**Data Science and Database Technology**

Politecnico di Torino

**Data warehousing in Oracle – Practice 1**

The practice purpose is to first build a data warehouse compliant with the specifications listed in the following points, using Oracle. You then write some queries, in extended SQL, to retrieve data from the design data warehouse.

The outline of the practice is as follows:

1. Problem specifications
2. Description of the OLTP database
3. Exercise: design of the data warehouse
4. Exercise: comparison with the logical schema of the data warehouse
5. Exercise (SQL Developer): query of the data warehouse

## 1. Problem specifications

A telephone company is interested in analyzing its own data to improve customer services. At present, the company has a database with call logs. For each call, the caller and receiver phone numbers, the duration, the type of charge (e.g., peak, off-peak rates), the start time (date, hour, minute, second) are known. The managers want to obtain very fast the information about the telephone traffic on the company lines and the daily income based on the caller location, the day and the phone rate.
In particular, the managers want to analyze the following situations:

- Monthly net income and number of calls for each caller city.
- Monthly net income and number of calls for each receiver city.
- Monthly net income and number of calls for each caller province and region.
- Monthly net income and number of calls for each receiver province and region.
- Daily net income and number of calls for each caller province.
- Yearly net income and number of calls for each caller province and region.
- Monthly net income and number of calls for each phone rate (type of charge).
- Net income and number of calls for each day of the week and phone rate.
- Daily number of calls for each caller region.
- Daily number of calls for each receiver region.

## 2. Description of the OLTP database

The OLTP database of the telephone company is reported in Table 1.

| Tables | Description |
|---|---|
| **DWABD.PHONERATES**<br>(<br>phoneRateType INT NOT NULL,<br>phoneRateName VARCHAR(20) NOT NULL,<br>phoneRate_CostPerSecond FLOAT NOT NULL,<br>PRIMARY KEY(phoneRateType)<br>); | Different phone rates<br><br>7 rows |
| **DWABD.PLACES**<br>(<br>Places_ID INT NOT NULL,<br>City VARCHAR(20) NOT NULL,<br>Province VARCHAR(20) NOT NULL,<br>Region VARCHAR(20) NOT NULL,<br>PRIMARY KEY(Places_ID)<br>); | Places<br><br>1500 rows |
| **DWABD.CALLS**<br>(<br>CallerPhoneNumber VARCHAR(20) NOT NULL,<br>ReceiverPhoneNumber VARCHAR(20) NOT NULL,<br>CallerLocation INT NOT NULL,<br>ReceiverLocation INT NOT NULL,<br>FullDate DATE NOT NULL,<br>StartTimeHour INT NOT NULL,<br>StartTimeMinute INT NOT NULL,<br>StartTimeSecond INT NOT NULL,<br>CallDuration FLOAT NOT NULL,<br>phoneRateType INT NOT NULL,<br>PRIMARY KEY(CallerPhoneNumber,ReceiverPhoneNumber,FullDate,StartTimeHour,StartTimeMinute,StartTimeSecond),<br>FOREIGN KEY(phoneRateType) REFERENCES PhoneRates(phoneRateType) ON DELETE CASCADE,<br>FOREIGN KEY(CallerLocation)REFERENCES Places(Places_ID) ON DELETE CASCADE,<br>FOREIGN KEY(ReceiverLocation) REFERENCES Places(Places_ID) ON DELETE CASCADE<br>); | Calls in 2003 and 2004<br><br>~ 1300000 rows |

**Table 1 – Source data base with single call information**

## 3. Exercise: design of the data warehouse

Design the conceptual scheme of a data warehouse for managing the issues discussed above. The designed scheme must be designed to allow:

- The analyzes requested by the mobile phone company
- The ETL (extraction, transformation, loading) phase to import the data from the OLTP base (Table 1) to the data warehouse.

The design of the conceptual schema can be done on paper or on any software you want.

