Hyperion®

CREATING ANALYTICAL DATA MARTS

TABLE OF CONTENTS

Given the pressures of the modern state of business, there is, perhaps, nothing more valuable than time. In respect to time, technology represents an inverse curve. At the same time as Business Intelligence requirements are becoming broader and far more sophisticated; batch processing windows are becoming smaller.

UNDERSTANDING THE VALUE OF TIME

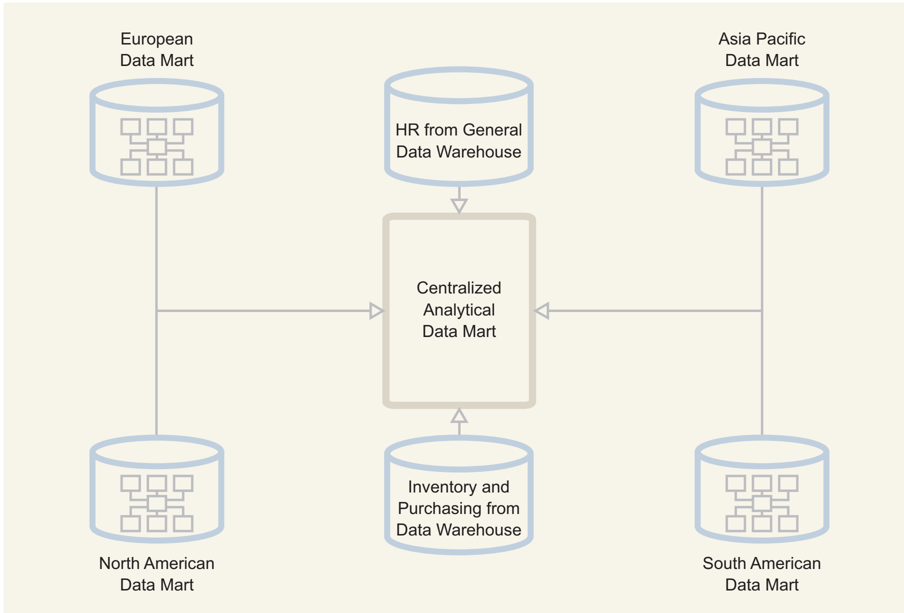
Driving this trend is the paradigm of the more vital and intricate the information is, the lengthier the time to derive it.

Consider the following business problem: you need to provide detailed cross functional access to a myriad of financial data, to a community of global users. The ultimate goal is to expand access to this data to as many users as possible to facilitate better resource management (where can you best spend your money). The crux of the challenge is in providing relevant data in a timely fashion. Realistically, there are two practical Information Technology choices to solve this: a Relational Database (RDBMS) data mart, or Online Analytical Processing (OLAP).

SETTING THE HISTORICAL PERSPECTIVE

When we hear the term "data mart," a certain perspective will usually come to mind. Specifically, a relational database containing a subset of a data warehouse, created for a specific analytical purpose (for example "sales analysis across a specific geographic region"). Compare a generalized definition of On-Line Analytical Processing (OLAP) database: "a database created, generally from a data warehouse, data mart, or other SQL source for a specific analytical purpose."

Given the similarity in the definitions one cannot help but question the logic of having two disparate but similar constructs in the same IT ecosystem. The ongoing deliberations by proponents of both OLAP and Data Warehouse technologies is similar to that of wine connoisseurs arguing as which region of the world produces the best grape. All wines, regardless of origin or type, all have a place in someone's glass. Much like the general distinction between red and white wines, the differential between the classical data warehouse and an OLAP model can be drawn along lines of scalability and functionality.



Data warehouses can hold massive quantities of data, given that the core technology used to construct them is the highly scalable transactionally-oriented RDBMS. For a data warehouse, data feeds are grouped and analyzed through a series of SQL procedures to answer a given analytical question, for example the distinct count and names of 1,000,000 customers buying a combination of a certain three products in a given month. While a query of that nature would not be altogether performant in comparison to a typical OLAP query, the sheer size of the data volume of data presents a problem for a standard OLAP engine. Even a classic *Hyperion® Essbase Analytic Services™* model (widely considered the best OLAP engine available) has an effective limit of 3,000,000 members.

Conversely, allocating material cost overruns across even1% (10,000) of those customers to determine a true cost of sale at the customer level is a metric much better derived in the OLAP engine, both in terms of complexity in defining the calculation and the speed to execute. In the data warehouse, the SQL required and the number of tables across which an SQL statement would have to navigate is far past the boundary of being replicatable.

While there are arguments to be made that given proper tuning either technology can handle either scenario presented, it is commonly agreed that the historical line can be drawn along economies of scale and calculation capability. Current developments with *Hyperion Analytic Services*, however, have succeeded in blurring the line between these two technologies.

