

Esercizi Esame – III

Business Intelligence
per Big Data



Politecnico
di Torino

Politecnico di Torino

Schema concettuale 1



Una compagnia di trasporti pubblici che opera in diverse città italiane intende adottare un sistema di data warehousing per raccogliere e analizzare in modo efficiente i dati relativi all'utilizzo quotidiano dei suoi servizi. L'obiettivo è comprendere meglio i flussi di passeggeri, individuare le fasce orarie più critiche, analizzare l'efficienza delle linee e migliorare la pianificazione e la qualità complessiva del servizio offerto.

Il sistema deve supportare analisi dettagliate basate su informazioni relative alle corse effettuate, alle caratteristiche delle linee, alle abitudini dei passeggeri, e a dati temporali (come giorno della settimana, fascia oraria e festività). Questo sistema permetterà di valutare le performance di ogni linea, ottimizzare la distribuzione dei mezzi e identificare eventuali situazioni di sovraccarico o sottoutilizzo.

Si vogliono analizzare le corse effettuate in base al tipo di mezzo pubblico (autobus, tram o metropolitana). Ogni corsa è associata ad uno specifico tipo di mezzo.

Si vogliono poi analizzare le corse in base al quartiere di partenza della corsa. Si memorizzano anche la città, provincia e regione associata al quartiere. Ogni quartiere è associato ad una sola città e ad una città sono associati più quartieri.

Il sistema ha accesso alle informazioni relativi ai/alle passeggeri/e della corsa e vuole analizzarne alcune caratteristiche. In particolare, il sistema deve analizzare le informazioni in base al tipo di titolo. Ci sono 4 tipo di tipo titolo: abbonamento mensile, abbonamento annuale, biglietto singolo, 'biglietto daily'. Ogni informazione delle corse è associata ad un solo tipo di titolo. Si registra inoltre la data della corsa e la fascia oraria (6-10, 10-14, 14-18, 18-22, 22-2, 2-6). Si registra inoltre se si tratta di un giorno festivo o feriale, il giorno della settimana, il mese e l'anno.

Schema concettuale 1



Le analisi devono essere effettuate al variare delle seguenti condizioni:

- tipo di mezzo pubblico
- quartiere di partenza della corsa
- città, provincia, regione
- tipo di abbonamento
- data di partenza della corsa
- giorno della settimana
- fascia oraria di partenza della corsa (6-10, 10-14, 14-18, 18-22, 22-2, 2-6)
- indicazione se si tratta di giorno festivo o feriale
- mese e anno della corsa

Le analisi devono permettere di ottenere indicatori utili a migliorare l'organizzazione del servizio, tra cui il numero medio di passeggeri/e per corsa.

Esempi di analisi richieste sono (i) la media dei passeggeri per corsa in funzione del giorno della settimana e della fascia oraria, il numero di passeggeri per tipologia di mezzo, il confronto del numero di passeggeri tra utenti abbonati e non abbonati, il numero di corse giornaliero separatamente per quartiere.

Schema concettuale 2



Una rete di centri sportivi diffusi su tutto il territorio nazionale vuole implementare un sistema di data warehousing per monitorare l'andamento delle iscrizioni, la partecipazione alle attività sportive, l'utilizzo delle strutture e il profilo degli utenti.

L'obiettivo è migliorare la gestione dei centri, ottimizzare la programmazione delle attività, comprendere le preferenze dei clienti e sviluppare strategie promozionali efficaci.

Si vogliono analizzare tali informazioni sulla base del centro sportivo. Ogni centro sportivo è identificato da un nome univoco, ed è situato in una città, provincia e regione. Ogni centro può offrire diverse attività sportive, ad esempio: nuoto, fitness, yoga, tennis, arti marziali, calcio. Tali attività sono varie e in numero ampio e possono essere proposte nuove attività nel tempo. Ogni centro può offrire una o più attività, e ogni attività sportiva può essere offerta da più centri.

