

Esercizio: Progettazione di un data warehouse

Descrizione del problema

Si vogliono analizzare alcune delle attività relative ad una ditta di trasporti europea. Nella base di dati della società sono memorizzate le informazioni di dettaglio sui viaggi svolti e sulle riparazioni effettuate sui mezzi di trasporto. La dirigenza della ditta è interessata ad analizzare i tempi di percorrenza e i costi associati ai viaggi in funzione di alcuni parametri di interesse descritti in seguito. La dirigenza è convinta che un'attenta analisi di tali informazioni permetterà all'ufficio logistica di migliorare la pianificazione dei trasporti, con un notevole risparmio economico per la ditta stessa. La dirigenza è inoltre interessata ad analizzare con attenzione le riparazioni effettuate e i costi di riparazione per i mezzi utilizzati. Tale analisi è ritenuta utile per pianificare gli acquisti di nuovi mezzi di trasporto e per decidere quali mezzi vendere o demolire tra quelli attualmente a disposizione. In merito ai viaggi la dirigenza della ditta è interessata ad analizzare la durata media (in minuti) per viaggio e il costo medio per viaggio in funzione:

- delle caratteristiche del luogo di partenza del viaggio (area geografica all'interno dello stato, stato, appartenenza alla comunità europea (si/no))
- delle caratteristiche del luogo di destinazione del viaggio (area geografica all'interno dello stato, stato, appartenenza alla comunità europea (si/no))
- tipologia del mezzo di trasporto utilizzato (furgone, autocarro, tir, ..)
- del mese, del trimestre e dall'anno in cui il viaggio è stato svolto

Relativamente alle riparazioni la ditta vuole poter analizzare il numero medio di riparazioni mensili effettuate e il costo medio per riparazione in funzione:

- del modello del mezzo di trasporto (Ducato, ...)
- della tipologia del mezzo di trasporto (furgone, autocarro, tir, ..)
- della casa produttrice del mezzo di trasporto
- dell'anno di immatricolazione del mezzo di trasporto
- del mese, del trimestre e dall'anno in cui è stata effettuata la riparazione

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

- a) Relativamente all'anno 2003, calcolare per ogni tripletta (Area Geografica di Partenza, Area Geografica di Arrivo, Tipologia Mezzo di trasporto) la durata media per viaggio e il costo medio per viaggio.
- b) Relativamente al primo trimestre del 2005, calcolare per ogni modello di mezzo di trasporto il costo medio per riparazione. Associare ad ogni modello di mezzo di trasporto un attributo di rank legato al costo medio per riparazione (l'attributo di rank assume il valore 1 per il modello con il costo medio per riparazione più elevato).
- c) Per ogni coppia (Casa produttrice del mezzo di trasporto, Anno di Immatricolazione) calcolare il costo medio per riparazione e il numero medio di riparazioni mensili effettuate nel 2007. Ordinare i risultati in base al numero medio di riparazioni mensili effettuate.
- d) Relativamente all'anno 2004, calcolare per ogni tripletta (Area Geografica di Partenza, Area Geografica di Arrivo, Tipologia Mezzo di trasporto) la durata media per viaggio.
- e) Per ogni Casa produttrice del mezzo di trasporto calcolare il costo medio per riparazione effettuate nel 2007.

Il data warehouse realizzato deve contenere le informazioni relative agli ultimi 3 anni. Al fine di una corretta progettazione del data warehouse sono state fornite le seguenti informazioni:

- Numero di mezzi di trasporto della ditta: ~50000
- Numero di tipologie diverse di mezzi di trasporto: ~10
- Numero di modelli diversi di mezzi di trasporto: ~50
- I mezzi di trasporto più vecchi sono stati immatricolati 10 anni fa
- Numero di case produttrici di mezzi di trasporto: ~10
- Numero di aree geografiche: ~200

Progettazione

- 1.1. Progettare lo schema del data warehouse per gestire le informazioni relative ai viaggi e alle riparazioni in modo da soddisfare le richieste descritte nella descrizione del problema.
- 1.2. Decidere come gestire la dinamicità (variazioni) dei dati all'interno delle dimensioni.
- 1.3. Rispondere alle interrogazioni a), b), c).
- 1.4. Considerando le caratteristiche del data warehouse realizzato e le cardinalità delle tabelle del data warehouse progettato, decidere se e quali viste materializzate o indici potrebbe essere utile definite al fine di ottimizzare i tempi di risposta delle interrogazioni proposte nella descrizione del problema

(considerare tutte le interrogazioni proposte e non solo quelle risolte in SQL al punto 1.3). Motivare le scelte fatte basandosi sulle cardinalità delle tabelle del data warehouse progettato.