With the successive releases of *Hyperion Analytic Services* versions 7X and 9.0, the concept of an Analytical Data Mart has now become a much more practical possibility. There are three specific enhancements in these releases that provide onus for this lien of thought:

- Introduction of Aggregate Storage technology (version 7X)
- Expansion of aggregate storage metadata volumes to 264 bits (version 9.0)
- 64-bit computing support with the Intel Itanium™ processor

Aggregate Storage databases provide the following key ingredients for an Analytic Data Mart to rival, and arguably surpass, the traditional RDMBS-based alternative:

- Virtually unlimited dimensionality and inter-dimensional complexity
- Outlines containing tens of millions of members
- Designed to efficiently handle large volumes of sparse data
- Dynamic aggregations that provide near-instantaneous upper-level data requests
- Highly advanced calculation engine

Recent versions of Analytic Services provide the possibility of creating specialized OLAP data marts fed either by data warehouses or smaller specialized data marts:

In the example above, the Analytic Services database is fed by four individual data marts, each containing regional sales information. In addition, Inventory, Purchasing and Human Resources information is sourced directly from the data warehouse into the analytical data mart.

When dealing directly with the regional sales information, there is a likely probability that the dimensionality across the four data marts is common. As such, users querying the analytic layer will, most likely have similar requirements. However, when we consider the addition of inventory and human resources information the commonality breaks down and the dimensional models will differ both in the granularity of data required but in the number of dimensions. This issue would be exacerbated in the scenario where you have data marts across departments:

- Marketing
- Inventory and Purchasing
- Sales
- Operations
- Human Resources

In this example you would have a large mass of users with disparate analytical needs accessing a single repository. Practical solutions in Essbase Enterprise Analytics for addressing these broad analytical needs of the user base are found in the use of attribute dimensions and metadata security.

ENHANCING TECHNOLOGY

Before jumping directly into the role of attribute dimensions and metadata security in the Analytic Data Mart solution, it is important to credit overall advances in Essbase Analytic Services technology which led to the overall capability. The most significant of these are as follows:

- The introduction of attribute dimensions in version 6.0 of Analytic Services
- The introduction of aggregate storage technology in version 7.0 of Analytic Services, and its ability to facilitate extra dimensionality
- Performance enhancements in the 9.0 product line which facilitate even more dimensional scalability

Other technology enhancements withstanding, attribute dimensions play a special role in aggregate storage databases. The aggregate storage paradigm is optimized for dynamic queries. Each upper level member in the outline, as well as those members containing formulae, is evaluated at query time or potentially upon the request of the Database Administrator (DBA) to pre-execute aggregations. Attributes simply fall into the same dynamic category.

Attribute queries across block storage databases require the dynamic movement of data blocks, causing query response times to degrade significantly. However, attribute queries in an aggregate storage database are no more or less expensive than a standard query across a base dimension.

Performance efficiencies aside, what makes attribute dimensions so useful in a large aggregate storage implementation is the fact that they are optional when issuing queries against the database. In this case, they provide the ability to easily segment your data.

For example, consider a database with six base dimensions:

- Measures
- Year
- Geography
- Customer
- Product
- Sales person

Along with these five base dimensions, consider the following ten attributes:

- Population size
- Average income
- Gender
- Age
- Marital status
- Children
- Sales region
- Salary range
- Discounted product
- Market type

By creating the additional ten dimensions as attributes, you provide access to disparate user queries in a single data source. For instance, Eric from Marketing might regularly query the product mix, in cities of less than 100,000 people for customers who are under 35 and married. Conversely, Allen in Operations) might want to query average net profit, for the eastern sales region, in large markets, where the sales rep makes more than $95,000 base salary. The resultant output grids would appear as follows:

Eric's grid:

	Product Mix	Under 35	Married	All Sales Staff						
		Jan			Feb			Mar		
		Stereos	DVD Players	Televisions	Stereos	DVD Players	Televisions	Stereos	DVD Players	Televisions
Less than 100,000	Union City									
	Saint Cloud									
	Casa Grande									
	Holbrook									
	Needles									

Allen's grid:

	Average Net Profit	Eastern Region	Comp > $95,000							
		Jan			Feb			Mar		
		Stereos	DVD Players	Televisions	Stereos	DVD Players	Televisions	Stereos	DVD Players	Televisions
Large Markets	New York									
	Chicago									
	Boston									
	Dallas									
	Los Angeles									

Both of these queries can be serviced by our example data source, but neither burdens the user with unnecessary base dimensionality as each user only queries on those attributes which are relevant to their business question.

The other aspect to consider in serving a heterogeneous user community is the granularity of the base dimensions. In the above example, Eric (the Marketing user) will want to query the Geography dimension down to the zip code level. However, Allen (the Operations user) will rarely require more than a city name in his queries. This aspect of isolation is handled by Essbase metadata security, where viewing the base dimensions can be restricted uniquely for each user or user role.

