

Esercizio: Progettazione di un data warehouse

Descrizione del problema

La società Autostrade Italiana è interessata ad analizzare le vendite e i profitti che derivano dalle pompe di benzina site negli Autogrill sulle autostrade in Italia. Gli autogrill possono essere molto diversi tra loro e possono o no offrire servizi quali: la presenza di un ristorante, il distributore di gas metano, etc. .

Ogni Autogrill è caratterizzato da un nome, univoco su tutta la rete autostradale, ed è interessante per l'analisi non solo il nome dell'autostrada sulla quale si trova ma anche il nome del casello che segue e del casello che precede l'Autogrill (es, Autogrill "Bon Voyage" sito sull'A4 Milano-Brescia tra il casello di Monza e Vimercate)

La dirigenza della società è interessata ad analizzare: il numero di auto che fanno rifornimento, il numero di litri di benzina erogati, il numero di litri di gasolio erogati e l'incasso effettuato in funzione:

- dell'autogrill, della provincia e regione in cui si trova l'autogrill
- della dimensione (in metri quadri) dell'area occupata dall'autogrill
- della presenza o meno di un ristorante all'interno dell'autogrill e del distributore di gas metano
- dell'autostrada dove è sito l'autogrill
- della lunghezza totale dell'autostrada e del numero totale di autogrill che vi sono presenti
- del nome del casello autostradale che precede l'autogrill e del nome del casello autostradale che segue l'autogrill
- della data, settimana, mese, trimestre, quadrimestre e anno
- del giorno della settimana con la possibilità di riconoscere i giorni festivi
- del tipo di mezzo in transito (auto, camion, motociclo, ...)

Il data warehouse realizzato deve contenere le informazioni relative agli anni 1998-2008. Al fine di una corretta realizzazione del data warehouse sono state fornite le seguenti informazioni (le informazioni ritenute necessarie ma non presenti in questa lista possono essere ipotizzate e stimate dal candidato):

- Numero di tipi di auto ~ 5
- Numero di autogrill ~ 1000

I dirigenti vogliono poter disporre delle seguenti informazioni:

- a) Per ogni giorno festivo del primo trimestre del 2008 (escludendo la domenica), visualizzare la data, il giorno della settimana e l'incasso medio per auto effettuato considerando solo gli autogrill con una dimensione superiore ai 50000 metri quadri. Inoltre, visualizzare la posizione in una graduatoria (rank) dal giorno con la media più alta a quello con la media più bassa.**
- b) Per ogni coppia regione-autogrill, trovare il numero di litri di benzina erogati e il totale dell'incasso nel 2005
- c) Per ogni giorno del Luglio 2008, trovare il numero di auto che hanno sostato negli autogrill del Piemonte
- d) Trovare tutti gli autogrill che a gennaio 2008 hanno venduto più litri di benzina che di gasolio ai mezzi di tipo "Camion". Visualizzare l'autogrill e il relativo incasso totale. Considerare solo gli autogrill che si trovano su autostrade più lunghe di 400 Km. Visualizzare i risultati ordinati dall'autogrill che ha incassato di più a quello che ha incassato di meno.**
- e) Per ogni anno, trovare la regione nella quale è sito l'autogrill che ha incassato di più.
- f) Trovare per ogni regione il numero medio mensile di litri di benzina venduti nel 2008 considerando solo gli autogrill che nel 2007 hanno avuto una media settimanale di incasso superiore a 50000 euro.**

Progettazione

- 1) Progettare il data warehouse necessario per 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 proposte nelle specifiche del problema.
- 2) Esprimere le interrogazioni frequenti (a), (d), (f) delle specifiche del problema utilizzando il linguaggio SQL esteso.

- 3) 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 al punto 2). Motivare le scelte fatte.
- 4) Decidere come gestire la dinamicità (variazione) dei dati all'interno delle dimensioni.