

Esercizio: Progettazione di un data warehouse

1. Descrizione del problema

Una grande catena di alberghi gestisce circa 500 alberghi, di diverse categorie, sparsi in tutto il mondo. Nella base di dati di ogni albergo sono memorizzate tutte le informazioni giornaliere relative ai clienti che soggiornano nell'albergo e in quali camere. Per ogni albergo sono note, sempre a livello giornaliero, non solo le informazioni su quali camere sono occupate, ma anche quali camere sono disponibili (libere) e quali non sono agibili perché si stanno facendo delle riparazioni alla camera o all'arredamento.

La dirigenza della catena di alberghi vuole disporre di un data warehouse che permetta di effettuare analisi di mercato relativamente agli incassi ottenuti dai vari alberghi in funzione:

- della zona geografica in cui si trova l'albergo (stato, regione, città)
- della categoria dell'albergo (5 stelle, 4 stelle, ..)
- delle caratteristiche della camera (numero di letti, dotata di televisore, dotata di vasca idromassaggio, ..)
- della data, del giorno della settimana, del mese e dell'anno

La dirigenza della catena di alberghi vuole inoltre sapere a livello giornaliero per ogni albergo:

- la frazione di camere occupate
- la frazione di camere libere
- la frazione di camere non agibili a causa di lavori di riparazione

Anche relativamente allo stato delle camere (occupata, libera, non agibile) le informazioni devono essere disponibili in funzione:

- della zona geografica in cui si trova l'albergo
- della categoria dell'albergo (5 stelle, 4 stelle, ..)
- delle caratteristiche della camera (numero di letti, dotata di televisore, dotata di vasca idromassaggio, ..)
- della data, del giorno della settimana, del mese e dell'anno

La dirigenza dell'albergo vuole disporre delle informazioni appena descritte sia a livello giornaliero, sia a livello mensile, sia a livello annuale. Si vuole inoltre poter capire come variano gli incassi in funzione del fatto che il giorno sia festivo oppure no.

Sono di seguito riportate alcune delle interrogazioni frequenti di interesse per la dirigenza della catena di alberghi:

- a) Relativamente all'anno 2005, selezionare per ogni coppia (stato, mese) la frazione mensile di camere occupate sulla totalità di camere, la frazione mensile di camere libere sulla totalità di camere e la frazione mensile di camere non agibili sulla totalità di camere.
- b) Relativamente all'anno 2005, selezionare per ogni stato la frazione di camere occupate sulla totalità di camere in ogni stato nel corso dell'anno 2005. Associare ad ogni stato un attributo di rank associato alla frazione di camere occupate sulla totalità di camere nello stato nel corso dell'anno 2005. L'attributo di rank deve assumere il valore 1 per lo stato con la maggiore frazione di camere occupate sulla totalità di camere nello stato nel corso dell'anno 2005.
- c) Relativamente all'anno 2005, selezionare per ogni coppia (stato, mese) l'incasso mensile calcolato considerando solo gli alberghi a 4 stelle e l'incasso cumulativo da inizio anno (sempre considerando solo gli alberghi a 4 stelle).
- d) Calcolare per ogni coppia (stato, anno) l'incasso totale effettuato nel corso dei giorni festivi.
- e) Calcolare per ogni albergo l'incasso totale effettuato nell'anno 2005 considerando solo le camere dotate di TV e vasca idromassaggio.

2. Progettazione

- 2.1. Progettare il data warehouse necessario per gestire le informazioni relative agli incassi degli alberghi e all'uso delle camere degli alberghi, in modo da soddisfare le richieste descritte al punto 1. Il data warehouse progettato deve permettere di rispondere in modo efficiente a tutte le interrogazioni frequenti proposte al punto 1 (interrogazioni a), b), c), d), e)).

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

- Numero di alberghi: ~500
- Numero di stati: ~40
- Numero di città: ~400
- Numero di caratteristiche delle camere (numero di letti, dotata di televisore, ..): ~8

- 2.2. Decidere come gestire la dinamicità (variazioni) dei dati all'interno delle dimensioni.
- 2.3. Rispondere alle interrogazioni frequenti (c) e (e) proposte nel punto 1 utilizzando il linguaggio SQL esteso.
- 2.4. 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 definite al fine di ottimizzare i tempi di risposta delle interrogazioni proposte al punto 1 (considerare tutte le interrogazioni proposte e non solo quelle risolte in SQL). Motivare le scelte fatte.