OLAP Operations

Helena Galhardas
DEI/IST

References

- A. Vaisman and E. Zimányi, Data Warehouse Systems: Design and Implementation, Springer, 2014 (chpt 3)
- J. Han and M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann, 2001 (chpt. 2)
- C. Ciferri, R. Ciferri, L.I. Gómez, M. Schneider, A.A. Vaisman, E. Zimányi, Cube algebra: a generic user-centric model and query language for OLAP cubes. Int. J. Data Warehousing Mining 9(2), 39–65, 2013
- A. Wichert, H. Galhardas, SAD slides, MEIC/IST

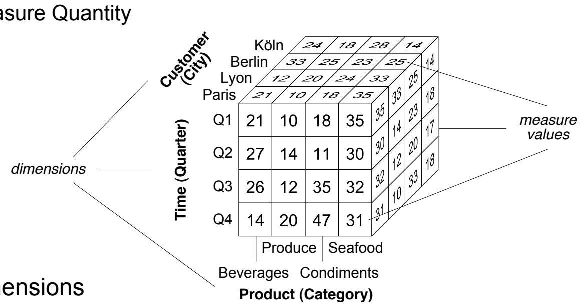
Recap. the multidimensional model

- Multidimensional model enables one to view data from multiple perspectives and at several levels of detail

Example

Dimensions: perspectives used to analyze the data

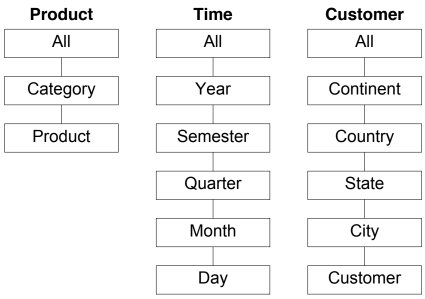
- **Example**: A 3-dimensional cube for sales data with dimensions Product, Time, and Customer, and a measure Quantity



- Attributes describe dimensions
  - Product dimension may have attributes ProductNumber and UnitPrice (not shown)
- Cells or facts have associated numeric values called measures
  - Each cell of the data cube represents Quantity of units sold by category, quarter, and customer's city

Hierarchies

- Allow viewing data at several granularities
  - Define a sequence of mappings relating **lower-level**, detailed concepts to **higher-level** ones
  - The lower **level** is called the child and the higher level is called the parent
  - The hierarchical structure of a dimension is called the dimension schema
  - A dimension instance comprises all members at all levels in a dimension
- Example
  - Hierarchies of the Product, Time, and Customer dimensions



