

# Master Degree Programme in Computer Science

## Enterprise Information Systems

### 11. Data Warehousing – Introduction to the Multidimensional Model



SAPIENZA  
UNIVERSITÀ DI ROMA

Prof. Ing. Claudio CILLI

cilli@di.uniroma1.it

<http://dsi.uniroma1.it/~cilli>

# Multidimensional model

**event** in the real world

**DB design:**

selection of all and only features  
having operational relevance

**datum** in the operational DB  
(representing the event)

**DW design:**

..., definition of aggregations (with a given granularity)  
of analytic interest to some users

**fact** in the enterprise DW  
(**measurement** in an **n-**  
**dimensional space**)

John, aged 28, buys  
a box of dish soap

one box of "DS"  
dish soap is sold

sold **N** units of article **X**  
at point od sale **Y**  
on day **Z**

# Events and Facts

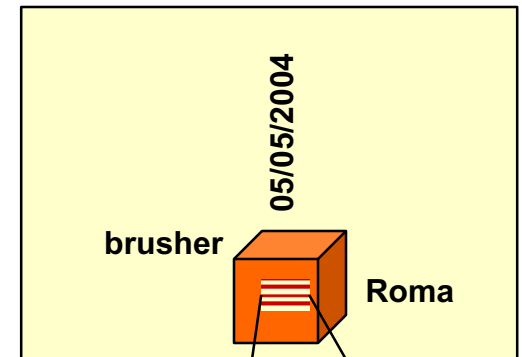
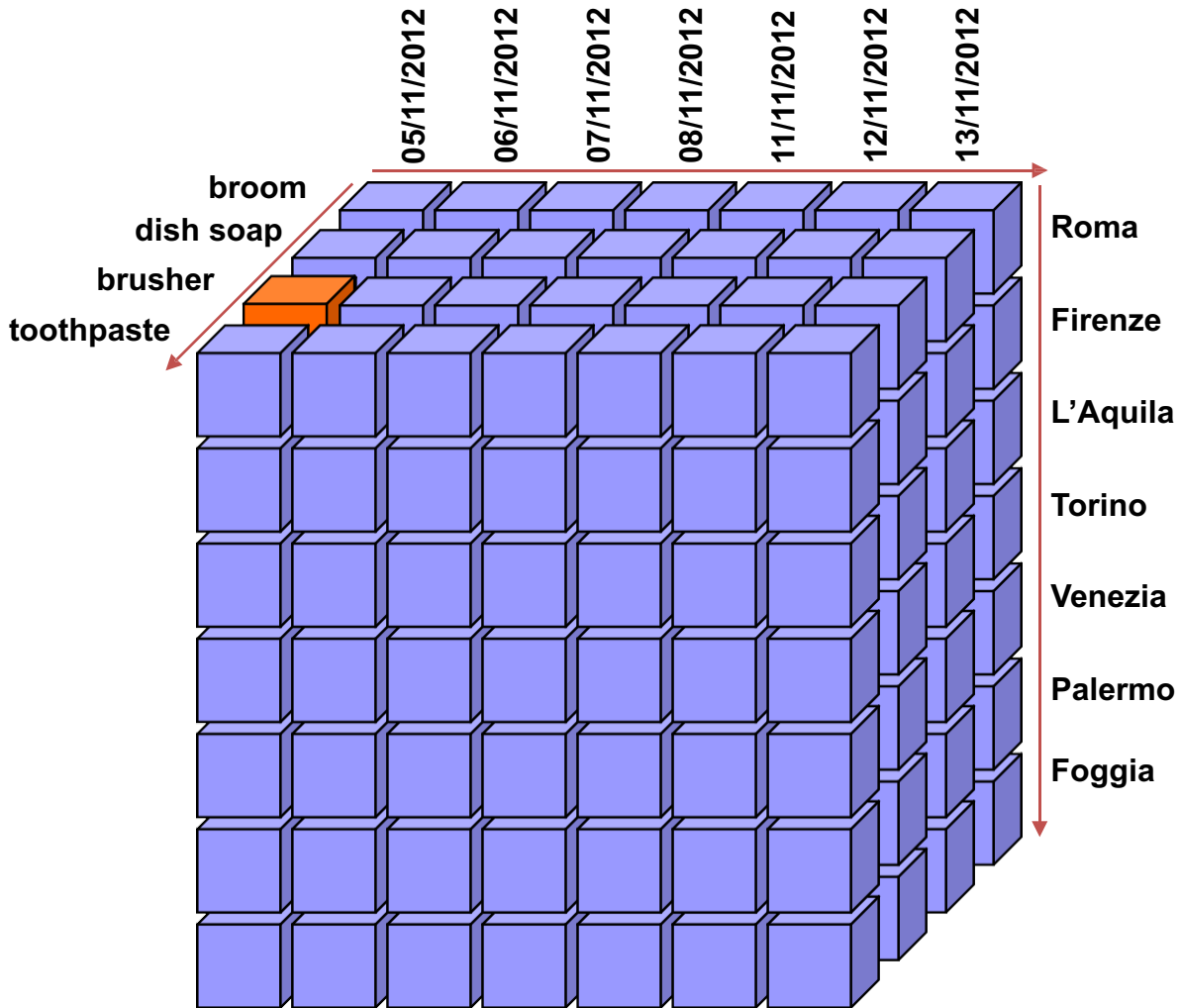
- **EVENT** (in the real world) and **FACT** (of interest) are terms whose meaning is determined by the granularity (level of detail) of the multidimensional representation of the Data Warehouse

