Paper #UC1448

The Practical Union of OLAP Analysis and Geographic Mapping

Gail Kramer, SAS, Cary, NC
Stuart Nisbet, SAS, Cary, NC

ABSTRACT

Whether it is product sales, insurance rates, or credit card fraud, multidimensional data often contains a geographic dimension that allows analysis based on varying locations. In many cases, these locations are political boundaries such as country, state, province, or city. In other cases, the locations, or regions, are defined based on business rules such as sales territory, wireless coverage plans, or population-based designated market areas (DMAs). In all cases, these defined regions provide a critical context for On-Line Analytical Processing (OLAP) exploration and reporting. Coupling the power and flexibility of ArcGIS Server 9 with modern OLAP reporting tools, the "A" in "OLAP" has never looked so good!

INTRODUCTION

This paper introduces some well-documented benefits of OLAP exploitation and suggests how these benefits can be extended to include geographic display and navigation. An overview of methodology and implementation details for linking levels of an OLAP hierarchy with layers in a geographic map is also described. Finally, the applicability of this functionality to a few vertical business domains is discussed.

WHAT IS OLAP?

No OLAP document would be complete without first defining what is meant by OLAP. In a formal sense, OLAP (OnLine Analytical Processing) is defined as fast access to large amounts of summarized data. Implied in this definition is the concept of dimensionality. For without dimensions, there would be nothing to summarize the data by. Thus, a more generalized definition might be the ability for users to quickly interrogate large amounts of data, at varying levels of detail, across a variety of combinations of business dimensions.

OLAP is full of terms and acronyms, which are often ill defined. Thus, before digging too deeply into the internals of OLAP, it is helpful to provide a few basic definitions of some of the OLAP terms that will be used in this paper:

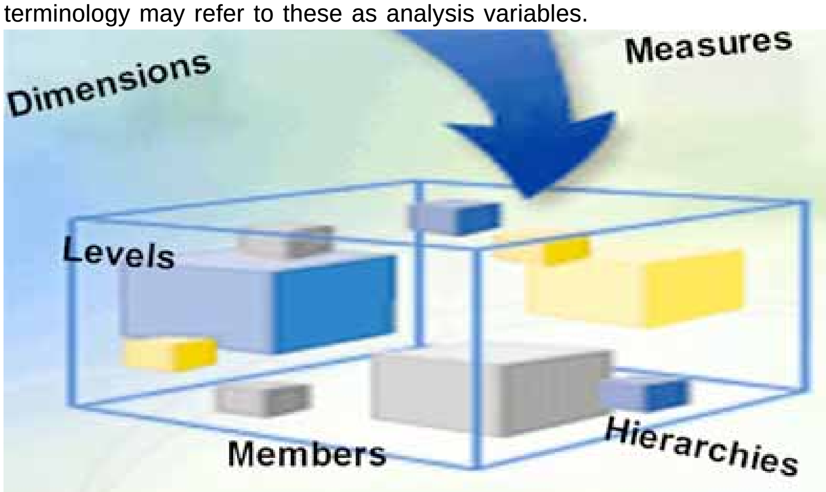
Dimension – A business perspective that is useful for analyzing data by or across. Often times referred to as a hierarchy. Examples are Time, Product or Geography.

Hierarchy - A Dimension may contain one or more hierarchies. Each hierarchy specifies an order for some or all of the levels in a dimension, which is then used to determine the navigation path. For example, "Country, Region, State" or "Region, City".

Level – Dimensions are often made up of various levels of detail. For example, the Time dimension may consist of Year, Quarter and Month levels. Some dimensions will only have one level, in which case the level is implied. When referring to the physical representation of data, a dimension level is sometimes referred to as simply a dimension, or for those familiar with SAS terminology, a class variable.

Member – A given value of a dimension level. For example, members of the Year level of the Time dimension could be “1998”, “1999”, and “2000”. The number of unique members at any dimension level is referred to as the cardinality of that dimension level.

Measure – The ultimate business measure that is being aggregated. For example, Sales or Profits. Measures also have statistics associated with them such as Sum, Count, Average, etc. Those familiar with SAS terminology may refer to these as analysis variables.



OLAP STORAGE TECHNOLOGY

Looking at the business understanding, OLAP delivers "fast, consistent, interactive access to a wide variety of possible views of information" - OLAP Storage Technology drives this delivery by summarizing detailed data and storing summaries for fast access.

Enterprises collect all sorts of business events using front-end business applications like ERP or CRM systems. These bits and pieces of information are stored in small transactions optimized for fast processing. In order to generate an overview on what is going on in business, all of these transactions need to be summed up, consuming lots of compute power, time and knowledge on how to access and summarize these transactions.