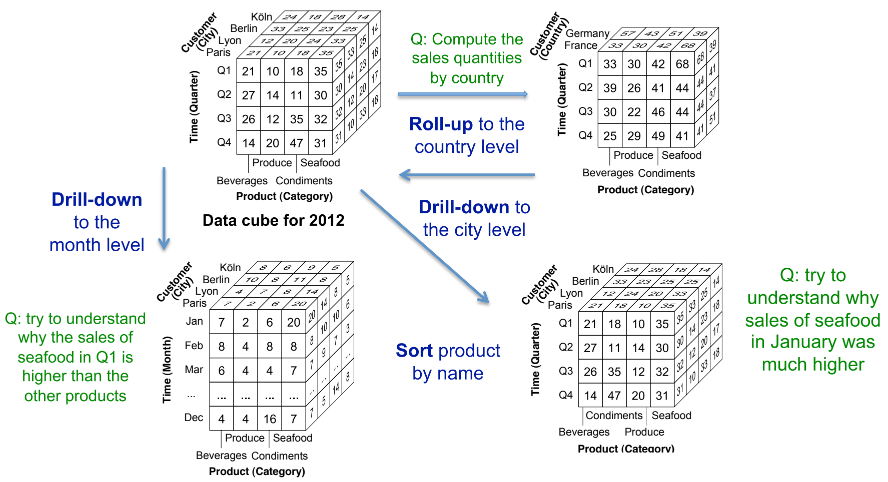
Outline

- OLAP operations

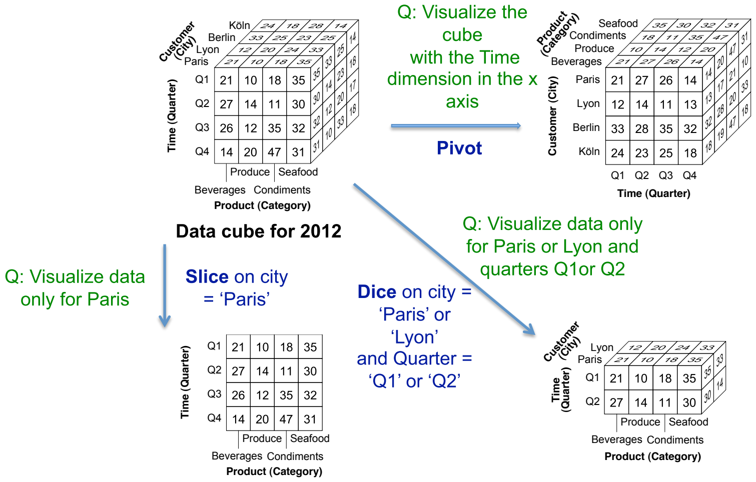
OLAP Operations: definition

- Allows these perspectives and several levels of detail to be materialized by exploiting dimensions and their hierarchies
- Provide an interactive data analysis environment
- Supported by OLAP modules
  - Ex: Saiku (Pentaho)

OLAP Operations (1)

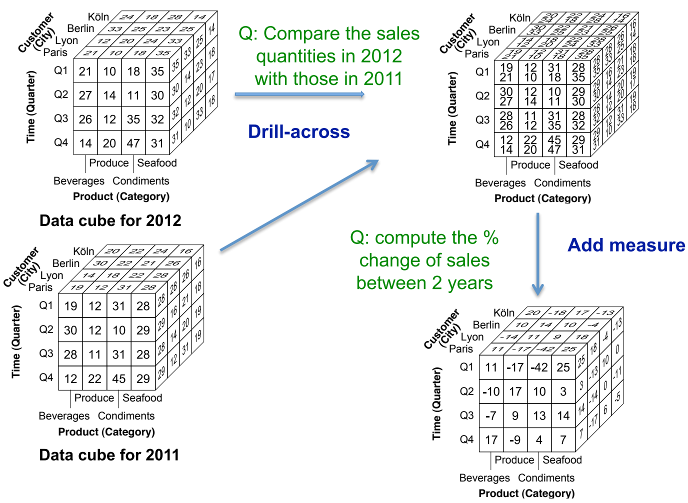


OLAP Operations (2)



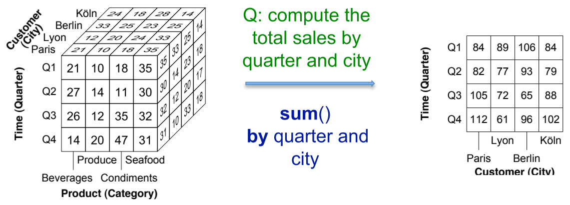
9

OLAP Operations (3)

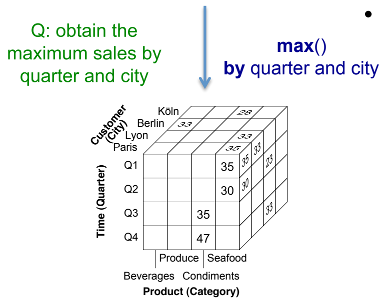


10

OLAP Operations (4)