Si vuole tener traccia della partecipazione degli e delle utenti iscritti/e al centro alle attività in funzione di alcune caratteristiche utente. Le analisi devono essere effettuate in funzione del genere e età (classificata in fasce: <18, 18-30, 30-50, 50-70, 70+) dell'utente. Si vuole inoltre memorizzare il tipo di abbonamento dell'utente associato all'ingresso al centro: mensile, trimestrale, annuale, ingresso singolo.

Si vuole inoltre memorizzare se l'ingresso al centro è associato ad un evento speciale (ad esempio (tornei, giornate open, workshop) o meno. Tale informazione è memorizzata solo come partecipazione a evento speciale o no, senza l'informazione dettagliata sul tipo di evento speciale.

Si vuole poi memorizzare la data associata all'ingresso al centro per lo svolgimento dell'attività sportiva. Si memorizza la data, il giorno della settimana, il mese, il bimestre e l'anno. Si vuole memorizzare inoltre le condizioni meteo associate alla data. Si ipotizza di memorizzare una sola condizione meteo per data (e.g., sole, pioggia, nuvoloso, ventoso).

Schema concettuale 2



Infine, si vuole memorizzare l'attività svolta per lo specifico ingresso al centro sportivo. Si ipotizzi che ogni ingresso al centro sia associato ad una sola attività (e.g., ingresso associato ad attività tennis).

Le analisi devono essere effettuate al variare di:

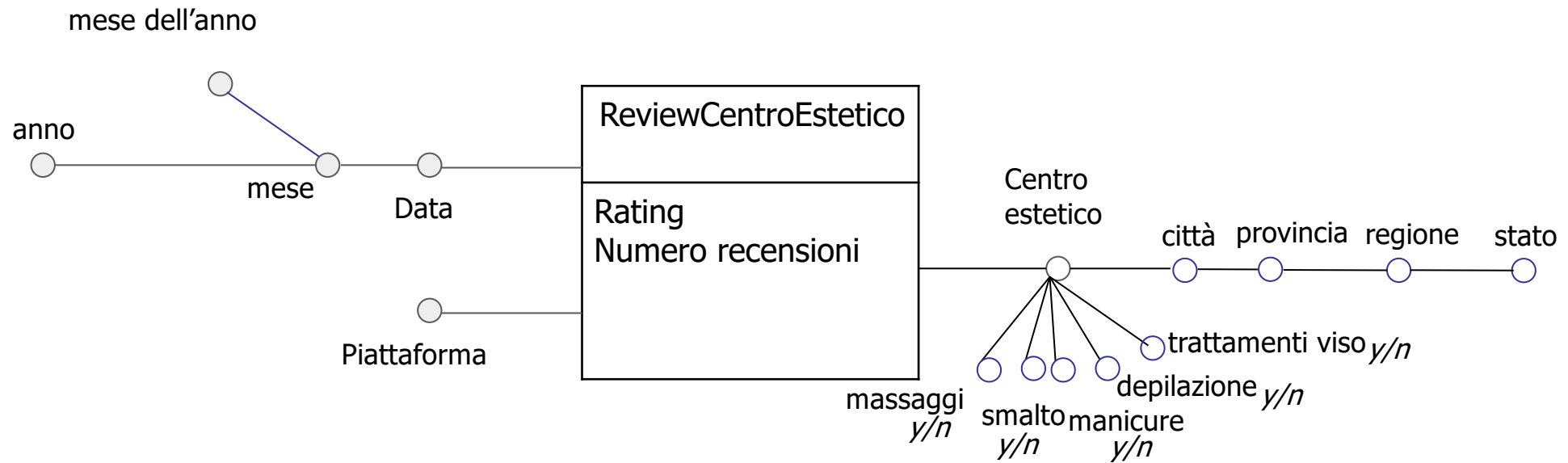
- centro sportivo
- città, provincia, regione del centro sportivo
- attività sportive offerte dal centro sportivo
- fascia di età dell'utente
- genere dell'utente
- tipo di abbonamento dell'utente
- se ingresso è legato a partecipazione a eventi speciali
- data dell'ingresso, mese, anno, bimestre, giorno della settimana
- condizioni meteo
- attività sportiva associata all'ingresso

Esempi di analisi di interesse sono il numero medio giornaliero di ingressi per centro sportivo, la durata media per ingresso, il confronto del numero di ingressi fra abbonati annuali e utenti a ingresso singolo, la percentuale di ingressi associati a eventi speciali.