## *Example of FACT:*

on **02/05/2012** in **“Rome 23”** PoS (Point of Sale) **278**  
boxes of **soap** have been sold, cashing **745** Euro

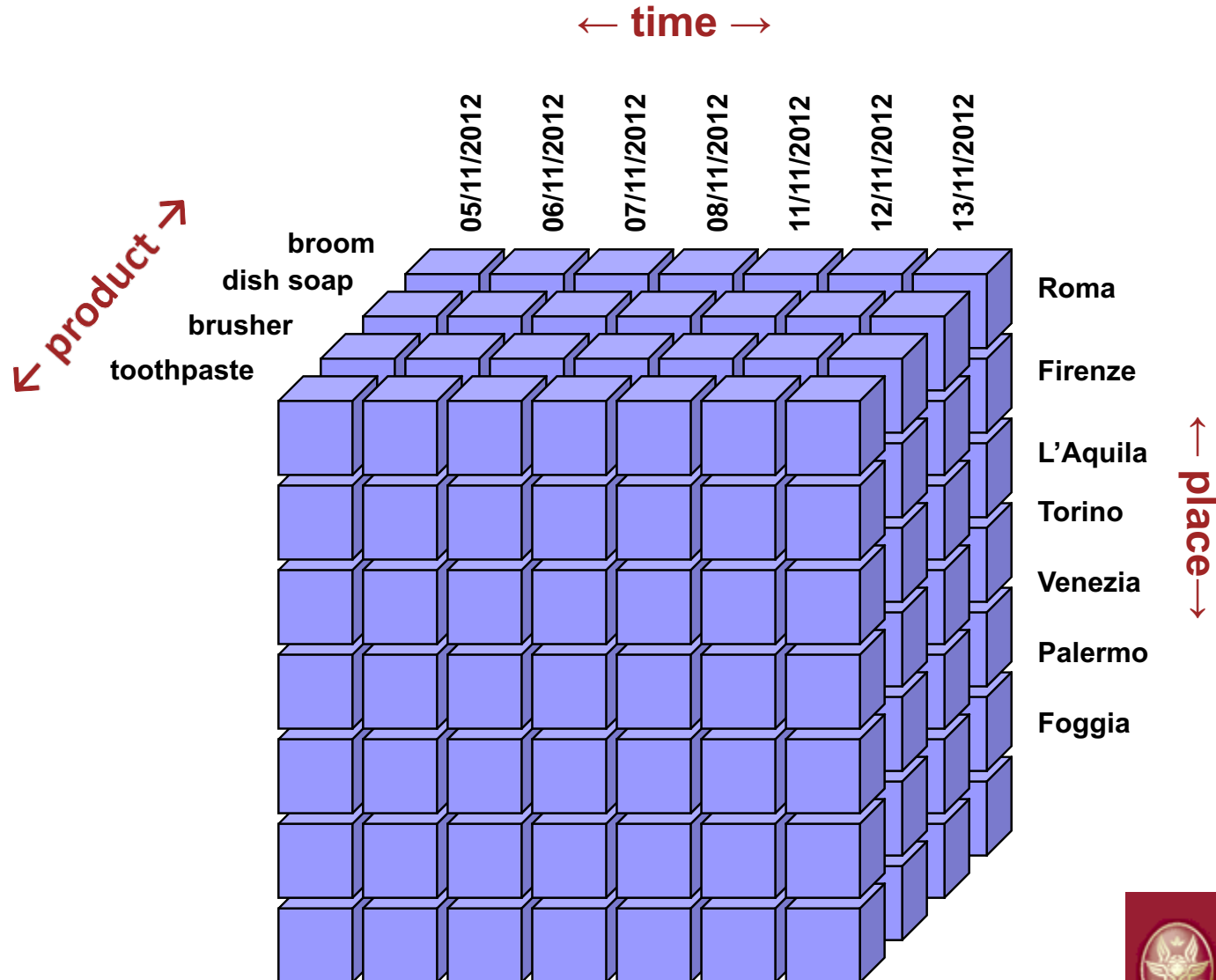


# Multidimensional Cube



units: 278  
sales: € 745

# Dimensions



# A relational schema for the multidimensional cube

TABLE(dimens-1, ..., dimens-k, measur-1, ..., measur-h)

Example:

SALES(product, city, date, numUnitsSold, sales)

**dimensions**

(classification attributes)

**measurements**

(variables, metrics, indicators, ...)



**IDENTIFICATIVE KEY**

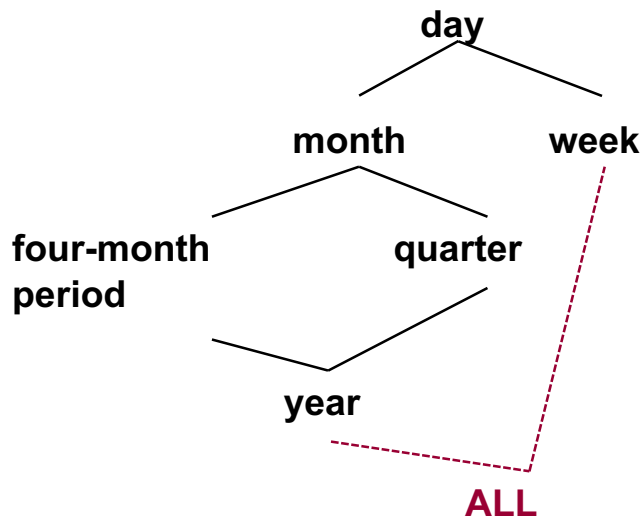
Functional dependence: **dimensions** → **measurements**

In the example: **product, city, date** → **numUnitsSold, sales**

# Dimensional hierarchy

- Each dimension is associated with a hierarchy that groups the values at different levels of aggregation (also orthogonally)
- Each node of the hierarchy is called DIMENSIONAL ATTRIBUTE

Example (time dimension):



Functional

Dependencies:

day → week

day → month

month → quarter

month → four-month-period

quarter → year

four-month-period → year



# Access to Data Warehouse

## Reporting

- predefined needs (poss., parametrized)
- automated extraction of information

## OLAP

- analysis needs that are not identifiable in advance (browsing on dimensions and indicators)
- interactive data exploration, searching for information of interest

## Data Mining

- analysis needs that are not identifiable in advance
- (semi)automatic data exploration





# Reporting

## Report with predefined structure and format

- query
  - typically based on restrictions and / or aggregations
- presentation
  - interleaved with text data
  - in tabular and / or graphic form
  - customized with environment variables (date, user, ...)
- generation
  - on user demand
  - periodic
  - under specified conditions
- distribution
  - preview/print, e-mail, web, ...