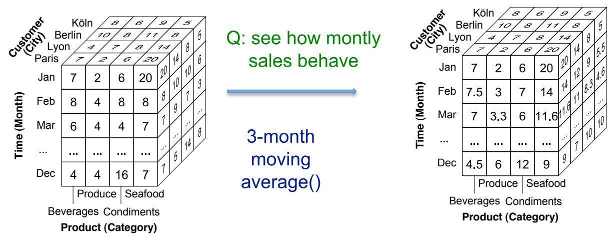
Data cube for 2012

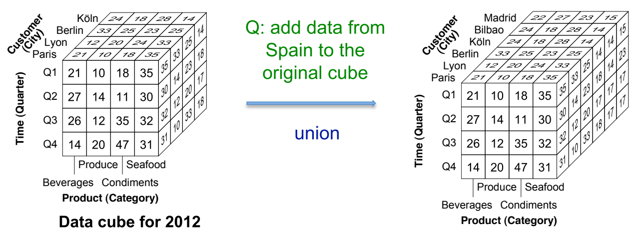


- According to the authors of the book, agg. functions can be classified as:
  - cumulative: compute the measure value of a cell from several other cells(e.g., SUM; COUNT, AVG)
  - filtering: filter the members of a dimension that appears in the result (MIN, MAX); must compute not only the aggregated value but also detemine the dimension members that belong to the result

11

OLAP Operations (5)





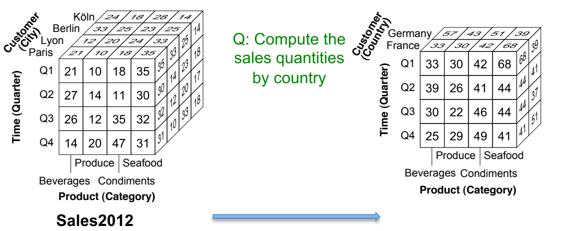
Data cube for 2012

12

Algebra of OLAP Operations

- There **is not yet a standard definition** of OLAP operations in a similar way to the relational algebra
- Many proposals of OLAP algebra in the literature
- We adopt the one proposed in **[Ciferri et al 2013]**

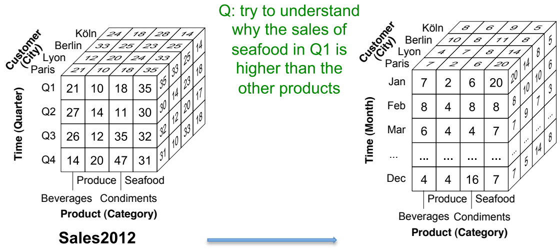
Algebra of OLAP Operations - rollup



ROLLUP(Sales2012, Customer ➔ Country, SUM(Quantity))

- **Roll-up**: aggregates measures along a dimension hierarchy (using an aggregate function) to obtain measures at a coarser granularity
  ROLLUP(CubeName, (Dimension ➔ Level), AggFunction(Measure))

Algebra of OLAP Operations – drill-down

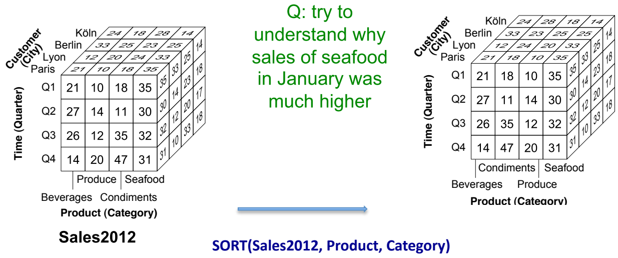


Q: try to understand why the sales of seafood in Q1 is higher than the other products

DRILLDOWN(Sales2012, Time ➔ Month)

- **Drill-down** moves from a more general level to a more detailed level in a hierarchy
  - **DRILLDOWN(CubeName, (Dimension ➔ Level))**

Algebra of OLAP Operations – sort

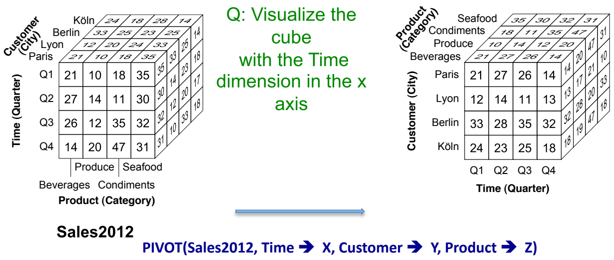


Q: try to understand why sales of seafood in January was much higher

SORT(Sales2012, Product, Category)

