Paper 144-25

SAS/MDDB® Server Software: Expanding Open Access to Your OLAP Data

Duane Ressler, SAS Institute Inc., Cary, NC
Mary Simmons, SAS Institute Inc., Cary, NC

ABSTRACT

With the introduction of the Open OLAP Server component of SAS/MDDB™ Server software, SAS Institute Inc. was one of the first companies to support an industry standard interface for access to OLAP data. Continuing to expand on that work, through the use of SAS® Integration Technologies we will be providing improved access to SAS' OLAP data. This paper will discuss our continuing efforts to provide open access to OLAP data via the SAS/MDDB Server software product.

INTRODUCTION

OLAP (Online Analytical Processing) products have historically used proprietary interfaces to transmit data between an OLAP server and an OLAP client. The introduction of Microsoft's OLE DB for OLAP specification established an industry standard for OLAP data interchange. This allowed many different OLAP and multidimensional tools, both client and server, to work together for the first time.

OLE DB for OLAP extended OLE DB by defining a set of objects and interfaces for accessing multidimensional data. In addition, it introduced a multidimensional query language, known as Multidimensional Expressions or MDX, which could be used to query data from an OLAP system.

SAS Institute Inc. is continuing is efforts with regard to the OLE DB for OLAP specification by combining features of SAS Integration Technologies with the Open OLAP Server component of SAS/MDDB Software to provide options for improved scalability.

WHAT IS SAS/MDDB SERVER SOFTWARE?

SAS/MDDB Server provides high performance OLAP capabilities within an integrated data warehouse environment. The multidimensional database server enables information technology teams to package warehouse data into multidimensional data structures (MDDB's), which deliver data to OLAP client software upon request.

SAS/MDDB Server software is an integrated part of SAS Institute's decision management architecture. This model for online analytical processing (OLAP) enables better decision making by giving business users quick, unlimited views of multiple relationships in large quantities of summarized data. SAS/MDDB Server software provides several options for storing and processing your OLAP data, including both multidimensional OLAP (MOLAP) and hybrid OLAP (HOLAP) solutions.

MOLAP uses a specialized data store, known as an MDDB, with pre-aggregated summaries to store the data. HOLAP bridges the MOLAP and relational OLAP (ROLAP) technologies, allowing you to use the best features of each technology.

WHAT IS AN MDDB?

An MDDB is a specialized data storage facility within a SAS data library that stores summarized data in a multidimensional format for fast and easy access. Users can quickly view large amounts of data as a value at any cross-section of business dimensions. A business dimension can be any vision of the data that makes sense, such as time, geography, or product.

The MDDB is built specifically to handle multidimensional queries and offers fast, efficient, and manageable access to multidimensional data. The MDDB is designed as a single file on the system with indexes built and designed for the best response when doing an OLAP query.

WHAT IS OPEN OLAP SERVER?

The Open OLAP Server enables users to access multidimensional data stored in SAS/MDDB Server software from an open client. The Open OLAP Server supports Microsoft Corporation's OLE DB for OLAP specifications and thus can surface OLAP data stored in the SAS® System to any client application supporting OLE DB for OLAP.

Like most OLAP servers, or providers as they are referred to in the Microsoft specification, the SAS Open OLAP Server will expose a subset of OLE DB for OLAP metadata objects and MDX capabilities. The OLE DB for OLAP specification is intended primarily to surface data to Microsoft specific client platforms, however, the specification allows for providers to connect to multidimensional servers in any platform or architecture. The SAS OLE DB for OLAP provider, running on Windows 95, Windows NT Workstation, and Windows NT Server, is designed to connect to any platform running the SAS System where the SAS/MDDB Server software product is licensed. This portable connection will allow you to easily convert existing applications into a thin client environment.

The Open OLAP Server supports the following consumers:

- Cognos (PowerPlay)
- Knosys (ProClarity)
- Microsoft Corporation (Excel 2000)
- Seagate Software (Crystal Reports)
- Seagate Software (Worksheet)

In Release 6.12 and Version 8, the Open OLAP Server component of SAS/MDDB Server software is based on a socket communication method. In future releases, it will be based on communication technology made available through the SAS Integration Technologies product.

WHAT IS INTEGRATION TECHNOLOGIES?

SAS Integration Technologies is a new SAS system product which enables users to make use of middleware technology standards to access SAS software's decision support, business intelligence, and data warehousing facilities. This allows deployment of thin client, Web, and n-tier distributed applications, as well as client development in commonly known programming languages such as Visual Basic, Java, and C++.

SAS Integration Technologies includes an integrated object model (IOM) that provides access to the SAS software procedural scripting language, data, file system, results content, formatting services, and OLAP data. This object model serves as the basis for access to SAS software's capabilities using both the Object Management Group's CORBA and Microsoft's COM/DCOM middleware technology standards. In addition, because COM/DCOM is limited to the Microsoft Windows platform and CORBA implementations on operating systems supported by the

SAS system are dependent on the installation of third party software products, IOM also includes it's own TCP/IP-based IOM bridge protocol as an additional connectivity option.

