Analytic Applications – The New Kids on the Block

In my April 2001 column, "Techno Wave – Mass Customization" I explained the development of customizable analytic applications by several software vendors. These new BI applications supply a rich environment, which includes "solutions" for 60 to 80 percent of generic business needs such as key performance indicators (KPIs), sales analyzers and retail analytics. The remaining 20 to 40 percent may be created through customizable points in the design and implementation processes.

As I described in that article, these applications have a number of advantages as follows:

- They are based on the many years of experience in providing the same kind of functionality.
- They reduce the costs of delivering the functionality.
- They are based on a standardized template, encompassing best practices across many industry types.
- They lower the barrier to entry into the BI world by making it easier for companies to acquire and leverage BI.

These vendors are in the process of taking something we have been doing for years and finding ways to do it better, cheaper and faster. The drivers behind their creation are improving productivity, decreasing costs and reducing waste in terms of time, resources and money. This is optimization at its finest. Examples of vendors offering these useful applications are Microsoft's Accelerator Toolkit, Hyperion's Business Performance Management Applications, Informatica's Business Analytics, SAS' Risk Management or Strategic Performance Management and Sand Technology's Analytical Server.

This column will focus on how to integrate these applications into your BI environment.

It should be noted that these applications, while useful and beneficial, do not replace or eliminate the need for an enterprise architecture. To show you what I mean, let's follow a natural progression that is common today in the enterprise resource planning (ERP) arena as an example of what happens without any architecture.

ERP systems such as SAP, PeopleSoft, J.D. Edwards and Oracle are similar to analytic applications in that they optimize the decades of operational system implementations that we have under our belts. These ERP systems promised similar benefits as those listed earlier. They are fully consistent and integrated sets of applications and data, as long as you stay within the boundaries of the systems. To complicate things, each ERP vendor had its own way of dealing with business processes such as order entry, billing, general ledger entries, human resource management and the supporting data and data relationships.

Many organizations took the best-of-breed approach in choosing their ERP systems. For example, they implemented SAP for their manufacturing processes, PeopleSoft for their HR requirements and Oracle for their financials. What's more, many organizations did not implement a single instance of each ERP system. Instead, they had multiple instances (each one different from the other) of SAP, PeopleSoft and Oracle.

While somewhat better than completely fractured operational systems, these companies had to create interfaces between their ERP applications, and then perform extensive and sometimes difficult ETL processes to populate their data warehouses.

What about the analytic applications? Each of these applications is integrated and consistent within its boundaries as well. Many have what they call a data warehouse or some form of staging area feeding their various multidimensional data marts or cube sets, which is in line with the Corporate Information Factory.

However, I am starting to see a pattern of implementations similar to that of the ERP systems. Companies are picking the best of breed among the analytic application vendors for their various BI needs – one vendor for KPI generation, another for sales and marketing analyses and yet another for financial analyses. While this is certainly one implementation approach, it may lead to problems down the road.

Figure 1 illustrates the ultimate resulting architecture that would include:

- Redundancy of source extractions, eventually having a very negative impact upon these critical operational systems.
- Inconsistency across the multiple integration and transformation processes. What is to keep each instance from extracting the data at different times, using different
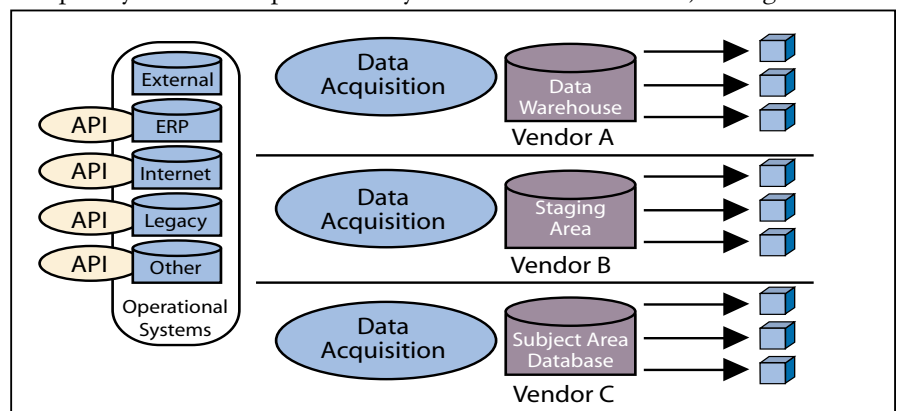

Figure 1: Analytic Applications from Various Vendors

integration and transformation logic, creating different aggregations or calculations, etc.?

- Inconsistency in informational reporting across the various marts, i.e., numbers in one report don't match similar numbers in another vendor's set of marts.
- Inconsistent or spotty meta data which is in accord with the differing ETL processes.