- **Sort** returns a cube where the members of a dimension have been sorted
  - **SORT(CubeName, Dimension, Expression [ASC | DESC])**
  - where the members of Dimension are sorted according to the value of Expression

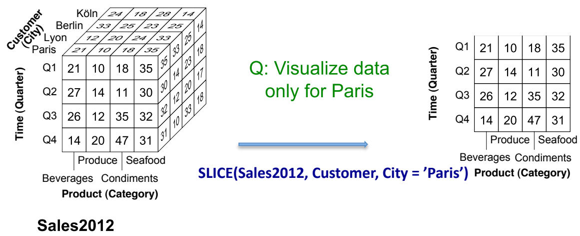
Algebra of OLAP Operations – pivot



- **Pivot (or rotate):** rotates the axes of a cube to provide an alternative presentation of data
  - **PIVOT(CubeName, (Dimension ➔ Axis)*)**
  - where the axes are specified as {X; Y; Z; X1; Y1; Z1; : : :}.

17

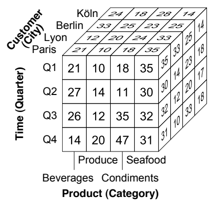
Algebra of OLAP Operations – slice



- **Slice**: removes a dimension in a cube so a cube of n-1 dimensions is obtained from a cube of n dimensions
  - **SLICE(CubeName, Dimension, Level = Value)**
- Dimension will be dropped by fixing a single Value in the Level; other dimensions unchanged
- Slice supposes that the granularity of the cube is at the specified level of the dimension

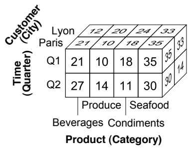
18

Algebra of OLAP Operations – dice



Sales2012

Q: Visualize data only for Paris or Lyon and quarters Q1or Q2

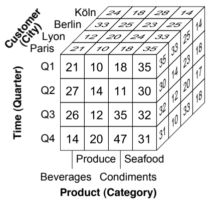


DICE(Sales2012, (Customer.City = 'Paris' OR Customer.City = 'Lyon') AND (Time.Quarter = 'Q1' OR Time.Quarter = 'Q2'))

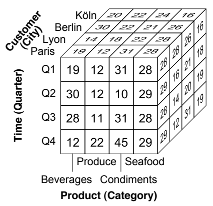
- **Dice**: keeps the cells of a cube that satisfy a Boolean condition Φ
  - **DICE(CubeName, Φ)**
- Φ is a Boolean condition over dimension levels, attributes, and measures.

Algebra of OLAP Operations – drill-across

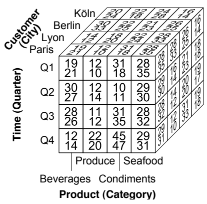
Sales2012



Sales2011



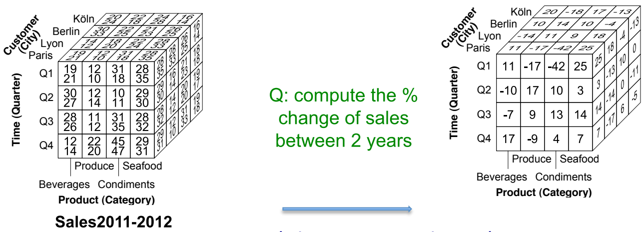
Q: Compare the sales quantities in 2012 with those in 2011



Sales2011-2012 ← DRILLACROSS(Sales2011, Sales2012)

- **Drill-across**: combines cells from two data cubes that have the same schema
  - **DRILLACROSS(CubeName1, CubeName2, [Condition])**

