# Visualising OLAP

*Jaturon Chattratichat*  *Yike Guo*  *Martin Köhler*
{j.chattratichat,y.guo,m.kohler}@doc.ic.ac.uk

Imperial College/Fujitsu Parallel Computing Research Centre
180 Queen's Gate, London SW7 2BZ

## Abstract

Mining Very Large-Scale Databases (VLDB) is an important task in data mining. With the help of parallel computers, information retrieval time could be substantially reduced. On-line Analytical Processing(OLAP), being a specialised data-warehousing tool, is likely to play a key role in this aspect. The process of sampling, pre-processing and output visualisation are integral parts of large-scale data mining. We introduce a new concept of visualising multi-dimensional data.

The traditional way of viewing OLAP data with a hierarchical structure is through spreadsheet-like tables with a maximum of three dimensions at one time. In our approach, we propose to display all dimensions graphically, giving a rich representation of the underlying OLAP database. We also propose an enhanced set of tools which would enable the user to perform traditional OLAP operations, such as "slice & dice" and "roll up & drill down".

## 1 Introduction

Decision support applications involve many complex queries on very large databases. The traditional OLTP systems can no longer support these complex queries and ensure good performance. Thus, a new way of storing and processing data, termed OLAP (On-Line Analystical Processing), is developed.

On-Line Analytical Processing (OLAP) is a category of software technology that enables analysts to gain insight into data through fast, consistent and interactive access. The technology transforms raw data to reflect the real dimensionality of the data.

An OLAP server is a multi-user data manipulation engine that is designed to support multi-dimensional data structures. It is designed for rapid, read-only retrieval. Thus, most "data warehouses" are based on an OLAP server. Most retrievals are often ad-hoc and thus make the OLAP server suitable for business analysis and decision support systems. Every data item is accessed based on the interaction of the dimension members that define the item. The OLAP server may have a specialised engine that stores pre-computed data structures and their aggregates, or may populate the data structure in real time from relational databases.

In this paper, we give a general introduction to the basic idea of an OLAP system. We also introduce our proposed method and give some usage examples.

## 2 Basic OLAP Features

### 2.1 OLAP Operations

In an OLAP system, some typical operations for selecting attributes in the multidimensional database are :

- *pivoting* - rotate the "cube" to show a particular face.
- *slice & dice* - select a subset of a cube/dimension.
- *roll up* - aggregating a dimension to the next level of the hierarchy.
- *drill down* - the converse of roll up. Displays detailed information for each aggregated point on a dimension. This operation

**Simple Relational Table**

| Product | Date | Store | Sale |
|---|---|---|---|
| Soap | Jan | London | £450 |
| Shampoo | Jan | London | £550 |
| Lotion | Jan | London | £770 |
| Soap | Feb | London | £400 |
| Shampoo | Feb | London | £500 |
| Lotion | Feb | London | £700 |
| Soap | Jan | Paris | £440 |
| ... | ... | Rome | ... |
| ... | ... | ... | ... |

**Mapping first row to the cube**

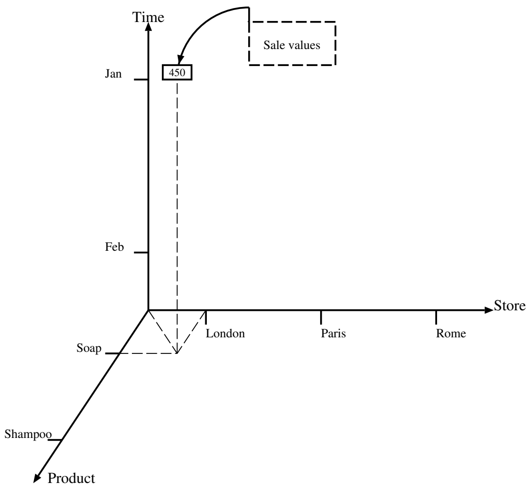


Figure 1: Product cube with product, time and store dimensions

is essential because users want to see aggregated data first and selectively see more detail.

