

Business Intelligence

Esercitazione di laboratorio n. 2

La finalità di questa esercitazione consiste nella realizzazione di un data warehouse conforme alle specifiche riportate nei punti seguenti e nel rispondere ad alcune query utilizzando Oracle (per i dettagli relativi alla connessione a Oracle fare riferimento alla precedente esercitazione di laboratorio). L'esercitazione può essere svolta usando Oracle Developer o Oracle XE.

1. Descrizione del problema

Una società di telefonia mobile è interessata ad analizzare i dati a sua disposizione per fornire un servizio mirato ai suoi clienti e per migliorare la distribuzione delle proprie apparecchiature sul territorio. Attualmente la società telefonica dispone di basi di dati contenenti tutte le informazioni relative alle chiamate effettuate dai suoi utenti. In particolare, per ogni chiamata sono noti il numero di telefono del chiamante, il numero di telefono del chiamato, la durata della telefonata, la tariffa applicata e l'istante di inizio della chiamata (data, ora, minuto, secondi).

La dirigenza della società vuole poter ottenere velocemente delle informazioni sul traffico telefonico effettuato sulle linee telefoniche dell'azienda e sui guadagni effettuati su base giornaliera in funzione della località del chiamante, del giorno e della tariffa.

In particolare alcune delle informazioni a cui i dirigenti sono interessati sono:

- incassi e numero di chiamate effettuate su base mensile in funzione della città nella quale si trova l'apparecchio del chiamante
- incassi e numero di chiamate effettuate su base mensile in funzione della città nella quale si trova l'apparecchio del chiamato
- incassi e numero di chiamate effettuate su base mensile in funzione della provincia e della regione nella quale si trova l'apparecchio del chiamante
- incassi e numero di chiamate effettuate su base mensile in funzione della provincia e della regione nella quale si trova l'apparecchio del chiamato
- incassi e numero di chiamate effettuate in base alla data nella quale si effettua la chiamata e della provincia nella quale si trova l'apparecchio del chiamante
- incassi e numero di chiamate effettuate su base annua in funzione della provincia e della regione nella quale si trova l'apparecchio del chiamante
- incassi e numero di chiamate effettuate su base mensile in funzione della tipologia di tariffa utilizzata
- incassi e numero di chiamate effettuate in funzione del giorno della settimana e della tipologia di tariffa utilizzata
- numero di chiamate effettuate in funzione della data e della regione nella quale si trova l'apparecchio del chiamante
- numero di chiamate effettuate in funzione della data e della regione nella quale si trova l'apparecchio del chiamato

2. Progettazione del data warehouse

- Progettare un data warehouse per la gestione delle problematiche illustrate in precedenza (progettazione concettuale e logica).

3. Interrogazione del data warehouse

Viene fornita una possibile soluzione per il problema proposto. Eseguire lo script associato al testo dell'esercitazione per caricare le tabelle e i dati della soluzione proposta. Usare tale base di dati come sorgente per i punti successivi. Risolvere usando il linguaggio SQL (versione estesa) le seguenti interrogazioni:

1. Selezionare per ogni coppia (tipo tariffa, anno) l'incasso totale. Selezionare inoltre l'incasso totale per ogni tipo di tariffa indipendentemente dall'anno, l'incasso totale per ogni anno indipendentemente dal tipo di tariffa e l'incasso totale (indipendentemente dall'anno e dal tipo tariffa).
2. Selezionare per ogni mese il numero di chiamate totali e l'incasso totale. Utilizzando la funzione RANK(), associare ad ogni record un numero che identifica la posizione del mese in funzione dell'incasso totale effettuato (1 per il mese che in cui è stato incassato di più).

3. Selezionare per ogni mese dell'anno 2003 il numero di chiamate totali. Utilizzando la funzione RANK(), associare ad ogni record un numero che identifica la posizione del mese in funzione del numero di chiamate totali (1 per il mese in cui sono state effettuate più chiamate).
4. Selezionare per ogni data del mese di luglio 2003 l'incasso totale, e la media degli incassi delle chiamate effettuate negli ultimi tre giorni.
5. Selezionare per ogni mese, l'incasso del mese e l'incasso cumulativo dall'inizio dell'anno.
6. Considerando solo le chiamate effettuate nel 2003 con tipo tariffa mattino (Tipo_Tariffa='Mattino'), visualizzare per ogni mese il numero di chiamate, l'incasso totale e l'incasso mensile cumulativo dall'inizio dell'anno, separatamente per ogni regione del chiamante.
7. Considerando solo la città del chiamato 6 (Citta='Citta6'), per ogni città del chiamante e per ogni mese, visualizzare l'incasso medio giornaliero, il numero totale di chiamate e l'incasso totale. Visualizzare inoltre il numero di chiamate totali, separatamente per città del chiamante (Indipendentemente dal mese).

