



Data Science e Tecnologie per le Basi Dati

Politecnico di Torino

Homework 1

La Divisione di analisi del trasporto urbano dell'Autorità nazionale dei trasporti pubblici è interessata ad analizzare i ricavi dei servizi di trasporto pubblico.

In particolare, vorrebbero che le analisi affrontassero le seguenti caratteristiche:

Una rete di trasporto pubblico comprende vari tipi di modalità di trasporto, come autobus, tram, metropolitane e treni suburbani. Ogni modalità di trasporto ha un codice di identificazione unico e opera in una o più città. Ogni città appartiene a una specifica provincia e regione. Vengono memorizzate anche le informazioni sulla provincia e sulla regione.

Le modalità di trasporto sono ulteriormente suddivise in percorsi unici. Ogni percorso può avere diverse fermate o stazioni lungo il suo percorso e ogni fermata ha un identificatore unico. Ogni percorso ha alcuni servizi disponibili (aria condizionata, WIFI, posti a sedere speciali).

I biglietti acquistati dai passeggeri vengono registrati. Sono disponibili 4 tipi di biglietti: "Corsa singola", "Pass giornaliero", "Pass settimanale" e "Abbonamento mensile". Ogni tipo di biglietto ha un prezzo e un periodo di validità diversi. Inoltre, i biglietti possono essere acquistati con sconti diversi in base al tipo di passeggero: "Adulto", "Studente" (età compresa tra 14 e 24 anni), "Anziano" (età pari o superiore a 65 anni) e "Bambino" (età inferiore a 14 anni).

Il sistema memorizza anche le informazioni sulle modalità di acquisto dei biglietti. I biglietti possono essere acquistati in diversi modi: online tramite il sito web o l'applicazione mobile, presso i distributori automatici di biglietti situati nelle stazioni, presso i punti vendita autorizzati (ad esempio, i chioschi) o direttamente dal conducente/conduttore.

I passeggeri possono passare da una modalità di trasporto all'altra utilizzando un unico biglietto e ogni viaggio (o gita) viene registrato nel sistema. La registrazione comprende l'ora di inizio, la fermata iniziale, la fermata finale, la durata del viaggio, le modalità di trasporto utilizzate e la data. Vengono registrati anche i trasferimenti tra linee o modalità diverse.

L'azienda è interessata a statistiche sul ricavo medio e sulla durata del viaggio.

L'analisi deve essere effettuata tenendo conto dei seguenti dettagli:

- Modalità di trasporto (ad esempio, autobus, metropolitana, tram, treno suburbano), percorso, fermate di partenza e di arrivo e servizi
- Città, provincia e regione in cui avviene il viaggio
- Tipo di biglietto (corsa singola, giornaliero, settimanale, mensile), modalità di acquisto (online, distributore automatico, punto vendita autorizzato, autista) e sconto sul biglietto (studente, normale, bambino, senior).

- Dettagli del viaggio: giorno, mese, bimestre, trimestre, anno, fascia oraria e se la fascia oraria è di punta o non di punta.

Homework

1. Progettare il data warehouse per soddisfare le specifiche e rispondere in modo efficiente a tutte le interrogazioni frequenti fornite. Disegnare lo schema concettuale del data warehouse e lo schema logico (tabelle dei fatti e delle dimensioni).
2. Scrivete le seguenti query frequenti utilizzando il linguaggio SQL esteso:
 - a. Separatamente per ogni modalità di trasporto e per ogni mese dell'anno, analizzare: il numero medio giornaliero di biglietti, il numero cumulativo di biglietti dall'inizio dell'anno e la percentuale di biglietti utilizzati per ogni modalità di trasporto rispetto al numero totale di biglietti in quel mese.
 - b. Considerando i viaggi a partire dal 2022, separatamente per ogni modalità e città, analizziamo: la durata media del viaggio, i ricavi totali generati da quella città, la percentuale dei ricavi totali apportati da ogni percorso per la modalità corrispondente in una città, e assegniamo una classifica a ogni percorso all'interno della sua modalità di trasporto in base ai ricavi totali generati in ordine decrescente.
3. Creare e aggiornare una vista materializzata con CREATE MATERIALIZED VIEW e CREATE MATERIALIZED VIEW LOG in ORACLE

Query frequenti di interesse:

- Analizzare separatamente per ogni modalità di trasporto e per ogni mese il numero medio giornaliero di biglietti.
- Separatamente per ogni modalità di trasporto e per ogni mese, analizzare il numero cumulativo di biglietti dall'inizio dell'anno.
- Separatamente per ogni modalità di trasporto e per ogni mese, analizzate il numero totale di biglietti venduti, l'incasso totale e l'incasso medio.
- Separatamente per ogni modalità di trasporto e per ogni mese, analizzate il numero totale di biglietti venduti, l'incasso totale e l'incasso medio per l'anno 2024.
- Analizzare la percentuale di biglietti relativi a ciascuna modalità di trasporto e a ciascun mese rispetto al numero totale di biglietti del mese per ciascuna modalità di trasporto.
 - a. Definire una vista materializzata con CREATE MATERIALIZED VIEW è utile per ridurre il tempo di risposta delle query frequenti riportate.
 - b. Definire i log delle viste materializzate con CREATE MATERIALIZED VIEW LOG per ogni tabella in cui lo si ritiene necessario. Per quali tabelle è utile tenere traccia dei log? Identificare tutte e solo le tabelle necessarie. Inoltre, per ogni

tabella identificare tutti e soli gli attributi per i quali è necessario tenere traccia delle variazioni.

- c. Specificare quali operazioni (ad esempio INSERT su una tabella specifica) causano l'aggiornamento della MATERIALIZED VIEW definita.
4. Aggiornamento e gestione delle viste tramite trigger. Supponendo che il comando CREATE MATERIALIZED VIEW non sia disponibile, create le viste materializzate definite nell'esercizio precedente e definite la procedura di aggiornamento a partire dalle modifiche sulla tabella dei fatti create tramite un trigger.
- a. Creare la struttura della vista materializzata con CREATE TABLE VM1 (...)
 - b. Specificare un esempio di istruzione per popolare la tabella VM1 con i record necessari utilizzando l'istruzione INSERT INTO VM1 (...) (SELECT ...)
 - c. Scrivere i trigger necessari per propagare le modifiche (inserimento di un nuovo record) apportate alla tabella FACTS nella vista materializzata VM1.
 - d. Specificare quali operazioni (ad esempio, INSERT) attivano il trigger creato al punto 4.c.

NOTA

Per qualsiasi problema, scrivete un'e-mail con i seguenti metadati:

A: daniele.regecambrin@polito.it, davide.napolitano@polito.it

Oggetto: [DSTBD] Bug Homework 1

Corpo: <Descrizione del problema>.

Le e-mail senza il formato precedente potrebbero andare perse.
