On-Line Analytical Processing (OLAP)

What is OLAP? OLAP is more than an acronym that means Online Analytical Processing. OLAP is a category of software tools that provides analysis of data stored in a database. With OLAP, analysts, managers, and executives can gain insight into data through fast, consistent, interactive access to a wide variety of possible views. Stated another way, OLAP is a category of applications and technologies for collecting, managing, processing, and presenting multidimensional data for analysis and management purposes. A widely adopted definition for OLAP used today in five key words is: Fast Analysis of Shared Multidimensional Information (FASMI).

- Fast refers to the speed that an OLAP system is able to deliver most responses to the end user.
- Analysis refers to the ability of an OLAP system to manage any business logic and statistical analysis relevant for the application and user. In addition, the system must allow users to define new ad hoc calculations as part of the analysis and report without having to program them.
- Shared refers to the ability of an OLAP system being able to implement all security requirements necessary for confidentiality and the concurrent update locking at an appropriate level when multiple write access is required.
- Multidimensional refers to a concept that is the primary requirement to OLAP. An OLAP system must provide a multidimensional view of data. This includes supporting hierarchies and multiple hierarchies.
- Information refers to all of the data and derived data needed, wherever the data resides and however much of the data is relevant for the application.
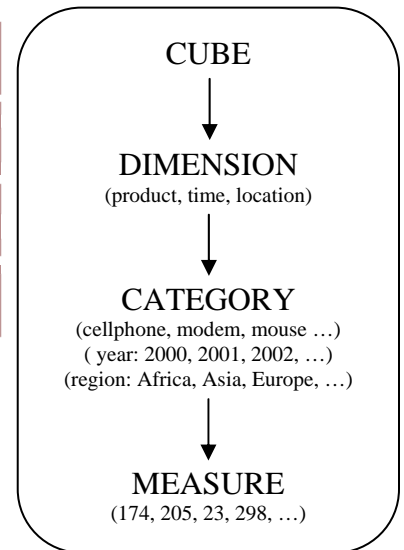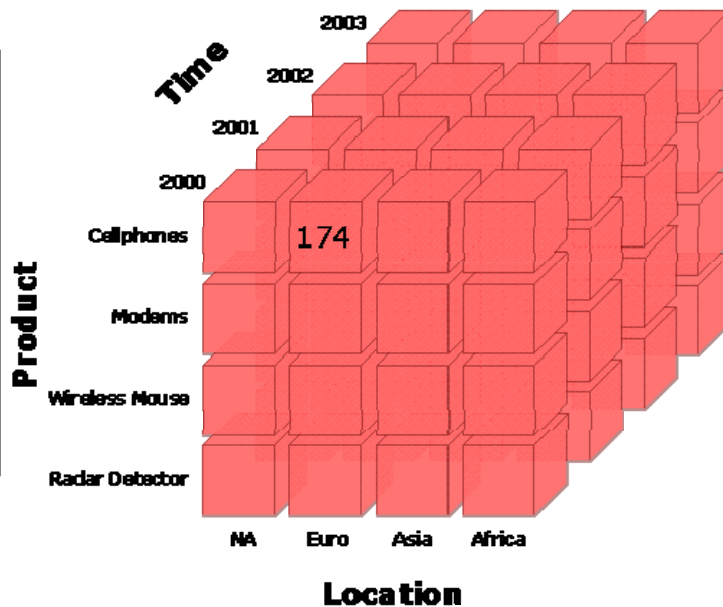
Cubes

What is an OLAP Cube? As you saw in the definition of OLAP, the key requirement is multidimensional. OLAP achieves the multidimensional functionality by using a structure called a cube. The OLAP cube provides the multidimensional way to look at the data. The cube is comparable to a table in a relational database.

The specific design of an OLAP cube ensures report optimization. The design of many databases is for online transaction processing and efficiency in data storage, whereas OLAP cube design is for efficiency in data retrieval. In other words, the storage of OLAP cube data is in such a way as to make easy and efficient reporting. A traditional relational database treats all the data in a similar manner. However, OLAP cubes have categories of data called dimensions and measures. For now, a simple definition of dimensions and measures will suffice. A measure represents some fact (or number) such as cost or units of service. A dimension represents descriptive categories of data such as time or location.

The term cube comes from the geometric object and implies three dimensions, but in actual use, the cube may have more than three dimensions.

The following illustration graphically represents the concept of an OLAP cube.