OLAP Storage Technology overcomes these issues by hiding the complexity of detailed data stores and by pre-calculating views on business that are likely to be requested by business users.

OLAP VIEWING COMPONENTS

Even though most of the "real" OLAP work is done in OLAP Storage, OLAP viewing components surface the information to end-users and are therefore most often associated with the term OLAP. There are different flavors of viewing components available from SAS. Basically, they can be grouped into web-based and desktop-based components providing static or dynamic views on the multidimensional information. Static web based reports are suitable to line-of-business users in order to keep them informed about business operations. Interactive desktop based reports are used to support decision-makers, providing them with a 360° view on their business and the capability to further analyze trends or exceptions in the information.

GEOGRAPHIC INTEGRATION WITH OLAP IS POWERFUL

Hopefully, you can see the natural hierarchies that present themselves in the business questions that are asked daily. Which products sold well? When were they sold? Who is selling the most? Where are the profitable stores? What demographic factors affect sales? These questions can be answered using OLAP tools and traditional reporting with hierarchical levels which might look like this:

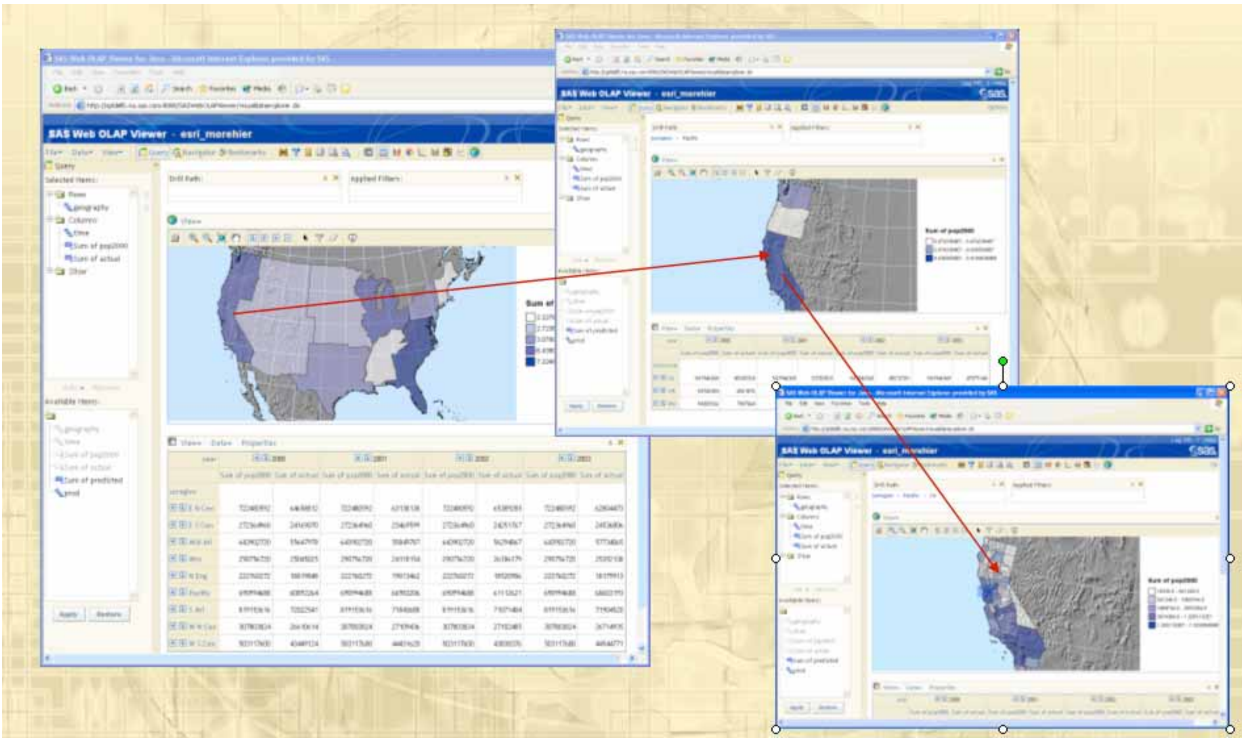
Question	Sample Hierarchy	Example
Which products sold well?	Product Category/Product Type/Item	Office->Furniture->Desk Office->Supplies->Pencils Home->Garden->Lawn Mower
When were they sold?	Year/Quarter/Month/Day	2004->First->January->25
Who is selling the most?	National Sales Director/Regional Sales Manager/Local Sales Representative	Betty Brown->Bob Smith->Sam Jones
Where are the profitable stores?	Region/State/Zip Code	Southeast->NC->27513
What demographic factors affect sales?	Age/Income/Education	Under 29->55K-100K->4-year degree