A full OLAP glossary can be found in [Cou95].

### 2.2 Multidimensional Features and Terminology

We introduce a simple database that contains sales information about some products from the relational database table Product as shown in figure 1.

The Product table contains four attributes: product, date, store, and sale. These attributes, or *dimensions* together form a key to the attribute that contain the sale value, commonly known as *measures*. The right hand side of the figure 1 is a three-dimensional representation of the first row the relational data. An interpretation of figure 1 is described as :
"At the intersection of "Soap" in product dimension, "London" in store dimension, and "Jan" in time dimension, a *cell* value of £450 (*measure*) is assigned".

Most literature use "attributes" and "dimensions" interchangeably. In this example, we note that sale is an attribute value in the relational table but becomes a measure. In most common business models, the sale measure is not the only value that analysts are interested in. There can be multiple measures such as profit, sale quantity, cost associated with the key attributes. Thus these measures can be grouped into a variable dimension. We will can interpret the cube in figure 1 as :
"At the intersection of "Soap" in product dimension, "London" in store dimension, "Jan" in time dimension, and "Sale" in variable dimension, the cell value is £450".

### 2.3 Hierarchies and Dimensional Grouping

One key feature of OLAP is the ability to organise data into hierarchies so that they fit the natural ordering of the attributes. Each dimension may have any number of *hierarchies* that specify the aggregation *levels* and granularity of the data.

For example, *Year* $\rightarrow$ *Quarter* $\rightarrow$ *Month* $\rightarrow$ *Day* is a four level hierarchy on the date dimension, where year is on the top-most level (also indicate that it is a *dimension* and that *Day* is on the bottom-most level 3.) Similarly, we think of geography in terms of *Continent* $\rightarrow$ *Country* $\rightarrow$ *State* $\rightarrow$ *Street*. In contrast, in a scenario dimension, we may have a *Forecast*, *Budget*, and *Actual* dimension. These dimensions are, however, not portrayed hierarchically. For concreteness, we extend the above example to include hierarchies in some dimensions. Note that we use three dimensional data for the sake of simplicity, but in reality, there can be $k$-dimensions.

Define the Product hierarchy as :
*Category* $\rightarrow$ *Type* $\rightarrow$ *ProductName*,

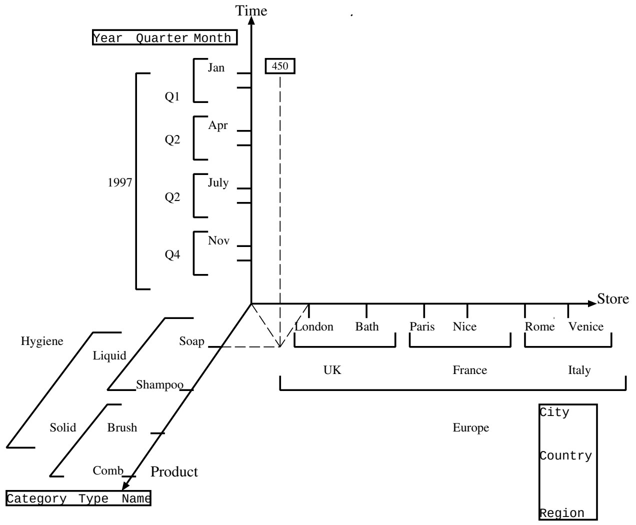


Figure 2: Products cube with product, data and store (hierarchical) dimensions

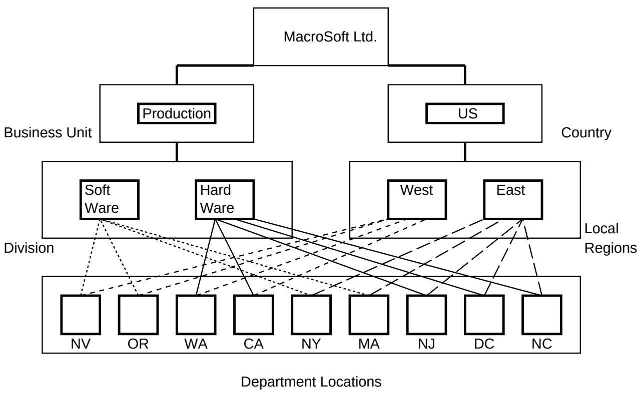


Figure 3: Members can be grouped into different business models

Time hierarchy as :
*Year* → *Quarter* → *Month*,
and Store hierarchy as :
*Region* → *Country* → *City*
The new cube is illustrated in the figure 2.

