensioni, le relazioni come quelle qui schematizzate:

| <u>KStudente</u> | <u>KCorso</u> | <u>KData</u> | Voto | ... |
|------------------|---------------|--------------|------|-----|
| 301              | 201           | 405          | 25   | ... |
| 301              | 202           | 406          | 28   | ... |
| 302              | 201           | 405          | 30   | ... |
| 302              | 203           | 407          | 22   | ... |
| ...              | ...           | ...          | ...  | ... |

| <u>KCorso</u> | Titolo    | Crediti | ... |
|---------------|-----------|---------|-----|
| 201           | Fisica I  | 6       | ... |
| 202           | Chimica   | 9       | ... |
| 203           | Geometria | 6       | ... |
| ...           | ...       | ...     | ... |

Supporre che si presentino le seguenti esigenze di modifica:

- i corsi cambiano nome nel tempo: per esempio, il corso nella prima ennupla potrebbe ad un certo punto cambiare nome da “Fisica I” in “Meccanica”; interessano selezioni e aggregazioni relative agli esami tanto con riferimento al nome del corso (al momento dell’esame) quanto alla sua identità (un codice che viene introdotto allo scopo, ma non sempre viene utilizzato, perché alcuni analisti preferiscono fare riferimento al nome corrente del corso); le modifiche sono rare, ma è possibile che ci siano corsi con vari cambiamenti di nome;
- per ogni corso, interessa rappresentare anche il docente, per supportare analisi sugli esami svolti da ciascun docente; i docenti cambiano nel tempo e passano da un corso all’altro (e possono anche tenere più corsi nello stesso momento, ma ogni corso ha, in un certo giorno, un solo docente); è disponibile l’informazione relativa ai docenti dei corsi nel tempo (per tutto il periodo, anche passato, di interesse).

Mostrare nuove versioni delle due tabelle che permettano di soddisfare le esigenze sopra citate (mostrare anche i dati, con riferimento a quelli presenti negli esempi sopra, aggiungendo nuovi dati ragionevoli, che permettano di comprendere le modifiche).

La tabella dei fatti può essere estesa aggiungendo una dimensione supplementare, che dipende da Corso e Data e può quindi essere facilmente aggiunta anche ad una tabella dei fatti esistente (i dati, come detto, sono disponibili e le ennuple rimangono le stesse, con un attributo in più; si noti che Kdocente dipende funzionalmente da KCorso e KData):

| <u>KStudente</u> | <u>KCorso</u> | <u>KData</u> | <u>KDocente</u> | Voto | ... |
|------------------|---------------|--------------|-----------------|------|-----|
| 301              | 201           | 405          | 701             | 25   | ... |
| 301              | 202           | 406          | 701             | 28   | ... |
| 302              | 201           | 405          | 701             | 30   | ... |
| 302              | 203           | 407          | 703             | 22   | ... |
| ...              | ...           | ...          | ...             | ...  | ... |

Per la dimensione può essere utile la tecnica della “slowly changing dimension” di tipo 2 (e contemporaneamente 3), con un nuovo elemento (quindi una ennupla nella relazione) per ogni modifica e, viste le specifiche, un codice aggiuntivo, che non cambia nel tempo, e due attributi per il nome, con quello attuale e quello “dell’epoca,” rispettivamente.

| <u>KCorso</u> | Codice | Titolo    | TitoloAttuale | Crediti | ... |
|---------------|--------|-----------|---------------|---------|-----|
| 201           | FIS01  | Fisica I  | Meccanica     | 6       | ... |
| 202           | CHIM01 | Chimica   | Chimica       | 9       | ... |
| 203           | GEOM01 | Geometria | Geometria     | 6       | ... |
| ...           | ...    | ...       | ...           | ...     | ... |
| 209           | FIS01  | Meccanica | Meccanica     | 6       | ... |

**Domanda 2** Una catena di negozi gestisce le attività utilizzando, in ciascun negozio, una base di dati con le seguenti relazioni:

- Prodotti(CodiceProdotto,Descrizione,Prezzo,Categoria)
- Categorie(Codice,Descrizione,MacroCategoria)
- MacroCategorie(Codice,Descrizione)
- Vendite(NumeroScontrino,Ora)
- DettaglioVendite(NumeroScontrino,CodiceProdotto,Quantità)

Si noti che

- Le informazioni relative alle vendite vengono mantenute solo nel corso della giornata.
- Il prezzo di un prodotto può variare da un giorno all'altro.