There are many such questions that can be answered with OLAP reporting, but they frequently fall into the categories of “who, what, where, when, and why”. Tabular reporting does a fine job of providing the facts in summarized form, but to answer the “where” questions in the most intuitive fashion, maps are needed. If you doubt this, imagine the 6 o’clock weather reports done with a tabular listing of states and counties along with their respective highs, lows, and expected rainfall! An alphabetized list, or even one sorted by measure (like temperature), loses a good deal of context if proximity to physically surrounding information is not available. OLAP reporting tools have often used graphical representations such as bar and pie charts to represent quantitative information more visually. They have not, however, leveraged mapping technologies to the same degree to help present location-based information in a form that is familiar to us all.

MAPS PROVIDE A NATURAL INTERFACE

The benefits of including geographic maps in an OLAP system do not end with more effective data visualization. Maps provide an equally effective tool for navigating the various levels of location-based hierarchies. Perhaps the most common action to take on an OLAP cube is to drill down or up between levels. If the level is TIME, this might mean drilling from year to quarter to month. If the level is GEOGRAPHY, this might mean drilling from state to county to zip code. Tabular drilling works, but drilling on the actual region of interest is much more intuitive. Assuming the OLAP table is linked to the map view of the data, an exact representation of the measure values accompanies the natural color-coded region.

This same principle applies to other common OLAP navigation tasks like selection, expand, collapse, and filter. If the map extents are recalculated based on the region of interest prescribed by the action, the effectiveness of the data presentation is also enhanced. Here is a visual example of drilling into the Pacific region of the US, and then drilling into California.



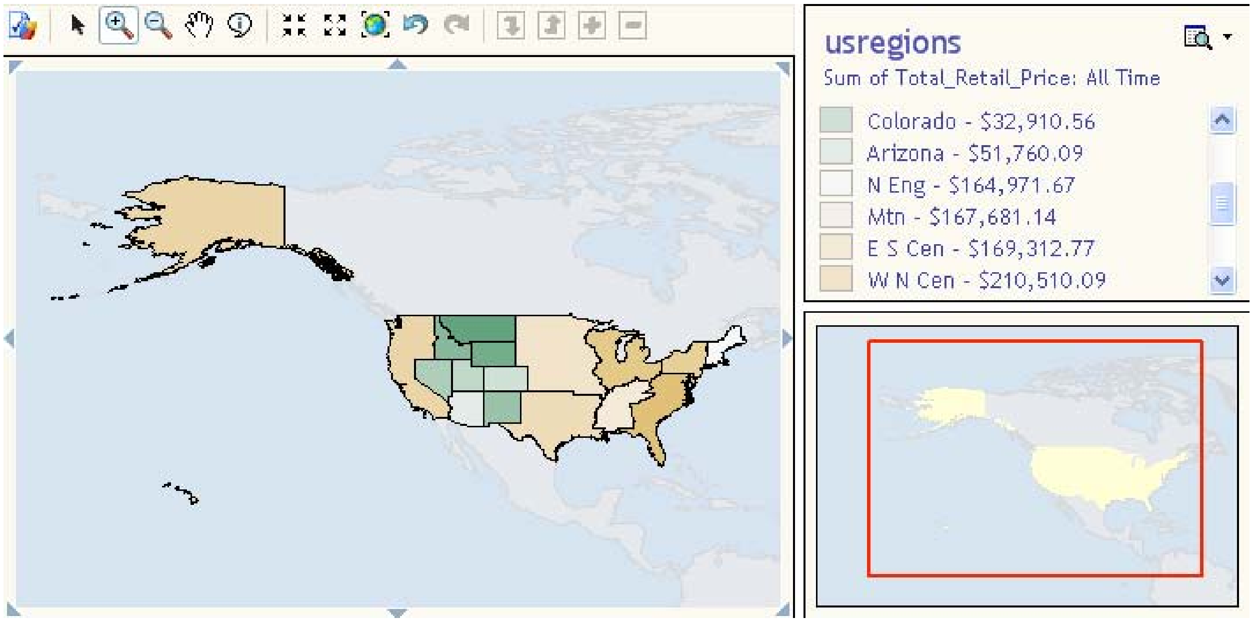
In each step, the map extents are recalculated to only include the region that is shown in the OLAP hierarchy. While there aren't universal terms for defining regional hierarchies within a country, the concept does exist. Whether you are looking at states, provinces, or divisions, the idea of one or more geographic hierarchies within a country is ubiquitous. In business cases, these hierarchies are frequently defined based on business rules. Sales territories may or may not follow political boundaries such as country or state. Shipping regions may be defined by travel time or distance from supplier or any other criteria that is important to a business.