Another related key feature of an OLAP system is the facility to let a user define a "business model". This business model is an essential part of the analytical process of OLAP. As illustrated in figure 3, there are many ways to aggregate members to fit different kinds of models. A regional sales manager would be interested in the sale figure of the "East" and "West" regions of the US. On the other hand, product manager may be interested in the "hardware" division of the company.

### 2.4 Aggregation

Aggregation or consolidation is one of the most important and frequently used OLAP functions. It involves aggregation of data over one or more dimension hierarchies or formulaic relationships within a dimension or among dimensions.

While "summations" is the most frequently used form, consolidation can involve complicated mathematical and statistical functions (e.g., variance, expected values, correlation). Decision makers use key performance metrics to evaluate the operations within their domain, and OLAP systems need to be capable of delivering these metrics in a user-customisable format. Commonly used metrics include [Mic95] :

- *Multidimensional Ratios* - Percent to Total
- *Comparisons* - Actual vs. Plan, This Period vs. Last Period
- *Ranking and Statistical Profiles* - Top N/Bottom N, 80/20, Quartiles
- *Custom Consolidations* - Financial Consolidations, Market Segments, Ad Hoc Groups

We consider the same example in section 2.2 where we have dimensions product, time, and store with sale as measures. To compute the "cube", we will compute sales grouped by all the subsets of these dimensions. That is, we will compute sales by "product, time, and store"; "sales by product and time"; "sales by product and store";"sales by time and store"; "sales by products" ; "sales by time"; "sales by store", and "overall sales". In a multidimensional system, all or a subset of these precomputations are done in response to a user query or part of a "load process".

## 3 OLAP Visualiser

### 3.1 Traditional method

Visualisation of aggregated, multi-dimensional data with hierarchical structure (OLAP data) is traditionally achieved through spreadsheet like tables which display the numerical values of measure dimensions in a 2-dimensional grid over two of the axes of data. A third dimension could be displayed using a method commonly known as

**Relational Table**

| Product | Time | Store | Cash | Credit |
|---|---|---|---|---|
| Soap | Jan | London | £450 | £445 |
| Shampoo | Jan | London | £550 | £340 |
| Lotion | Jan | London | £770 | £440 |
| Soap | Jan | London | £400 | £405 |
| Shampoo | Jan | London | £500 | £340 |
| Lotion | Jan | London | £700 | £400 |
| ... | ... | Paris | ... | ... |
| ... | ... | ... | ... | ... |
| ... | Feb | ... | ... | ... |

**London *slice* of the hyperCube**

| Page : London | | | | |
|---|---|---|---|---|
| | **Jan** | | **Feb** | |
| | Cash | Credit | Cash | Credit |
| Soap | £450 | £445 | £400 | £405 |
| Shampoo | £550 | £340 | £500 | £340 |
| Lotion | £770 | £440 | £700 | £400 |

Figure 4: Crosstabulation- A traditional way of viewing OLAP

"pivoting". An example of such a spreadsheet method is presented in figure 4. The top table shows a relational table that is translated into a hypercube. A slice of the hypercube is illustrated using a spreadsheet. In this case, the spreadsheet is enhanced to support dimension groupings. Product type is assigned to the rows, a hierarchical grouping of Time → PaymentMethod is assign to the column, and Store location is assigned to the page or pivot. The user is able to switch between pages of different store locations or an aggregated "ALL" for all the cities in the database.

Commercial vendors have enhanced the spreadsheet tools to include many other functionalities such as dimension swapping, model-definition tool, and click-and-drill-down. However, the user is always restricted to viewing three dimensions at a time.