Utilizzando tali informazioni, la catena vuole realizzare un data mart relativo alle vendite dei prodotti, avente come misure le quantità vendute e gli incassi, che permetta di effettuare analisi di tipo temporale (incluse, oltre ai giorni, anche le fasce orarie della giornata, ad esempio 9-10, 10-11 e così via, oppure mattina e pomeriggio) e su prodotti (singoli e per categoria) e sui negozi. Allo scopo:

1. specificare un possibile dettaglio del data mart; al riguardo, si supponga che la quantità delle vendite sia tale che si è deciso di non utilizzare il massimo livello di dettaglio, ma solo quello strettamente indispensabile (in altri termini, la grana non deve essere il singolo dettaglio di vendita, ma una opportuna aggregazione; **specificare esplicitamente la grana scelta**)
2. indicare quali dati debbono provenire da fonti diverse rispetto alle basi di dati con lo schema sopra mostrato
3. specificare la struttura dell'interrogazione SQL necessaria per produrre, giornalmente, le nuove entuple da inserire nella tabella dei fatti

Schema dimensionale

- FattiVendite(KData, KFasciaOraria, KProdotto, KNegozio, Quantità, Importo)
- Prodotti(KProdotto, CodiceProdotto, Descrizione, CodCategoria, DescrizioneCategoria, CodMacroCategoria, ...)
- Negozi(KNegozio, ...)
- FasciaOraria(KFasciaOraria, ...)
- Data(KData, ...)

Commenti:

- la grana è l'insieme delle vendite per prodotto, negozio, data e fascia oraria
- in alternativa, si potrebbe usare un'unica dimensione per data e fascia oraria
- sono indicate chiavi ad hoc per le dimensioni
- *KNegozio* e *KData* vengono inseriti nella fase di caricamento quotidiana (e vengono da dati presumibilmente nella staging area)

Per l'interrogazione SQL (o, comunque, per il processo di alimentazione), sono necessari due tipi di attività, uno relativo alla conversione delle chiavi e l'altro al calcolo delle misure, con riferimento alle chiavi. Si può pensare, ad esempio, di specificare un'interrogazione che calcola quantità e incasso per ogni prodotto e ora (supponendo che il campo Ora della relazione Vendite contenga il solo valore dell'ora; se invece contenesse ore e minuti si dovrebbe fare una conversione, con la funzione HOUR di SQL)

```
SELECT Ora, CodiceProdotto, Sum(Quantità) AS Quantità, SUM(Quantità*Prezzo) As Importo
FROM Vendite V JOIN DettaglioVendite D ON D.NumeroScontrino=DV.NumeroScontrino
      JOIN Prodotti P ON D.CodiceProdotto=P.CodiceProdotto
GROUP BY Ora, CodiceProdotto
```

Per la gestione delle chiavi nel datamart, si può pensare di mantenere (aggiungendo record quando necessario) "tabelle di corrispondenza," che contengano almeno la chiave usata nella base di dati e quella del data mart, ad esempio

- CorrProdotti(KProdotto, CodiceProdotto)
- CorrFasciaOraria(KFasciaOraria, Ora)

A questo punto, l'interrogazione per la generazione dei fatti può, per ogni negozio e ogni giorno, prevedere il join della relazione ottenuta come risultato per l'interrogazione sopra mostrata con le tabelle di corrispondenza, con due opportune costanti per Kdata e KNegozio.