

The Power of Hybrid OLAP in a Multidimensional World

Ann Weinberger, SAS Institute Inc., Cary, NC

Matthias Ender, SAS Institute Inc., Cary, NC

ABSTRACT

Version 8 of the SAS® System brings powerful new features for managing a Hybrid OLAP (HOLAP) or Distributed Multidimensional Data environment. The HOLAP component of the SAS/MDDB® Server software enables you to include SAS Multidimensional databases (MDDB), SAS files, and relational (RDBMS) databases into a single, powerful OLAP reporting environment. Support for HOLAP data groups is fully integrated into a number of SAS Institute's OLAP solutions.

INTRODUCTION

In Version 6, the HOLAP extensions to SAS/EIS® software have become a widely implemented back-end for OLAP reporting applications. In Version 8, SAS/MDDB Server software features the ability to access and retrieve multidimensional data stored in a variety of formats on local or remote servers. One logical MDDB can be stored in many different formats and locations. Working with distributed multidimensional data requires two facilities that are part of SAS/MDDB Server software: the Distributed Multidimensional Metadata (DMM) facility and the Multidimensional Data Provider (MDP). The DMM enables you to define logical MDDBs, while the MDP enables you to retrieve the data. These components include and extend the feature set of the Version 6 HOLAP Extensions. In addition, the HOLAP or MDP component is now fully integrated into a number of OLAP solutions offered by SAS Institute.

WHAT IS OLAP?

OLAP (Online Analytical Processing) is a technology that offers high-performance access to large amounts of summarized data for complex multidimensional analysis and easy reporting. With OLAP, managers and analysts can rapidly and easily examine key performance data and perform complex comparisons on large volumes of data. Until recently, companies have implemented OLAP using two different technologies: Relational OLAP (ROLAP) and Multidimensional OLAP (MOLAP). Each technology has its own strengths and weaknesses.

Relational OLAP (ROLAP) uses relational databases (RDMS) to store the data, sometimes using a star schema or snowflake schema. ROLAP tools offer sophisticated SQL and navigational methods on top of traditional relational databases. Relational OLAP tools offer scalable, manageable technologies for very large data. ROLAP databases can easily handle dimensions with high cardinality. However, the processing required to answer OLAP queries can put a heavy load on the server.

Multidimensional OLAP (MOLAP) uses a specialized data store with pre-aggregated summaries to store the data. The MOLAP data store is built specifically to handle multidimensional queries and offers fast, efficient, and manageable access to multidimensional data. MOLAP databases are one file on the system with indexes built and designed for the best response when doing an OLAP query. The limitation with MOLAP is that it is not as scalable as ROLAP.

Hybrid OLAP (HOLAP) bridges the technology gap between ROLAP and MOLAP, enabling you to use both multidimensional data stores (MDDB) and RDBMS data stores.

WHY DO I NEED HOLAP?

Hybrid OLAP (HOLAP) gives you the best of both the MOLAP world and the ROLAP world. In general, MOLAP and ROLAP can be characterized as follows:

MOLAP (Multidimensional OLAP)

- Fast response time
- Easier maintenance
- Less scalable

ROLAP (Relational OLAP)

- Extremely scalable
- Uses current technologies
- Generally slower
- More maintenance and design

HOLAP supports both MOLAP and ROLAP storage formats, enabling you to mix and combine them. By using both ROLAP and MOLAP data stores, Hybrid OLAP encompasses the strengths of both. It gives you full flexibility in dealing with questions of

- Size and scalability
- Access to existing ROLAP databases
- Query performance tuning
- Storage size reduction
- Non-standard aggregations.

A Hybrid OLAP (HOLAP) approach should not be used for every OLAP application. While HOLAP offers flexibility in designing and accessing OLAP data sources, it does add a layer of abstraction, the metadata, which must be maintained. Using a HOLAP approach can be beneficial if any of the following situations exists:

- if the amount of data is too large to handle within one MDDB
- if there is a performance bottleneck when accessing data on a server
- if you want to make use of existing summarized data sources
- if you want to secure your investment in OLAP applications by
 - providing for scalability as data volumes grow
 - maintaining flexibility in your choice of data storage format and locations.