## 3.2 OLAPViz Overview

The aim of data visualisation is used to convert large amounts of information into a meaningful and interpretable visual representation. The challenge in visualising an OLAP system is to be able to display all dimensions simultaneously.

Here we propose a new method of visual display for OLAP system - the OLAPViz. The system consists of three main components : Model Builder, Control Tools, and the MetaCube display .

In the subsections below, we describe the specifications of the three components of the OLAPViz system. In the next section, we give some illustrations of how OLAPViz can be used for trend analysis.

### 3.2.1 Model Builder

The Model Builder is a tool that enables the user to customise a business model. It allows a user to group dimensions and to build a hierarchy such as the model shown in figure 3.

Initially, the builder would read data from the relational database. Then, based on the model specified, a new cube and its meta-data would be created. Aggregations (summations) are pre-computed and stored within the cube.

We adopt the "eager" approach by precomputing all the aggregates for all dimensions. For example, a cube with dimensions Time, Location, Scenario (t,l,s) over the measure Sales_Value will require aggregates for (*,l,s), (t,*,s), (t,l,*), (*,*,s), (*,l,*), (*,*,s), and (*,*,*) where "*" denotes the "ALL" notation. If other metrics such as average or top-10 quartiles are needed, they are computed on the fly.

### 3.2.2 MetaCube Display

The MetaCube is a three dimensional plot in which the data and the relationship among them are plotted in 3D space. The display structure is based on a series of vertical planes arranged alongside each another. Each plane is a two-dimensional user-defined plot of `measures` vs. `dimension`. The combination of the (`measures`, `dimension`) pair that forms a plane is dependent on the user.

In general, the number of planes $\#K$ in a MetaCube is equal to the number of dimensions $\#N$in the model. When $\#K <$#N, the dimensions that are *not* present in the MetaCube is assumed to be at the highest (top-most) level in the hierarchy.

The horizontal "floor" of the MetaCube would consists of columns of dimensions while the vertical "back-wall" has columns of dimensions. Given a model with dimensions A, B, C and D, and a MetaCube has A, B and C defined as its 3 planes. Since D is not included in the MetaCube, each of its data refers to a cell at $(a_i, b_j, c_k, D_{ALL})$ for $0 \leq i \leq |A|, 0 \leq j \leq |B|, and 0 \leq k \leq |C|$ on the model cube. Although each plane is defined independently, they are derived from the same model. Thus, a drill-down/roll-up operation in one plane will affect every other planes, since the intersection of the dimensions are different. The user is able to rotate and zoom on the MetaCube to get a clearer view. An illustration of the MetaCube is given in the next sections.

#### 3.2.3 Control Tools

The operations in OLAPViz are done via the control tools. The control tools will have the following features :

- *drill-down* - unfold along a hierarchy. More detail is revealed in the MetaCube.
- *roll-up* - replace current selection ( or group of selections) with that of a parent along the defined hierarchy. This has the effect of reducing the details and replacing them with aggregated values.
- *Dimension Brushing* - choose one or some of the current level's dimensions. This similar to a "select" operation on the the dimension attributes.
- *Range Brushing* - define a range for one or some "measures". This operation will remove data values which doesn't satisfy the desired range of measures.

### 3.3 System Usage Examples

By modifying the OLAP data model defined in 2.3, we define a table of data names and its dimension grouping in figure 5. The leftmost set is the top level, and is a parent of the set following the "→" notation.

| Name | Hierarchy |
|---|---|
| Time | {Year96, Year97,Year98} → {{Quart1, Quart2, Quart3,Quart4}, {Quart1, Quart2, Quart3,Quart4}, {Quart1, Quart2, Quart3,Quart4}} |
| Location | {Americas, Europe} → { {Canada, USA}, {UK, France, Italy}} |
| Product | {Liquid, Solid} → {{Soap, Shampoo},{Brush, Comb}} |
| Scenario | {Budget, Actual} |
| Measures | Cost_Value, Sale_Value, Profit, Stock_Unit |

Figure 5: Data Model Table

The goal is to view the data in this ad-hoc model in a graphical way. We will describe how a graphical view of OLAP data can be constructed. Suppose we are interested in the `cost_value` against `time` in `1997`. We plot a (`cost_value`,`time`) pair in a 3 dimensional space, as illustrated in left side figure 6. The "measure" is always placed on the y-axis while the "dimension" on the z-axis. Each coloured dot in space corresponds to the cost value for `Quart1, Quart2, Quart3,` and `Quart4` in `1997` respectively.

