



Just Plug in the Data Appliance and GO!

Many of us have experienced the rapid growth and acceptance of our data warehouse and business intelligence environments. The environments have quickly acquired a must-have, mission-critical status, making their performance a significant factor to be considered.

How do you improve or maintain the performance this critical component of your enterprise's strategic decision-making capability? There are four possibilities:

- Limit/remove certain functionalities.
- Limit/remove the depth or breadth of the information available for analysis.
- Increase the hardware/software horsepower.
- Adopt a new paradigm for the technological solution; use a data appliance.

The first two alternatives are impractical in today's competitive world – restricting the business community's ability to perform certain analyses (e.g., data mining or exploration) that require massive amounts of data is simply not acceptable. Business problems today have big data requirements that cannot simply be ignored.

Competitive companies have no choice today; they must be able to analyze detailed data efficiently and effectively with reasonable response times. The response time is mandatory because the next query will be based on the results of the first query. Stale results caused by nonresponsive systems are useless or, worse, misleading.

That leaves the last two alternatives: loading up on traditional technologies or moving to a completely different model for BI technology – the data appliance.

Let's look at each of these approaches, paying particular attention to the new paradigm – the data appliance.

Many companies have implemented their BI environments using general-purpose hardware (e.g., IBM,

Sun, HP, etc.) and DBMS software (Oracle, SQL Server, DB2, etc.). These combinations can be further clustered to improve performance, using various forms of configurations for high-end computing (e.g., symmetric multiprocessing, massively parallel processing, and non-uniform memory access).

For many, the solution to increasingly poor performance and unhappy users in the traditional approach is to add more machine power at the problem. This usually takes the form of:

- Purchasing more CPUs and memory.
- Purchasing bigger, faster disks and controllers.
- Purchasing upgrades to servers and new DBMS versions.
- Replicating or redistributing databases across multiple platforms.

With each new purchase, there are additional maintenance costs and associated administrative expenditures that ensure the ongoing performance of the new components. Fortunately, most enterprises standardize on one or possibly two of the platform combinations. This standardization can help mitigate overall upgrade and resource costs; however, even with standardization, these costs can become quite eye-catching.

Doug Laney, vice president at META Group, states, "Throwing incremental hardware and DBMS software at analytic performance problems may be not more than an expensive stopgap solution."¹ He continues, "Incremental or add-on solutions to address business performance frequently cannot provide the ROI that alternate/specialized approaches do."

One alternate/specialized approach is a data appliance. To explain a data appliance, a general definition of appliances is helpful. The *American Heritage Dictionary* defines an appliance as, "A device or instrument designed to perform a specific function, especially an electrical device, such as a toaster." The inner workings of the appliance

are irrelevant to the ultimate user. The bottom line is that an appliance is designed to do one thing very well.

My favorite analogy of a data appliance is a stereo amplifier. Most of us have no idea how it boosts performance and no clue as to the inner workings inside the metal box. More importantly, we do not care. An amplifier is a black box into which we plug other stereo components (DVD, CD, video player, etc.). We do not configure it. We do not have to tweak it once everything is plugged in. Additionally, if we want to upgrade the amplifier, it is a simple process of buying another one and dropping it into the slot where the old one resided. Talk about plug-and-play!

The characteristics of a good appliance are that it: is transparent to the user; yields an obvious performance boost; is easy to install and administer; and has very low maintenance costs.

Such a device comes about as a result of the maturation of the technology. The amplifier of today is the result of years of research and development and was built with a specific purpose in mind. This "purpose-built" specification means that it has been thoroughly optimized to perform a well-defined and documented purpose.

This appliance was also made possible by the creation of many electronic industry standards. Things plug into it, and it can be easily replaced because of these standards. By taking advantage of mature industry standards, a purpose-built appliance such as the amplifier can easily integrate into anyone's home entertainment center.

Finally, an appliance such as the amplifier is much cheaper to purchase than the individual pieces that would be required if we were to build the appliance ourselves. The synergy of controlling all the "stuff" in the black box and configuring it for user has greatly reduced the overall cost of the appliance.

Appliances in the computer world

have also been around for many years. In fact, they are so ubiquitous that we often don't realize we are using them. An appliance in our world is defined as an integrated box that can retrieve information at the request of external applications. Similar to the amplifier, its inner workings are hidden to maintain simplicity and ease of use. Examples of purpose-built technology appliances are the network router, hubs and switches. These devices require minimal configuration, are basically plug-and-play and greatly increase the performance of network transport, but are relatively transparent to the user.

