

Data Science e Tecnologie per le Basi di Dati

Politecnico di Torino

Data warehousing in Oracle - Esercitazione 1

Lo scopo di questa esercitazione è realizzare un data warehouse conforme alle specifiche elencate di seguito, utilizzando Oracle. Successivamente, si scriveranno delle query in SQL esteso per recuperare i dati dal data warehouse progettato.

La traccia dell'esercitazione è la seguente:

- 1. Descrizione del problema
- 2. Descrizione del database OLTP
- 3. **Esercizio:** progettazione del data warehouse
- 4. **Esercizio:** confronto con lo schema logico del data warehouse
- 5. **Esercizio:** interrogazione del data warehouse

1. Descrizione del problema

Una compagnia telefonica è interessata ad analizzare i propri dati per migliorare i servizi ai clienti. Attualmente, l'azienda dispone di un database con i registri delle chiamate. Per ogni chiamata sono noti i numeri di telefono del chiamante e del destinatario, la durata, il tipo di addebito (ad esempio, tariffe di punta, non di punta), l'ora di inizio (data, ora, minuti, secondi). I gestori vogliono ottenere molto velocemente le informazioni sul traffico telefonico sulle linee aziendali e il reddito giornaliero in base alla posizione del chiamante, al giorno e alla tariffa telefonica.

In particolare, i manager vogliono analizzare le seguenti situazioni:

- Reddito netto mensile e numero di chiamate per ogni città chiamante.
- Reddito netto mensile e numero di chiamate per ogni città ricevente.
- Reddito netto mensile e numero di chiamate per ogni provincia e regione del chiamante.
- Reddito netto mensile e numero di chiamate per ogni provincia e regione ricevente.
- Reddito netto giornaliero e numero di chiamate per ogni provincia chiamante.
- Reddito netto annuo e numero di chiamate per ogni provincia e regione del chiamante.
- Reddito netto mensile e numero di chiamate per ogni tariffa telefonica (tipo di addebito).
- Reddito netto e numero di chiamate per ogni giorno della settimana e tariffa telefonica.
- Numero giornaliero di chiamate per ogni regione chiamante.
- Numero giornaliero di chiamate per ogni regione del destinatario.

2. Descrizione del database OLTP

Il database OLTP della compagnia telefonica è riportato nella Tabella 1.

Tabelle			Descrizione	
DWABD.TARIFFE			Tipologie di	
Tip also sin To siffe		NOT NUUL	tariffe	
TipologiaTariffa NomeTariffa	INT	NOT NULL,	Cardinalità: 7	
CostoAlSecondo	VARCHAR(20)	NOT NULL,	tuple	
CostoAlSecolido	FLOAT	NOT NULL,	tapio	
PRIMARY KEY (TipologiaTariffa)				
);				
DWABD.LUOGO			Luogo	
(
CodLuogo	INT	NOT NULL,	Cardinalità:	
Città	VARCHAR(20)	NOT NULL,	1500 tuple	
Provincia	VARCHAR(20)	NOT NULL,		
Regione	VARCHAR(20)	NOT NULL,		
PRIMARY KEY (CodLuogo)				
);				
DWABD.TELEFONATE			Telefonate	
TelChiamante	\/ADCHAD(20)	NOT NULL,	effettute negli anni 2023 e	
TelChiamato	VARCHAR(20) VARCHAR(20)	NOT NULL,	2024	
LuogoChiamante	INT	NOT NULL,		
_	INT	NOT NULL,	Cardinalità:	
LuogoChiamato	DATE	NOT NULL,	~ 1300000	
Data	INT	NOT NULL,	tuple	
Ora	INT	NOT NULL,		
Minuti	INT	NOT NULL,		
Secondi	FLOAT	NOT NULL,		
DurataTelefonata	INT	NOT NULL,		
TipologiaTariffa	2	,		
PRIMARY				
PRIMARY	inuti Canadi)			
KEY(TelChiamante, TelChiamato, Data, Ora, M FOREIGN KEY(TipologiaTariffa) REFERENCES T				
CASCADE,				
FOREIGN KEY(LuogoChiamante) REFERENCES Luogo(CodLuogo), ON DELETE CASCADE,				
FOREIGN KEY(LuogoChiamato) REFERENCES Luogo(CodLuogo) ON DELETE CASCADE				
);				

Tabella 1 – Database sorgente con informazioni sulle singole telefonate

3. Esercizio: progettazione del data warehouse

Progettare lo schema concettuale di un data warehouse per la gestione delle problematiche discusse in precedenza. Lo schema progettato deve essere concepito in modo da consentire:

- Le analisi richieste dalla compagnia di telefonia mobile
- La fase ETL (estrazione, trasformazione, caricamento) per importare i dati dal database OLTP (Tabella 1) al data warehouse.

La progettazione dello schema concettuale può essere eseguita su carta o su qualsiasi software tu voglia.