What is the question to which "174 million dollars" is the answer?

_____

_____

_____

CUBE
↓
DIMENSION
(product, time, location)
↓
CATEGORY
(cellphone, modem, mouse …)
(year: 2000, 2001, 2002, …)
(region: Africa, Asia, Europe, …)
↓
MEASURE
(174, 205, 23, 298, …)

Time, product, and location represent the dimensions of the cube, while 174 represents the measure. Recall that a dimension is a category of data and a measure is a fact or value.

Three important concepts associated with analyzing data using OLAP cubes and an OLAP reporting tool are slicing, dicing, and rotating.

Cubes

Slicing

A slice is a subset of a multidimensional array corresponding to a single value for one or more members of the dimensions not in the subset. For example, if the member Actuals is selected from the Scenario dimension, then the sub-cube of all the remaining dimensions is the slice that is specified. The data omitted from this slice would be any data associated with the non-selected members of the Scenario dimension, for example Budget, Variance, Forecast, etc. From an end user perspective, the term slice most often refers to a two- dimensional page selected from the cube.
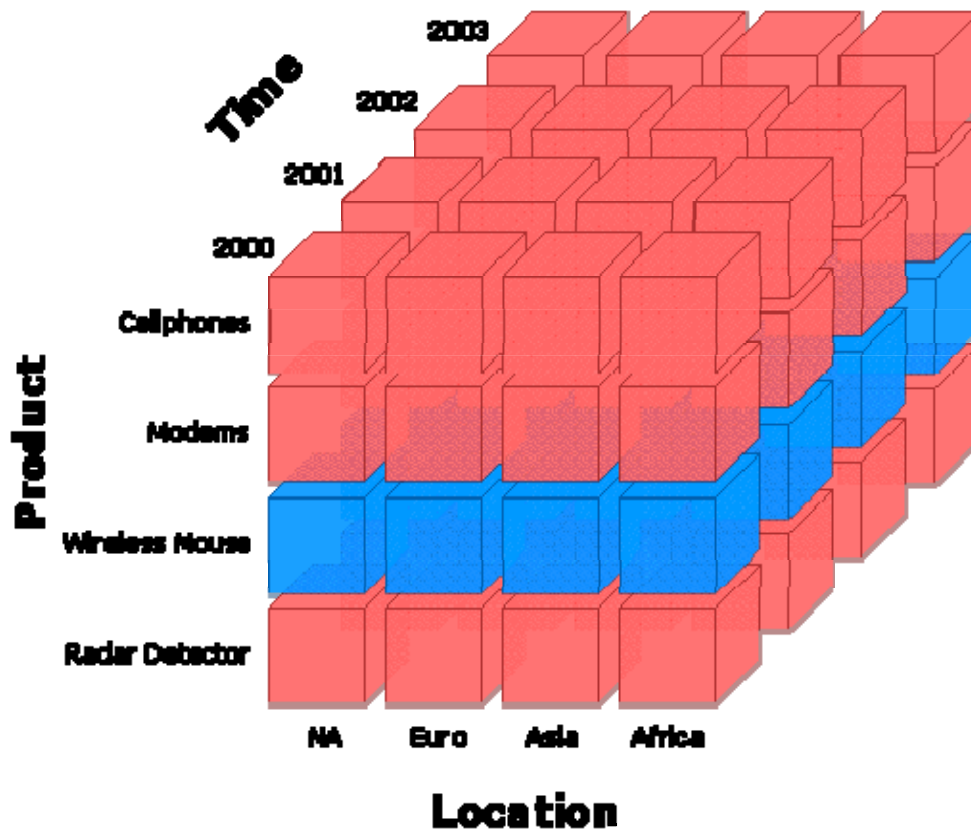