HOLAP SPECIFICS

MULTIPLE STORAGE FORMATS

The HOLAP or MDP component of the Version 8 SAS/MDDB Server software enables you to treat data in different formats and different locations as one logical unit. In detail, the MDP component can access aggregated data stored in

- MDDBs
- PROC SUMMARY output data sets
- Relational data (SAS data sets and SAS views - including SAS/ACCESS® software views)
- Star Schemas

In fact, HOLAP enables OLAP reporting on any data

sources that can be surfaced within the SAS System.

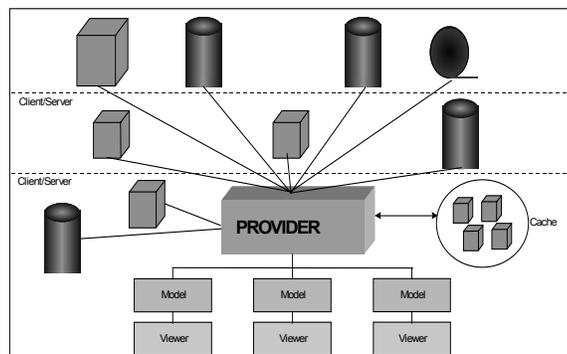


Figure 1 HOLAP Data Model

STACKING

Stacking is the ability to store individual aggregation levels in separate files (and on separate servers, if desired). Stacking can be used to improve performance by storing the most requested aggregations in an MDDB stored on the local network with other aggregations stored in data sets or relational tables on a remote server. For instance, summaries by country, region, product group, product and year are stored in a local MDDB and the monthly summaries are stored in a relational table on a remote server.

RACKING

Racking enables you to store parts of individual aggregation levels in separate files by the partitioning the data across values of one or more dimensions. Racking provides you with the flexibility to design and maintain your OLAP data. For instance, an MDDB is created to contain each month's sales summaries. Every month a new MDDB can be created and registered in the HOLAP metadata. Another example of racking is to have individual country or region summaries stored in separate MDDBs. By defining a HOLAP data group that contains all the MDDBs, you can report across all countries.

REMOTE COMPUTE SERVICES

The remote compute services of the MDP enables dynamic rollups, calculation of derived columns, and application of sorting and cutting criteria. This feature of the MDP shifts some of the processing of the data from the client to the server. Beyond shifting the processing to the server, the remote computed services of the MDP also reduce the amount of data that is transferred over the network.

Dynamic rollup is the ability to calculate, at query time, aggregation levels that are not stored. When the MDP receives a request or query for data, it tries to choose the smallest set of data that will satisfy the request. Often, the smallest set is not an exact match for the request. When that is true, the MDP rolls up or aggregates the next largest subtable or aggregation. The rollup is done on the server and only the query result is sent back to the client.

Derived or computed columns are additional analysis columns that have been defined in the metabase. The values for these columns are not stored in any of the HOLAP group's data sources and must be calculated at query time. Using the MDP, the computed columns are evaluated on the server and included in the query result that is sent back to the client.

Since OLAP applications involve the interrogation of your data, many times you need to filter or sort the data to get the desired results. The MDP sorts and filters the requested data based on the criterion set by the application. This not only reduces the processing that is done on the client, it also reduces the amount of data that is transferred over the network.

These abilities provide you great flexibility in designing, accessing and tuning your HOLAP data groups. It enables you to find an appropriate solution for every scenario you might encounter.