4. Esercizio: confronto con lo schema logico del data warehouse

Confrontare lo schema concettuale progettato nell'esercizio precedente con lo schema logico proposto nella Tabella 2. Verificare di aver scelto correttamente le misure e il livello di aggregazione dei dati. Esaminando lo schema logico nella Tabella 2, rispondere alle seguenti domande:

- Quali sono le misure scelte per il data warehouse?
- Qual è il livello minimo di aggregazione nel data warehouse? Corrisponde a ciò che è stato progettato nello schema concettuale?

La progettazione dello schema logico può essere fatta su carta o su qualsiasi software tu voglia.

Tabelle			Descrizione
DWABD.TEMPO (CodTempo Data GiornoSettimana Mese	INT DATE CHAR(15) CHAR(15)	NOT NULL, NOT NULL, NOT NULL, NOT NULL,	Dimensione tempo. Cardinalità: 10 tuple
Anno PRIMARY KEY(CodTempo));	NUMBER	NOT NULL,	
DWABD.TARIFFA (CodTar Tipo_Tariffa PRIMARY KEY(CodTar));	INTEGER VARCHAR(20)	NOT NULL, NOT NULL,	Dimensione tariffa. Cardinalità: 7 tuple

DWABD.LUOGO			
(Dimensione
CodLuogo	INTEGER	NOT NULL,	luogo
Citta	VARCHAR(20)	NOT NULL,	Cardinalità:
Provincia	CHAR(20)	NOT NULL,	1500 tuple
Regione	CHAR(20)	NOT NULL,	
PRIMARY KEY(CodLuogo)			
);			
DWABD.FATTI			
(Tabella dei
CodTempo	INTEGER	NOT NULL,	fatti.
CodTar	INTEGER	NOT NULL,	Cardinalità:
CodLuogo_Chiamante	INTEGER	NOT NULL,	7809 tuple
CodLuogo_Chiamato	INTEGER	NOT NULL,	
Prezzo	FLOAT	NOT NULL,	
NumChiamate	INTEGER	NOT NULL,	
PRIMARY KEY(CodTempo, CodTar, CodL	.uogo Chiamante, CodLuogo Ch	iamato),	
FOREIGN KEY(CodTempo) REFERENCE	· · · · ·	<i>"</i>	
FOREIGN KEY(CodTar) REFERENCES T	1 /-		
FOREIGN KEY(CodLuogo Chiamante) RI		o),	
FOREIGN KEY(CodLuogo Chiamato) RE	, ,	, ·	
);	. 5	•	

Tabella 2 - Soluzione proposta - Tabelle del data warehouse

5. Esercizio (SQL): Esecuzione di query sul data warehouse

Per l'esercizio seguente è necessario utilizzare una delle seguenti opzioni:

- Oracle FreeSQL: Servizio Oracle che permette di istanziare un Database SQL in remoto (istruzioni più dettagliate sulla pagina del corso). È necessario creare un account Oracle. Per questa opzione sono necessari gli script SQL del dataset.
- <u>SQLite su Colab</u>: Notebook in Python per instanziare un Database. Poiché SQLite non è Oracle, non sono disponibili alcune funzionalità avanzate proprie di Oracle. Per questa opzione sono necessarie le tabelle in CSV del dataset.

Le tabelle corrispondenti allo schema nella Tabella 2 e contengono alcuni dati di esempio. Utilizzare queste tabelle come origine per le query seguenti.

- 1. Seleziona il reddito annuo per ogni tariffa telefonica, il reddito totale per ogni tariffa telefonica, il reddito annuo totale e il reddito totale.
- 2. Seleziona il numero mensile di chiamate e il reddito mensile. Associa il RANK() a ogni mese in base al suo reddito (1 per il mese con il reddito più alto, 2 per il secondo, ecc., l'ultimo mese è quello con il reddito più basso).

- 3. Per ogni mese del 2023, selezionare il numero totale di chiamate. Associa il RANK() a ciascun mese in base al suo numero totale di chiamate (1 per il mese con il maggior numero di chiamate, 2 per il secondo, ecc., l'ultimo mese è quello con il minor numero di chiamate).
- 4. Per ogni giorno del mese di luglio 2023, selezionare il reddito totale e il reddito medio degli ultimi 3 giorni.
- 5. Seleziona il reddito mensile e il reddito mensile cumulativo dall'inizio dell'anno.
- 6. Consideriamo l'anno 2023. Separatamente per tariffa telefonica e mese, analizza (i) il totale delle entrate, (ii) la percentuale di entrate rispetto alle entrate totali considerando tutte le tariffe telefoniche, (iii) la percentuale delle entrate rispetto alle entrate totali considerando tutti i mesi.
- 7. Per ogni regione del chiamante, selezionare il numero mensile di chiamate e il numero cumulativo di chiamate mensili dall'inizio dell'anno.
- 8. Consideriamo l'anno 2023. Analizza il reddito totale per (i) separatamente per ogni mese e (ii) separatamente per ogni mese, tariffa telefonica e regione del chiamante e (iii) separatamente per ogni mese, tariffa telefonica e regione del destinatario.