Figure 2: Slicing-Wireless Mouse

Figure 2 illustrates slicing the product Wireless Mouse. When you slice as in the example, you have data for the Wireless Mouse for the years and locations as a result. Stated another way, you have effectively filtered the data to display the measures associated with the Wireless Mouse product.
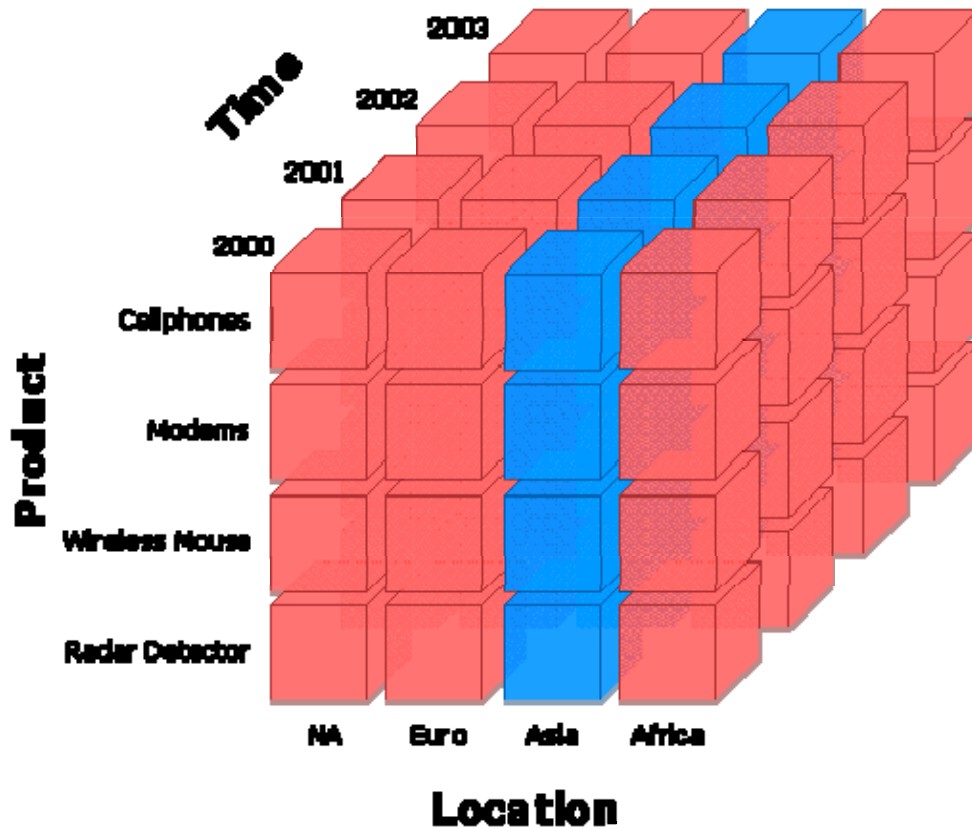
Figure 3: Slicing-Asia

Figure 3 illustrates slicing the location Asia. When you slice as in the example, you have data for Asia for the product and years as a result. Stated another way, you have effectively filtered the data to display the measures associated with the Asia location.

Cubes

Dicing

A related operation to slicing is dicing. In the case of dicing, you define a subcube of the original space. The data you see is that of one cell from the cube. Dicing provides you the smallest available slice.
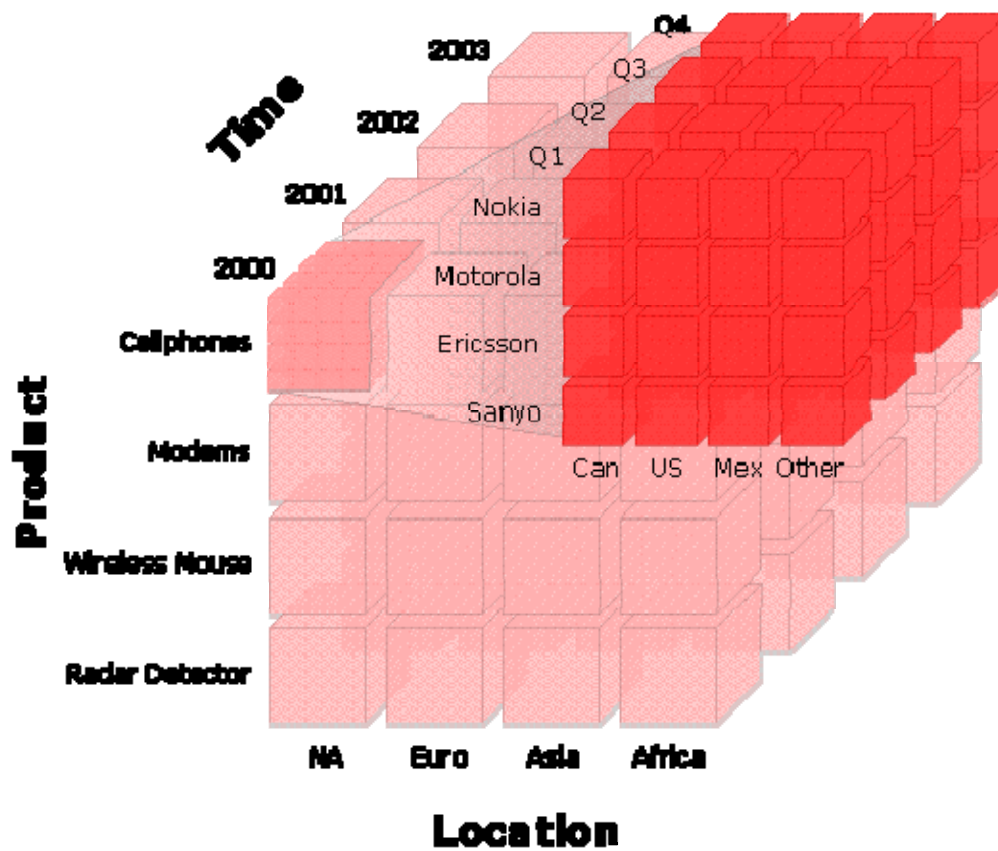
Figure 4 provides a graphical representation of dicing.