In this example, the Geography dimension will look like this for Eric (as shown above):

- Union City
- Saint Cloud
- Casa Grande
- Holbrook
- Needles

And like this for Allen (as shown above):

- New York
- Chicago
- Boston
- Dallas
- Los Angeles

This concept could be taken further on a regional basis, where you would not want to "clutter" one region's view of a dimension with the details of another region. For example, users in Europe will not see the sub-city details in Asia or the zip codes in North America if the Analytical Data Mart designer does not want to allow it. Again, this enables the use of a single large OLAP solution to serve the needs of several user groups, in contrast with the traditional pattern of creating several smaller databases.

TECHNOLOGY ADDING VALUE

Up to this point, we have discussed the potential of an Analytical Data Mart from a purely theoretical perspective. However, we will now take a look at a recent implementation. While the specific identity and dimensionality of the client will remain confidential, the business case around the Analytic Services database supports the Analytical Data Mart concept.

The customer in question had the business problem introduced at the beginning of this document: provide detailed analytic data to a group of global users with disparate analytical needs. The source data for the solution is 20 million rows of relational data. Historically, it took the customer 20-24 hours to run 7 dependant calculations and provide the relevant analytical data using a relational data mart solution.

When considering the value of time, 24 hours introduces latency into the analysis process, and is capital better spent elsewhere. This is what led the customer to an initial Essbase Analytic Services solution. In the first phase, the Analytic Services model contained 8 base dimensions and 12 attribute dimensions and allowed a subset of the user base to perform the analysis they required. To expand the footprint and therefore the overall user base, the client needed to expand the dimensionality. In this case it meant simply adding additional attribute dimensions. From a certain perspective, as you add more attributes to an Essbase Analytic Services model, the more relational it becomes.

The keys to adding the additional dimensionality were the recent enhancements to Essbase Analytic Services mentioned above. As the client took advantage of the advancements in Analytic Services technology they added more attribute dimensions and more users.

The database in its previous state contained 8 base dimensions and 34 attribute dimensions. The specific business case centers on the need to perform financial segmentation analysis for 600 world-wide users, with the ultimate goal being a more effective strategy for the deployment of human and IT capital.

The client's first incarnation was a traidtional OLAP model that contained the 8 base dimensions and 12 initial attributes. With the advent of aggregate storage technology combined with additional enhancements in *Hyperion System 9*™ , the client has moved through a series of build cycles, gradually adding attribute dimensions and by doing so expanding the analytic capability in overall user community accessing the application.

While the sheer scale of the database (considering the dimensionality) is impressive, the wide variance in queries that it handles is even more so. What this application represents for the client is a single source for analytic data. Because the 8 base dimensions are consistent across every query, users with disparate analytic requirements can go to this single repository and completely understand the basic model. This dramatically lowers overall training costs by not requiring a user to learn the varying nature of several data sources. In addition, the ability to house this data in a single analytic application provides opportunity for users to perform analysis that they generally would not consider nor have access to.

On average, users perform queries using 4 of the possible 34 attributes, and the specific business nature of these queries varies by user need. For instance, a user may want to perform regional analysis based on overall profitability and the specific global presence of a particular customer. Another user, for example, might be focused on industry analysis and query against equipment type and delivered energy consumption projections. A third user may be interested in specific coverage analysis and issue a query that is a combination of the two. Of prime importance is the fact that these three distinct users can gather analytic data from a single source, meaning lower IT infrastructure costs, quicker data availability, and broader scope of analysis.

With the advent of aggregate storage technology combined with additional enhancements in *Hyperion System 9*™, the client has moved through a series of build cycles, gradually adding attribute dimensions and by doing so expanding the analytic capability in overall user community accessing the application.

