

Sistemi di gestione di basi di dati

Esercitazione di laboratorio n. 6

Data warehouse in Oracle

La finalità di questa esercitazione consiste nella scrittura di alcune interrogazioni, in linguaggio SQL (nella versione estesa), utilizzate per interrogare frequentemente il data warehouse disponibile in Oracle conforme alle specifiche riportate nel punto 1. Sarà inoltre necessario creare alcune viste materializzate ritenute utili per rispondere velocemente a tali interrogazioni. Il server Oracle a cui collegarsi è installato sulla macchina cclix4.polito.it (fare riferimento al punto 3 dell'esercitazione per i dettagli sulla connessione a Oracle).

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

La base di dati (OLTP) della società telefonica in cui vengono memorizzate le singole telefonate è riportata in Figura 1.

2. Interrogazione del data warehouse

In Figura 2 è riportata una possibile soluzione per il problema proposto. Le tabelle della soluzione proposta sono già state create e contengono al loro interno dei dati d'esempio. Usare tale base di dati come sorgente per i punti successivi. Fare riferimento al punto 3 per ciò che riguarda la connessione ad Oracle.

- Risolvere utilizzando il linguaggio SQL (versione estesa) le seguenti interrogazioni:
 1. Selezionare per ogni coppia tipo tariffa, anno l'incasso totale. Selezionare inoltre l'incasso totale, l'incasso totale per ogni tipo di tariffa indifferentemente dall'anno, e l'incasso totale per ogni anno indifferentemente dal tipo di tariffa.
 2. Selezionare per ogni mese il numero di chiamate totali e l'incasso totale. Utilizzando la funzione RANK(), associare ad ogni mese un numero che identifica la posizione del mese all'interno dei mesi in funzione dell'incasso totale effettuato (1 per il mese che ha incassato di più, 2 per il secondo mese, ecc.)

3. Selezionare per ogni mese dell'anno 2003 il numero di chiamate totali. Utilizzando la funzione RANK(), associare ad ogni mese un numero che identifica la posizione del mese all'interno dei vari mesi dell'anno 2003 in funzione del numero di chiamate totali (1 per il mese con più telefonate, 2 per il secondo mese, ecc.)
 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. Partizionare i dati per anno.
- 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 e verificare se le viste create vengono effettivamente utilizzate dal DBMS analizzando il piano di esecuzione generato per ogni interrogazione.

3. Connessione ad Oracle

Connessione alla macchina cclix4.polito.it

Per potersi connettere al database Oracle è necessario collegarsi alla macchina cclix4.polito.it su cui è installato Oracle 10g. Le interrogazioni possono essere eseguite per mezzo dell'interfaccia web di Oracle 10g.

In particolare:

1. Collegarsi alla seguente pagina web:
 - <http://cclix4.polito.it:8080/apex>
2. Per autenticarsi collegandosi all'interfaccia Web, inserire i seguenti parametri.
 - Nome utente: bdati[scegliere un valore compreso tra 1-100]
 - Password: orac[scegliere un valore compreso tra 1-100]

Ad esempio, collegandosi dalla macchina numero 23 del laboratorio, usare come username **bdati23** e come password **orac23**.

Una volta eseguito il login verrà visualizzata una finestra indicante il seguente menu.



3. Eseguire le interrogazioni e creare le viste materializzate per mezzo dell'interfaccia grafica

Tabelle	Descrizione
TARIFFE (TipologiaTariffa INT NOT NULL, NomeTariffa VARCHAR(20) NOT NULL, CostoAlSecondo FLOAT NOT NULL, PRIMARY KEY(TipologiaTariffa));	Tipologie (categorie) di tariffe esistenti Cardinalità: 7 tuple
LOCALITA (CodLocalita INT NOT NULL, Citta VARCHAR(20) NOT NULL, Provincia VARCHAR(20) NOT NULL, Regione VARCHAR(20) NOT NULL, PRIMARY KEY(CodLocalita));	Località Cardinalità: 1500 tuple
TELEFONATE (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));	Telefonate effettuate negli anni 2003, 2004 Cardinalità: 422416 tuple

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 INT 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, PREZZO FLOAT NOT NULL, 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