4. Creazione delle viste materializzate mediante Oracle XE

- Basandosi sulle cardinalità delle tabelle del data warehouse (tabelle dei fatti e tabelle delle dimensioni), decidere quali viste materializzate possono essere definite al fine di ridurre i tempi di risposta per le interrogazioni appena risolte.
- Provare a creare alcune delle viste materializzate ritenute utili per almeno due interrogazioni e riscrivere tali interrogazioni SQL utilizzando la vista definita.

```
CREATE MATERIALIZED VIEW Nome
BUILD IMMEDIATE
REFRESH COMPLETE ON DEMAND
AS
Query
```

Tabelle	Descrizione
<p>TARIFFE</p> <pre>(TipologiaTariffa INT NOT NULL, NomeTariffa VARCHAR(20) NOT NULL, CostoAlSecondo FLOAT NOT NULL, PRIMARY KEY(TipologiaTariffa));</pre>	<p>Tipologie (categorie) di tariffe esistenti</p> <p>Cardinalità: 7 tuple</p>
<p>LOCALITA</p> <pre>(CodLocalita INT NOT NULL, Citta VARCHAR(20) NOT NULL, Provincia VARCHAR(20) NOT NULL, Regione VARCHAR(20) NOT NULL, PRIMARY KEY(CodLocalita));</pre>	<p>Località</p> <p>Cardinalità: 1500 tuple</p>
<p>TELEFONATE</p> <pre>(TelChiamante VARCHAR(20) NOT NULL, TelChiamato VARCHAR(20) NOT NULL, LocalitaChiamante INT NOT NULL, LocalitaChiamato INT NOT NULL, Data DATE NOT NULL, Ora INT NOT NULL, Minuti INT NOT NULL, Secondi INT NOT NULL, DurataTelefonataSecondi FLOAT NOT NULL, TipologiaTariffa INT NOT NULL, PRIMARY KEY(TelChiamante, TelChiamato, Data, Ora, Minuti, Secondi), FOREIGN KEY(TipologiaTariffa) REFERENCES TARIFFE(TipologiaTariffa) , FOREIGN KEY(LocalitaChiamante) REFERENCES LOCALITA(CodLocalita), FOREIGN KEY(LocalitaChiamato) REFERENCES LOCALITA(CodLocalita));</pre>	<p>Telefonate effettuate negli anni 2003, 2004</p> <p>Cardinalità: 422416 tuple</p>

Figura 1 – Base di dati sorgente contenente le informazioni sulle singole telefonate

Tabelle	Descrizione
DWABD.TEMPO (ID_TEMPO INT NOT NULL, DATA DATE NOT NULL, GIORNO CHAR(15) NOT NULL, MESE CHAR(15) NOT NULL, ANNO INT NOT NULL, PRIMARY KEY(ID_TEMPO));	Dimensione tempo Cardinalità: 30 tuple
DWABD.TARIFFA (ID_TAR INT NOT NULL, TIPO_TARIFFA CHAR(20) NOT NULL, PRIMARY KEY(ID_TAR));	Dimensione tariffa Cardinalità: 7 tuple
DWABD.LUOGO (ID_LUOGO INT NOT NULL, CITTA CHAR(20) NOT NULL, PROVINCIA CHAR(20) NOT NULL, REGIONE CHAR(20) NOT NULL, PRIMARY KEY(ID_LUOGO));	Dimensione luogo (località) Cardinalità: 1500 tuple
DWABD.FATTI (ID_TEMPO INT NOT NULL, ID_TAR INT NOT NULL, ID_LUOGO_CHIAMANTE INT NOT NULL, ID_LUOGO_CHIAMATO INT NOT NULL, INCASSO FLOAT NOT NULL, NUM_CHIAMATE INT NOT NULL, PRIMARY KEY(ID_TEMPO, ID_TAR, ID_LUOGO_CHIAMANTE, ID_LUOGO_CHIAMATO), FOREIGN KEY(ID_TEMPO) REFERENCES TEMPO(ID_TEMPO), FOREIGN KEY(ID_TAR) REFERENCES TARIFFA(ID_TAR), FOREIGN KEY(ID_LUOGO_CHIAMANTE) REFERENCES LUOGO(ID_LUOGO), FOREIGN KEY(ID_LUOGO_CHIAMATO) REFERENCES LUOGO(ID_LUOGO));	Tabella dei fatti Cardinalità: 7809 tuple

Figura 2 –Possibile soluzione del punto 2 - Tabelle del data warehouse