Now let's turn our attention to BI data appliances. We are fortunate today to have more than a decade of building BI environments behind us. Just as in the electronics industry, we developed standards for use in creating BI systems, including proven and stable architectures such as the Corporate Information Factory (CIF).

In addition to standardized architectures, we also have our own mature industry standards such as ODBC, JDBC, XML and SQL. The confluence of these two factors, a solid architecture and established industry standards, has made it possible for a viable data appliance to be built for the BI environment.

The data appliance for BI is a purpose-built database machine specifically used to manage analytical data and retrieve the results from massive data analyses with impressive performance – a matter of seconds or minutes instead of hours or even days. The data appliance for BI is a combination of hardware, software, DBMS and storage, all under one umbrella – a black box that yields high performance in both speed and storage – making the BI environment simpler and more useful to the users.

Well designed data appliances use the best aspects from both SMP and MPP configurations, incorporating them into a new technical architecture that processes queries in the most optimal fashion possible. The result is a streamlined environment that is efficient, inexpensive, and simple to use and maintain.

As stated earlier, there is great demand for in-depth, complicated

queries requiring massive amounts of data. These types of queries yield more sophisticated insight and business intelligence than the simpler “slice-and-dice” queries supported by multi-dimensional databases and designs. With a data appliance created specifically to support very large data warehouses, very complex queries and sophisticated analysts, we can now satisfy these critical business needs easily.

It also appears that these data appliances mimic their toaster cousins by adhering to the same set of features as any appliance:

Transparent to the user: The administrator of this black box does not need to constantly optimize or tweak the separate pieces. The appliance comes specifically tuned for the BI environment.

Yield a measurable performance boost: Benchmarks comparing traditional infrastructures against data appliances have demonstrated a considerable improvement in performance from the data appliance results. Complex queries are run orders of magnitude faster using the data appliance environment.

Easy to install and administer: Similar to a generic appliance, the data appliance for BI was designed to do one thing well. It has been optimized for that purpose and is highly efficient. There is nothing to “manage.” Secondly, it increases the plug-and-play nature of the overall environment. The data appliance should be compatible with all common open-source database and operating system standards, making it easy to integrate into any existing BI environment.

Low maintenance costs: With the amount of data in data warehouses easily doubling every year, an IT department either has to double its staff or double its productivity every year – neither approach is welcomed. A significant reduction in the total cost of ownership can come from the simplicity of a data appliance for BI, not just the cost of the products. This simplicity translates into a reduced requirement for systems and database administrators.

The challenges to this new technology can be classified as:


Resistance to the new paradigm: Administrators of BI environments

who are accustomed to constantly tuning and tweaking the technology will need to make a similar leap of faith. The belief that this black box is fully optimized without human intervention requires a new way of thinking.

Limited offerings from vendors: There are few vendors offering such data appliances for BI. The best-known today are Teradata and Netezza. Certainly, the strides made by these two companies in furthering the industry's understanding of data appliances have been considerable. Competition should be welcomed, knowing that it leads to lower prices, accelerated technology advances and a greater diversity of approach.

Point of failure: The data appliance's unique integration of hardware, DBMS software and storage shares the same problems that other architectures do in that it has points of failure in a BI environment. This can be overcome by mirroring or clustering the technology and the usage of a SMP and MPP architecture.

The double-edged sword of standards: The use of ANSI SQL standards has the benefit of making the data appliance fully compatible with existing BI tools and applications, as mentioned earlier. However, it also means that access may be slower than the native formats for each interface. You should look for data appliances using drivers such as ODBC or JDBC so that the interface is fast enough to make the limiting factor only the network connection or the speed of the client machine.

Data appliances are becoming a viable option as a part of a coherent enterprise BI program. They are popular today due to their effective means of delivering critical intelligence along with the key advantages of simple installation and troubleshooting ease. The benefits of this simplicity, lowered costs and performance boosts should not be ignored. 

References:

1. “Accelerating Analytics: Alternatives Abound,” November 6, 2002, META Group.

Claudia Imhoff, Ph.D., is the president and founder of Intelligent Solutions (www.IntelSols.com), a leading consultancy on CRM and business intelligence technologies and strategies. She is a popular speaker and internationally recognized expert and serves as an advisor to many corporations, universities and leading technology companies on these topics. She has co-authored five books and more than 50 articles on these topics. She may be reached at cimhoff@intelsols.com.