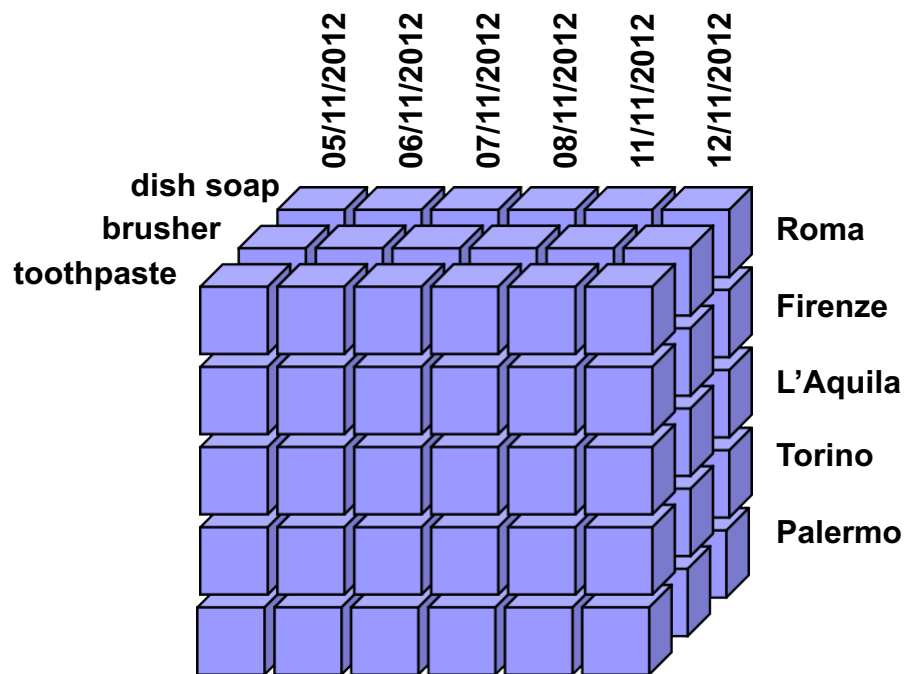
# OLAP

---

- implementation:
  - analysis session, divided into a series of steps, each dependent on the results obtained previously
- typical users:
  - domain experts, not necessarily computer experts



# OLAP: example



# OLAP – Basic operations in the multidimensional model

## **RESTRICTION** (filtering based on values)

- selection on specific values of dimensional attributes
- in case of equality with a single value, the operator is called SLICING (one dimension disappears)

