

# Dimension

Pedersen, Torben Bach

*Published in:*
Springer Encyclopedia of Database Systems

*Publication date:*
2009

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

# Dimension

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

## SYNONYMS

None

## DEFINITION

A *dimension* is a hierarchichally organized set of *dimension values*, providing categorical information for characterizing a particular aspect of the data stored in a multidimensional cube.

## MAIN TEXT

As an example, a 3-dimensional cube for capturing 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, etc.). The Time dimension captures information about the time of the sale, at the Date level or finer, as well as groupings of time such as Week, Month, Weekday, Quarter and Year. It may also contain application-specific time-related information, e.g., what the temperature was on the particular day (interesting for ice cream sellers) or whether there was a special event in town on that day, e.g., a big sports event. The Store dimension captures information about stores (Name, Size, Layout), as well as various groupings of Stores (City, State, Sales District).

The notion of a dimension is an essential and distinguishing concept for multidimensional cubes, where dimensions are a first-class object. Dimensions are used for two purposes: the *selection* of data and the *grouping* of data at a desired level of detail. A dimension is organized into a containment-like hierarchy composed of a number of *levels*, each of which represents a level of detail that is of interest to the analyses to be performed. The instances of the dimension are typically called *dimension values*. Each such value belongs to a particular level.

In some multidimensional models, a dimension level may have associated with it a number of *level properties* that are used to hold simple, non-hierarchical information. For example, the Weekday of a particular date can be a level property in the Date level of the Time dimension. This information could also be captured using an extra Weekday dimension. Using the level property has the effect of not increasing the dimensionality of the cube.

Unlike the linear spaces used in matrix algebra, there is typically no ordering and/or distance metric on the dimension values in multidimensional models. Rather, the only ordering is the containment of lower-level values in higher-level values. However, for some dimensions, e.g., the Time dimension, an ordering of the dimension values is available and is used for calculating cumulative information such as "total sales in year to date."

When implemented in a relational database, a dimension is stored in one or more *dimension tables* using either a so-called star schema (one table per dimension, with a surrogate key and one column per dimension level or level property) or a so-called snowflake schema (one table per dimension level, each with a surrogate key and an attribute for the textual name of the dimension value, as well as one attribute per level property).

## CROSS REFERENCE*

Cube, Hierarchy, Multidimensional modeling, Snowflake Schema, Star Schema.

## RECOMMENDED READING

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

[1] R. Kimball, L. Reeves, M. Ross, and W. Thornthwaite. *The Data Warehouse Lifecycle Toolkit.* Wiley Computer Publishing, 1998.

[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.