Data Warehouse Models - Exercise Sheet

LSDM

1 Furniture Sales

A furniture sales company wants to create a data warehouse. Consider the following information:

- The data warehouse should allow analysis with respect to the furniture sold, customers, and time.
- A piece of furniture is associated with a type (chair, table, display cabinet, etc.), a category (kitchen, living room, bathroom, etc.), and a material (wood, marble, etc.).
- A customer should be associated with a location (city, region, and country).
- The company is interested in the number of sales, revenue, and discounts on completed sales.

Question 1. Develop a *multidimensional model* for a data warehouse corresponding to the above requirements. Indicate *the facts, dimensional hierarchies, and measures.*

Question 2. Develop a *star schema* for a data warehouse corresponding to the above requirements.

Question 3. Develop a *snowflake schema* for a data warehouse corresponding to the above requirements.

2 Railway Company

A railway company's data warehouse contains information about train segments. It includes *six dimensions*: departure station, arrival station, route, train, arrival time, and departure time, and *three measures*: number of passengers, duration, and number of kilometers.

Question. Propose a snowflake schema for this data warehouse.

3 Maritime Transport Company

A maritime transport company wants to create a data warehouse:

- The warehouse should allow analysis of containers transported on each route, according to customers, ports, time, and container characteristics.

- Each route is associated with a *departure port* and an *arrival port*, for which the *city and country* are known (one port maximum per city). The departure and arrival dates are also known (day, month, and year).
- A container is described by a *type* (electronics, chemicals, foodstuffs, or textiles), and a *category* (dangerous, fragile, risk-free): electronics are fragile, chemicals are dangerous, and the rest are risk-free.

  The name of the customer requesting its transport is also recorded. The container type determines the *format* of the container (given by a string).
- The company is interested in the *number of containers* and the *revenue*.
- The customer is identified by their *name*. No other information is available about them.
- **The warehouse does not record any information uniquely identifying each container. Therefore, it is not possible to track the journey of a container across multiple steps (routes).**

Question 1. Propose a multidimensional model for this warehouse; specify the hierarchy schemas, give an example instance for the container hierarchy, and the facts and measures analyzed.

Question 2. How many granularities can be considered on the cube of this warehouse? *(Hint: pay attention to certain dimension(s), for example: there is a departure port and an arrival port).*

Question 3. Propose a star schema for this warehouse.

Question 4. Explain which elements of your schema you would change if the warehouse also had to model the capacity of ports to handle each type of container. The goal is to record:

- how many containers of each type can be stored in each port
- at what rate (number of units per hour) they can be unloaded (this rate depends on the container type and the port).

Is your schema still a star schema? If yes, why? If not, what is this type of schema called?

4 University

Consider the data warehouse of a university that contains information about teaching and research activities.

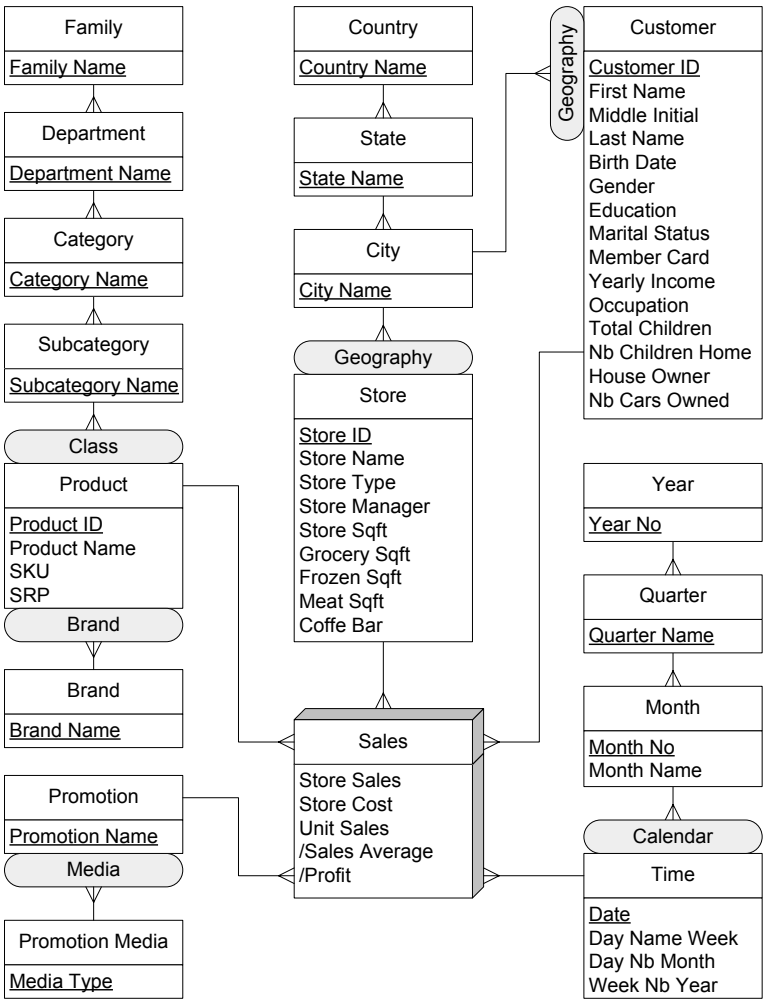
Information about *teaching activities* is linked to the dimensions department, professor, course, and time, the latter at the granularity of the academic semester. The measures for teaching activities are the number of hours and the number of credits.

Information about *research activities* is linked to the following dimensions: professor, funding agency, project, and time (start and end date, granularity of a day). Professors are linked to the department to which they are affiliated. The measures for research activities are the number of person-months and the amount.

Question. Develop a *galaxy* (or constellation) schema for the data warehouse, taking into account the different granularities of the time dimension.

5 OLAP Queries

Consider the following *Foodmart* diagram:



Write the OLAP operations needed to answer the following queries:

a. All measures in the *Store* hierarchy.

b. All measures in the *Store* hierarchy for stores in the cities of `California` and *Washington*, aggregated at the state level.

c. Average sales in 2012 by store state and by store type.

d. Profits by store in Q1 of each year.