## 4. Exercise: comparison with the logical schema of the data warehouse

Compare the conceptual scheme designed in the previous exercise with the logical scheme proposed in Table 2. Check you have correctly chosen the measures and the level of data aggregation. Looking at the logic schema in Table 2, answer the following questions:

- What are the measures chosen for the data warehouse?
- What is the minimum level of aggregation in the data warehouse? Does it correspond with what was designed in the conceptual schema?

The design of the logical schema can be done on paper or on any software you want.

| Tables | Description |
|---|---|
| **DWABD.TIMEDIM**<br>(<br>ID_time INT NOT NULL,<br>DayDate DATE NOT NULL,<br>DayOfWeek CHAR(15) NOT NULL,<br>DateMonth CHAR(15) NOT NULL,<br>DateYear INT NOT NULL,<br>PRIMARY KEY(ID_time)<br>); | Time dimension<br><br>10 rows |
| **DWABD.PHONERATE**<br>(<br>ID_phoneRate INTEGER NOT NULL,<br>phoneRateType VARCHAR(20) NOT NULL,<br>PRIMARY KEY(ID_phoneRate)<br>); | Phone rate dimension<br><br>7 rows |
| **DWABD.LOCATION**<br>(<br>ID_location INTEGER NOT NULL,<br>City VARCHAR(20) NOT NULL,<br>Province CHAR(20) NOT NULL,<br>Region CHAR(20) NOT NULL,<br>PRIMARY KEY(ID_location)<br>); | Place dimension<br><br>1500 rows |
| **DWABD.FACTS**<br>(<br>ID_time INTEGER NOT NULL,<br>ID_phoneRate INTEGER NOT NULL,<br>ID_location_Caller INTEGER NOT NULL,<br>ID_location_Receiver INTEGER NOT NULL,<br>Price FLOAT NOT NULL,<br>NumberOfCalls INTEGER NOT NULL,<br>PRIMARY<br>KEY(ID_time,ID_phoneRate,ID_location_Caller,ID_location_Receiver),<br>FOREIGN KEY(ID_time) REFERENCES timeDim(ID_time),<br>FOREIGN KEY(ID_phoneRate) REFERENCES phoneRate(ID_phoneRate),<br>FOREIGN KEY(ID_location_Caller) REFERENCES location(ID_location),<br>FOREIGN KEY(ID_location_Receiver) REFERENCES location(ID_location)<br>); | Fact table<br><br>7809 rows |

**Table 2 – Proposed solution - Data warehouse tables**

## 5. Exercise (SQL): Querying the data warehouse

For the following exercise you have to use Oracle Live or Oracle SQL Developer.
You can find the instructions in the zip file associated to Lab1.

The tables corresponding to the schema in Table 2 have already been created in Oracle and they contain some sample data. Use these tables as source for the following queries.

1. Select the yearly income for each phone rate, the total income for each phone rate, the total yearly income and the total income.
2. Select the monthly number of calls and the monthly income. Associate the RANK() to each month according to its income (1 for the month with the highest income, 2 for the second, etc., the last month is the one with the least income).
3. For each month in 2003, select the total number of calls. Associate the RANK() to each month according to its total number of calls (1 for the month with the highest number of calls, 2 for the second, etc., the last month is the one with the least number of calls).
4. For each day in July 2003, select the total income and the average income over the last 3 days.
5. Select the monthly income and the cumulative monthly income from the beginning of the year.
6. Consider the year 2003. Separately for phone rate and month, analyze (i) the total income, (ii) the percentage of income with respect to the total revenue considering all the phone rates, (iii) the percentage of income with respect to the total revenue considering all the months.
7. For each caller region, select the monthly number of calls and the cumulative monthly number of calls from the beginning of the year.
8. Consider the year 2003. Analyze the total income for (i) separately for each month and (ii) separately for each month, phone rate and caller region and (iii) separately for each month, phone rate and receiver region.