In future releases of SAS/MDDB Server software, the Open OLAP Server will make use of the new IOM technology made available through SAS Integration Technologies. This will allow Open OLAP Server to provide additional capabilities in communication protocols and server scalability.

WHAT IS OLE DB FOR OLAP?

Microsoft has a long history of defining data access interfaces including DAO, ODBC, and now OLE DB and ADO. OLE DB is a low level interface for establishing a client/server connection and moving rectangular data through this connection. The OLE DB architecture allows it to be extended to accommodate many types of data, and OLE DB for OLAP represents the first such extension, made to accommodate OLAP, or multidimensional, data.

THE SPECIFICATION

The OLE DB for OLAP specification defines a set of Component Object Model (COM) objects, as well as the interfaces for those objects. The specification addresses both the server, or provider, as well as the client, or consumer. An OLE DB for OLAP provider implements the defined objects and their interfaces. A consumer uses the objects and interfaces to request data from a provider. Because of the clearly defined interfaces in the OLE DB for OLAP specification, an OLAP consumer can request and receive data from any provider implementation. Likewise, an OLAP provider can supply data to any consumer. Because SAS Institute has chosen to implement an OLE DB for OLAP provider, any consumer that complies with the OLE DB for OLAP specification can consume OLAP data from the SAS system.

The OLE DB for OLAP specification is an extension of Microsoft's OLE DB specification. The OLE DB specification describes how a consumer contacts a data source, acquires metadata about the data source, and sends commands to the data source to retrieve specific data. With an OLE DB data source, these commands are typically SQL statements and the retrieved data is usually a table with rows and columns. Because the OLE DB specification can be easily extended to accommodate other types of data, Microsoft defined the OLE DB for OLAP extension to handle multidimensional data.

The OLE DB for OLAP specification describes a new command language called Multidimensional Expressions (MDX). Instead of returning data in a tabular form, these commands return the data in a multidimensional form. To extend the OLE DB specification to accommodate OLAP data, one additional COM object was added to the specification with two new interfaces. The Dataset object represents data returned from execution of an MDX command. Consumers use the new IMDDataset and IMDFind interfaces to navigate the returned data and present it to the user.

SAS Institute has been closely involved in the development of the OLE DB for OLAP specification and made every effort to ensure the specification will support the needs and requirements for OLAP data that are most important to our customers. These efforts include providing important feedback to Microsoft on technical aspects of the specification, as well as working with other software vendors on proof of concept efforts on the OLE DB for OLAP specification. In fact, at Microsoft's invitation, SAS was the first company to demonstrate a working implementation of OLE DB for OLAP and successfully connected to multiple early consumer implementations.

SPECIFICATION PRIMER

In order to begin to understand the concepts behind the OLE DB for OLAP specification, it is important to look at the following topics in more detail:

- Multidimensional schemas
- Multidimensional expressions (MDX)
- Datasets

MULTIDIMENSIONAL SCHEMAS

OLE DB for OLAP provides interfaces for multidimensional data providers (MDPs) to expose a set of multidimensional cubes, or schemas, and for multidimensional consumers to view these schemas. OLE DB for OLAP also provides interfaces that allow users to browse a set of metadata objects for information about multidimensional data.

The primary metadata object defined by OLE DB for OLAP is the cube. A cube is a set of related dimensions, similar to the hierarchies or drillpaths in SAS/EIS® software. These dimensions define a multidimensional data space. Each data point in this data space is uniquely identified by a set of coordinates. Each coordinate is a member of one component dimension.

A cube is made up of a set of dimensions and each dimension consists of a set of members. Just as a dimension is similar in concept to a hierarchy or drillpath in SAS/EIS software, the members of a dimension are the union of the values of the classification variables in that drillpath. In the OLE DB for OLAP definition, the members of a dimension can be aggregated along a hierarchy, with some dimensions having more than one hierarchy. Each hierarchy has levels, equivalent to classification variables in the SAS multidimensional model.

A multidimensional consumer can examine several pieces of metadata from an MDP, including:

- cubes
- schemas, a collection of cubes
- catalogs, a collection of schemas, with no relationship to a SAS catalog
- dimensions
- measures, a special type of dimension
- hierarchies
- levels
- properties (for each dimension level)
- members

For example, consider the ProdSale sample MDDB supplied with SAS EIS software. Examining it from an OLE DB for OLAP perspective, this cube includes four dimensions: Geographic, ProductLine, Time, and the special Measures dimension. Each of these dimensions contains a single hierarchy consisting of all the levels for the dimension. The Measures dimension contains a single level, whose members consist of a combination of the analysis/statistic pairs in the MDDB. The Geographic dimension contains three levels: Country, Region, and Division. The ProductLine dimension contains two levels: ProductType and Product. The Time dimension contains three levels: Year, Quarter, and Month.

