



Data Science and Database Technology

Politecnico di Torino

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, 3-month period, 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:

- In the first trimester of 2013, 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).
- In 2014, 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).
- In the first 6 months of 2014, select the percentage of free surface for each storehouse and date.
- In 2013, select the average daily total value of products for each storehouse and month.
- In 2015, select the average daily total value of products for each region.
- In 2014, select the average percentage of daily free surface for each month and region.

CAN BE SOLVED 12 WAYS
WATCH BELOW

Design

1. Design the data warehouse needed to manage the needs of the appliance company to meet the requirements described in the problem specification. The designed data warehouse must also be able to efficiently respond to **all** of the frequent queries proposed in the problem specification.

The data warehouse must contain information for the last 10 years. To properly implement the data warehouse, the following information has been provided:

- Product types: ~100
- Product categories: ~10
- Storehouses: ~500
- Storehouses in Turin: ~5
- Cities: ~200
- Regions: ~20
- Provinces: ~90

2. Write **all** the frequent queries of the “problem specifications” using the extended SQL language.
3. Define the materialized view that could be used to efficiently answer all the following queries:

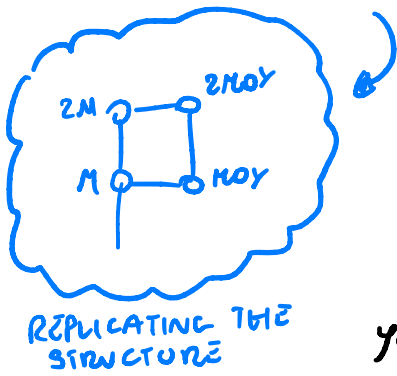
- In the first half of the years 2015 and 2016, considering only the warehouses in the Piedmont region, select the total value of products and the total number of products for each storehouse, province and month.
- In the year 2013, select the average monthly value of products for each province of the storehouse and semester.
- In the year 2014, select the total value of products for each region and 3-month period.

```
SELECT province, month,
       SUM(tot_value),
       SUM(tot_number)
FROM Products P, Storehouse S, Time T
WHERE P.Tid=T.Tid AND P.SID=S.SID
AND (year=2015 or year=2016) AND region='Piedmont'
GROUP BY province, month;
```

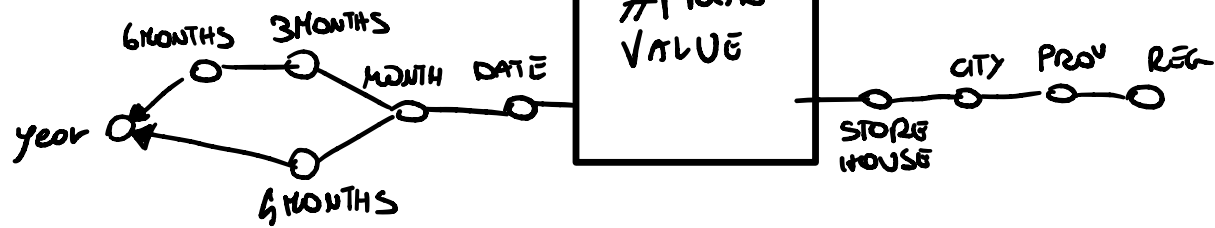
```
SELECT province, 6M,
       SUM(tot_value)/count(distinct month)
FROM Products P, Storehouse S, Time T
WHERE P.Tid=T.Tid AND P.SID=S.SID
AND year=2013
GROUP BY province, 6M;
```

```
SELECT region, 3M
       SUM(tot_value)
FROM Products P, Storehouse S, Time T
WHERE P.Tid=T.Tid AND P.SID=S.SID
AND year=2014
GROUP BY region, 3M;
```

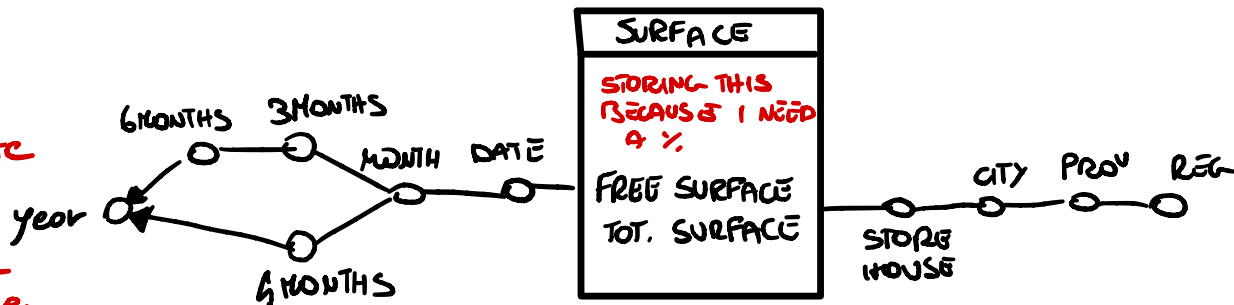
IN CASE WE HAVE
TO LOOSE THE YEAR INFO



if a 2 months period
was present it would
be connected to both
6 MONTHS and 6 MONTHS



THIS new fact
table is separate
from the one
before because
the product type
would be insignificant for the surface values



SOL

DIMENSIONS

TIME(-TID-, Date, Month, 3M, 6M, Year)

STOREHOUSE (-SID-, Storehouse, City, Province, Region)

TYPES (-TPID-, type, category)

FACTS

PRODUCTS (-TID-, -SID-, -TPID-, tot-number, tot-value)

SURFACE (-TID-, -SID-, free-surface, tot-surface)

SOLVING QUERY ① ①

```
SELECT DISTINCT Storehouse, Month  
      AVG(SUM(tot_value)) OVER (PARTITION BY SID, Month)  
FROM products P, Storehouse S, Time T  
WHERE //join conditions AND year = '2013'  
GROUP BY Storehouse, Month, SID, Date
```

SOLVING QUERY ① ②

NOT SO COMPUTATIONAL HEAVY
LIKE THE SOLUTION
BEFORE

```
SELECT Storehouse, Month  
      SUM(tot_value)/COUNT(DISTINCT Date)  
FROM products P, Storehouse S, Time T  
WHERE //join conditions AND year = '2013'  
GROUP BY Storehouse, Month, SID
```

MAT VIEW, QUESTION ③

```
SELECT province, region, month, 3M, 6M, year  
      SUM(tot_value),  
      SUM(tot_number)  
FROM Products P, Storehouse S, Time T  
WHERE P.Tid=T.Tid AND P.SID=S.SID  
GROUP BY province, region, month, 3M, 6M, year
```

$$100 \cdot 300 \cdot 365 \cdot 10 = 18250 \text{ K}$$

STANDARD

$$90 \cdot 12 \cdot 10 = 10800$$

MAT. VIEW