EXAMPLES:

- Instead of just one physical unit, stored on one server, you have the choice of splitting your cube up, providing the most requested subcubes in an MDDB stored on the local network, and less popular parts on a remote server in data sets. You might end up with a greatly reduced storage volume and greatly improved query response times.
- A Star Schema has already been built within a data warehouse. With HOLAP, there is no need to load it into an MDDB. You can assign a SAS/ACCESS software Libref to your star schema files and define MDP metadata to point them.
- As your data grows over time, or changes with organizational restructuring, you can add and remove files and update your MDP metadata instead of re-building your MDDB.
- You have data aggregates that follow non-standard summarization techniques, for example, that cannot be rolled up by the set of statistics provided with the SAS/MDDB Server software. You can store the fully aggregated data in a data set and define a HOLAP data group within the MDP metadata that points to it.
- Dimensions with high cardinality are typically best stored in relational stores. However, small or medium size dimensions are best suited for MDDBs. With HOLAP, you can optimize your data sources by using the most appropriate data store.

HOLAP FEATURES

In addition to the flexibility in designing and tuning your HOLAP data groups, the MDP component of the Version 8 SAS/MDDB Server software provides other features to improve application performance.

CACHING

Caching is the ability to save the results of a query so that it can be reused at a later time. The MDP automatically caches any data retrieved in a query. This avoids data access on subsequent queries to the same aggregation level. You can control the cache by setting the absolute number and maximum size of the queries to hold in the cache. You can also choose to make your cache permanent to persist it between sessions.

LOGGING

Logging creates a file that contains information about each query that is made against the MDP. This file can then be analyzed to help with optimizing the structure of your HOLAP data group. For instance, you may find that more requests require dynamic roll-up while fewer requests are accessing the stored aggregations. On your next update of the MDDBs, you might choose to create different aggregations for those that are accessed frequently and drop the aggregations that are not requested.

EXTENSIBILITY – A POWERFUL DATA MODEL

The HOLAP component is surfaced as a set of SAS/AF® software classes. This object-oriented implementation provides a convenient way for you to extend and modify the data provider's behavior programmatically. The data provider, SASHELP.MDDB.HOLAP_DP.CLASS, has public attributes that enable you to control caching, logging, validation, format handling and many other aspects of HOLAP functionality. Methods are provided to support computed columns, sorting, cutting (top/bottom n), and dynamic aggregation of class values. You can override these methods and attributes to extend and change the functionality of the data provider.

In addition to the data provider class, the actual data access components are individual classes with a clearly

defined interface. The MDP component of SAS/MDDB Server software provides data access components for SAS multidimensional databases (MDDBs), SAS data sets, SAS views and star schemas. You can extend the set of data access components by subclassing the abstract class SASHELP.MDDB.DSOURCE.CLASS and implementing the appropriate methods.

The HOLAP data provider or MDP can also be integrated into your own SAS/AF software applications. The MDP is used for data retrieval. You can request to return the query results as a data set, an MDDB, or as an in-memory SLIST. You must provide a model to track dimensional and navigational information.

HOLAP METADATA

The glue that holds together all the disparate data sources of a HOLAP data group is a set of metadata. You can edit the HOLAP metadata via the **MDMDDB** command or from the SAS/EIS Metabase Window by selecting **Edit > Distributed Multidimensional Metadata**. Definitions of logical cubes are stored in the metadata as data groups. A data group is made up of one or more data sources. You use the Define Data Group window to add or change the data sources in a data group (Figure 2).

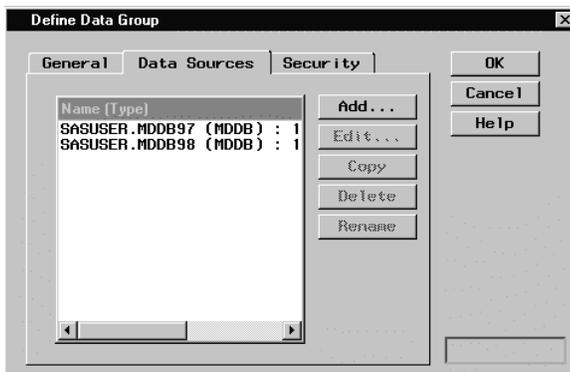


Figure 2 Data Sources for a HOLAP Data Group

Metadata about a specific data source contains the location of the data source, the analysis columns/statistics and the subtables or aggregations (Figure 3). If the data source is password-protected, you can specify the password in this window to avoid prompting the user when the data is accessed at application runtime. If it is a non-local data source, select **Server** to open a window in which you specify an MDP server. For existing data sources, the information about aggregation levels and analysis columns/statistics is automatically filled in for you.

