

Progettazione di Data Warehouse

L'Unione Europea nell'ambito della valutazione dei risultati della ricerca scientifica finanziata con fondi pubblici ha deciso di analizzare le pubblicazioni prodotte dai ricercatori delle università europee.

I ricercatori e i professori delle università europee presentano i risultati delle proprie attività di ricerca scrivendo degli articoli scientifici, denominati genericamente pubblicazioni. Ogni pubblicazione ricade in una tipologia specifica (es. articolo a conferenza, articolo di giornale, capitolo di libro, ecc.) ed è caratterizzata da una data specifica di pubblicazione, da uno o più autori e da un luogo di pubblicazione (es. conferenza, workshop, rivista, libro, ecc.), il quale ne determina la tipologia specifica. Il luogo di pubblicazione (rivista, conferenza, ecc.) è tipicamente caratterizzato da una o più edizioni, delle quali si è interessati all'anno (es. conferenza internazionale delle Basi di Dati edizione 2015, conferenza internazionale delle Basi di Dati edizione 2016, ecc.) e da un editore (es. Elsevier, ecc.).

Tra gli autori della pubblicazione è identificato un autore principale, il quale aderisce ad uno specifico dipartimento universitario. Ogni dipartimento appartiene ad un campus universitario e ciascuna università può essere formata da uno o più campus. Inoltre, le università sono suddivise per dimensione (piccole, medie, grandi, in base al numero di ricercatori che vi lavorano). Ogni singolo dipartimento è infine caratterizzato da un settore scientifico-disciplinare di afferenza. Per esempio, il Dipartimento di Automatica e Informatica afferisce al settore scientifico-disciplinare ING-INF/05, fa parte del campus Cittadella Politecnica, a sua volta appartenente all'università Politecnico di Torino.

L'Unione Europea è interessata ad analizzare il numero di pubblicazioni al variare di:

- mese, bimestre, trimestre, semestre, anno della pubblicazione;
- anno accademico (da settembre ad agosto);
- mese di sospensione delle attività didattiche (luglio e agosto di tutti gli anni);
- mese dell'anno;
- dipartimento dell'autore principale, relativo campus e università;
- dimensione dell'università, stato europeo dell'università,
- settore scientifico-disciplinare di afferenza del dipartimento;
- numero di autori (da 1 a 10 con incrementi singoli, e una categoria per più di 10 autori);
- luogo in cui è avvenuta la pubblicazione (conferenza, rivista, ecc.), edizione (l'anno in cui si è tenuta la conferenza o è stata pubblicata la rivista), editore, e tipo della pubblicazione (articolo a conferenza, articolo di giornale, capitolo di libro, ecc.).

Progettazione

1. 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.
2. Esprimere le due interrogazioni seguenti utilizzando il linguaggio SQL esteso. Indicare la risposta nello spazio sottostante ciascuna domanda.
 - a. Per ogni università europea, per ogni tipo di pubblicazione, e per ogni anno, calcolare il numero medio mensile di pubblicazioni, e il totale cumulativo annuale delle pubblicazioni. Si considerino solo i mesi in cui c'è stata almeno una pubblicazione.
 - b. Per ogni anno, calcolare la percentuale di pubblicazioni del dipartimento rispetto al totale della sua università. Assegnare inoltre un rank globale ai dipartimenti di tutte le università separatamente per ogni anno per numero di pubblicazioni totali (il primo dipartimento è quello con il maggior numero di pubblicazioni).

3. Creare e mantenere aggiornate una vista materializzata con i comandi CREATE MATERIALIZED VIEW e CREATE MATERIALIZED VIEW LOG di ORACLE.
 - a. Considerare le seguenti query di interesse:
 - i. Considerando solo le università site in Italia e Francia e il periodo di pubblicazione 2013-2015, visualizzare per ogni Campus universitario e semestre il numero totale di pubblicazioni.
 - ii. Considerando solo il settore scientifico disciplinare "sistemi di elaborazione delle informazioni", per ogni coppia (anno, Università) visualizzare il numero totale di pubblicazioni.
 - iii. Visualizzare per ogni trimestre e Stato il numero totale di pubblicazioni separatamente per settore scientifico disciplinare.
 - b. Definire la query in SQL per una vista materializzata definita con CREATE MATERIALIZED VIEW, in modo da ridurre il tempo di risposta delle query elencate al punto 3a. Utilizzare la sintassi riportata di seguito


```
CREATE MATERIALIZED VIEW VM1
BUILD IMMEDIATE
RERESH FAST ON COMMIT
ENABLE QUERY REWRITE
AS
  Query
```
 - c. Definire i log della vista materializzata con CREATE MATERIALIZED VIEW LOG (esempio di istruzione riportato di seguito), per ogni tabella in cui lo si ritiene necessario. Per quali tabelle è utile tenere traccia dei log? Si individuino tutte e sole le tabelle necessarie. Inoltre, per ogni tabella si individuino tutti e soli gli attributi per cui è necessario tener traccia delle variazioni.


```
CREATE MATERIALIZED VIEW LOG on NomeTabella
WITH SEQUENCE, ROWID
(Lista di attributi)
INCLUDING NEW VALUES;
```
4. Aggiornamento e gestione delle viste tramite Trigger supponendo che il comando CREATE MATERIALIZED VIEW non sia disponibile. Creare la vista materializzata definita al punto 3 e definire la procedura di aggiornamento a partire da modifiche sulla tabella dei fatti realizzata tramite trigger. Nello specifico,
 - a. Scrivere l'istruzione SQL per popolare opportunamente la tabella creata con l'istruzione CREATE TABLE VM1 (...) nella seguente istruzione


```
INSERT INTO VM1 (...) ( SELECT ... ... )
```
 - b. Scrivere il trigger necessario per propagare le modifiche (inserimento di un nuovo record) effettuate nella tabella dei FATTI alla vista materializzata VM1.