HOW OLAP LINKING WORKS

First, let's consider an OLAP hierarchy which includes levels of REGION, STATE, and ZIPCODE with aggregated total sales at each level. Using traditional OLAP analysis and reporting tools, it would be easy to a see a tabular representation of this data at any of the specified levels. Depending on the OLAP tool being used, it might look like this:

	MeasuresLevel	Sum of Total_Retail_Price
Sub Region	*State*	
E N Cen		$482,838.80
E S Cen		$169,312.77
Mid Atl		$455,884.06
Mtn	**Mtn**	$167,681.14
	Arizona	$51,760.09
	Colorado	$32,910.56
	Idaho	$6,892.70
	Montana	$1,225.50
	Nevada	$25,222.15
	New Mexico	$14,414.68
	Utah	$29,685.80
	Wyoming	$5,569.66
N Eng		$164,971.67
Pacific		$446,858.96

A geographic map representing the same data would look like this:



As you might imagine, a single application attempting to provide a synchronized view of a single OLAP cube would need to ensure the measure value (Total Retail Price, in our example) is linked to the correct map area at each level of the OLAP hierarchy. When this data is managed in an OLAP storage mechanism, traditional OLAP reporting tools use queries to retrieve data values and client-based or web-based tools to display them. Similarly, if the data were extracted and merged with the attribute data of an ESRI layer or shape file, the ESRI toolset provides many options for geographic display. The steps you would take to merge this data into an attribute table are very similar to the steps a single application must do “under the covers” to ensure accurate linking. Namely, a unique key value is needed for each value of an OLAP level which matches with the value of a specific field in an ESRI map layer.

MAPPING CAPABILITIES WITHIN OLAP

While the OLAP-to-Map linking is very powerful, it is often useful to use traditional map navigation and query tools to interact with the map irrespective of the OLAP data. For example, if you are drilling through a cube with STATE/COUNTY/ZIP CODE levels, you may not have any data below the Zip Code level aggregation. You still may want to zoom in further to the map to see census tracts, block groups, or street-level information. The capability to zoom, pan, and turn layers on/off is still available. You would not expect any corresponding changes to the OLAP table view, but the synchronization would begin again when an OLAP event took place (drill, expand, etc.). Also, there may be additional information stored in the map attribute table that is available through an identify action. The OLAP data may contain sales information by region, but the ESRI attributes may include population, income, or any other demographic information. All of the information that would traditionally be available through an ESRI identify action is still available in an OLAP/ESRI enabled environment. This is a very powerful feature that truly extends functionality rather than replacing one set of features with another.

WHERE DO WE GO FROM HERE

The ideas and implementations described here are currently available in products from SAS and ESRI, but they are just the beginning of what is possible. It is hard to imagine an industry which wouldn’t benefit from a tighter integration of mapping and advanced analytics. Marketing Automation, Demand Planning, Supply Chain Optimization, and Fraud Detection are only a few business problems which provide opportunities for this integration. Using drive-time calculations or distance from irregular geographic features (lakes and rivers for water supply, for example) as input to data mining applications adds an extremely valuable component to analytic models which results in better business decisions.

CONCLUSION

Traditional decision support systems, including those using OLAP technology, have always benefited from including graphical visualization techniques like bar and pie charts. Most of these systems also rely on data sources which contain location information, and this information is as critical as any other dimension of the data. The most logical way to understand and interact with this aspect of data is through maps. Because of this, there is a tremendous opportunity to combine OLAP analysis with geographic mapping to unveil new insights into business operations and to make better decisions on future business directions.

ACKNOWLEDGMENTS

The authors would like to acknowledge the contributions of the following colleagues at SAS who have provided valuable feedback and assistance in producing this paper. Their work in the fields of OLAP processing and geographic mapping provided the foundation for the concepts presented here. Lina Clover, Jeff Diamond, Tammy Gagliano, Duane Hamilton, Greg Henderson, Will Ivey, Jim Metcalf, Duane Ressler, Jim Sheedy, Eric Short, Ann Weinberger and Michael Yilma.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Gail Kramer
SAS
SAS Campus Drive
Cary, NC 27513
Work Phone: 919.677.8000
Fax: 919.677.4444
Email: Gail.Kramer@sas.com
Web: http://www.sas.com

Stuart Nisbet
SAS
SAS Campus Drive
Cary, NC 27513
Work Phone: 919.677.8000
Fax: 919.677.4444
Email: Stuart.Nisbet@sas.com
Web: http://www.sas.com