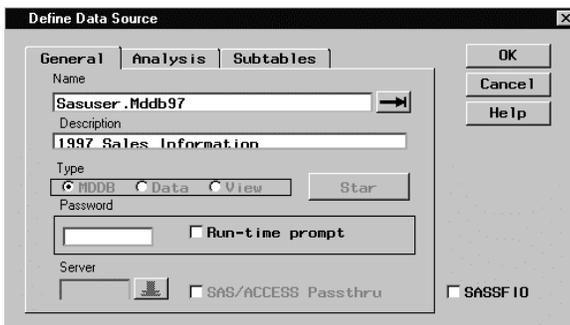


Figure 3 Data source metadata window

After you specify general information about your data source, select the **Subtables** tab to specify or to verify the aggregation levels that your data source contains. A subtable (or aggregation level) is defined by the combination of class variables it represents. If a subtable exists in more than one data source, the subtables will be merged at query time. This enables you to store parts of individual aggregations in

different data sources ("racking").

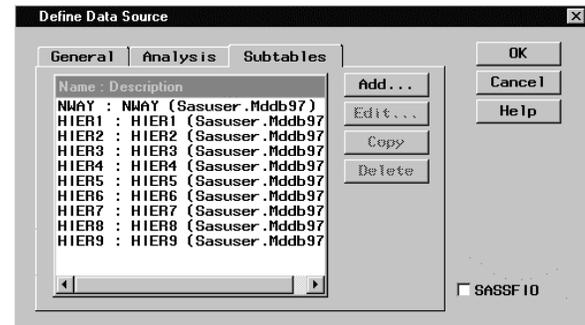


Figure 4 Subtables tab of the Define Data Source window

In addition to the HOLAP Data Groups themselves, you can also define servers within the metadata (Figure 5). Server definitions enable you to not only specify the location of the data servers, but also add post-logout and pre-logout code and options for controlling remote processing load. You can also choose to let the HOLAP data provider control the signing on/off to the servers at query time.

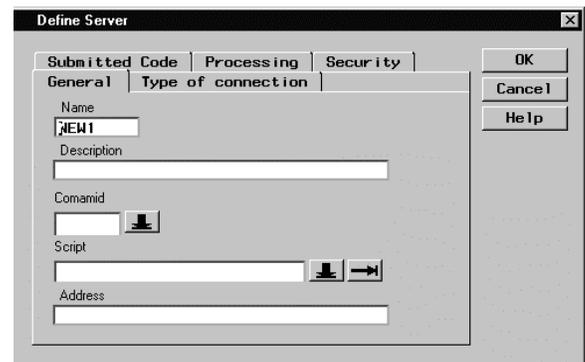


Figure 5 Server definition window

Once you define the metadata for a HOLAP Data Group, you can use it within your own SAS/AF® software applications or you can export it to the SAS/EIS® software metabase for use in SAS/EIS software applications (Figure 6). When exporting, a proxy or skeleton MDDB is created and assigned the HOLAP attribute. Additional attributes, such as colors and ranges, and computed columns can then be added to the proxy registration. (NOTE: This is a different procedure than what was required to use the 6.12 HOLAP Extensions. In 6.12, you added an attribute to an existing SAS/EIS software registration. In Version 8, you define the HOLAP metadata and then export it.)

