

Esercizio di progettazione di un data warehouse

Progettare un data warehouse per la gestione delle problematiche illustrate nei punti seguenti relative ai magazzini di una ditta di elettrodomestici italiana con sedi in tutta Italia.

Descrizione del problema

Una grande ditta di elettrodomestici ha magazzini sparsi in tutta Italia. La dirigenza dell'azienda ha la necessità di analizzare l'attuale uso dei magazzini al fine di capire se e quanto i vari magazzini sono effettivamente utilizzati in funzione della loro dislocazione geografica. La finalità dell'azienda è quella di decidere quali magazzini ampliare, quali ridimensionare e quali chiudere o affittare, anche solo parzialmente, a terze parti.

La dirigenza della ditta stipula periodicamente delle polizze assicurative per coprire i danni dei furti che si verificano nei magazzini. I prezzi delle polizze contro i furti sono legati alla quantità media giornaliera di prodotti presenti nei magazzini e al loro valore. Esistono diverse tipologie di polizze e la dirigenza vuole capire quali sono le polizze più adeguate per i propri magazzini analizzando quali modelli di prodotti e in quali quantità sono mediamente presenti a livello giornaliero nei magazzini. Alcune polizze sono specifiche per un singolo magazzino mentre altre sono relative ad insiemi di magazzini che si trovano nella stessa città, regione o provincia. In particolare i premi assicurativi delle polizze sono calcolati considerando, su base media giornaliera:

- il numero di prodotti presenti nei magazzini per ogni modello di prodotto presente
- il valore dei prodotti presenti nei magazzini per ogni modello di prodotti.

Viste le problematiche appena descritte la ditta ha deciso di creare un apposito data warehouse per gestire le informazioni relative all'uso dei magazzini e al loro contenuto.

Relativamente all'analisi dell'uso dei magazzini (superficie utilizzata) la dirigenza è interessata ad analizzare la percentuale giornaliera di superficie libera rispetto alla superficie totale disponibile in ogni magazzino in funzione del magazzino e del luogo in cui esso si trova (città, provincia, regione). La superficie totale disponibile per ogni magazzino può variare di giorno in giorno in quanto alcune aree dei magazzini possono essere affittate a terzi o essere temporaneamente inagibili (queste aree non fanno parte della superficie totale disponibile). L'analisi deve essere possibile a livello di singola data, ma anche a livello di mese, trimestre, quadrimestre, semestre e anno.

Per stimare i costi delle possibili polizze assicurative la ditta deve poter analizzare il numero medio giornaliero di prodotti presenti nei magazzini e il loro valore complessivo giornaliero a livello di singola data, a livello di mese, di trimestre, quadrimestre, di semestre e di anno. Tali informazioni devono essere disponibili in funzione del modello dei prodotti e della loro categoria (scopa elettrica, impastatrice, ..), del magazzino e del luogo in cui si trova (città, provincia, regione).

Sono di seguito riportate **alcune** delle interrogazioni frequenti di interesse per la dirigenza della ditta:

- a) Relativamente al primo trimestre dell'anno 2003, considerando solo i magazzini della città di Torino, trovare per ogni coppia (magazzino,data) il valore complessivo di prodotti presenti in tale data nel magazzino e il valore complessivo medio giornaliero di prodotti presenti nel magazzino nel corso della settimana precedente la data in esame (data in esame inclusa).
- b) Relativamente all'anno 2004, trovare per ogni coppia(città,data) la percentuale di superficie libera giornaliera nella città. Associare ad ogni coppia un attributo di rank legato alla percentuale di superficie libera giornaliera nella città (1 per la coppia con la più bassa percentuale di superficie libera giornaliera).
- c) Relativamente ai primi sei mesi dell'anno 2004, trovare per ogni coppia (magazzino,data) la percentuale di superficie libera giornaliera.
- d) Relativamente all'anno 2003, trovare per ogni coppia (magazzino,mese) il valore complessivo medio giornaliero di prodotti presenti.
- e) Relativamente all'anno 2003, trovare per ogni regione il valore complessivo medio giornaliero di prodotti presenti nella regione.
- f) Relativamente all'anno 2004, trovare per ogni coppia(mese, regione) la percentuale di superficie libera giornaliera nella regione.

Progettazione

1. Progettare il data warehouse necessario per gestire le necessità della ditta di elettrodomestici 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 proposte nelle specifiche del problema.
Il data warehouse realizzato deve contenere le informazioni relative agli ultimi 2 anni. Al fine di una corretta realizzazione del data warehouse sono state fornite le seguenti informazioni:
 - Numero di modelli diversi di prodotti: ~100
 - Numero di categorie diverse di prodotti: ~10
 - Numero di magazzini: ~100
 - Numero di magazzini di Torino: ~5
 - Numero di città: ~90
 - Numero di regioni: ~40
 - Numero di province: ~10
2. Esprimere le interrogazioni frequenti (a), (b), (d) 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 o indici 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.