## **AGGREGATION** (level of abstraction increases)

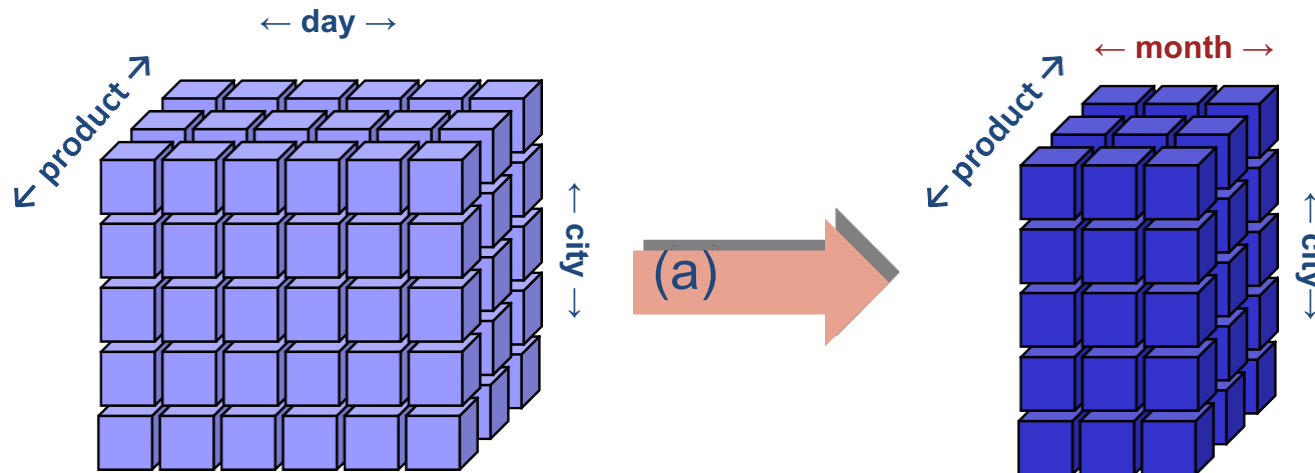
- considered one dimension, group one dimensional attribute to a higher level of granularity
- joining up to the highest level (the most generic, consisting of a single value), a dimension disappears

# OLAP: Roll-Up Operator

- Aggregation along one dimension
  - use of an aggregate operator on a measurement attribute

Examples:

- sales of each product by city and month
- daily sales per city (total on all products)
- daily sales by product (North / Central / South)



# OLAP: Aggregative Operators

**DISTRIBUTIVE:** computable from:

1. partial aggregates
  - Examples: SUM, MAX, MIN, ...

**ALGEBRAIC:** computable from:

1. partial aggregates
2. finite set of support information
  - Examples : weighted AVG, standard deviation, ...