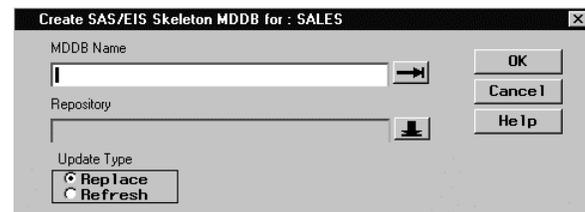


Figure 6 Create SAS/EIS Skeleton MDDB window

With Version 8, there is an experimental batch interface for reading and writing MDP Metadata from and to a flat file, as a means to programmatically update the metadata as data sources are added, updated, or removed. Instructions on how to use the interface can be downloaded from the SAS Institute Technical Support web page (see References).

INTEGRATION WITH SAS SOLUTIONS

In SAS/EIS Software

With Version 8, the multidimensional data model of SAS/EIS software fully supports the HOLAP Data Provider. No additional setup is required other than creating the HOLAP metadata and exporting the metadata to the SAS/EIS software metabase. Once you have created the SAS/EIS software registration, you can build any of the following SAS/EIS software multidimensional reports using the registration:

- 3D Business Graph
- Bubble Chart
- Map
- Multidimensional Business Trends
- Multidimensional Chart
- Multidimensional Pie Chart
- Multidimensional Report
- Organizational Chart
- Graphical Variance Report
- Critical Success Factors (CSF) and Dimension Navigators of the Application Screen Builder.

Additionally, the Setup window in SAS/EIS software provides an interface for customizing the multidimensional environment (Figure 7). Through the Setup interface, you can specify which HOLAP Data Provider class you want to use, either the one shipped from SAS or a subclass that you have created. You can also provide custom classes for the Distributed Multidimensional Metadata user interface.

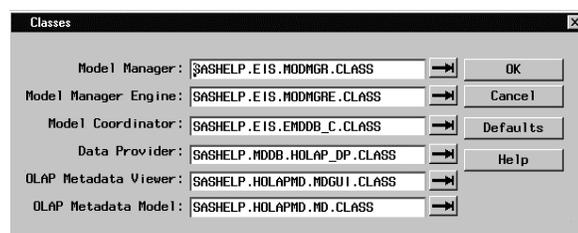


Figure 7 SAS/EIS Multidimensional Environment window

The SAS/EIS software metabase also provides a new table attribute, COMPDRV. This attribute enables you to take advantage of the remote computing capabilities of the MDP without having to define a HOLAP Data Group. The attribute is only available for MDDBs that reside on a remote server.

Finally, by using the Access Control feature of SAS/EIS software with a HOLAP or MDP Data Group, you can control access to the cubes and columns, as well as setting subsets and navigation states.

In MDDB Report Viewer

The MDDB Report Viewer enables users who do not have access to SAS software (or who do not want to invoke SAS software) to view the data in an MDDB. This capability eliminates the need to have SAS software running on all users' machines and provides access to the MDDB reports and graphs in a Web environment. The MDDB Report Viewer provides full support for HOLAP data groups by using the same multidimensional model that ships with SAS/EIS software.

In AppDev Studio™

AppDev Studio is a complete suite of application development tools for building thin-client SAS applications. Two of its core components, webAF™ and webEIS™ provide support for the MDP component of SAS/MDDB Server software.

In Open OLAP Server

The Open OLAP Server grants OLE DB for OLAP consumers access to multi-dimensional data stored in SAS. Consumers can access either SAS/MDDB Server MDDBs or HOLAP data groups.

SAS/Warehouse Administrator™ software now directly integrates more complete support for HOLAP through its new OLAP Groups, OLAP Tables, and OLAP MDDBs. These new OLAP metadata objects provide better and more flexible support for OLAP-specific attributes, such as crossings (aggregations), dimensions, and hierarchies. Export of metadata for all of these new OLAP objects into the SAS/MDDB Server software MDP or HOLAP metadata and the SAS/EIS software metabase is supported. You'll find more information on this in the paper, *New Features in SAS/Warehouse Administrator* (see References).

