



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 (SQL Developer): 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.

Tabella	Descrizione
<p>DWABD.TARIFFE (TipologiaTariffa INT NOT NULL, NomeTariffa VARCHAR(20) NOT NULL, CostoAlSecondo FLOAT NOT NULL, PRIMARY KEY (TipologiaTariffa));</p>	<p>Tipologie di tariffe</p> <p>Cardinalità: 7 tuple</p>
<p>DWABD.LUOGO (CodLuogo INT NOT NULL, Città VARCHAR(20) NOT NULL, Provincia VARCHAR(20) NOT NULL, Regione VARCHAR(20) NOT NULL, PRIMARY KEY (CodLuogo));</p>	<p>Luogo</p> <p>Cardinalità: 1500 tuple</p>
<p>DWABD.TELEFONATE (TelChiamante VARCHAR(20) NOT NULL, TelChiamato VARCHAR(20) NOT NULL, LuogoChiamante INT NOT NULL, LuogoChiamato INT NOT NULL, Data DATE NOT NULL, Ora INT NOT NULL, Minuti INT NOT NULL, Secondi INT NOT NULL, DurataTelefonata FLOAT NOT NULL, TipologiaTariffa INT NOT NULL, PRIMARY KEY(TelChiamante, TelChiamato, Data, Ora, Minuti,Secondi), FOREIGN KEY(TipologiaTariffa) REFERENCES Tariffe(TipologiaTariffa) ON DELETE CASCADE, FOREIGN KEY(LuogoChiamante) REFERENCES Luogo(CodLuogo), ON DELETE CASCADE, FOREIGN KEY(LuogoChiamato) REFERENCES Luogo(CodLuogo) ON DELETE CASCADE);</p>	<p>Telefonate effettuate negli anni 2003 e 2004</p> <p>Cardinalità: ~ 1300000 tuple</p>

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 INT NOT NULL, Data DATE NOT NULL, GiornoSettimana CHAR(15) NOT NULL, Mese CHAR(15) NOT NULL, Anno NUMBER NOT NULL, PRIMARY KEY(CodTempo));			Dimensione tempo. Cardinalità: 10 tuple
DWABD.TARIFFA (CodTar INTEGER NOT NULL, Tipo_Tariffa VARCHAR(20) NOT NULL, PRIMARY KEY(CodTar));			Dimensione tariffa. Cardinalità: 7 tuple
DWABD.LUOGO (CodLuogo INTEGER NOT NULL, Citta VARCHAR(20) NOT NULL, Provincia CHAR(20) NOT NULL, Regione CHAR(20) NOT NULL, PRIMARY KEY(CodLuogo));			Dimensione luogo Cardinalità: 1500 tuple

DWABD.FATTI			Tabella dei fatti. Cardinalità: 7809 tuple
(
CodTempo	INTEGER	NOT NULL,	
CodTar	INTEGER	NOT NULL,	
CodLuogo_Chiamante	INTEGER	NOT NULL,	
CodLuogo_Chiamato	INTEGER	NOT NULL,	
Prezzo	FLOAT	NOT NULL,	
NumChiamate	INTEGER	NOT NULL,	
PRIMARY KEY(CodTempo, CodTar, CodLuogo_Chiamante, CodLuogo_Chiamato),			
FOREIGN KEY(CodTempo) REFERENCES TEMPO(CodTempo),			
FOREIGN KEY(CodTar) REFERENCES TARIFFA(CodTar),			
FOREIGN KEY(CodLuogo_Chiamante) REFERENCES LUOGO(CodLuogo),			
FOREIGN KEY(CodLuogo_Chiamato) REFERENCES LUOGO(CodLuogo)			
);			

Tabella 2 - Soluzione proposta - Tabelle del data warehouse

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

Per l'esercizio seguente è necessario utilizzare Oracle Live (**Consigliato**) o Oracle SQL Developer. Puoi trovare le istruzioni nel file zip associato a Lab1.

Le tabelle corrispondenti allo schema nella Tabella 2 sono già state create in Oracle 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 2003, 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 2003, 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 2003. 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 2003. 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.