Data Warehousing: Building the Foundation

By Dr. Ramon Barquin

President of Barquin International
&
BeyeNETWORK Government Expert

Data Warehousing: Building the Foundation

By Dr. Ramon Barquin

For too long, many enterprises have been data rich and information poor — technologically condemned to be information mazes. Data warehousing promises to change all that by becoming the centerpiece of new information architectures.

However, how can the promises of the hardware and software providers be put into proper perspective? How can organizations decide whether data warehousing is a real potential solution to their problems or just the latest fad in an industry that produces one every month? These are tough questions to answer, but they are at the heart of the problem that many enterprises face.

Where to start? First of all, surveys tell us that about 90 percent of Fortune 500 companies are currently engaged in some form of data warehousing activity or are soon planning to be. Second, within the federal government, there are numerous initiatives already under way at such agencies as the Department of Transportation, the Postal Service and the Federal Aviation Administration. The National Science Foundation already has a full production data warehouse in operation.

The truth is that many organizations have been moving toward the creation of data warehouses without fully realizing it. The typical problems that data warehousing tries to address — multiple entry points to data, lack of integrated systems, ambiguous and multiple definitions, and the need for analytical processing that doesn't disturb operational systems — have been hounding management for some time.

Over the last few years, some organizations have been developing different types of database constructs for analytical purposes, usually by extracting data from their legacy systems and placing them in separate storage with its own distinguishing characteristics. Voila! Data warehousing.

These approaches may or may not fit the classical definition, but they certainly try to provide the type of decision support environment that is characteristic of the practice.

The true difference between where data warehousing is now and the discrete attempts of the past lies in the proliferation of new and powerful tools in nearly every relevant area of the process. Hence, today you can rely on several excellent data extraction, cleansing and transformation tools that substantially reduce the pain of loading a data warehouse.

Likewise, we have seen a number of solid new tools emerge that move, build and manage meta data repositories. The advances in DBMS capabilities, especially in bitmap indices, have been substantial. Intelligent storage systems have started to appear in the open systems world with a strong positive impact. This induces the emergence of powerful data mining techniques and the appearance of relational online analytical processing, or ROLAP tools that can obtain multi-dimensional views from relational databases.

In addition, the new tools make the production process much less laborious. While we are not yet at the point of ordering a shrink wrapped data warehouse from a catalog. We can now plan design and build data warehouses knowing we will have a full set of appropriate tools to do it.

Many organizations, in their pursuit of the newest and latest technology, often run the risk of putting the cart ahead of the horse. They start to choose tools and build data warehouses without first having done the necessary homework to ensure that they don't just wind up developing a brand new layer of potentially incompatible stovepipes.

The key issue for most organizations is to take stock of where they are now, and then decide on a data warehousing strategy. The strategy should be developed by understanding the following domains:

- The business domain. What are the basics of your business? How is it structured? What kind of information do you need for decision making?
- The data domain. What data does your organization collect? How is it stored? Who owns it and what is its quality? What formal databases do you have?
- The information system domain. What does your IS environment look like? What platforms, languages, and protocols exist? What kind of information security do you have?
- The decision support domain. Is there an executive information system for your organization? Do your end-users understand basic decision support system concepts? What decision support tools are in place?
- The people domain. Who are your end-users? Are they computer literate? What is their level of training? Where are they located?

Unless an organization is extremely complex, this exercise can usually be completed in 30 to 120 days. The process should create a better understanding of the infrastructure and decision making priorities that will drive your data warehousing effort.

About the Author

Dr. Barquin is the President of Barquin International, a consulting firm, since 1994. He specializes in developing information systems strategies, particularly data warehousing, customer relationship management, business intelligence and knowledge management, for public and private sector enterprises. He has consulted for the U.S. Military, many government agencies and international governments and corporations.

He had a long career in IBM with over 20 years covering both technical assignments and corporate management, including overseas postings and responsibilities. Afterwards he served as president of the Washington Consulting Group, where he had direct oversight for major U.S. Federal Government contracts.

Dr. Barquin was elected a National Academy of Public Administration (NAPA) Fellow in 2012. He serves on the Cybersecurity Subcommittee of the Department of Homeland Security's Data Privacy and Integrity Advisory Committee; is a Board Member of the Center for Internet Security and a member of the Steering Committee for the American Council for Technology-Industry Advisory Council's (ACT-IAC) Quadrennial Government Technology Review Committee. He was also the co-founder and first president of The Data Warehousing Institute, and president of the Computer Ethics Institute. His PhD is from MIT.

Dr. Barquin can be reached at rbarquin@barquin.com.

Editor's note: More articles from Dr. Barquin are available in the BeyeNETWORK's Government Channel.