MULTIDIMENSIONAL EXPRESSIONS (MDX)

Once a multidimensional data consumer connects to a data provider, the consumer can submit a query to request data from the multidimensional provider. A query may request data from one or more cubes and extracts the data into a single, multidimensional cube. Typically, the cubes requested by a query contain no more than 4 dimensions and are often no more than 2 or 3, depending on the display capabilities of the data consumer on the client. These queries are similar to relational SQL queries that request data from one or more tables and return

the results in a single table.

An MDX query must include:

- One or more cubes that set the scope of the query
- The number of axes
- The dimension(s) projected on each axis and level of nesting at which each dimension appears on the axis.
- The members from each dimension to include on the axis and their sort order
- The members included on the filter axis.

An MDX statement takes the form of:

```
SELECT <axis specification> [,<axis
specification>…]
FROM <cube specification>
WHERE <filter specification>
```

When making a query, the user or application divides the dimensions into two groups: axis dimensions, which specify the dimensionality of the cube and filter dimensions for which data is retrieved for a single member. For example, the following query specifies members of a single dimension on each of two axes and a single member of two dimensions on the filter axis.

```
SELECT {[1993], [1994]} ON COLUMNS,
{[CANADA], [EAST_CANADA], [U.S.A.],
[WEST_U.S.A.]} ON ROWS
FROM [SAMPLE]
WHERE ([PRODUCT LINE:-ALL-],[PREDICT_SUM])
```

Notice that this query specifies each dimension: Time on the columns, Geographic on the rows, and ProductLine and Measures on the filter, or subset, axis.

There is a definite similarity between MDX and SQL In fact, the OLE DB for OLAP specification includes information on mapping MDX queries into SQL queries to allow multidimensional access to relational data.

While MDX in itself is a very powerful query language, the true power of MDX comes through the use of client applications that submit MDX to a provider. These client applications serve to insulate the user from the MDX language, allowing them to query data using a point and click environment.

DATASETS

Execution of an MDX query returns a dataset. Unlike SAS data sets, which store tabular data, an OLE DB for OLAP dataset object returns multidimensional data. A dataset consists of a set of axes, with the points along each axis being called coordinates. If an axis consists of a single dimension, then the coordinates are a subset of the members of the dimension. If an axis contains more than one dimension, then each coordinate consists of a member from each of the dimensions displayed on the axis.

So, looking at the ProdSale query shown above, the returned dataset would look like this (represented in a simple table form):

		1993	1994
Canada		$119,329.00	$113,690.00
	East	$62,426.00	$58,220.00
U.S.A.		$123,763.00	$117,959.00
	West	$60,484.00	$60,651.00

OPEN OLAP SERVER AND INTEGRATION TECHNOLOGIES

In future releases of Open OLAP Server, we will be shifting to use SAS Integration Technologies and IOM as the basis of the SAS server components of Open OLAP Server. The socket communication method will be replaced by a server communication mechanism based on integrated object model, providing greater flexibility in communication and server scalability.

The server side components of Open OLAP Server are currently based on an object-oriented implementation using SAS/AF® software. These server side components are being revised to:

- Use the integrated object model of SAS Integration Technologies as the communication protocol from the server to the client side components of the Open OLAP Server
- Provide faster response in processing metsdata and MDX queries
- Expanding server scalability by making use of additional features of MVA architecture.

Client applications that communicate with the Open OLAP Server following the OLE DB for OLAP specification will be able to continue to communicate with the new releases of the server without modification. This will allow you to transparently migrate clients to the new version of the server in much the same way as you can migrate clients from SAS/EIS software and WebEIS software to new releases of the SAS System.

The initial release of this new Open OLAP Server, which will be used by SAS Enterprise Guide™ software, will not provide support for HOLAP and the Multidimensional Data Provider (MDP) but will support these features in future releases.

CONCLUSION

SAS/MDDB Server software and the Open OLAP Server are providing SAS customers with open access to OLAP data stored in the SAS system. This open access is based on the OLE DB for OLAP specification developed in conjunction with Microsoft Corporation.

The Open OLAP Server component of SAS/MDDB Server software is continuing to provide new capabilities for accessing your OLAP data, while protecting your investment in client-side applications. This continuing work is based on functionality provided by the new SAS Integration Technologies product.

REFERENCES

SAS Integration Technologies Overview, Steve Jenisch, SAS Institute Inc. White Paper

OLE DB for OLAP: SAS Institute Leading the Way, Duane Ressler, Randy Pierce, and Mark Moorman, SAS Institute Inc.

ACKNOWLEDGEMENTS

The authors wish to gratefully acknowledge the assistance of Mark Gass and Steve Jenisch of SAS Institute Inc.

CONTACT INFORMATION

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

Duane Ressler
SAS Institute Inc.
SAS Campus Drive
Cary, NC 27513
Work Phone: 919.677.8000 Ext. 6950
Fax: 919.677.4444
Email: Duane.Ressler@sas.com

Mary Simmons
SAS Institute Inc.
SAS Campus Drive
Cary, NC 27513
Phone: 919.677.8000 Ext. 7751
Fax: 919.677.4444
Email: Mary.Simmons@sas.com