

Politecnico di Torino

Sistemi per la gestione di basi di dati

28 Febbraio 2012

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

E-SERVICE(E-ServiceID, Nome, URL, Società)
TASK(TaskID, E-ServiceID, Tipo, Priorità)
SERVER(ServerID, #Core, Memoria, Hard_Disk, IP_address)
SCHEDULING(Data, Timestamp, TaskID, ServerID, Tempo_CPU, Uso_memoria)

Si ipotizzino le seguenti cardinalità per le tabelle:

- $\text{card}(\text{E-SERVICE}) = 10^4$ tuple,
valori distinti di Società $\simeq 100$,
- $\text{card}(\text{TASK}) = 10^6$ tuple,
valori distinti di Tipo $\simeq 10$,
valori distinti di Priorità $\simeq 5$
- $\text{card}(\text{SERVER}) = 10^3$ tuple,
 $\text{MIN}(\#Core) = 1$, $\text{MAX}(\#Core) = 16$,
 $\text{MIN}(\text{Memoria}) = 2$ GB, $\text{MAX}(\text{Memoria}) = 32$ GB,
 $\text{MIN}(\text{Hard_Disk}) = 100$ GB, $\text{MAX}(\text{Hard_Disk}) = 1$ TB
- $\text{card}(\text{SCHEDULING}) = 10^{10}$ tuple,
 $\text{MIN}(\text{Data}) = 1/1/2011$, $\text{MAX}(\text{Data}) = 31/12/2011$.

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

- $\text{having avg}(\text{Uso_memoria}) < 100$ MB $\simeq \frac{1}{10}$.

Si consideri la seguente query SQL:

```
select Nome, URL
from TASK T, E-SERVICE E
where E.E-ServiceID=T.E-ServiceID
      Priorità <> 'Bassa' and Società = 'ItalBank'
      and TaskID in (select TaskID
                    from SERVER S, SCHEDULING SC
                    where S.ServerID=SC.ServerID
                      and Data  $\geq$  1/8/2011 and Data  $\leq$  30/8/2011
                      and Memoria = 32 GB and #Core  $\leq$  4
                    group by TaskID
                    having avg(Uso_memoria) < 100 MB)
```

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 hanno l'asterisco):

RIUNIONE(CodR, Argomento, NumeroInvitati, DataRiunione*, CodS*, NumeroPartecipanti*)
DATE_PROPOSTE_PER_RIUNIONE (CodR, Data, NumeroDisponibili)
SALA (CodS, NomeSala, Locazione, Capienza)
DISPONIBILITÀ_PERSONA (CodImpiegato, CodR, Data)

Si vogliono gestire alcune attività relative alla organizzazione delle riunioni presso una società. La tabella RIUNIONE riporta le informazioni sulle riunioni organizzate presso la società. La tabella SALA contiene le sale disponibili presso la società.

Selezione della data e della sala per una riunione. Si scriva il trigger per definire la data di una riunione, la sala in cui si terrà la riunione, e il numero di persone che parteciperanno alla riunione nella data selezionata (attributi DataRiunione, CodS, NumeroPartecipanti nella tabella RIUNIONE).

Per ogni riunione che si vuole organizzare presso la società sono proposte alcune date possibili (tabella DATE_PROPOSTE_PER_RIUNIONE). Ciascuna persona invitata ad una riunione esprime la propria disponibilità per *una sola data* tra quelle proposte per la riunione. Quando una persona esprime la sua disponibilità (inserimento nella tabella DISPONIBILITÀ_PERSONA), si devono svolgere le seguenti operazioni. (a) Si deve aggiornare il numero di disponibilità per quella data (attributo NumeroDisponibili). (b) Si deve verificare se tutte le persone invitate alla riunione (attributo NumeroInvitati) hanno espresso la loro disponibilità per una delle date proposte per la riunione. In questo caso si devono selezionare la data e la sala in cui si terrà la riunione. In particolare, tra le date proposte per la riunione viene selezionata la data per cui hanno espresso la disponibilità almeno 80% delle persone invitate. Viene quindi selezionata una sala che non sia già occupata per altre riunioni nella data selezionata ed abbia una capienza tale da poter ospitare tutte persone che hanno espresso la disponibilità per la data selezionata. Si consideri che ci sia sempre al più una sola data ed una sola sala che verifichino tutte queste condizioni. Si deve quindi aggiornare la tabella RIUNIONE. Nel caso non sia possibile selezionare una data con un numero sufficiente di disponibilità, o una sala libera con capienza adeguata, il trigger termina senza svolgere ulteriori attività.

3. Progettazione Data Warehouse

Una società internazionale che distribuisce pezzi di ricambio automobilistico (per esempio lampadine, sospensioni, freni, etc.) vuole valutare l'efficienza e l'efficacia di distribuzione delle proprie filiali sparse nel mondo, analizzando il fatturato di ogni filiale e le loro politiche di distribuzione. Un ricambio può essere utilizzato per un solo modello di automobile, mentre lo stesso modello di automobile può utilizzare più ricambi. Inoltre un ricambio può essere distribuito da più filiali e la stessa filiale può distribuire più ricambi.

Attualmente ogni filiale ha una base di dati indipendente per gestire la propria attività economica, ma la società è interessata a progettare e realizzare un data warehouse che offra una visione d'insieme che consenta di analizzare il fatturato medio giornaliero e il fatturato medio per ricambio delle proprie filiali.

L'analisi deve essere effettuata in funzione di:

- data, mese, trimestre, quadrimestre, semestre e anno in cui è stato realizzato il fatturato,
- giorno festivo, giorno della settimana, giorno del mese,
- filiale,
- modalità di pagamento, che include il periodo di tempo entro il quale viene effettuato (e.g., a 30 giorni, a 60 giorni, a 90 giorni, etc.) e il tipo di pagamento (e.g., bonifico, contrassegno, assegno, etc.)
- ricambio,
- modello di automobile (e.g., Musa, Cinquecento, Touran) su cui può essere utilizzato il ricambio,
- marca di automobile (e.g., FIAT, Peugeot, Renault) su cui può essere utilizzato il ricambio,
- tipo e marca del ricambio,
- città, regione, stato e continente in cui è sita la filiale.

Il data warehouse conterrà le informazioni relative agli anni 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011. 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) Considerando solo le filiali site in Italia, visualizzare per ogni filiale e modello di automobile, il fatturato mensile, il fatturato medio giornaliero per ogni mese, e il fatturato totale mensile cumulativo dall'inizio dell'anno, separatamente per ogni tipo di pagamento.
- (b) Per ogni ricambio, visualizzare il fatturato totale annuo di ogni filiale cinese.
- (c) Per ogni ricambio per le automobili FIAT e considerando solo il fatturato delle filiali europee del 2011, visualizzare per ogni città in cui è sita una filiale il fatturato mensile, il numero di ricambi distribuiti ogni mese, il fatturato totale per ogni regione in cui è sita la filiale.
- (d) Per ogni trimestre del 2006 e 2007, visualizzare il fatturato totale e il fatturato medio per ricambio distribuito, separatamente per ogni tipo di pagamento.

Il data warehouse conterrà le informazioni relative agli anni 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011.

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.