**Aalborg Universitet**

AALBORG UNIVERSITY
DENMARK

# Hierarchy

Pedersen, Torben Bach

*Published in:*
Springer Encyclopedia of Database Systems

*Publication date:*
2009

*Document Version*
Early version, also known as pre-print

# Hierarchy

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

## SYNONYMS

None

## DEFINITION

A *hierarchy* is a structure specifying the containment relationships of a number of *values* (or *nodes* or *elements*). A hierarchy has a single *root* (or *top* or *ALL*) value. A value in a hierarchy may be linked to one or more *children* and a non-top value is linked to one or more *parents*. The links between values thus specify the containment structure. Hierarchies are essential for specifying dimensions in multidimensional cubes.

## MAIN TEXT

The notion of a dimension is an essential and distinguishing concept for multidimensional cubes. Dimensions are used for two purposes: the *selection* of data and the *grouping* of data at a desired level of detail. As an example, a 3-dimensional cube for capturing product sales may have a Product dimension, a Time dimension, and a Store dimension. The Product dimension captures information about the product sold, such as textual description, color, weight, etc., as well as groupings of products (product groups, product families, departments, etc.). The root value, representing "All products" then has store departments such as "Food" and "Electronics" as children, "Food" has product families such as "Dairy", "Meat", etc., as children and so on.

A hierarchy is typically organized into a a number of *levels*, each of which represents a level of detail that is of interest to the analyses to be performed. In the example above, the levels could be ALL Products, Departments, Product Families, Product Groups and Products. The instances of the dimension, e.g., "Food" are typically called *dimension values.* Each such value then belongs to a particular level. The hierarchy is used intensively, e.g., when performing On-Line Analytical Processing (OLAP). Here, data is explored by either moving up in the hierarchy to get a better overview (*rollup*) or moving down in the hierarchy to get more details (*drilldown*).

In some cases, it is advantageous for a dimension to have *multiple hierarchies* defined on it. For example, a Time dimension may have hierarchies for both *Fiscal Year* and *Calendar Year* defined on it. Multiple hierarchies share one or more common lowest level(s), e.g., Day and Month, and then group these into multiple levels higher up, e.g., Fiscal Quarter and Calendar Quarter to allow for easy reference to several ways of grouping. Most multidimensional models allow multiple hierarchies. A dimension hierarchy is defined in the metadata of the cube, or the metadata of the multidimensional database, if dimensions can be shared.

Most models require dimension hierarchies to form *balanced trees.* This means that the dimension hierarchy must have uniform height everywhere, e.g., all departments, even small ones, must be subdivided into Product families (and so on, all the way down to Products). If the hierarchy is not balanced like this, it is referred to as a *non-onto* or *unbalanced* hierarchy [2, 3]. Additionally, direct links between dimension values can only go between immediate parent-child levels, and not jump two or more levels. For example, all cities are first grouped into states and then into countries, so cities cannot be grouped directly under countries (as is the case in Denmark which has no states). If such non-immediate links occur in the hierarchy, it is called a *non-covering* or *ragged* hierarchy [2, 3], Finally, each non-top value has precisely one parent, e.g., a product must belong to exactly one product group. This may not always be desirable, e.g., it would be natural to put skimmed milk into both the "Diet" and "Dairy" product groups. If the hierarchies do not form balanced trees, this affects the so-called summarizability of the data, which means that special care must be taken to obtain correct aggregation results [1].

## CROSS REFERENCE*

Dimension, Multidimensional modeling, On-Line Analytical Processing, Summarizability.

## RECOMMENDED READING

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


[1] H. Lenz and A. Shoshani. Summarizability in OLAP and Statistical Data Bases. In *Proceedings of SSDBM*, pages 39–48, 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.