The illustration on the bottom is an extension of the one on top in figure 6 . Suppose, then, that we are interested in the relationship between `cost_value` and `sale_value` for `1997` and all scenarios. We would extend the MetaCube to include another `sale_value` against `scenario` plane. Note that for `Quart1` in the first plane, there are now two points on the vertical dotted lines, one for *actual* cost value and another *budgeted* cost value. On the other hand there are exactly four points describing the *actual* sale value, one for each quarter(`Quart`). In essence, this plot describes the points for $(\{Quart1 \cup Quart2 \cup Quart3 \cup Quart4\}^{Time}, \{Actual \cup Budget\}^{Scenerio}, ALL^{Location}, ALL^{Product})$ dimension with `Cost_Value` and `Sale_Value` as measures. Note also that in the MetaCube, these two points are joined by a straight line. "Line A" in figure 6 could be described as "The relationship between Cost and Sale for the `Time=Quart4` and `Scenario=Budget`".

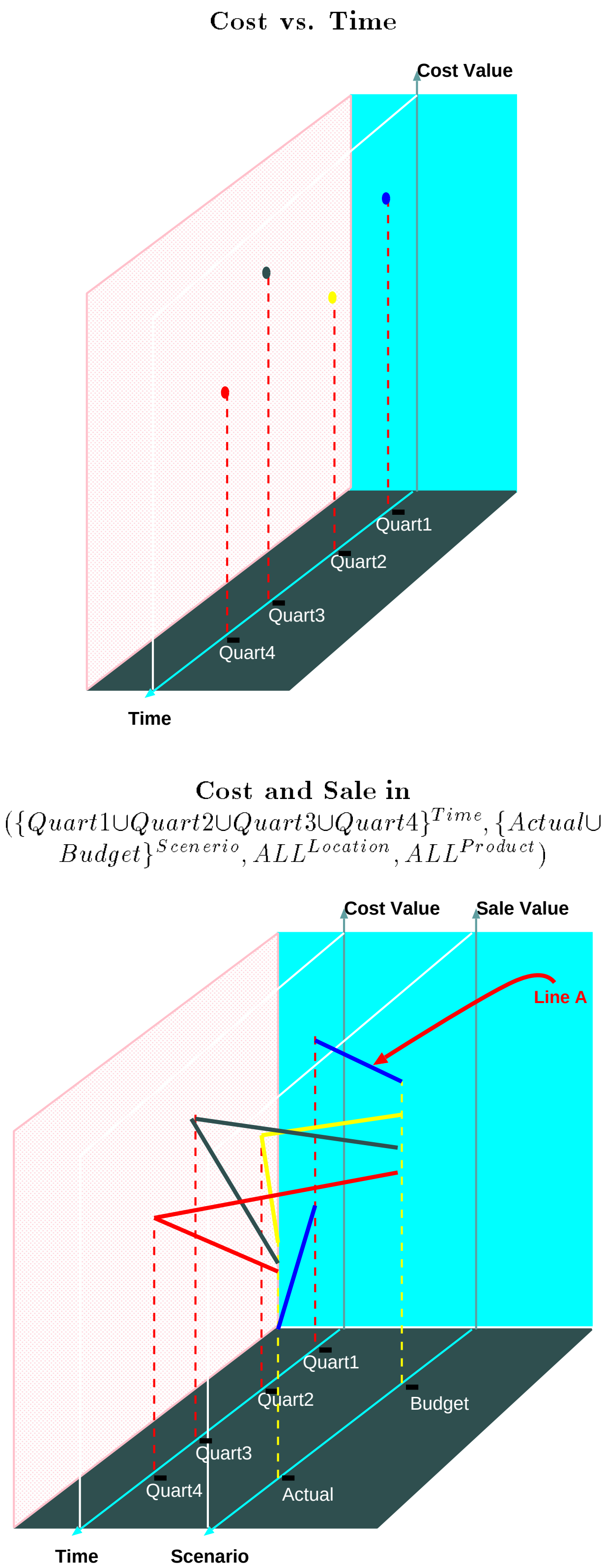


Figure 6: Simple MetaCube

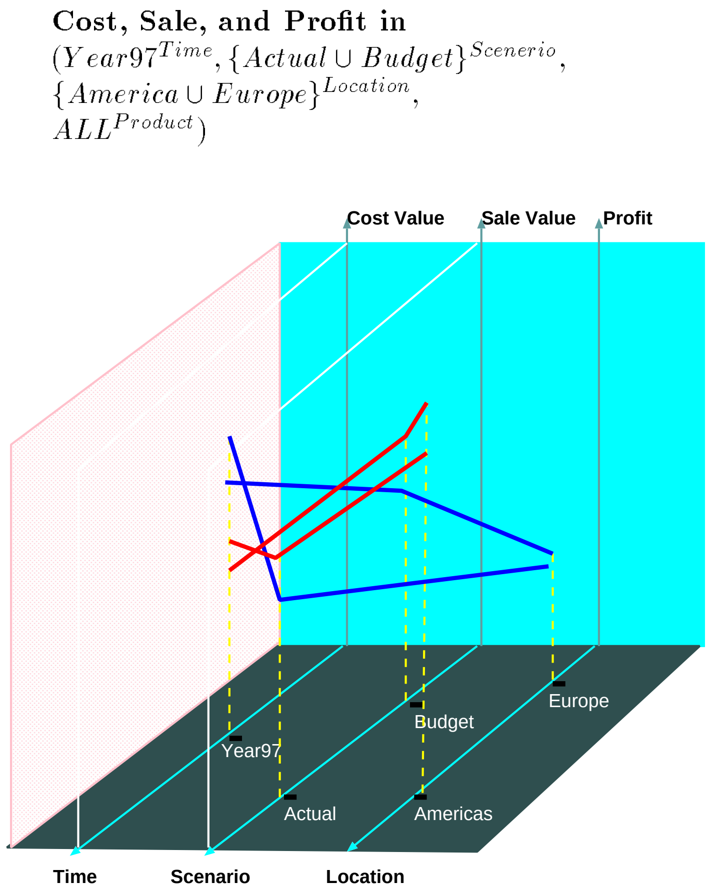


Figure 7: Extended Meta-Cube

Multiple planes can be added to the Meta-Cube. In figure 7, a additional Profit vs. Location plane was added. Also, we have rolled up the first plane (i.e. Cost Value vs. Time) by one level. A colour scheme has also been assigned. For example, all red lines are associated with `Location=Americas` while all blue lines are associated with `Location=Europe`. A general analysis of the MetaCube display reveals that the profit in Americas are higher than that of Europe for the year 97 in both scenarios. Notice that the Product dimension has been left out, so we can assume that the MetaCube is describing all products.

# 4 Conclusion

In this paper, we have presented a novel method of viewing multidimensional data in a graphical way. The focus is however not only on the MetaCube display but also on fact that the user could build the OLAP model and manipulate the display to weed out uninteresting data. The use of a 3D paradigm have proven to be an effective method of viewing hidden data.

## References

[Cha97] Jaturon Chattratichat. Issues in olap. Technical report, Department of Computing, Imperial College, London, 1997. unpublished.

[Cou95] The OLAP Council. Olap and olap server definitions. *Datamation*, May 1995.

[Mic95] MicroStrategy. The case for relational olap. *White Paper*, 1995.

**Acknowledgement.** This work was supported by the Fujitsu European Centre for Information Technology.