**OLISTIC:** computable from:

1. original data
  - Example: rank of an array



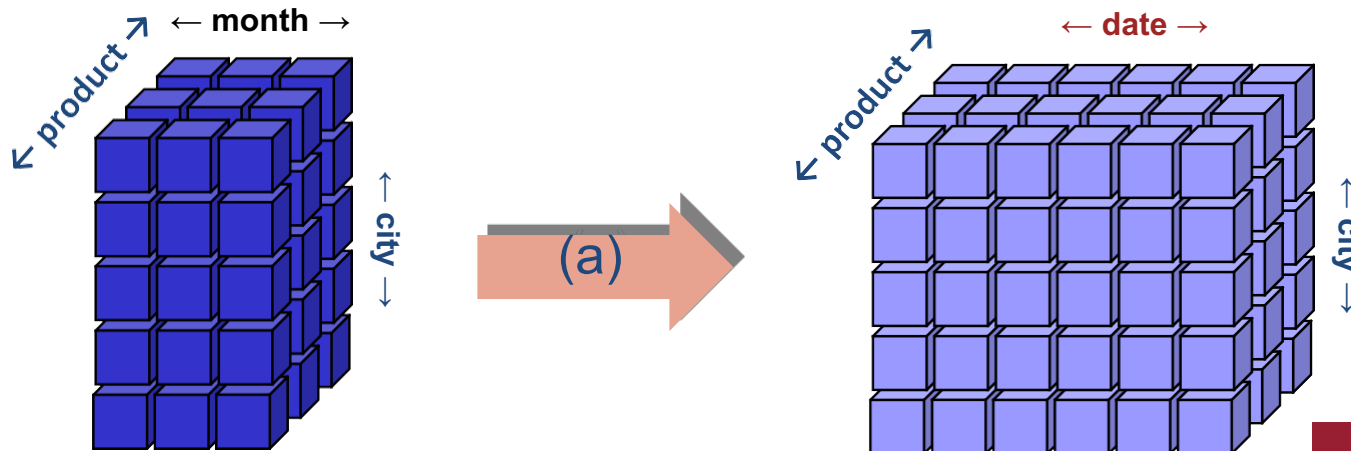
# OLAP: Drill-Down Operator

Disaggregation over one dimension

→ disaggregated data are assumed to be available

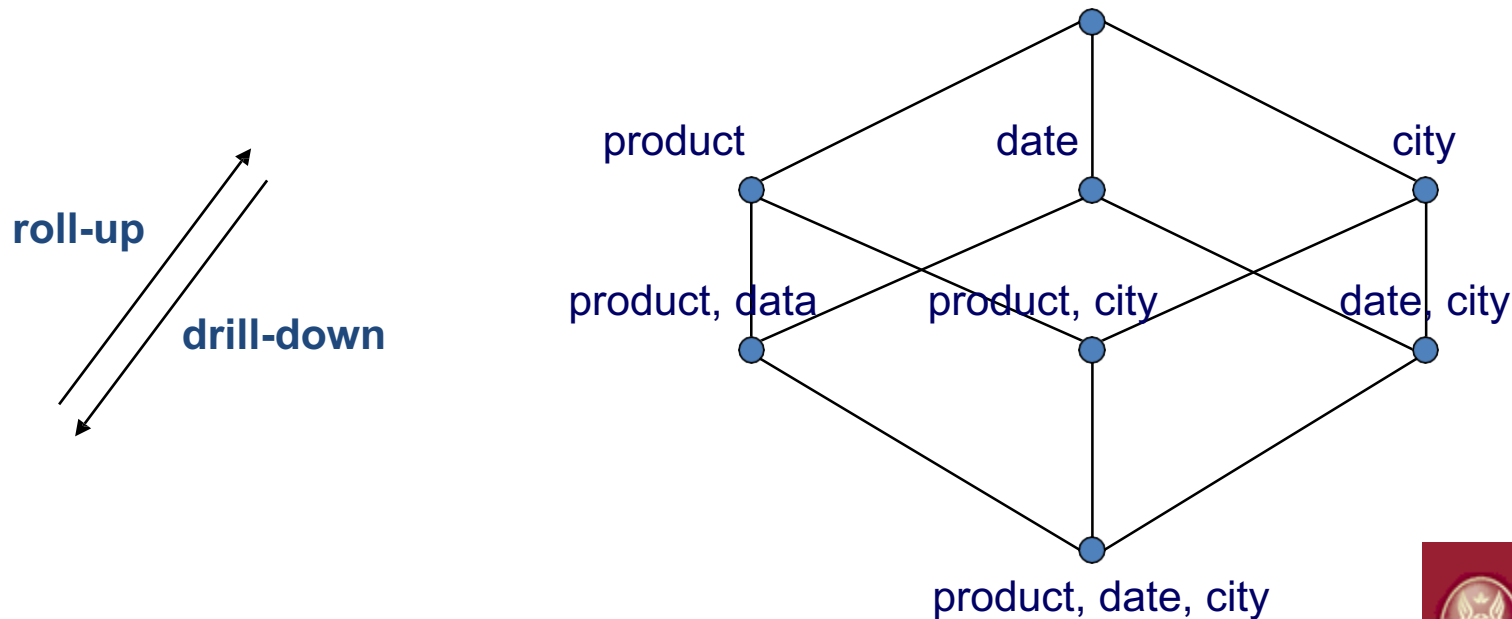
Examples:

- from “sales of each product by city and month” to “daily sales”
- from “monthly sales by city (total on all products)” to “daily sales by city and product”



