

Esercizio di progettazione di un data warehouse

Progettare un data warehouse per la gestione delle problematiche illustrate nei punti seguenti relative ad un sito per la pubblicazione di annunci relativi all'affitto di immobili.

Descrizione del problema

Il sito internet www.cerca_la_tua_casa.it permette di fare ricerche in tutte le città italiane per trovare gli immobili in affitto. I privati o le agenzie possono pubblicare su questo sito gli annunci relativi agli immobili disponibili, specificando per ogni immobile le sue caratteristiche principali: zona della città in cui si trova (quartiere), tipo di immobile (attico, mansarda, villa a schiera, rustico,...), prezzo di affitto mensile, superficie in metri quadri, numero di stanze, numero di bagni, piano a cui è sito l'immobile, eventuale presenza dell'ascensore o della cantina. In più è possibile segnalare l'eventuale arredamento già presente nell'immobile: presenza di tavolo/sedie, frigorifero, forno, stufa, ventilatore, letto, lavatrice, lavastoviglie, tv, ecc.

Gli annunci vengono aggiornati settimanalmente, ogni lunedì vengono eliminati gli annunci relativi agli appartamenti che sono stati occupati e vengono aggiunti i nuovi annunci relativi agli appartamenti liberi.

Gli utenti del sito possono aggiungere alla sezione personale "aggiungere ai preferiti" tutti gli immobili ai quali sono interessati, in modo da poterli poi visionare in un'unica pagina e poterli confrontare più agevolmente.

La società che gestisce il sito internet è interessata ad analizzare la situazione degli affitti in Italia. I parametri presi in considerazione per questa analisi sono: la disponibilità di immobili in affitto (n° di immobili in affitto), la media del prezzo di affitto per immobile, e la media del prezzo di affitto al metro quadro. Questi parametri si devono poter analizzare in funzione

- della settimana, del mese, del bimestre, del trimestre, del quadrimestre, del semestre e dell'anno
- della zona d'Italia (nord, centro, sud, isole), della regione, della provincia, della città, della zona della città (quartiere) in cui è situato l'immobile
- della presenza di università nella città in cui è situato l'immobile
- del tipo di immobile e delle tipologie di arredi eventualmente presenti (sedia, tavolo, frigorifero, ecc.).
- del numero di stanze

Si è anche interessati a fare delle statistiche in base a quali sono gli immobili "preferiti" dagli utenti, e quindi a valutare il numero medio di utenti interessati ad un immobile in funzione:

- dell'anno e della stagione (primavera, estate, autunno, inverno)
- dalla zona d'Italia (nord, centro, sud, isole), della regione, della provincia, della città, della zona della città (quartiere)
- della presenza di università nella città in cui è situato l'immobile
- del tipo di immobile
- del range di costo dell'affitto (100-200€, 200-300€,...) e del range di dimensione (0-50 mq, 50-100 mq,...) dell'immobile
- del numero di stanze

Sono di seguito riportate **alcune** delle interrogazioni frequenti di interesse per la società che gestisce il sito contenente gli annunci di affitto:

- a) Relativamente al 2004, considerando solo gli immobili situati in città nelle quali sono presenti delle università, trovare per ogni coppia (città, mese) il costo medio di affitto per immobile e il costo medio di affitto per immobile da inizio anno nella città in esame.
- b) Relativamente al mese di settembre 2004, considerando solo gli immobili situati in provincia di Torino, trovare per ogni coppia (città, settimana) il numero totale di immobili disponibili, il rapporto tra il numero totale di immobili disponibili relativi alla coppia in esame e il numero totale di immobili disponibili nella settimana in esame. Associare ad ogni coppia un attributo di rank legato al numero totale di immobili disponibili. Associare il valore 1 alla coppia con il maggior numero di immobili disponibili. Ordinare i dati in funzione dell'attributo di rank.
- c) Relativamente alla stagione estiva del 2005, considerando solo le mansarde dotate di letto, frigorifero e tavolo situate presso la città di Rimini, trovare per ogni coppia (zona città, range affitto) il numero medio di utenti interessati per immobile e il numero medio di utenti interessati per immobile nella zona della città della coppia in esame. Ordinare i dati in funzione della zona della città e del numero medio di utenti interessati per immobile.
- d) Considerando solo gli immobili situati in città nelle quali sono presenti delle università e dotati di letto e tavolo, trovare per ogni tripletta (città, mese, anno) il costo medio di affitto per immobile, il costo

medio di affitto al metro quadro, il costo medio di affitto per immobile da inizio anno nella città in esame.

- e) Considerando i mesi di settembre, ottobre e novembre 2004 e gli immobili situati presso la regione Piemonte, trovare per ogni città il prezzo medio di affitto per immobile e il prezzo medio di affitto per immobile relativo alla provincia della città in esame.
- f) Relativamente al 2004, considerando solo gli immobili situati in città nelle quali sono presenti delle università e dotati di letto e tavolo, trovare per ogni coppia (città, mese) il costo medio di affitto per immobile e il costo medio di affitto al metro quadro.

Progettazione

1. Progettare il data warehouse necessario per gestire le necessità della società che gestisce il sito contenente gli annunci di affitto 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 zone d'Italia ~ 4
 - Numero di provincie ~ 100
 - Numero di città ~ 8000
 - Numero di zone città ~ 10000
 - Numero di città presso cui ci sono delle università ~ 100
 - Numero di tipologie di immobili ~ 5
 - Numero di tipologie di arredi ~ 10
 - Numero di stanze ~ da 1 a 5
 - Numero di range di costo d'affitto ~ 10
 - Numero di range di dimensioni immobili ~ 10
2. Esprimere le interrogazioni frequenti (b), (d), (e) 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.