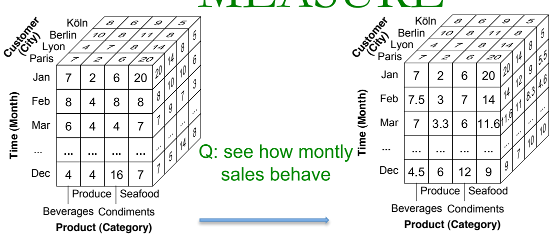
Algebra of OLAP Operations – ADD-MEASURE



ADDMEASURE(Sales2011-2012, PercChange = (Quantity2011-Quantity2012)/Quantity2011)

- **Add Measure**: adds new measures to a cube
  - **ADDMEASURE(CubeName, (NewMeasure = Expression)*)**
- **Drop measure**: Deletes a measure from a cube schema
  - **DROPMEASURE(CubeName, Measure*)**

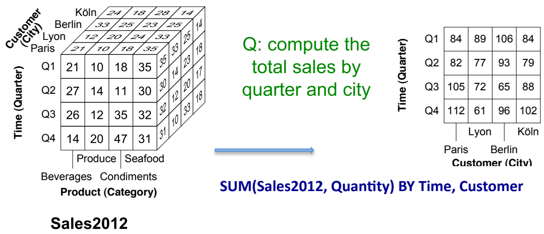
Algebra of OLAP Operations – ADD-MEASURE



ADDMEASURE(Sales2012, MovAvg = AVG(Quantity) OVER Time 2 CELLS PRECEDING)

- Another ex:
  - Computes the value of a cell by aggregating the measures of several nearby cells

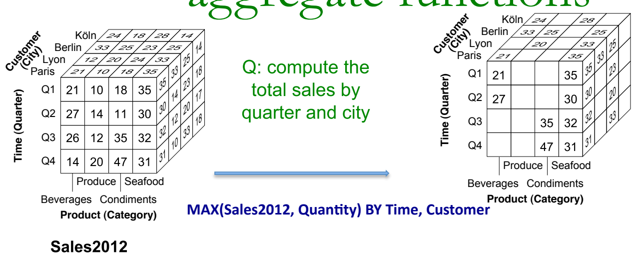
Algebra of OLAP Operations – aggregate functions



- **Aggregation functions** in OLAP are also needed at the current granularity, that is without performing roll-up.
  - **AggFunction(CubeName, Measure) [BY Dimension*]**
  - Cumulative: compute the measure value of a cell from several other cells; examples are SUM, COUNT, and AVG
  - Filtering: Filters the members of a dimension that appear in the result; examples are MIN and MAX. Filtering functions compute not only the aggregated value, but also the members of the dimension that belong to the result

23

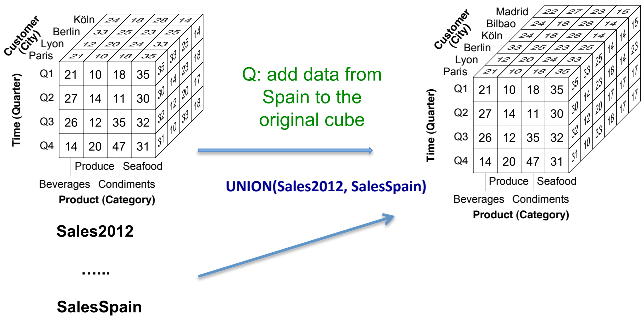
Algebra of OLAP Operations – aggregate functions



- Another example: max sales by quarter and city

24

Algebra of OLAP Operations – union, difference, drill-through



- **Union** merges two cubes having the same schema but disjoint instances
- Ex: If CubeSpain is a cube having the same schema as the original cube but containing only the sales to Spanish customers, we can perform:
- **Difference** removes the cells in a cube that belong to another one; the two cubes must have the same schema
- **Drill-through** allows to move from data at the bottom level in a cube to data in the operational systems from which the cube was derived; Could be used when trying to determine the reason for outlier values in a data cube

Next Lecture

- Conceptual Data Warehouse Design