DEPARTAMENTO DE ELETRÓNICA, TELECOMUNICAÇÕES E INFORMÁTICA UNIVERSIDADE DE AVEIRO

EXPLORAÇÃO DE DADOS / DATA MINING

(Part I) Laboratory Assignment 02: Multidimensional data models

In the following, you must use the *Domain Fact Model (DFM)* notation *[1]* to design a conceptual model for the data warehouse/mart described in the following exercises.

Tip: use the *Business Intelligence Modeler* tool (http://www.bimodeler.com/) to draw your diagrams and refer to https://www.youtube.com/watch?v=oMnqZKevGHw for a 6 minutes tutorial.

1. A mobile telecommunications company wants to have regular information about traffic events and recharges made using prepaid cards, namely:

   • The consumed values and the amount billed by traffic event (phone call, sms or mms), event type, service type and card.

   • The origin (cell) of the traffic events and the destination prefix.

   • The date and time of the traffic events.

   • The recharges by card, date and payment type.

   The multidimensional data models designed to represent this information are:
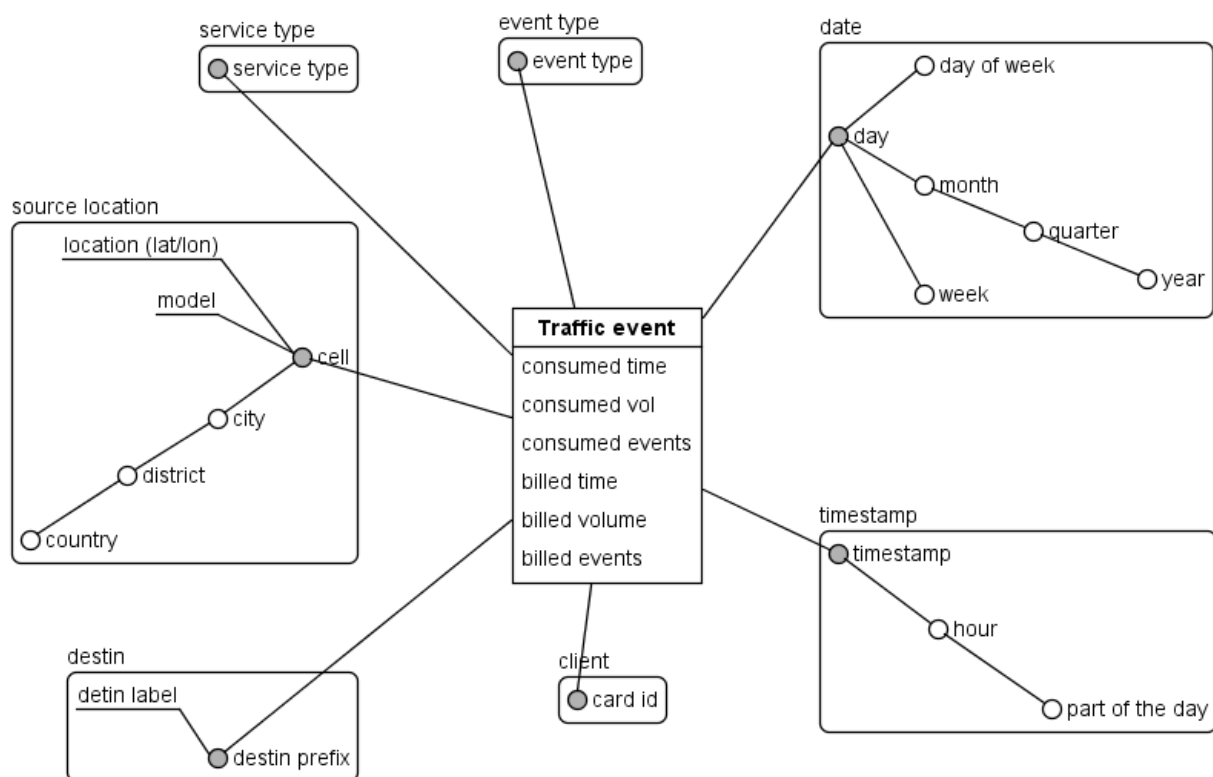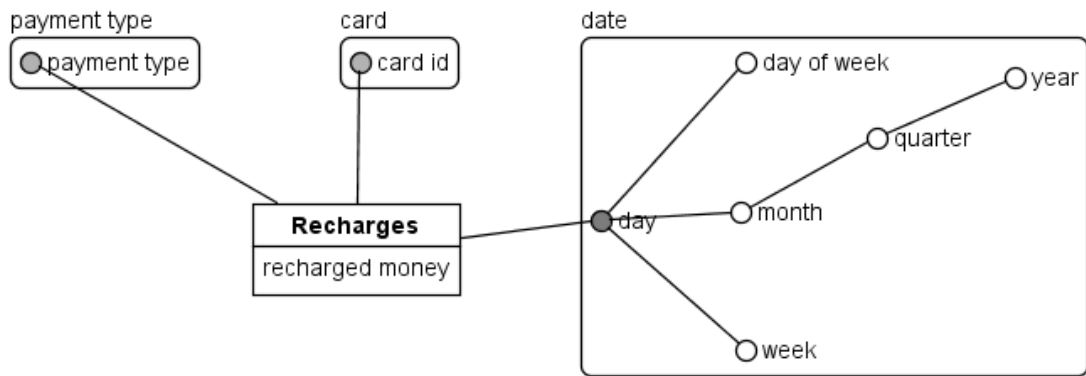
a. Give the names of the fact tables, the dimensions and the measures in each diagram.

