

Esercizio: Progettazione di un data warehouse

Progettare un data warehouse per analizzare le attività del fornitore di Telepass della rete autostradale europea, considerando le seguenti specifiche.

Descrizione del problema

Il servizio Telepass è composto da un dispositivo portatile a radio frequenza installato sui veicoli. Quando un veicolo munito di Telepass attraversa un casello autostradale (cioè all'ingresso e all'uscita dell'autostrada), avvengono due eventi:

- il dispositivo Telepass è automaticamente riconosciuto dal suo identificativo, e di conseguenza al suo possessore è fatturato il pedaggio dovuto (solo per i passaggi in uscita);
- attraverso una telecamera, il numero di targa del veicolo è riconosciuto, in modo da poter ottenere informazioni sul veicolo transitato.

L'attuale sistema informativo del servizio Telepass è frammentato ed eterogeneo. La direzione, pertanto, ha deciso di creare un data warehouse centralizzato per l'analisi del servizio Telepass sull'intera rete autostradale europea. In particolare, la direzione è interessata ad analizzare il pedaggio medio per km percorso sulla rete autostradale e il pedaggio medio per viaggio effettuato.

I caselli possono essere dotati solo di casse automatiche (es. corsie per il Telepass, corsie per pagamento con ViaCard, FastPay, ecc.), oppure possono avere solo casse con pagamento manuale (in questo caso l'identificativo del Telepass e la targa sono inseriti manualmente dall'operatore), oppure possono avere entrambe le opzioni. Ogni casello appartiene ad una sola autostrada, che percorre una determinata nazione. Per ogni autostrada, il massimo numero di corsie disponibili è noto. Dalla targa dei veicoli è possibile risalire ad alcune informazioni: il modello del veicolo (es. Ford Focus), il tipo (es. automobile) e la marca (es. Ford). Inoltre, esistono diverse versioni di Telepass, in base alle caratteristiche tecniche e agli aggiornamenti di firmware.

La direzione è interessata ad analizzare i pedaggi, i viaggi e i km percorsi per ogni casello di ingresso e di uscita, per i diversi metodi di pagamento disponibili nei caselli, per ogni autostrada e nazione. L'analisi deve essere svolta per ogni data, settimana, mese, bimestre, trimestre, semestre, anno, oltre che per giorno della settimana e per i giorni di vacanza.

Si noti che lo stesso dispositivo Telepass può essere usato su diversi veicoli e che lo stesso veicolo può avere diversi Telepass per differenti viaggi. Ciò si verifica quando un utente cambia contratto, quindi il suo dispositivo Telepass è sostituito, oppure quando lo stesso veicolo è usato sia per viaggi personali sia per viaggi di lavoro, quindi il Telepass personale o dell'ufficio sono rispettivamente usati. Infine si noti che ogni viaggio è associato alla data di uscita dalla rete autostradale.

Sono di seguito riportate alcune interrogazioni frequenti di interesse per la direzione.

- a) Per ogni anno e per ogni tipo di veicolo, considerando solo i viaggi in Italia, selezionare il pedaggio giornaliero medio per ogni mese e il totale mensile cumulativo dall'inizio dell'anno.**
- b) Considerando solo i viaggi effettuati all'interno della stessa autostrada e il dispositivo Telepass in versione "Oro", per ogni anno selezionare il pedaggio medio per km percorso per ogni autostrada e la percentuale di pedaggi (in €) pagati presso ciascuna autostrada rispetto al totale annuale.**
- c) Ordinare e assegnare un rank ai caselli di uscita per pedaggio medio per km percorso (il primo casello è quello col più alto pedaggio medio per km percorso e deve avere rank pari a 1). Per ogni casello di uscita, selezionare anche il numero medio di passaggi e il pedaggio medio per passaggio.**

- d) Per ogni casello di ingresso, selezionare il numero medio di passaggi al giorno, la somma dei pedaggi giornalieri e la percentuale di pedaggi pagati in ciascun giorno rispetto al totale mensile.
- e) Considerando solo le autostrade con il valore massimo di corsie pari a 3, selezionare il pedaggio medio per numero di viaggi semestrali di ogni autostrada di uscita, e il pedaggio medio semestrale per km percorso.
- f) Per ogni giorno della settimana, escludendo i giorni di vacanza, selezionare il numero totale di viaggi giornalieri, la somma dei pedaggi pagati e il totale dei chilometri percorsi sulle autostrade Italiane.

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:
 - Nazioni: ~20
 - Numero massimo di corsie: da 2 a 8
 - Autostrade: ~1000
 - Caselli: ~100 000
 - Versioni di Telepass: ~10
 - Modelli di veicoli: ~400
 - Marche di veicoli: ~20
 - Tipi di veicoli: ~20
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.