Figure 4: Dicing

Cubes

Rotating

Rotating changes the dimensional orientation of the report from the cube data. For example, rotating may consist of swapping the rows and columns, or moving one of the row dimensions into the column dimension, or swapping an off-spreadsheet dimension with one of the dimensions in the page display (either to become one of the new rows or columns), etc. You also may hear the term pivoting. Rotating and pivoting are the same thing.
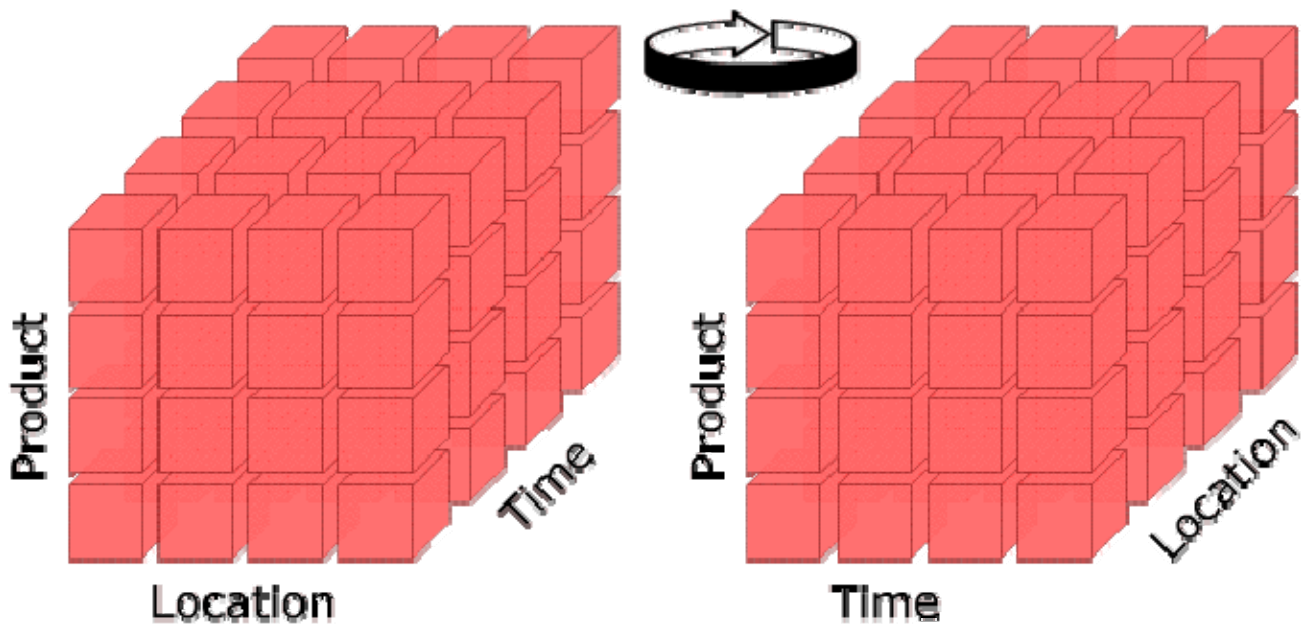
Figure 5: Rotating

Dimensions

Recall, that a dimension represents descriptive categories of data such as time or location. In other words, dimensions are broad groupings of descriptive data about a major aspect of a business, such as dates, markets, or products. The value of OLAP in reporting data is having levels within the dimensions. Each dimension includes different levels of categories. Dimension levels allow you to view general things about your data and then look at the details of your data.