WHAT'S NEW SINCE THE VERSION 6.12 HOLAP EXTENSIONS

In addition to the coveted remote computing capabilities and the splitting out of individual aggregations ("stacking"), already available with the Version 6.12 HOLAP Extensions, the Version 8 MDP component provides you with the ability to split data sources by class value ("racking" – the combination of stacking and racking is called "stracking"). Version 8 MDP also supports multiple statistics per analysis variable. The MDP component also has the capability of processing derived analysis values ("computed columns"), thus passing local processing of these on to the server.

More generally, the HOLAP component has moved from being an extended attribute of SAS/EIS software metabase registrations to a stand-alone part of the SAS/MDDB Server software and is now supported by many different clients. You can also use the HOLAP data provider or MDP as part of your SAS/AF software applications.

MIGRATION FROM VERSION 6.12

Version 6.12 SAS/EIS software applications using the HOLAP extensions can be moved to Version 8 with little effort. Both the metabase registrations and the applications need to be updated for Version 8.

METABASE REGISTRATIONS

As you convert your Version 6 Metabase into a Version 8 Repository (as described in the Version 8 installation instructions), any registrations with the `_ASSDATA` attribute will be converted as follows:

- A new MDP Data Group will be created in the same repository.
- From this new MDP Data Group, the metadata registration will be updated and a new table attribute, HOLAP, will be added, pointing back to the MDP Data Group.
- You access and modify the information previously stored in the `_ASSDATA` attribute through the Distributed Multidimensional Metadata interface described previously.

You may also need to add server definitions for any remotely accessed data sources.

APPLICATIONS

Applications that used the Version 6.12 HOLAP extensions needed to use an alternative model and, for application screen builders, an alternative resource and viewers. You can use the utility, `sashelp.eistools.happlconvert.scl`, to programmatically convert the attributes of these applications. The syntax for this utility is included in the source code header (see References).

SYSTEM REQUIREMENTS

- SAS/MDDB Server software on each platform that has data accessed by the MDP.

- SAS/MDDB Server software on the client platform (which might be the application server platform)
- SAS/EIS software, if used through an EIS application
- SAS/IntrNet™ software, if using the MDDB Report Viewer
- AppDev Studio, if using webEIS

CONCLUSION

The HOLAP component of the Version 8 SAS/MDDB Server software provides flexibility in designing, accessing, and maintaining HOLAP data groups. HOLAP enables the integration of both multidimensional and relational data stores that can be used in concert to address issues with scalability and performance. It is clearly the most flexible and powerful OLAP solution out there.

REFERENCES

Mark Moorman, *The Art of Designing HOLAP Databases*, Paper 139, Proceedings of the 24th SAS Users Group International Conference

Marge Scerbo, *The ABC of MDDBs*, Paper 121, Proceedings of the 24th SAS Users Group International Conference

Ken Wright, *New Features in SAS/Warehouse Administrator*, Paper 114, Proceedings of the 25th SAS Users Group International Conference

General information on OLAP and OLAP in SAS:

<http://www.sas.com/software/olap/>

Instructions on using the batch interface for reading/writing MDP metadata from/to flat files can be found at:

<http://ftp.sas.com/techsup/download/technote/ts596.pdf>

Information on how to migrate Version 6.12 SAS/EIS software applications to Version 8 can be found by selecting the file, eis_v8_migration.zip, found at:

<http://ftp.sas.com/techsup/download/affsp>

SAS Institute Inc (1999), *SAS/MDDB® Server Administrator's Guide, Version 8-0*, Cary, NC: SAS Institute Inc.

SAS Institute Inc. (1999), *SAS/EIS® Software: Administrator's Guide, Version 8-0*, Cary, NC: SAS Institute Inc.

ACKNOWLEDGMENTS

Thanks to Matthias Ender for providing the basis for this paper.

CONTACT INFORMATION

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

Ann Weinberger
SAS Institute Inc.
SAS Campus Drive
Cary NC 27513
Work Phone: 919.677.8000 x6607
Email: Ann.Weinberger@sas.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.