

Analisi di basi di dati

Esame del 4 maggio 2009

Nome	
Cognome	
Matricola	

Descrizione del problema

Una società che assembla e vende personal computer vuole analizzare i costi e i ricavi che derivano dall'acquisto dei componenti e dalla vendita all'ingrosso dei propri pc.

La società assembla i pc comprando dalle ditte fornitrici quattro dei componenti principali: la ram, il disco rigido (HD), il processore (CPU) e la scheda video. Ogni componente ha un proprio codice alfanumerico che dipende da chi l'ha prodotto e dalle caratteristiche del componente stesso. Ogni componente è prodotto da una sola azienda. Una volta acquisiti, i componenti vengono montati con una serie di "configurazioni" standard per venire incontro alle richieste delle diverse fasce del mercato. Una configurazione è quindi determinata univocamente dalla combinazione dei quattro codici dei componenti. I pc così assemblati vengono poi venduti a grossi blocchi alle catene commerciali (grossisti) che si occupano di distribuirli presso i loro punti vendita.

Per quanto riguarda la fase di produzione gli analisti della società sono interessati ad analizzare il numero dei componenti acquistati, il costo dei componenti e il costo di trasporto dei pezzi ordinati in funzione:

- della ditta produttrice dalla quale vengono comprati i componenti
- della posizione geografica della ditta (regione e zona d'Italia (nord/centro/sud/isole))
- del tipo di componente (RAM, CPU, HD, scheda video)
- di alcune caratteristiche relative ai componenti;
 - o codice del componente
 - o per la RAM: capacità, tipo, velocità di accesso
 - o per l'HD: capacità e dimensioni fisiche
 - o per la CPU: tipo e velocità
 - o per la scheda video: risoluzione massima.
- del mese, bimestre, quadrimestre, semestre, anno in cui sono stati comprati i componenti

Per quanto riguarda la distribuzione, la dirigenza dell'azienda è interessata ad analizzare per ogni configurazione prodotta il costo di produzione (determinato dal costo dei componenti più il costo di assemblaggio), il prezzo di vendita ai grossisti, il numero di prodotti venduti ai grossisti e il numero di prodotti difettosi che sono stati restituiti durante il periodo di garanzia in funzione:

- della configurazione
- del codice del componente RAM, HD, CPU e scheda video presenti in una configurazione
- delle ditte produttrici dei componenti di RAM, HD, CPU e scheda video presenti in una configurazione
- della catena di distribuzione alla quali sono stati venduti
- del numero di punti vendita che gestisce la catena di distribuzione
- dal nome della catena di distribuzione
- del mese e il bimestre in cui sono stati venduti
- del trimestre e dall'anno in cui sono stati venduti

Il data warehouse realizzato deve contenere le informazioni relative agli anni 1998-2007. Al fine di una corretta realizzazione del data warehouse sono state fornite le seguenti informazioni (le informazioni ritenute necessarie ma non presenti in questa lista possono essere ipotizzate e stimate dal candidato):

- o Numero di componenti ~ 1000
- o Numero di ditte ~ 20
- o Numero di catene ~ 15
- o Numero di punti vendita medio per catena ~ 500
- o Numero di configurazioni prodotte/vendute ~ 30
- o Numero di articoli ~ 100 000

Sono di seguito riportate **alcune** delle interrogazioni frequenti di interesse per la fondazione di ricerca:

- a) Per ogni quadrimestre del 2005 e del 2006 visualizzare il numero di CPU comprate separatamente per tipo di CPU dalla ditte del nord Italia
- b) Per ogni anno visualizzare la configurazione con il miglior rapporto "costo di produzione"/"prezzo vendita". Considerare solo le configurazioni in cui tutti la RAM e la CPU sono state prodotte dalla stessa ditta.
- c) Visualizzare per ogni componente il numero di pezzi acquistati a marzo 2008 e il numero totale di pezzi acquistati nel 2008
- d) Relativamente all'anno 2002, per ogni configurazione in cui è presente la RAM acquistata dalla ditta "IntelligenceDevice", visualizzare la percentuale di prodotti difettosi rispetto al totale venduto.
- e) Nell'anno 2007, per le memorie RAM, visualizzare per ogni quadrimestre la spesa totale sostenuta (costo componenti + costi di trasporto) e la spesa cumulativa dall'inizio dell'anno separatamente per ogni zona d'Italia e tipo di memoria RAM .
- f) Separatamente per ogni catena di distribuzione e per ogni mese del 2008, visualizzare il numero di configurazioni diverse che sono state vendute in quel mese a quella catena.
- g) Per ogni zona d'Italia e anno, visualizzare la somma totale delle spese per il trasporto dei componenti acquistati.
- h) Considerando solo le ditte dalle quali nel 2003 sono stati acquistati più di 100000 pezzi in totale, visualizzare per ogni ditta e tipo di componente (RAM, CPU, HD, schede video) il numero di pezzi acquistati e il costo totale dei componenti nel secondo bimestre del 2003.

Progettazione

1. (12 PUNTI) Progettare il data warehouse necessario per analizzare le pubblicazioni 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.
2. (15 PUNTI) Esprimere le interrogazioni frequenti **(d)**, **(e)**, **(h)** delle specifiche del problema utilizzando il linguaggio SQL esteso.
3. (2 PUNTI) 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. (1 PUNTO) Decidere come gestire la dinamicità (variazione) dei dati all'interno delle dimensioni.