

1 febbraio 2008

Appello di Analisi di Basi di dati

Progettazione

Progettare un data warehouse per l'analisi dell'attività commerciale gestita da una ditta di spedizioni internazionali che ha circa 200 filiali nel mondo.

Descrizione del problema

Una grande ditta di spedizioni internazionali ha filiali sparse in tutto il mondo. Le filiali organizzano le spedizioni di merci per conto terzi. Attualmente il sistema informativo aziendale è completamente decentralizzato, quindi ogni filiale ha una propria base dati per la gestione della propria attività commerciale. La responsabilità di organizzare l'intera spedizione ricade su una sola filiale. Una spedizione è costituita da una o più tratte, ciascuna caratterizzata da luogo di partenza e luogo di destinazione, in termini di città, provincia, regione e stato. La dirigenza dell'azienda ha la necessità di analizzare il volume di traffico e il fatturato associati alle spedizioni gestite dalla diverse filiali in modo da analizzare l'efficienza di organizzazione delle singole filiali rispetto alle altre, decidere quali filiali ampliare, quali ridimensionare e quali chiudere.

L'azienda vuole effettuare l'analisi del traffico della merce, sia in termini di volume che di peso (espresso in kg), e del fatturato associato alle spedizioni. L'analisi deve poter evidenziare il fatturato realizzato dalle singole filiali rispetto al fatturato complessivo dei distretti di appartenenza (delle filiali stesse). Inoltre è necessario valutare la redditività delle spedizioni per categoria merceologica e per tipo di spedizione (per esempio via aerea, via mare, via terra, ...).

La dirigenza aziendale è inoltre interessata ad analizzare le merci spedite in termini di volume, peso e fatturato in funzione delle diverse tratte di spedizioni per decidere in quali nuove aree geografiche potrebbe essere utile aprire nuove filiali.

Infine la direzione aziendale è interessata ad analizzare il peso medio, il volume medio, e il fatturato medio considerando le spedizioni effettuate in una singola data, giorno della settimana, mese, bimestre, trimestre, quadrimestre, semestre e anno.

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

- a) Relativamente alle tratte italiane, per ogni categoria di merce e per ogni anno, visualizzare il fatturato medio giornaliero in ogni mese e il fatturato totale al passare dei mesi dall'inizio dell'anno per ciascuna categoria.
- b) Relativamente alle spedizioni via aerea, visualizzare per ogni provincia di destinazione il fatturato medio annuo per unità di volume, la percentuale di tale fatturato rispetto al fatturato medio annuo per unità di volume dello stato di destinazione.
- c) Relativamente all'anno 2006, visualizzare per ogni tratta, in termini di regioni di partenza e di destinazione, il fatturato medio mensile per unità di peso (in kg) delle merci spedite e il fatturato medio giornaliero delle merci spedite per quella tratta in ciascun mese.
- d) Per ogni distretto e tipo di spedizione visualizzare il fatturato totale in ogni mese e il volume totale spedito in ogni mese. Associare ad ogni record un valore di rank legato al volume totale mensile (1 per il record con il più alto volume totale mensile).
- e) Relativamente all'anno 2005, trovare per ogni tratta, in termini di città di partenza e di destinazione, il fatturato medio annuo per unità di kg delle merci spedite e il fatturato medio giornaliero delle merci spedite per ciascuna tratta.
- f) Relativamente alle spedizioni marittime del 2005 e 2006, visualizzare per ogni filiale e regione di partenza il fatturato semestrale per unità di volume e il fatturato semestrale per unità di peso (in Kg).
- g) Per ogni città di partenza e per ogni distretto, visualizzare il fatturato quadrimestrale e il volume totale spedito in ogni quadrimestre.

Progettazione

1. Progettare il data warehouse necessario per gestire le necessità della ditta di spedizioni internazionali 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 4 anni. Al fine di una corretta realizzazione del data warehouse sono state fornite le seguenti informazioni:

- Numero di filiali:~200
 - Numero di distretti:~50
 - Numero di categorie diverse di merce: ~20
 - Numero di tipo spedizione: ~5
 - Numero di città: ~1000
 - Numero di province: ~300
 - Numero di regioni: ~100
 - Numero di stati: ~30
2. Esprimere le interrogazioni frequenti (a), (c), (d) 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.