DATABASE CHARACTERISTICS

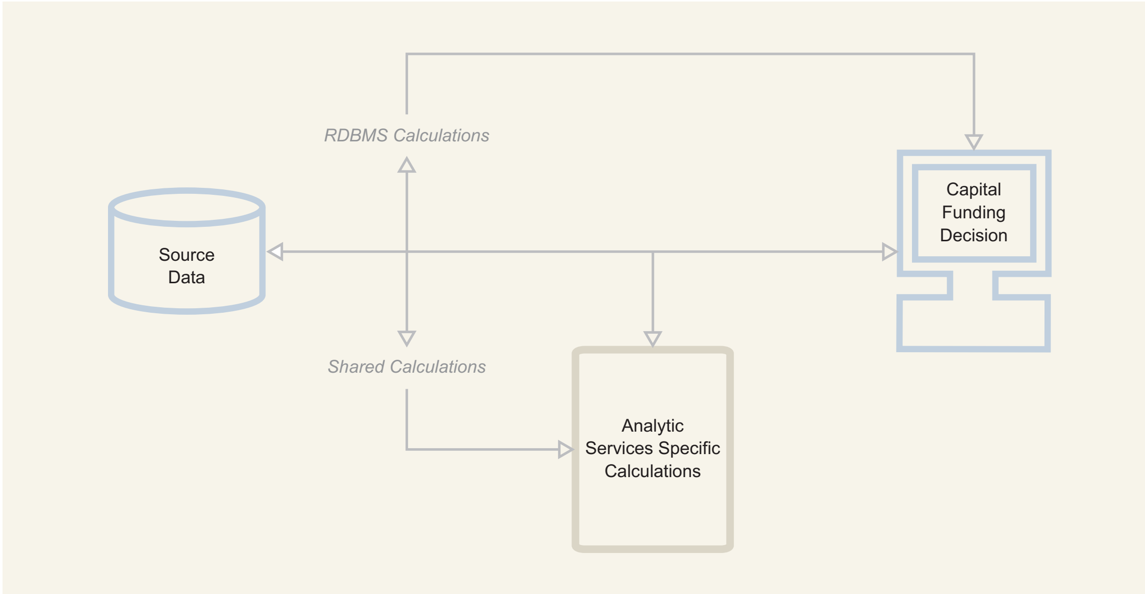
The Analytic Services implementation specifics are as follows:

- The Analytic Services database is fed from a relational database that was built in-house over a period of 50 person-months.
- The Essbase Analytic Services database was created in-house over a period of 10 person-months.
- The Essbase Analytic Services database is aggregate storage.
- Ongoing relational maintenance requires 2 FTE's in support and an additional 2 in development.
- Ongoing Analytic Services maintenance requires half an FTE in support and half an FTE in development.

From an overall implementation perspective, the analytical application took significantly less resources to create and maintain. Of course, much of this can be credited to the initial work put into the relational source, without which the solid foundation the Analytic Data Mart would not exist. What is important to consider, however, is the fact that to add deep and broad analytic capability to the environment took only one-fifth as much time and to create and requires 75% less resources to maintain. The incremental cost is minimal and the overall positive impact very large.

SELLING THE VALUE OF TIME

Simply put, there is nothing more valuable than time. In the aforementioned scenario, the current standing is a combination of the relational data mart and an evolving OLAP data mart. From a graphical perspective, the data delivery processs can be represented as follows:



The source data is moved through a series of sequential stages. First, there are the RDBMS and shared calculations to prepare data for analysis in Essbase Analytic Services and for general distribution for production reporting. Next data is calculated in Essbase Analytic Services to derive advanced analytic values and perform capital expenditure analysis. The data is then moved to the decision maker.

The incredible value of analytics, be it a data mart implementation, an OLAP implementation, or a combination of the two cannot be disputed. The ultimate goal is to supply the analytic data to your user community in the most efficient way possible. The introduction of aggregate storage technology and additional advancements with *Hyperion System 9 BI + Analytic Services*™ provides a vehicle to provide both the broad access indicative of a data warehouse and the raw analytic ability of the an OLAP solution.

In this instance, the shared calculations are a given, certain groups of users and secondary systems need the staged data at this point. The portion of the data delivery process that takes the most time is the RDMBS calculations that are prepping data for Analytic Services. As we add more capabilities into the Analytic Services model (through the addition of attributes and calculation functionality) we eliminate a portion of this flow and effectively shorten the time it take from data collection to capital decision, in a very true sense, selling the value of time.

PRODUCT INFORMATION TEL: 1.800.286.8000 (U.S. ONLY)
CONSULTING SERVICES E-MAIL: NORTHAMERICAN_CONSULTING@HYPERION.COM TEL: 1.203.703.3000
EDUCATION SERVICES E-MAIL: EDUCATION@HYPERION.COM TEL: 1.203.703.3535
WORLDWIDE SUPPORT E-MAIL: WORLDWIDE_SUPPORT@HYPERION.COM

PLEASE CONTACT US AT WWW.HYPERION.COM/CONTACTUS FOR MORE INFORMATION.

Hyperion®

HYPERION SOLUTIONS CORPORATION WORLDWIDE HEADQUARTERS
5450 GREAT AMERICA PARKWAY, SANTA CLARA, CA 95054
TEL: 1.408.744.9500 FAX: 1.408.588.8500

© 2005 Hyperion Solutions Corporation. All rights reserved. “Hyperion,” the Hyperion “H” logo, “Business Intelligence” and Hyperion’s product names are trademarks of Hyperion. References to other companies and their products use trademarks owned by the respective companies and are for reference purpose only. 5434_1005