b. Give an example of a one-to-one relationship and of a one-to-many relationship in the first diagram.

c. Draw a multidimensional data model to represent the state transitions (card is enabled, locked, expired or disabled) by card and by date.

d. Generate the scripts to create the data mart for this multidimensional data model.

    − Run the BI Modeler application and open the document *telecommunications.bim*.

    − Change the view mode to Relational Model and then execute the command Relational Model → Forward engineering model. Have a look to the Glossary Properties.

    − At this point you should drag all tables in the relational model to the design pane and see the database schema generated by BI Modeler. You should define the data types and the size of the attributes in each table according to the specifications of your target DBM.

    − Execute the command Relational Model → New Physical Model to define the properties to be used in the creation of the data mart. If you are using Postgres, it is recommended to choose Oracle 10g/11g in the DBMS combo box. Give a name to your Physical Model and then edit the Default build properties as follows (SQL Server in the left or Postgres/Oracle in the right):
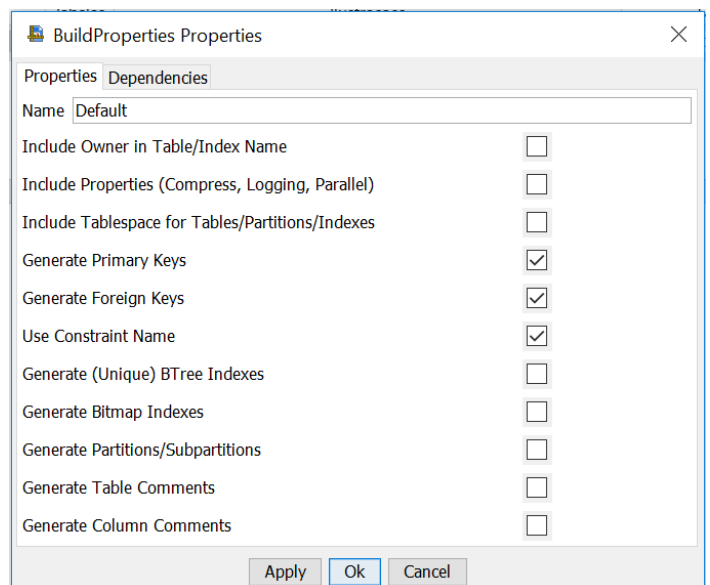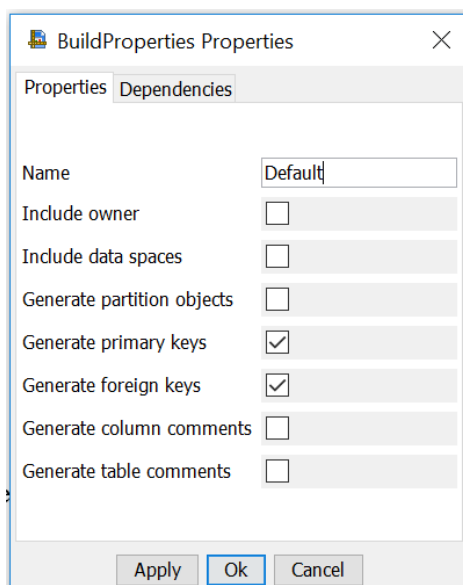

    − Execute the command Generate DDL, add all tables and generate the script to create the data mart in your target DBMS.

        Notes: There is only a subject area. You will learn about partitions later in this course. You must consider that you are creating the data mart for the first time, hence you are free to recreate the tables

or clear their contents whenever you need. This would not be the case if you were migrating data into a data mart already in production.

