



AALBORG UNIVERSITY
DENMARK

Aalborg Universitet

Cube

Pedersen, Torben Bach

Published in:
Springer Encyclopedia of Database Systems

Publication date:
2009

Document Version
Early version, also known as pre-print

[Link to publication from Aalborg University](#)

Citation for published version (APA):
Pedersen, T. B. (2009). Cube. In Springer Encyclopedia of Database Systems (2009 ed., pp. 538-539). Berlin/Heidelberg: Springer.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- ? Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- ? You may not further distribute the material or use it for any profit-making activity or commercial gain
- ? You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Cube

Torben Bach Pedersen, Aalborg University, Denmark, <http://www.cs.aau.dk/~tbp>

SYNONYMS

Hypercube

DEFINITION

A *cube* is a data structure for storing and analyzing large amounts of multidimensional data, often referred to as On-Line Analytical Processing (OLAP). Data in a cube lives in a space spanned by a number of hierarchical dimensions. A single point in this space is called a *cell*. A (non-empty) cell contains the values of one or more measures.

MAIN TEXT

As an example, a 3-dimensional cube for capturing sales may have a Product dimension *P*, a Time dimension *T*, and a Store dimension *S*, capturing the product sold, the time of sale, and the store it was sold in, for each sale, respectively. The cube has two measures: DollarSales and ItemSales, capturing the sales price and the number of items sold, respectively. In a cube, the combinations of a dimension value from each dimension define a *cell* of the cube. The measure value(s), e.g., DollarSales and ItemSales, corresponding to the particular combination of dimension values are then stored in the corresponding cells.

Data cubes provide true multidimensionality. They generalize spreadsheets to any number of dimensions, indeed cubes are popularly referred to as “spreadsheets on steroids.” In addition, hierarchies in dimensions and formulas are first-class, built-in concepts, meaning that these are supported without duplicating their definitions. A collection of related cubes is commonly referred to as a *multidimensional database* or a *multidimensional data warehouse*.

In a cube, dimensions are first-class concepts with associated domains, meaning that the addition of new dimension values is easily handled. Although the term “cube” implies 3 dimensions, a cube can have any number of dimensions. It turns out that most real-world cubes have 4–12 dimensions [3]. Although there is no theoretical limit to the number of dimensions, current tools often experience performance problems when the number of dimensions is more than 10–15. To better suggest the high number of dimensions, the term “hypercube” is often used instead of “cube.”

Depending on the specific application, a highly varying percentage of the cells in a cube are non-empty, meaning that cubes range from *sparse* to *dense*. Cubes tend to become increasingly sparse with increasing dimensionality and with increasingly finer granularities of the dimension values. A non-empty cell is called a *fact*. The example has a fact for each combination of time, product, and store where at least one sale was made.

Generally, only 2 or 3 dimensions may be viewed at the same time, although for low-cardinality dimensions, up to 4 dimensions can be shown by nesting one dimension within another on the axes. Thus, the dimensionality of a cube is reduced at query time by *projecting* it down to 2 or 3 dimensions via *aggregation* of the measure values across the projected-out dimensions. For example, to view sales by Store and Time, data is aggregated over the entire Product dimension, i.e., for all products, for each combination of Store and Time.

OLAP SQL extensions for cubes were pioneered by the proposal of the data cube operators CUBE and ROLLUP [1]. The CUBE operator generalizes GROUP BY, crosstabs, and subtotals using the special “ALL” value that denotes that an aggregation has been performed over all values for one or more attributes, thus generating a subtotal, or a grand total.

CROSS REFERENCE*

Dimension, Hierarchy, Measure, Multidimensional modeling, On-Line Analytical Processing.

RECOMMENDED READING

Between 5 and 15 citations to important literature, e.g., in journals, conference proceedings, and websites.

- [1] J. Gray, S. Chaudhuri, A. Bosworth, A. Layman, M. Venkatrao D. Reichart, F. Pellow, and H. Pirahesh. Data cube: A relational aggregation operator generalizing group-by, cross-tab and sub-totals. *Data Mining and Knowledge Discovery*, 1(1):29–54, 1997.
- [2] T. B. Pedersen, C. S. Jensen, and C. E. Dyreson. A Foundation for Capturing and Querying Complex multidimensional data. *Information Systems*, 26(5):383–423, 2001.
- [3] E. Thomsen. *OLAP Solutions: Building Multidimensional Information Systems*. Wiley, 1997.