

Esercizio di data warehouse design

Specifiche del problema

Una squadra di data scientist è stata incaricata di una serie di analisi sui dati raccolti da un'azienda di consegna di cibo. Per l'analisi devono considerare i ristoranti associati alle consegne. Per ogni ristorante si conosce il nome e la città in cui è situato. Inoltre, i ristoranti possono essere caratterizzati da una o più categorie (es. "Indiano", "Italiano", "Pizza", ..., "Hamburger", "Sandwich"). Il numero delle categorie è 10 e se ne conosce l'elenco completo. In base al ristorante, l'ordine associato alla consegna è caratterizzato da una tipologia che può assumere i valori "Trasporto o invio", "Ordine partner", "Ordine non partner". Ad esempio, gli ordini dei partner sono associati ai ristoranti partner dell'azienda di consegna di cibo.

È necessario considerare la zona di destinazione della consegna e la modalità di pagamento (es. "Bancomat", "Carta di credito", ...). L'analisi deve essere effettuata considerando la data e la fascia oraria delle consegne. Ogni fascia oraria dura esattamente 1 ora (es. "9-10", "10-11").

Ad ogni consegna è associato un rider. Per le analisi è necessario conoscere il mezzo di trasporto utilizzato dal ciclista (es. "bici", "auto").

L'azienda è interessata alle statistiche sul numero medio di chilometri percorsi dai ciclisti, sul tempo medio di consegna (in minuti) e sul ricavo medio per la consegna. L'analisi deve essere effettuata sulla base di:

1. nome del ristorante, categoria del ristorante, città, provincia, regione
2. tipo di ordine
3. consegna quartiere, città, provincia, regione
4. modalità di pagamento
5. mezzo di trasporto del corridore
6. data di consegna, mese, 2 mesi, trimestre, 4 mesi, 6 mesi, anno, giorno lavorativo
7. fascia oraria

Design

Progettare il data warehouse per soddisfare le specifiche e per rispondere in modo efficiente alle domande frequenti fornite. Disegnare lo schema concettuale del data warehouse e lo schema logico (tabelle dei fatti e delle dimensioni).

Scrivi le seguenti query frequenti utilizzando il linguaggio SQL esteso.

- a) Considerare gli ordini con tipologia "Ordine partner". Separatamente per mezzo di trasporto e trimestre, analizzare: il tempo medio di consegna, il numero medio di consegne per fascia oraria (il numero medio di consegne effettuate in un'ora), assegnare una classifica ai trimestri in base al numero decrescente di chilometri percorsi in media in un minuto, separatamente con mezzo di trasporto.
- b) Considerare i ristoranti che hanno la "pizza" tra le categorie associate. Effettua l'analisi separatamente per modalità di pagamento, città di consegna e mese. Analizzare: il ricavo mensile cumulativo dall'inizio di ogni trimestre, il ricavo medio per consegna, la percentuale del ricavo rispetto al ricavo totale considerando tutti i metodi di pagamento