Think of the levels of categories as a hierarchy. For example, your OLAP cube could have a time dimension. The time dimension then could have year, quarter, and month as the levels, as in Figure 6. Another example is location as a dimension. For the location dimension, you could have region, country, and city as the levels (shown in Figure 6). An important concept to OLAP is drilling. Drilling refers to the ability to drill-up or drill-down. These levels of categories (hierarchies) are what provide the ability to drill-up or drill-down on data in an OLAP cube. When you drill-down on a dimension, you increase the detail level of viewing the data. For example, you start with the year and view that data, but you want to see the data by each quarter of the year. You would drill-down to see the quarterly data. You could drill-down again to see the monthly data within a specific quarter.
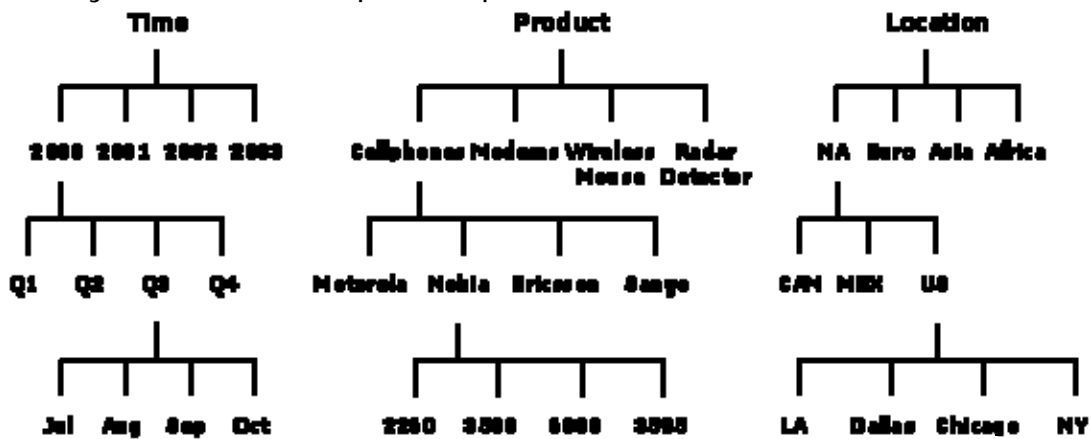

Figure 6: Dimensions

Categories

Dimensions have members or categories. A category is an item that matches a specific description or classification such as years in a time dimension. Categories can be at different levels of information within a dimension. You can group any category into a more general category. For instance, you can group a set of dates into a month, a set of months into quarters, and a set of quarters into years. In this example, years, quarters, and months are all categories of the time dimension.

Categories have parents and children. A parent category is the next higher level of another category in a drill-up path. For example, 2003 is the parent category of 2003 Q1, 2003 Q2, 2003 Q3, and 2003 Q4. A child category is the next lower level category in a drill-down path. For example, January is a child category of 2003 Q1.