Esercizio Progettazione Logica



Una rete internazionale di centri estetici desidera effettuare analisi dettagliate sulle recensioni fornite online dai clienti per migliorare la qualità dei servizi e ottimizzare le operazioni.



La dimensione 'Piattaforma' memorizzare il tipo di piattaforma dalla quale è stata effettuata la recensione che viene monitorata dalla rete di centri estetici. Vengono monitorate 3 piattaforme: Google ('GO'), Trustpilot ('TP'), Tripadvisor ('TA').

Dato il seguente schema logico relazione di un sistema di data warehouse (le chiavi primarie sono sottolineate e in grassetto)

SENSORI-IOT

(**ID Sensore**, Tipo Sensore, Modello Sensore, Categoria Sensore, Stato Operativo, Data Installazione)

ZONE-URBANE (**ID Zona Urbana**, Nome Zona, Tipo Zona, Quartiere, Municipio, Città, Regione)

TEMPO (**ID Tempo**, Data, Giorno Settimana, IsGiornoFeriale, IsWeekend, Settimana, Mese, Trimestre, Semestre, Anno)

RILEVAMENTI-SMART-CITY

(**ID Sensore**, **ID Zona Urbana**, **ID Tempo**, Valore Misurato, Numero Rilevamenti, Costo Operativo, Consumo Energetico)

SQL Esteso 1



Si assume che un municipio sia essere associato a una sola città, ma una città possa avere più municipi, allo stesso modo un quartiere è associato a un solo municipio e un municipio può avere più quartieri.

- Separatamente per ogni Municipio e mese, analizza:
 - la media mensile dei valori misurati per Quartiere all'interno del Municipio;
 - la proporzione dei valori misurati che ciascun Municipio ha registrato rispetto al totale del mese su tutta la città ad esso associata;
 - la classifica (rank) dei municipi, all'interno di ciascun mese, sulla base della media dei valori misurati per quartiere, dal più grande al più piccolo.

SQL Esteso 2



Con riferimento agli anni dal 2020 al 2024. Per ciascun municipio e per ogni categoria di sensore, seleziona i seguenti indicatori:

- il costo operativo medio trimestrale per ogni rilevamento;
- il numero cumulativo di rilevamenti registrati dal primo trimestre dell'anno fino al trimestre corrente;
- la media mobile dei rilevamenti sugli ultimi 3 trimestri (incluso quello corrente).

SQL Esteso 3



Sapendo che un tipo di sensore è associato a una singola categoria, che a sua volta può avere più sensori associati, per ciascun mese registrato, tipo di zona urbana e per ogni tipo di sensore, analizza:

- la media giornaliera dei valori misurati;
- la percentuale dei consumi energetici rispetto al totale mensile registrato in tutte le zone urbane;
- la cumulata mensile dei consumi energetici misurati per ciascuna categoria di sensore dal primo mese dell'anno fino al mese corrente;

Vista materializzata



A partire dalle seguenti query di interesse

- per ogni città situata nelle regioni Piemonte, Lombardia e Veneto, visualizzare valore medio misurato per rilevamento, per categoria di sensore e anno.
- per ciascuna categoria di sensore e semestre degli ultimi 10 anni, calcolare il consumo energetico medio per rilevamento.
- visualizzare per ogni tipo di zona urbana il costo operativo totale, per ogni trimestre.

a) Definire una vista materializzata con CREATE MATERIALIZED VIEW, in modo da ridurre il tempo di risposta delle query riportate. In particolare si specifichi la query SQL associata al Blocco A nella seguente istruzione:

```
CREATE MATERIALIZED VIEW ViewSensori  
BUILD IMMEDIATE  
REFRESH FAST ON COMMIT  
AS  
Blocco A
```

b) Individuare la combinazione minimale di attributi che costituisce un identificatore per la vista materializzata