

Business Intelligence

3 Settembre 2015

Descrizione del problema

La società OrganizzaFiere gestisce l'organizzazione di Fiere in diversi settori/categorie (industriale, commerciale, servizi, ...) in Europa e per diverse categorie di utenti (aziende, privati, enti pubblici, ...). La fiera è un evento internazionale a cui partecipano diversi espositori, di solito aziende, per pubblicizzare i propri prodotti e/o servizi offerti, mentre diverse categorie di utenti (visitatori) sono invitati a partecipare pagando un biglietto di ingresso. Ogni fiera è caratterizzata dal nome, dall'anno in cui viene organizzata, dalla durata (data inizio e data fine) e dal luogo in cui viene svolto. La stessa fiera può essere organizzata più volte in anni diversi e eventualmente anche in città diverse, ma una singola edizione della stessa fiera in un anno. Una fiera è organizzata in un'area dedicata all'organizzazione delle fiere, denominata centro-fiere, che include diversi padiglioni. Ogni padiglione ha al proprio interno diverse aree espositive di dimensione variabile. Ogni azienda che desidera partecipare ad una fiera in qualità di espositore deve prenotare uno o più spazi espositivi e pagare la rispettiva quota di partecipazione. Un'azienda può allocare spazi espositivi di dimensione e caratteristiche diverse a seconda della data in cui si svolge la fiera. Ogni spazio espositivo è caratterizzato da una dimensione e da un elenco di servizi (e.g., accesso wi-fi, presenza di frigo-bar, ...) variabile a seconda della fiera che viene organizzata. Per esempio per la Fiera del libro Edizione 2015 a Torino erano disponibili 100 spazi espositivi. In ogni spazio era stata allocata una sola azienda, mentre la stessa azienda poteva allocare più spazi nella stessa fiera. Nello spazio espositivo X per la Fiera del libro 2014 erano disponibili 2 servizi: l'accesso Wi-Fi e il caffè americano.

I visitatori (utenti) che vogliono partecipare ad una fiera devono iscriversi sul corrispondente sito web della fiera indicando alcune informazioni (nome, cognome, data di nascita, sesso, indirizzo e-mail, stato civile, ...) e specificando la data (anche più di una) nella quale parteciperanno alla fiera e il mezzo pubblicitario (newsletter, mail, pubblicità televisiva, ...) con il quale sono venuti a conoscenza dell'evento. Inoltre i visitatori possono prenotare delle riunioni individuali con un referente di un'azienda che espone in fiera per avere maggiori dettagli sul prodotto/servizio offerto. Per ogni riunione si tiene traccia della durata effettiva.

Attualmente la società OrganizzaFiere ha diverse sedi distribuite sul territorio europeo e ogni sede ha una propria base di dati per gestire la propria attività. Ogni fiera è organizzata da una sola agenzia, mentre la stessa agenzia può organizzare diverse edizioni della stessa fiera, o fiere differenti. La società OrganizzaFiere è interessata a progettare e realizzare un data warehouse che offra una visione d'insieme della propria attività economica. L'azienda, durante il processo di ETL arricchisce il proprio data warehouse con altri dati interessanti disponibili in altre sorgenti dati.

Il data warehouse realizzato dovrà consentire di analizzare:

1. l'efficienza di organizzazione degli spazi espositivi per le diverse fiere;
2. la partecipazione degli utenti alle diverse fiere organizzate.

Per il punto (1) la società OrganizzaFiere è interessata ad analizzare il numero di aziende che espongono nelle diverse fiere, i metri quadrati totali degli spazi espositivi allocati, il numero medio di metri quadrati allocati per ogni azienda, il numero di installatori degli spazi espositivi richiesti in funzione di:

- data, mese, bimestre, trimestre, quadrimestre, semestre e anno in cui è stata organizzata la fiera;
- giorno della settimana, giorno festivo/feriale e settimana in cui è stata organizzata la fiera;
- fiera, categoria della fiera;
- centro-fiere in cui è organizzata la fiera;
- agenzia che organizza la fiera;
- città, provincia, regione e stato dell'agenzia che organizza la fiera;
- città, provincia, regione e stato in cui è sito il centro-fiere;
- dimensione (piccola, media, grande) dell'azienda che espone, categoria dell'azienda e tipologia di prodotto/servizio pubblicizzato;
- caratteristiche dello spazio espositivo.

