

Data warehouse design exercise

Problem specifications

Design a data warehouse to manage the storehouses of an Italian household appliance company, addressing the following issues. The company has storehouses all over Italy. The management needs to analyze the current storehouse usage to decide which ones to expand, reduce, or partially rent. The management periodically takes out theft insurance against robbery in its storehouses. The cost of the insurance depends on the average daily amount of products stored in the storehouses and on their value. Different insurances exist and the management needs to decide which one is the best. Some insurance policies are specific to a particular storehouse, while others are common to different storehouses of the same city, province or region. The insurance costs depend on the following daily measures:

- Number of products stored in each storehouse for each type of product
- Value of the products stored in each storehouse for each type of product

Regarding the usage of the storehouses (i.e., surface taken by goods), the management needs to analyze the daily percentage of free surface with respect to the total available surface of each storehouse, according to the storehouse location (city, province, and region). The total available storehouse surface may change, since some areas of the storehouse may be rented or be temporary unavailable. The analysis will be performed for each day, month, trimester, 4-month period, semester, and year.

To assess the insurance costs, the management needs to analyze the average daily number of products stored in each storehouse and the daily total value of the products for each date, month, trimester, 4-month period, semester, and year. This information must be available for each product type, product category, storehouse, and storehouse location (city, province, and region).

The following are **some** of the frequent queries the management is interested in:

- a) In the first trimester of 2003, regarding the storehouses in Turin, select the total value of the products stored in each storehouse at any given date, and select the **average daily total value of the products** in each storehouse during the previous week (including the current date).
- b) In 2004, for each city and date, select the percentage of daily free surface of the storehouses. Give a rank to the results (rank 1 is the lowest percentage).
- c) In the first 6 months of 2004, select the percentage of free surface for each storehouse and date.
- d) In 2003, select the average daily total value of products for each storehouse and month.
- e) In 2003, select the average daily total value of products for each region.
- f) In 2004, select the average percentage of daily free surface for each month and region.

Design

The data warehouse will store information of 2003 and 2004. The following cardinalities are known:

- Product types: ~100
 - Product categories: ~10
 - Storehouses: ~100
 - Storehouses in Turin: ~5
 - Cities: ~90
 - Regions: ~40
 - Provinces: ~10
1. Design the data warehouse to address the described issues. In particular, the designed data warehouse must allow efficient execution of **all** the queries described in the specifications.
 2. Write **all** the frequent queries of the “problem specifications” using the extended SQL language.
 3. Considering the designed data warehouse and its cardinalities, decide whether and which materialized views and/or indexes are convenient to improve response time of the frequent queries (consider **all** the frequent queries). Explain reasons for your choices.