

Business Intelligence

Esercizio: Progettazione di un data warehouse

Descrizione del problema

La società SkiPass gestisce diversi comprensori sciistici localizzati nelle località montane europee e vorrebbe analizzare i dati relativi all'utilizzo dei propri impianti e ai ricavi derivanti dalla loro gestione.

Gli impianti di risalita di un comprensorio sono caratterizzati da tipologia (seggiovia, skilift, . . .), lunghezza e dislivello. Ogni comprensorio sciistico può essere caratterizzato da diversi impianti della stessa topologia, caratterizzati da diversa lunghezza, e/o diverso dislivello. Per avere diritto di utilizzo degli impianti è necessario acquistare un biglietto. I biglietti, validi per un solo comprensorio, possono essere: biglietti a punti, giornalieri e stagionali.

La società vorrebbe analizzare i propri dati al fine di valutare gli incassi derivanti dalla gestione dei propri impianti e il relativo utilizzo da parte dei clienti. Si noti che ogni incasso derivante dalla vendita dei biglietti è associata alla data di vendita del biglietto.

Per l'analisi degli incassi, la società SkiPass è interessata ad analizzare l'incasso totale, e l'incasso medio per biglietti venduti in funzione di:

- comprensorio sciistico
- città, provincia, regione e stato in cui è sito il comprensorio sciistico
- catena montuosa in cui è sito il comprensorio sciistico
- zona geografica della regione in cui è sito il comprensorio sciistico
- data, mese, bimestre, trimestre, quadrimestre, semestre, ed anno in cui è stato realizzato l'incasso
- tipo di giorno (feriale o festivo) e stagione
- modalità di pagamento (e.g., contanti, carta di credito, ...)
- tipo di biglietto

Per l'analisi dell'utilizzo degli impianti, tutti i possibili biglietti danno il diritto di utilizzare più volte lo stesso impianto, e più volte impianti diversi siti nello stesso comprensorio sciistico. ogni volta che un cliente utilizza un impianto del comprensorio viene obliterato il biglietto e di conseguenza registrato il passaggio. La società SkiPass è interessata ad analizzare il numero di passaggi medi giornalieri per valutare se e' necessario ampliare le proprie strutture, oppure chiuderne alcune. L'analisi sarà effettuata in funzione di:

- mezz'ora del giorno, ora del giorno
- data, giorno (se è festivo o no), giorno della settimana, mese e anno
- periodo della giornata: mattino (8:30-13:30), pomeriggio (13:30-18:30), sera (18:30, 23:30)
- tipologia, lunghezza e dislivello degli impianti di risalita
- comprensorio sciistico,
- città, provincia, regione e stato in cui è sito il comprensorio sciistico
- catena montuosa e zona geografica della regione in cui è sito il comprensorio sciistico.

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

- a) Per i comprensori sciistici siti nella catena montuosa delle Alpi che nel gennaio 2007 hanno incassato complessivamente più di 20000 €, trovare per ogni coppia (comprensorio sciistico, tipo di biglietto), l'incasso mensile del 2007, e l'incasso totale mensile cumulativo dall'inizio dell'anno.**
- b) Ordinare e assegnare un rank ai comprensori sciistici per numero medio giornaliero di passaggi (il primo comprensorio è quello col più alto numero medio giornaliero di passaggi e deve avere rank pari a 1). Per ogni comprensorio sciistico, selezionare anche il numero medio giornaliero di passaggi, e il numero totale di passaggi.**
- c) Per ogni comprensorio sciistico sito in Italia e considerando solo gli incassi del 2008, visualizzare per ogni mese e per ogni tipologia di biglietti l'incasso medio giornaliero e il numero medio giornaliero di biglietti venduti.**

- d) Per ogni stagione e per ogni anno, trovare il numero complessivo di biglietti venduti in ogni comprensorio sciistico e il corrispondente incasso totale.
- e) Per ogni coppia comprensorio sciistico, tipo di biglietto visualizzare l'incasso mensile e il numero totale di biglietti venduti.
- f) Per ogni mese del 2006 e 2007, visualizzare l'incasso bimestrale, separatamente per ogni modalità di pagamento.
- g) Per ogni comprensorio sciistico sito nell'Italia del nord, visualizzare, il numero mensile di passaggi effettuati su ogni tipologia di impianti.

Progettazione

1. Il data warehouse conterrà le informazioni relative agli anni 2004, 2005, 2006, 2007 e 2008. Al fine di una corretta realizzazione del data warehouse sono state fornite le seguenti informazioni:
 - Numero di comprensori sciistici:~800
 - Numero di città: ~800
 - Numero di province: ~350
 - Numero di regioni: ~100
 - Numero di stati:~10
 - Numero di zone geografiche:~5
 - Numero di tipi di impianti:~5
 - Numero di dislivelli diversi:~5
 - Numero di lunghezze:~10
 - Numero di tipi di biglietto:~3
 - Numero di metodi di pagamento:~4
2. Esprimere le interrogazioni frequenti (a), (b), (c) 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 (variazioni) dei dati all'interno delle dimensioni. Motivare le scelte fatte.