Keep in mind that these "data warehouses," "staging areas" or "subject-area databases" are largely multidimensional in design. This precludes their use as sources for nondimensional analyses such as exploration, statistical or some data mining capabilities. I would say that they are more like "super marts" supporting a defined set of smaller data marts.

What should you do? Obviously these applications offer significant benefit, but how do you reap the benefits without also obtaining the baggage of a poor architectural design?

You may not be surprised to hear that I favor using the Corporate Information Factory architecture (see Figure 2) as the starting point when implementing these applications. If you create your own version of this flexible and proven architecture and then develop a set of standards that supports the proper placement of these applications, you will be able to enforce consistency across various analytic application offerings.

In any event, your architecture must include a warehouse design that supports all forms of analytic functionalities (multidimensional, statistical, exploratory and whatever comes next) and must ensure that:

- The "heavy lifting" of ETL (extraction, integration, transformation, data cleansing, referential integrity checks, data error handling and so on) occurs only once, making the creation of these analytic applications much easier. Then the implementation phase of the applications can focus exclusively on data delivery rather than worrying about the sources of the data and all that entails.
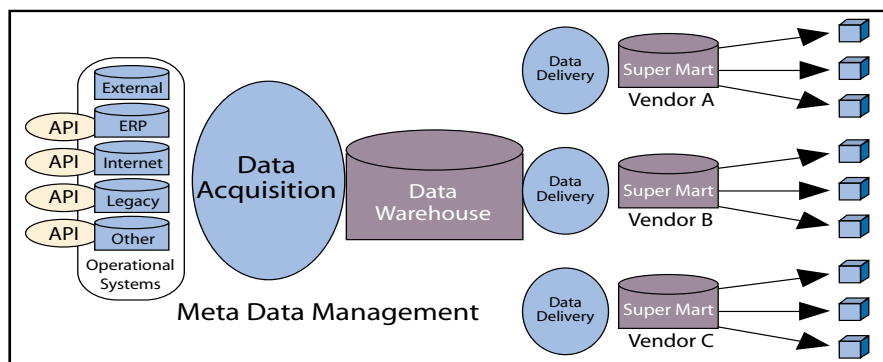- The impact on the operational systems is greatly reduced. We perform the

Figure 2: The Corporate Information Factory and Analytic Applications

ETL process only once *and* there is a clear and distinct separation between data acquisition and data delivery.

- The reconciliation of various vendor analyses is possible and performed quickly. The degree of separation between each set of applications is only one – you only need to go back to the data delivery process to determine why there are differences between the analytic results.
- Perhaps most importantly, many different forms of BI analyses can be supported from the same source of data (the warehouse). By creating a more "normalized" data model (in the Codd and Date sense) for the warehouse, you have not eliminated any significant forms of analyses by biasing the data format. Rather, you can now support multidimensional analysis, data mining, exploration, statistical analysis and even simple querying and reporting capabilities. You can satisfy your entire BI community's needs.

For standards, I suggest that the XML for Analysis standards supported by Microsoft, Hyperion and SAS may be a good start. It is my hope that the marketplace will eventually move in the direction of some form of clearing house for the vendors, making exchange of data a breeze.

Second, recognize the limitations of these applications. They do not answer all your analytic needs, but they do give you significant functionality within their domains. You will always need analyses that are not found within the data or technology supported by one vendor.

Finally, make sure that the vendors you choose have the following:

- A philosophy that supports the con-

cept of the Corporate Information Factory or another architectural road map of your choice.
- Open interfaces to *and* from their "super marts" to other data marts and various access tools.
- Thorough documentation that contains, at a minimum, implementation tips and techniques (in particular, tips for working with other vendors' applications or a separate data warehouse), best practices for the realization of maximum value, the data model(s) used in the creation of the super marts and ultimate marts, audit trails and quality checks as the data moves from the warehouse to the super mart to the mart and so on.
- A flexible, customizable environment in which you can modify the analytic applications to more closely match your own rather than one in which you are forced to match the vendor's "ideal" set of business processes.
- Adherence to the leading standards for analysis.

I am very heartened by the progress made to date with the various analytic applications. The vendors have done a very nice job of optimizing what we have learned over the past decade or more in BI. The benefits from this optimization are certainly very clear; and, as long as they're built upon a strong foundation, they will deliver on their promises. It remains your job to ensure that there is a proper foundation upon which you can implement their offerings. **DM**

Claudia Imhoff is the president of Intelligent Solutions, Inc. Imhoff is a popular speaker and internationally recognized expert on the Corporate Information Factory, business intelligence and CRM. She has coauthored four books and more than 40 articles on these topics. She may be reached at cimhoff@intelsols.com.