# OLAP: Cuboids (dimensions)

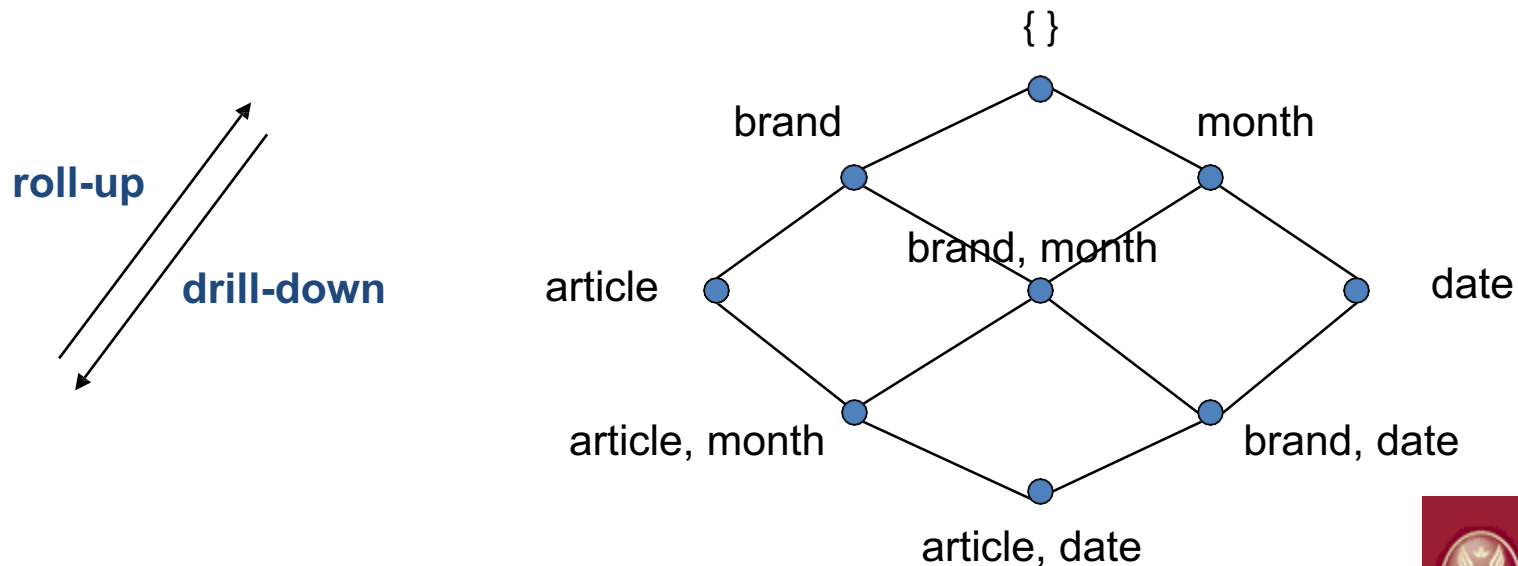
- given a data cube, you can perform different roll-up operations up to the disappearance of a dimension
- given a k-dimensional data cube, there exist  $2^k$  possible cuboids





# OLAP: Cuboids (aggregation levels)

- given a data cube, you can perform different roll-up operations, increasing the aggregation level (up to the disappearance of a dimension)
- given a k-dimensional data cube, resp. with  $n_1, n_2, \dots, n_k$  aggregation levels, there exist  $n_1 \times n_2 \times \dots \times n_k$  possible cuboids