Figure 7 below shows you the time, product, and location dimensions with the parent and children categories. In the time dimension, you see the parent Year Category and the children Quarter Category and Month Category.
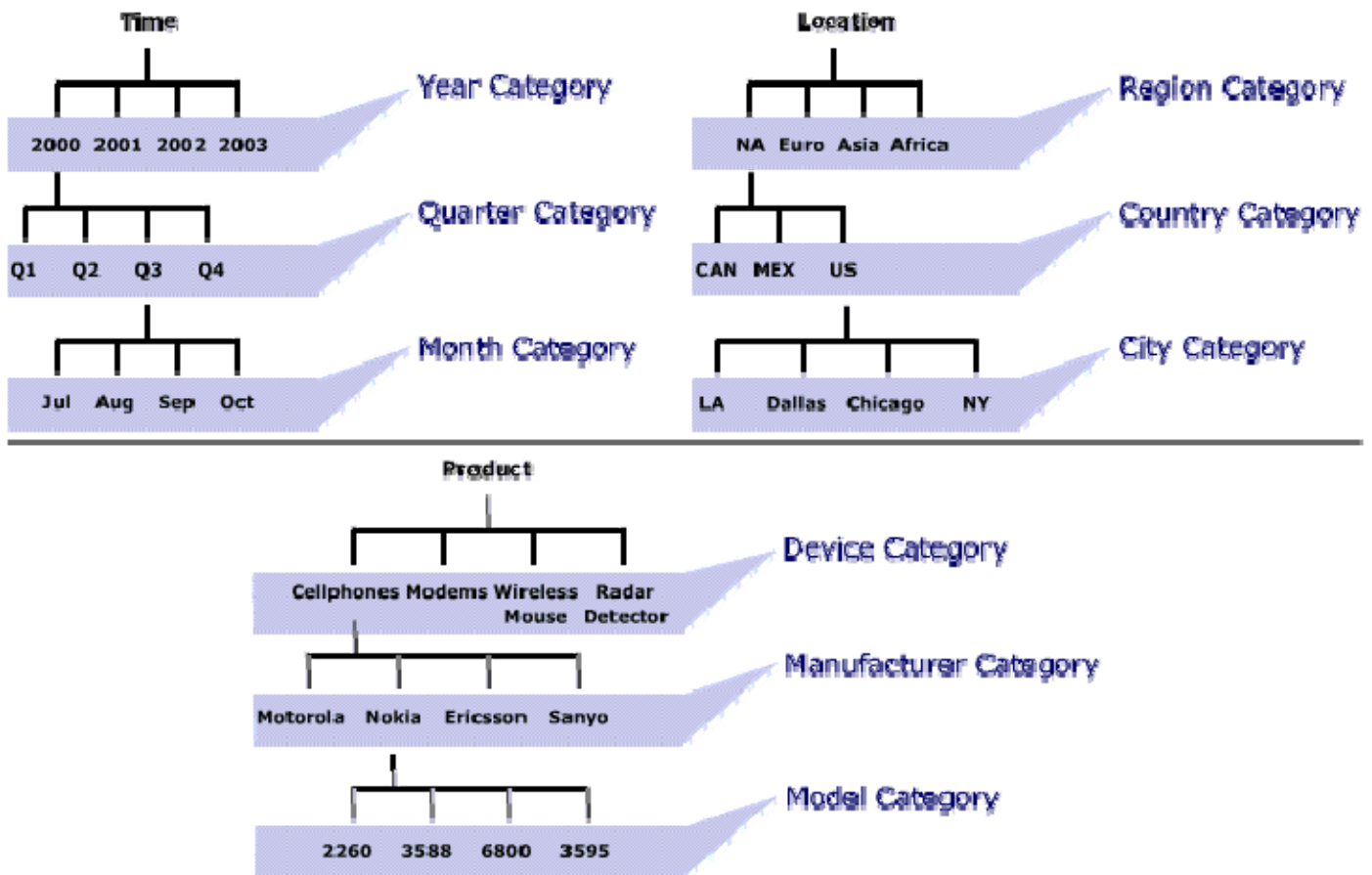

Figure 7: Categories

Measures

The tutorial to this point has provided you with a definition of OLAP, a description of cubes, the meaning of dimensions, and a description of categories. The focus now is an explanation of measures.

The measures are the actual data values that occupy the cells as defined by the dimensions selected. Measures include facts or variables typically stored as numerical fields, which provide the focal point of investigation using OLAP. For instance, you are a manufacturer of cellular phones. The question you want answered is how many xyz model cell phones (product dimension) did a particular plant (location dimension) produce during the month of January 2003 (time dimension). Using OLAP, you found that plant a produced 2,500 xyz model cell phones during January 2003. The measure in this example is the 2,500.

Additionally, measures occupy a confusing area of OLAP theory. Some believe that measures are like any other dimension. For example, one can think of a spreadsheet containing cell phones produced by month and plant as a two-dimensional picture, but the values (measures)—the cell phones produced—effectively form a third dimension. However, although this dimension does have members (e.g. actual production, forecasted production, planned production), it does not have its own hierarchy. It adopts the hierarchy of the dimension it is measuring, so production by month consolidates into production by quarter, production by quarter consolidates into production by year.

The example in Figure 8 shows the measure Volume of Product. Each value in Figure 8 is the measure of Volume of Product per year listed by product. From Figure 8, ABC Company produced 84,000 modems in 2003.

Volume of Product (Numbers in 1000)	CellPhones	Modems	Wireless Mouse	Radar Detector
2003	101	84	98	75
2002	94	60	70	60
2001	88	65	80	85
2000	105	82	88	93

Figure 8: Measures