Per il punto (2) la società OrganizzaFiere è interessata ad analizzare il numero medio di visitatori per spazio espositivo, il numero totale di riunioni effettuate con i visitatori, e il numero medio di ore per riunione in funzione di:

- caratteristiche del visitatore (per esempio, range di età, sesso, stato di residenza, ...);
- mezzo pubblicitario con il quale il visitatore è venuto a conoscenza della fiera;

- data della fiera, mese, bimestre, semestre e anno;
- spazio espositivo nel contesto di una fiera;
- padiglione in cui è sito lo spazio espositivo;
- fiera;
- centro-fiere;
- città, provincia, regione e stato in cui è sito il centro-fiere;
- azienda sita nello spazio espositivo nel contesto di una fiera;
- dimensioni dell'azienda (piccola, media, grande), categoria dell'azienda e tipologia di prodotto/servizio offerto.

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

- Considerando solo le aziende commerciali (Categoria=Commerciale) di media dimensione, visualizzare per ogni fiera e mese i metri quadri medi giornalieri occupati e il numero medio di installatori per azienda. Calcolare inoltre il rapporto tra i metri quadrati medi giornalieri occupati per ogni fiera in ogni mese rispetto ai metri quadrati medi giornalieri occupati da tutte le fiere organizzate nello stesso centro-fiere in ogni mese.**
- Considerando solo i visitatori di sesso femminile che sono venuti a conoscenza delle fiere mediante newsletters, visualizzare per ogni città e anno, i centro-fiere dove sono stati presenti il numero medio di visitatori per spazio espositivo pari al massimo di visitatori.**
- Considerando solo i visitatori di sesso maschile di età compresa nel range 18-25, separatamente per le aziende grandi, piccole e medie, calcolare per ogni mese e spazio espositivo il numero medio giornaliero di riunioni fissate e le ore totali dedicate alle riunioni. Calcolare inoltre il rapporto tra il numero medio di visitatori per spazio espositivo considerando tutti i centri-fiera nella stessa città e il numero medio di visitatori per spazio espositivo considerando tutti i centri-fiera nella stessa regione.**
- Considerando solo gli le fiere organizzate nel 2013 e 2014, per ogni coppia (fiera, dimensione azienda), visualizzare i metri quadrati medi giornalieri occupati in ogni trimestre e il numero medio di installatori per azienda.
- Per ogni coppia (città del centro-fiere, semestre) visualizzare il numero medio di visitatori per spazio espositivo nel contesto di una fiera e il numero totale di riunioni fissati.
- Visualizzare per ogni città in cui è sito almeno un centro-fiere la media dei metri quadrati occupati per azienda. Effettuare l'analisi separatamente per le aziende grandi, piccole e medie.
- Per ogni centro-fiere e per ogni mese visualizzare il numero totale di riunioni fissati. Effettuare l'analisi separatamente per i visitatori di sesso maschile e di sesso femminile.
- Per ogni centro-fiere in cui è stata allestita almeno una fiera nel 2012 visualizzare la media dei metri quadrati occupati per azienda e il numero di installatori medi per azienda. Effettuare l'analisi considerando solo gli spazi espositivi in cui non è disponibile l'accesso wi-fi e separatamente per ogni anno.

Progettazione

- Il data warehouse conterrà le informazioni relative agli anni 2009, 2010, 2011, 2012, 2013, 2014.
- Esprimere le interrogazioni frequenti **(a)**, **(b)**, **(c)** delle specifiche del problema utilizzando il linguaggio SQL esteso.
- 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.
- Decidere come gestire la dinamicità (variazioni) dei dati all'interno delle dimensioni. Motivare le scelte fatte.