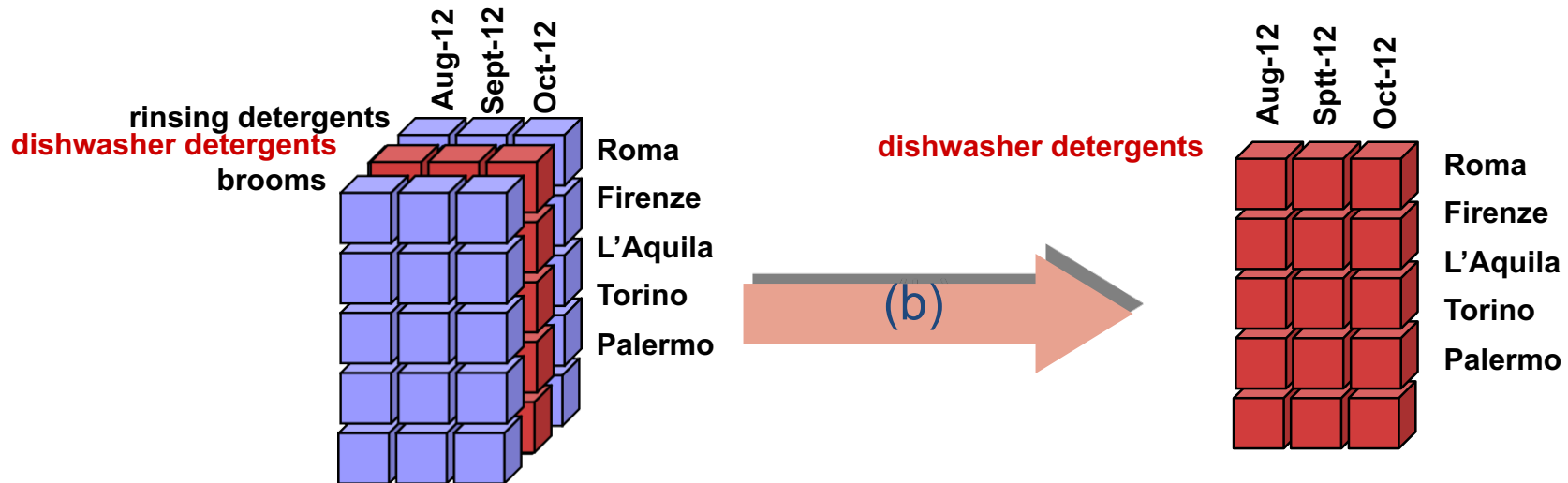
# OLAP: Slice-and-Dice Operators

## Selection (up to “slicing”)

Examples:

from “monthly sales of products by city” to:

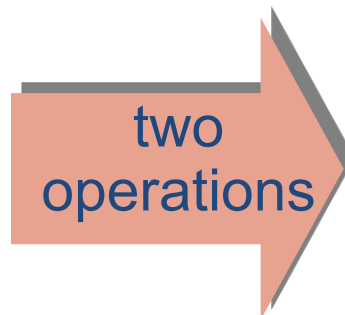
- monthly sales of products by city in central Italy
- monthly sales by city of dishwasher detergents



# OLAP: Pivoting Operators

- Rotating dimensions on a bidimensional table
  - Example:
    - from:** sales of (products by month) **by** (city)
    - to:** sales of (products by city) **by** (month)

		RM	FI	AQ	TO	PA
rinsing det.	aug-12	34	23	12	56	65
	sep-12	56	45	23	44	67
	oct-12	76	34	34	55	45
dishw. det.	aug-12	57	46	35	79	88
	sep-12	79	68	46	67	90
	oct-12	99	57	57	78	68
brooms	aug-12	46	35	24	68	77
	sep-12	68	57	35	56	79
	oct-12	88	46	46	67	57



		aug-12	sep-12	oct-12
rinsing det.	RM	34	56	76
	FI	23	45	34
	AQ	12	23	34
	TO	56	44	55
	PA	65	67	45
dishw. det.	RM	57	79	99
	FI	46	68	57
	AQ	35	46	57
	TO	79	67	78
	PA	88	90	68
brooms	RM	46	68	88
	FI	35	57	46
	AQ	24	35	46
	TO	68	56	67
	PA	77	79	57

# OLAP: Drill-Through Operator

---

- access data at a level of aggregation lower than the Data Warehouse
- access to reconciled data (three-level architecture) or to the operational DB
- a sort of extreme Drill-Down

# OLAP: Drill-Across Operators

- Correlate data from two or more cubes, based on the dimensional values, returning the value in a new cube

Example:

Starting from

- sales by product, date, city
- employees by city

calculate:

- average sales by product, employee

# Books on Datawarehouses

---

- Matteo Golfarelli, Stefano Rizzi «Data Warehouse Design – Modern Principles and Methodologies», McGraw-Hill, 2009
- Matteo Golfarelli, Stefano Rizzi «Data Warehouse – Teoria e Pratica della Progettazione» (2a ed.), McGraw-Hill, 2006

