

Politecnico di Torino

Sistemi per la gestione di basi di dati

7 settembre 2012

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

UTENTE (UID, Nickname, Nome, Cognome, Nazionalità, Data_Nascita, E-Mail)
ARTICOLO_IN_BLOG(BID, Titolo, Testo, UID, Data_Pubblicazione, Lingua)
COMMENTO(BID, UID, Timestamp, Testo, Voto)
ACCESSI_GIORNALIERI(BID, Data, #Accessi)

Si ipotizzino le seguenti cardinalità:

- $\text{card}(\text{UTENTI}) = 10^4$ tuples,
valori distinti di Nazionalità $\simeq 10$,
 $\text{MIN}(\text{Data_Nascita}) = 1/1/1945$, $\text{MAX}(\text{Data_Nascita}) = 31/12/1994$,
- $\text{card}(\text{ARTICOLO_IN_BLOG}) = 10^7$ tuples,
 $\text{MIN}(\text{Data_Pubblicazione}) = 1/1/2009$, $\text{MAX}(\text{Data_Pubblicazione}) = 31/12/2011$,
valori distinti di Lingua $\simeq 10$,
- $\text{card}(\text{COMMENTO}) = 10^{10}$ tuples,
 $\text{MIN}(\text{TO_DATE}(\text{Timestamp})) = 1/1/2009$, $\text{MAX}(\text{TO_DATE}(\text{Timestamp})) = 31/12/2011$,
 $\text{MIN}(\text{Voto}) = 1$, $\text{MAX}(\text{Voto}) = 10$,
- $\text{card}(\text{ACCESSI_GIORNALIERI}) = 10^9$ tuples,
 $\text{MIN}(\text{Data}) = 1/1/2009$, $\text{MAX}(\text{Data}) = 31/12/2011$,
 $\text{MIN}(\#\text{Accessi}) = 1$, $\text{MAX}(\#\text{Accessi}) = 1000$.

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

- $\text{having AVG}(\text{Vote}) > 7 \simeq \frac{1}{10}$.

Si consideri la seguente query SQL:

```
select Titolo
from ARTICOLO_IN_BLOG B, ACCESSI_GIORNALIERI A
where B.BID=A.BID and
      A.Data_Pubblicazione >= 1/1/2011 and Lingua='Italiano' and
      exist (select * from COMMENTO C, UTENTE U
            where C.UID=U.UID and
                  and TO_DATE(Timestamp) ≥ 1/1/2011
                  and U.Nazionalità = 'Italia' and Data_Nascita ≥ 1/1/1985
                  and U.UID=B.UID
            group by UID
            having AVG(Vote)>7);
```

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 *):

ATTIVITÀ (CodA, NomeA, TipoAttività, Data, FasciaOraria, FasciaEtà, CodS, NumeroEducatori)
SALA (CodS, NomeS, Capienza)
INTERVALLO_FASCIA_ETÀ (FasciaEtà, EtàMinima, EtàMassima)
INTERVALLO_FASCIA_ORARIA (FasciaOraria, OraInizio, OraFine)
BAMBINO (CodFiscale, NomeB, Età)
ISCRIZIONE (CodFiscale, CodA)
RICHIESTA_ISCRIZIONE(CodR, CodFiscale, CodA)

Si scrivano i trigger per gestire le seguenti attività relative ad un baby parking. La tabella ATTIVITÀ elenca le attività previste. Le attività si tengono in ogni giorno in diverse fasce orarie (attributo FasciaOraria) e variano in base all'età dei bambini (attributo FasciaEtà).

(1) *Iscrizione ad una attività.* Viene richiesta l'iscrizione ad una attività per un bambino (inserimento nella tabella RICHIESTA_ISCRIZIONE). L'iscrizione è possibile solo se sono soddisfatte tutte le seguenti condizioni. (a) Il bambino soddisfa i requisiti di età minima e massima richiesti per quella attività. (b) Il numero di iscrizioni già effettuate per quella attività (tabella ISCRIZIONE) è inferiore alla capienza della sala in cui si terrà l'attività (tabella SALA). (c) Il bambino non è già iscritto ad una attività nella stessa data e fascia oraria. Se almeno una delle condizioni non è verificata, la richiesta di iscrizione deve essere annullata.

(2) *Vincolo di integrità sulla pianificazione delle attività.* Ciascun tipo di attività può essere proposto, per ogni fascia di età, al più 3 volte nella stessa giornata. Ogni operazione di modifica della tabella ATTIVITÀ che causi la violazione del vincolo deve essere annullata. Si scriva il trigger che gestisce il vincolo d'integrità.

3. Progettazione Data Warehouse

L'Ente Nazionale dell'Aviazione Civile (ENAC) effettua i monitoraggi di tutti i voli aerei. L'Ente è interessato a valutare l'efficienza e l'efficacia delle diverse compagnie aeree analizzando i tempi di volo, il numero di passeggeri e i relativi incassi per ogni percorso aereo. Per effettuare tali analisi si deve realizzare un data warehouse in grado di memorizzare tutte le informazioni utili per analizzare in modo efficiente:

- in numero medio di passeggeri per volo,
- il tempo medio di viaggio per volo,
- l'incasso medio giornaliero per volo.

L'analisi deve essere effettuata in funzione di:

- data, giorno della settimana, giorno dell'anno (1-365), mese, bimestre, trimestre, semestre, anno,
- compagnia aerea,
- modello di aereo e casa costruttrice,
- configurazione di servizi di comfort di ogni modello (i.e., TV, poltrona di prima classe, etc.)
- aeroporto, città, provincia, regione, stato, continente sia di partenza sia di destinazione.

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 voli effettuati nel 2011 con aerei BOEING 747, per ogni compagnia aerea, visualizzare il numero mensile di voli, il numero mensile di passeggeri, e l'incasso mensile cumulativo dall'inizio dell'anno, separatamente per ogni stato di partenza.
- (b) Considerando solo gli aeroporti europei, per ogni città di partenza localizzata in Italia, per ogni città di arrivo e per ogni mese, visualizzare l'incasso medio giornaliero, il numero medio di passeggeri per volo, il tempo medio di viaggio per volo, l'incasso totale, e il numero totale di passeggeri per città di partenza, separatamente per ogni compagnia aerea.
- (c) Per ogni modello di aereo con poltrone di prima classe, visualizzare il numero totale di voli e il numero totale di passeggeri, separatamente per ogni compagnia aerea. Ordinare e assegnare un rank in ordine crescente di numero totale di voli.

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 (b) utilizzando il linguaggio SQL esteso.