

# Business intelligence

## Esercizio di progettazione di un data warehouse

### Descrizione del problema

L'ente per la promozione del turismo in Italia vuole analizzare le presenze di turisti nelle località italiane per capire dove è più conveniente fare promozioni. L'ente è interessato a capire quali sono i posti preferiti dai turisti delle varie nazionalità nei diversi periodi dell'anno, quali sono le strutture (alberghi, campeggi,...) dove preferiscono soggiornare e qual è il ritorno economico delle promozioni effettuate sul territorio.

In particolare si è interessati ad analizzare il numero di turisti presenti nelle strutture e la loro età media in funzione di:

- località di villeggiatura
- provincia e regione della località
- classificazione della località (in base alla posizione geografica una località può essere classificata come "marittima", "montana", "collinare" o "altro")
- posizione geografica della regione di villeggiatura (nord, centro,sud, isole)
- tipologia di struttura scelta (albergo, campeggio, casa in affitto,...)
- classificazione della struttura (in base a determinati parametri le strutture possono essere classificate come: economiche, medie, lusso)
- settimana in cui il turista è presente
- mese, bimestre, trimestre, anno in cui il turista è presente
- regione, stato e continente di provenienza del turista
- stati appartenenti o no all'Unione Europea
- posizione della regione di provenienza (nord, centro,sud, isole), solo nel caso il turista sia Italiano.

Inoltre per misurare il ritorno economico della pubblicità si è interessati ad analizzare l'investimento effettuato, il numero di turisti, la frazione di turisti italiani sul totale per ogni regione (e sua posizione: nord, centro, sud, isole) e anno.

Il data warehouse realizzato deve contenere le informazioni relative agli anni 2001-2010. Al fine di una corretta realizzazione del data warehouse sono state fornite le seguenti informazioni (le informazioni ritenute necessarie ma non presenti in questo elenco possono essere ipotizzate e stimate dal candidato):

- Numero di località ~ 10000
- Numero tipi di struttura ~ 5
- Numero stati di provenienza dei turisti ~ 50

Sono di seguito riportate **alcune** delle interrogazioni frequenti:

- Visualizzare per ogni anno l'investimento totale, l'investimento cumulativo (dal 2001 al 2010) e la frazione di turisti italiani sul totale, separatamente per ogni regione visitata.**
- Per ogni località turistica marittima calcolare il numero di turisti nei mesi di luglio e agosto 2007
- Per ogni regione del centro Italia e per ogni anno, calcolare la media mensile di turisti considerando solo gli alberghi di lusso.**
- Per ogni bimestre del 2009 visualizzare il numero di turisti totali e il numero cumulativo di turisti da inizio anno, separatamente per tipologia di località.
- Considerando solo i campeggi, visualizzare l'età media dei turisti a settembre 2008, che hanno soggiornato in Liguria
- Relativamente a Gennaio 2005, considerando solo le regioni che hanno avuto complessivamente più di 100 mila turisti in località di tipo "montagna", visualizzare per ogni località di tipo diverso da "montagna" il numero di turisti e la loro età media.**

### 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), (c), (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.