

Sistemi per la Gestione delle Basi di Dati

23 febbraio 2016

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

```
CLIENTE(CIId, Nome, Cognome, Email, Città, Regione, Data_Nascita)
ACQUISTO(Timestamp, CIId, Negozio, Importo, Cashback, #Oggetti)
DONAZIONE(Timestamp, CIId, Importo, IIId)
INIZIATIVA(IIId, Nome, Descrizione, Città, Regione, Categoria)
```

Si ipotizzino le seguenti cardinalità:

- $\text{card}(\text{CLIENTE}) = 10^5$ tuple,
valori distinti di città = 500,
valori distinti di regione = 21,
 $\text{MIN}(\text{Data_Nascita}) = 1/1/1945$, $\text{MAX}(\text{Data_Nascita}) = 31/12/1994$,
- $\text{card}(\text{ACQUISTO}) = 10^8$ tuple,
 $\text{MIN}(\text{TO_DATE}(\text{Timestamp})) = 1/1/2014$, $\text{MAX}(\text{TO_DATE}(\text{Timestamp})) = 31/12/2015$,
numero di negozi = 10^3 ,
 $\text{MIN}(\text{Importo}) = 1$ Euro, $\text{MAX}(\text{Importo}) = 1000$ Euro
 $\text{MIN}(\text{Cashback}) = 0$ Euro, $\text{MAX}(\text{Cashback}) = 10$ Euro
- $\text{card}(\text{DONAZIONE}) = 10^4$ tuple,
 $\text{MIN}(\text{Importo}) = 1$ Euro, $\text{MAX}(\text{Importo}) = 100$ Euro
 $\text{MIN}(\text{TO_DATE}(\text{Timestamp})) = 1/1/2014$, $\text{MAX}(\text{TO_DATE}(\text{Timestamp})) = 31/12/2015$,
- $\text{card}(\text{AZIONE-DI-SOSTEGNO}) = 5 \cdot 10^2$ tuple,
valori distinti di città = 50,
valori distinti di regione = 21,
numero di categorie = 10

Inoltre, ipotizzare il seguente fattore di riduzione per la condizione group by:

- $\text{having count}(\ast) \leq 100 \simeq \frac{1}{10}$.

Si consideri la seguente query SQL:

```
select Regione, SUM(Importo)
from DONAZIONE D, INIZIATIVA I
where D.IIId = I.IIId and Categoria = 'Animali'
and ((TO_DATE(D.Timestamp) ≥ 1-12-2014 and TO_DATE(D.Timestamp) ≤ 31-12-2014)
or (TO_DATE(D.Timestamp) ≥ 1-12-2015 and TO_DATE(D.Timestamp) ≤ 31-12-2015))
and D.CId NOT IN (select C.CId
from ACQUISTO A, CLIENTE C
where A.CId = C.CId and A.Cashback ≤ 1 and C.Data_Nascita ≥ 1-1-1985
group by C.CId
having count(\ast) ≤ 100)
group by Regione;
```

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 *).

```
LIBRO(ISBN, Editore, DataPubblicazione)
AUTORE(CodAutore, Nome, Cognome, DataNascita)
TOTALE_VENDITE_LIBRERIA(CodLibreria, ISBN, TotaleNumCopieVendute, IncassoComplessivo)
INCASSO_AUTORE(ISBN, CodAutore, ContributoAutore, IncassoAutore)
VENDITA_LIBRO(CodLibreria, ISBN, Data, NumCopieVendute, Incasso)
LIBRO_PIÙ_VENDUTO(CodLibreria, Timestamp, ISBN, IncassoComplessivo, NumeroAutori)
```

Si vogliono gestire in modo automatico le informazioni relative alle vendite di libri. La tabella `LIBRO` elenca i libri in vendita presso le diverse librerie. Per ciascuna libreria, la tabella `TOTALE_VENDITE_LIBRERIA` riporta il numero di copie complessivamente vendute per ciascun libro e l'incasso conseguito. Ogni libro può avere uno o più autori. Per ogni libro la tabella `INCASSO_AUTORE` riporta l'elenco degli autori del libro. A ciascun autore del libro spetta un incasso proporzionale al suo contributo al libro (attributo `ContributoAutore` nella tabella `INCASSO_AUTORE`, attributo numerico con valori tra 0 e 1). L'attributo `IncassoAutore` indica l'incasso ottenuto dall'autore sul libro, ed è calcolato come prodotto dell'incasso conseguito sul libro moltiplicato per il contributo dell'autore al libro.

Si scrivano i trigger per eseguire le seguenti operazioni.

(1) *Aggiornamento delle vendite di un libro in una libreria e dell'incasso per i co-autori del libro.* Ogni giorno sono memorizzate le informazioni sulle vendite giornaliere di un libro in una libreria nella tabella `VENDITA_LIBRO` (inserimento di un record). Le seguenti operazioni devono quindi essere eseguite.