2. Draw a multidimensional data model for the following database schema.
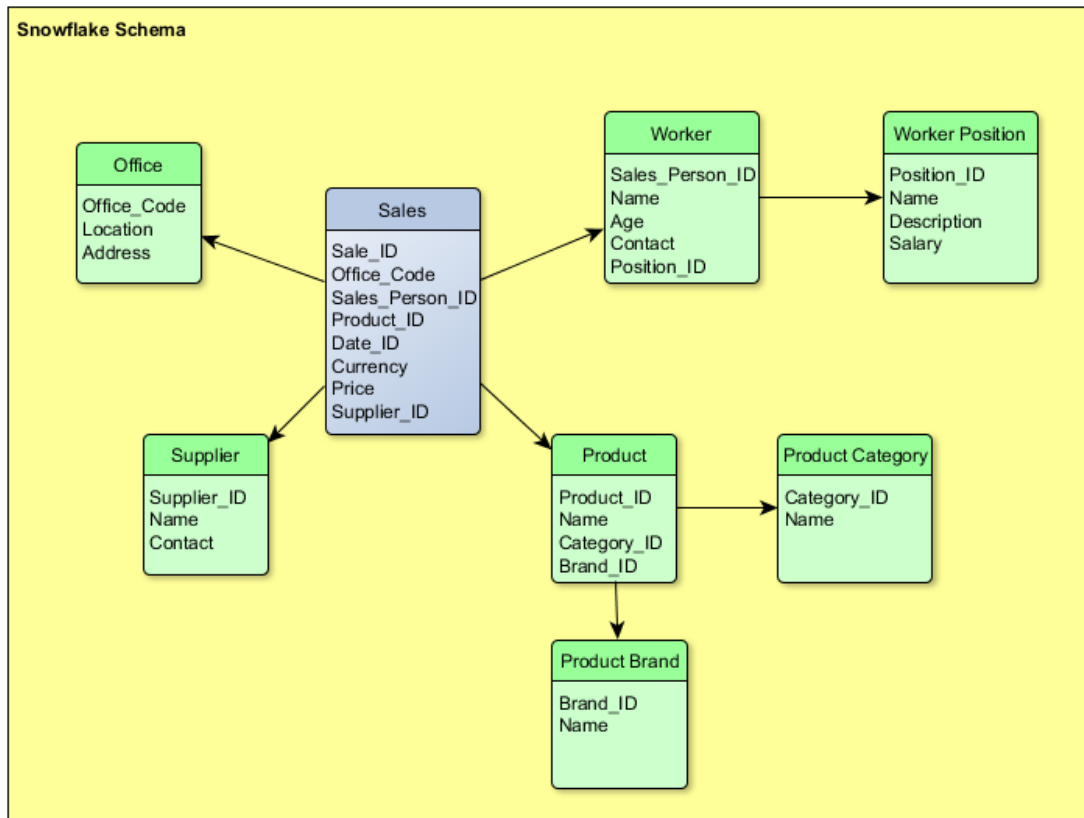
3. A supermarket chain wants to implement a data mart to analyze the sales (quantity sold and amount of revenues) by product, supermarket, promotion and time:

- The products are identified by code and the attributes are name, category, subcategory, brand, weight and supplier;

- The supermarkets are identified by code and the attributes are name, city, region, size and layout of the supermarket (e.g., one floor, two floors, and so on).

- The promotions are identified by code and the attributes are name, type and discount percentage.

- The date and related attributes representing the day of sale within the week (Sunday to Saturday), the day of the month (1 to 31) and the year (1 to 366); the week within the year (1 to 52); the month within the year; the season; the preholiday flag, which indicates whether the sale happens on a day before a holyday; and the holyday flag, which indicates whether the sale happens during a holyday.

Next, consider that the supermarket chain also wants to keep a daily inventory level to analyze the number of products in stock by time, store, category, subcategory, brand and supplier.

Design the multidimensional data models. Also classify the facts as event or snapshot, and the measures as additive, semi-additive, or non-additive. Explain briefly your choices.

4. A company that administrates several medical centers maintains an operational system for registration of medical appointments at each center. To improve their business, they want to implement a decision support system and they have identified the following needs:

- To understand billing by medical specialty or by location (city), considering the gender and the age group of the patients.
- To analyze the data by month, quarter and year.
- To understand the evolution of drug prescriptions (quantity) by clinic, physician, gender and age group.

Design a multidimensional data model for this system. Also classify facts as event or snapshot. Explain briefly your choices.

Would it be interesting for data analytics purposes to replace the age group (infant, children, teenager, young, adult or senior) by age (0-120)? Explain briefly your answer.

References

[1] Stefano Rizzi. *Conceptual modeling solutions for the data warehouse. In Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications, J. Wang (Ed.), Information Science Reference, pp. 208-227, 2008.*

You must save the solutions of exercises 1 and 4 because you will need to submit them on November 4 (date to be confirmed).

Tabelas Facto:
- Consulta
- Prescrição Medicamento