

Business intelligence

Esercizio di progettazione di un data warehouse

Descrizione del problema

Un network televisivo gestisce diversi canali tematici e vorrebbe analizzare i dati sugli ascolti nei programmi per ottimizzare il palinsesto e i profitti della pubblicità. I canali sono caratterizzati da un nome ("CartoonNetwork", "TuttoTennis", "Natura&Co.",...) e trattano un tema specifico (film, sport, politica, economia,...).

Il network vorrebbe, incrociando i dati del palinsesto dei suoi canali con i dati forniti dall'Auditel, capire quali sono i momenti della giornata in cui è più proficuo inserire le pubblicità pagate dagli sponsor.

L'Auditel fornisce ad intervalli di 15 minuti, per ogni canale il numero massimo e minimo di ascoltatori sintonizzati. In più, per ogni singolo programma tv, l'Auditel fornisce lo share (percentuale di spettatori sul totale) del programma e una statistica sulla provenienza geografica (provincia) degli ascoltatori.

La società è interessata ad analizzare gli andamenti del numero massimo e minimo di spettatori e i minuti di pubblicità trasmessi in funzione di:

- l'ora del giorno, il giorno (e se è festivo o no), il giorno della settimana e il mese,anno
- il quarto d'ora del giorno e la mezz'ora del giorno
- il periodo della giornata: notte (dalle 0 alle 6:30), mattino (6:30-12:30), pomeriggio (12:30-19:30), sera (19:30, 0)
- della fascia oraria consigliata: adulti (dalle 0 alle 7), bambini (7-15), tutti (15-0)
- del canale televisivo e del tema del canale

In più per ogni programma tv si è interessati ad analizzare lo share ottenuto, i minuti totali di pubblicità trasmessi e il numero di spot trasmessi in funzione di:

- nome del programma e il canale che lo trasmette
- tipo di programma (notiziario, talk show, evento, telefilm,...)
- giorno e mese
- provincia e regione degli ascoltatori

Il data warehouse realizzato deve contenere le informazioni relative agli ultimi 5 anni. Al fine di una corretta realizzazione del data warehouse sono state fornite le seguenti informazioni:

- Numero di canali:~40
- Numero di programmi:~500
- Numero di tipi di programmi: ~50
- Numero di temi:~15
- Numero di province: ~100
- Numero di regioni: ~20

Gli analisti vogliono poter disporre delle seguenti informazioni:

- 1) Relativamente alla regione Piemonte nel mese di Gennaio 2008, considerando solo i programmi di tipo "notiziario" con una media giornaliera di almeno il 10% di share, trovare la durata media degli spot trasmessi.
- 2) Trovare la media del numero massimo di spettatori sintonizzati sui canali sportivi tra le 12 e le 13 dei giorni festivi.
- 3) Nel 2007 trovare il totale dei minuti di pubblicità trasmessa di lunedì in ogni canale.
- 4) Trovare il "periodo della giornata" con il minor numero di spettatori del canale "TuttoTennis".
- 5) Relativamente ai canali tematici "sportivi" nel mese di gennaio 2008, trovare la classifica delle ore del giorno con più minuti di pubblicità (dall'ora del giorno con più pubblicità a quella con meno).
- 6) Nel 2007, trovare per ogni canale il massimo numero di spettatori sintonizzati contemporaneamente.
- 7) Trovare il programma nel quale il 1 Gennaio 2008 sono stati trasmessi più spot.

Progettazione

- Progettare lo schema concettuale e logico relazionale del data warehouse necessario per gestire le informazioni richieste dalla dirigenza, valutando anche eventuali uniformazioni di dimensioni.
- Scrivere le interrogazioni in SQL corrispondenti alle interrogazioni frequenti 1), 5) e 6).
- Considerando le caratteristiche del data warehouse realizzato e la cardinalità dei dati memorizzati nel data warehouse, decidere se e quali viste materializzate potrebbe essere utile definire al fine di

ottimizzare i tempi di risposta delle interrogazioni proposte nelle specifiche del problema (considerare **tutte** le interrogazioni proposte e non solo quelle risolte in SQL. Motivare le scelte fatte.

- Decidere come gestire la dinamicità (variazione) dei dati all'interno delle dimensioni.