(a) Si deve aggiornare nella tabella `TOTALE_VENDITE_LIBRERIA` il numero complessivo di copie del libro vendute presso la libreria e l'importo complessivamente incassato dalla libreria per quel libro. Per ciascuna libreria è già presente nella tabella `TOTALE_VENDITE_LIBRERIA` il record relativo ad un libro solo per i libri per cui è stata precedentemente venduta almeno una copia.

Si deve poi registrare (inserimento nella tabella `LIBRO_PIÙ_VENDUTO`) il libro che al momento ha conseguito l'incasso maggiore nella libreria. Si ipotizzi che ci sia al più un libro che ha conseguito tale valore di incasso nella libreria. Per tale libro si deve registrare l'incasso complessivo conseguito nella libreria e il numero di autori del libro. Il valore dell'attributo `Timestamp` è assegnato mediante la variabile di sistema `sysdate`.

(b) Si deve aggiornare l'incasso complessivo per tutti gli autori del libro (aggiornamento dell'attributo `IncassoAutore`) tenendo conto delle nuove copie vendute.

(2) *Vincolo di integrità sul contributo degli autori di un libro.* La somma dei contributi (attributo `ContributoAutore`) dei diversi autori ad uno stesso libro deve sempre essere minore o uguale a 1. Ogni operazione di modifica della tabella `INCASSO_AUTORE` che causi la violazione del vincolo non deve essere eseguita.

3. Progettazione Data Warehouse

Il Ministero dell'Università e della Ricerca vuole svolgere un'analisi del tempo dedicato dai ricercatori ai vari progetti di ricerca. Ogni ora dedicata alle attività di un progetto di ricerca da parte di un ricercatore ha un corrispettivo costo orario nominale.

Ogni progetto di ricerca è costituito da insiemi disgiunti di attività denominati Work Package. Le attività di un progetto di ricerca appartengono ad uno dei seguenti tre tipi: Management, Ricerca & Sviluppo, Altro. Ogni Work Package raggruppa attività omogenee tra loro per tipo.

I progetti di ricerca sono suddivisi per argomento (es. Big Data, Smart City); ogni argomento appartiene ad una disciplina (es. informatica, urbanistica). Ogni progetto di ricerca è identificato da un nome, ha una durata prefissata (uno o più anni), e nasce in risposta ad un bando di finanziamento. Ogni bando ha una specifica modalità di finanziamento (es. fondo perduto, cofinanziamento, venture capital), è pubblicato da uno specifico ente finanziatore (es. Comunità Europea, Ministero, fondazioni), e appartiene ad uno specifico settore (es. ICT, Nanotechnologies, Energy).

I singoli ricercatori compilano ogni mese un resoconto, denominato timesheet, con l'indicazione di quante ore di attività hanno svolto per ogni Work Package di ogni progetto nel quale sono coinvolti. I singoli ricercatori appartengono a delle Unità di Ricerca (es. i dipartimenti universitari), le quali a loro volta appartengono ad uno specifico Ente di Ricerca (es. Politecnico di Torino, Istituto Superiore Mario Boella).

Il Ministero vuole analizzare il numero totale di ore, il costo totale, e il costo medio orario, al variare di:

- progetto e relativi Work Packages;
- tipo di attività, argomento del progetto e disciplina coinvolta;
- durata del progetto (i valori di interesse sono 1, 2, 3, 4, 5, o più anni);
- bando e settore del bando;
- modalità di finanziamento ed ente finanziatore;
- Unità di Ricerca ed Ente di Ricerca;
- mese, mese dell'anno, bimestre, trimestre, semestre, anno, anno accademico (inizia a settembre, finisce ad agosto).

Progettazione

- (a) (6 Punti) Progettare il data warehouse (modello concettuale e tabelle dei fatti e delle dimensioni) in modo da soddisfare le richieste descritte nelle specifiche del problema. Il data warehouse progettato deve inoltre permettere di rispondere in modo efficiente alle seguenti interrogazioni.
- (b) (8 Punti) Esprimere le due interrogazioni seguenti utilizzando il linguaggio SQL esteso. Indicare la risposta nello spazio sottostante ciascuna domanda.

(a) Per ogni bimestre e per ogni progetto, calcolare il costo medio orario e il numero medio di ore per Work Package. Calcolare inoltre il cumulativo per ogni anno accademico dei costi totali bimestrali, separatamente per ogni progetto.

(b) Calcolare il costo orario medio per ogni Work Package, la percentuale di ore di ogni Work Package rispetto al totale di ogni progetto, e la percentuale di ore che ogni progetto ha dedicato ai vari tipi di attività rispetto al totale di ogni progetto.