

Data Warehousing: A Moderated Panel Discussion

Summary of discussion

Moderator: Robin Way, NW Natural, Portland OR

Introduction

Rationale

The Data Warehousing panel represents a new session format for the PNWSUG annual conference. In contrast to traditional paper presentation sections and plenary sessions, the moderated panel discussion allows for more give-and-take between panelists and audience members. Desired outcomes from this format are a different type of education and learning taken away by the audience, and greater insights and feedback taken away by the panelists.

Data warehousing was a logical choice for the context of this panel discussion. Quite a few large and medium-scale organizations in the Pacific Northwest are engaged in data warehouse implementation, and many more are involved in the planning and design phases. This discussion drew a sizable audience and healthy discussion. The PNWSUG conference organizers plan to repeat this format in their 1998 conference.

The discussion was organized around three themes:

1. Tracing the historical roots of data warehousing
2. The application of data warehousing in contemporary organizations
3. The likely futures of data warehousing as a driver for business success

The flow of discussion consisted of 15-minute prepared remarks by each panelist, followed by approximately one hour of questions from the floor and from the moderator. The moderator's role is to facilitate, rather than lead the discussion.

Panelists

Paul Oldenkamp has worked at the Boeing Company since 1989 as a SAS programmer and internal consultant. He has been a SAS user since 1985. Currently he is working on a prototype data warehouse for Airplane Safety analysis.

Mark Thompson is founder and president of Forefront Economics. An agricultural and natural resource economist by training, Mark has experience with econometric modeling and forecasting in the fisheries, transportation and utilities industries. As Manager of Economic Forecasting at Union Pacific Railroad, he was responsible for tracking and forecasting daily and monthly rail traffic in several products. He joined Portland General Electric in 1987 where he was responsible for forecasting sales of electricity and a variety of market research projects. Mark founded Forefront Economics in 1993, a consulting firm specializing in information management, and statistical and econometric analysis. Forefront Economics is a SAS Quality Partner located in Beaverton Oregon.

Randy Betancourt is a Systems Support Manager for SAS Institute. He provides consultation to Information Systems departments of Fortune 500 organizations. These discussions include considerations for client-workstations, network and database design for decision support applications. Previously, Randy was Program Manager for data warehousing. In this role he worked with end-users to provide feedback into software engineering. He also worked with analysts and industry consultants to articulate SAS Institute's direction in data warehousing. Randy has been with SAS Institute for 13 years and has been programming decision support applications since 1980.

Robin Way, who served as the discussion moderator, has 13 years of experience with SAS software and the SAS user community, and serves on the Executive Committee of the Pacific Northwest Regional SAS Users Group

(PNWSUG). Mr. Way is also Director of Market Research for Northwest Natural Gas Company, a natural gas distribution company serving 450,000 customers in northern Oregon and southwestern Washington. He is accountable for all of the company's market planning, forecasting, and research activities and provides leadership in the development and integration of marketing information systems.

Prepared remarks

Paul Oldenkamp

Paul traced the roots of data warehousing back 15 years, beginning with an influential article in Harvard Business Review on the topic of Executive Information Systems. The article's influence eventually led Paul to presenting a paper at SUGI ten years ago. In that paper, he presented a model of a multidimensional database and described efforts to implement it in version 5.16 SAS/AF code. Five years ago SAS Communications published an article on data warehousing at Group Health Cooperative in Seattle, where Paul was involved in their data warehousing projects as far back as 1987. Paul has witnessed a striking level of similarity between the issues he was tackling then and the ones he is facing in his current post at Boeing.

Mark Thompson

Mark addressed the role of data warehousing in business applications, specifically the construction of data warehousing tailored to the needs of corporate marketing organizations.

Mark demonstrated the growing demand for consumer insights among marketing analysts and researchers, and how data warehousing is an integral part of meeting these demands. Applications including database marketing, customer segmentation, sales forecasting and market basket analysis are among the most popular to leverage successful data warehousing implementations.

However, Mark noted, the reliance on information generated within the enterprise may not sufficiently address all the needs of marketers, who increasingly look for strategic partnerships with other firms who complement

their business offerings. (Editor's note: refer to the partnership of Working Assets Long Distance and Ben & Jerry's Ice Cream, for instance). Marketers are beginning to warehouse and leverage information generated outside the organization as well. This includes demographic and firmographic data, census data, data on competitors' customers, and market and sales tracking data. In concert with internal data stores, such as customer information systems, marketers can assemble more comprehensive views of the market environment.

To accomplish this objective, data cleansing techniques including standardization and transformation routines are critical to building and maintaining a high-quality marketing data warehouse. While these techniques and their applications are technically detailed and often tedious, they are a critical component of the overall design. Considering the potential value riding on results of marketing analyses, quality and accuracy in data warehouse construction is a vital task indeed.

Finally, Mark described some of the trends in the care and feeding of marketing data warehouse construction and maintenance that his firm encounters in client projects. He is encountering more and better secondary data sources, such as census and geographic data. He is finding more tools for standardization, such as customer address standardization packages that make linking of disparate data sources more practical. Finally, Mark believes the market of expertise for developing subject-oriented data warehouses is growing, which will deliver greater benefits to organizations engaged in data warehousing.

Randy Betancourt

Data warehousing requires a structured approach for construction and deployment. Designers and implementers can draw on a number of excellent resources to help understand how such a structured approach can be used in their organizations.

Data warehousing applications **must** use business-oriented benchmark as a basis for justification and evaluation. An example of a business-oriented benchmark is the following: "Currently, inventory turn rates are 2.1 per quarter, and after our data warehouse

deployment, we expect an inventory turn-rate of 3.0 per quarter.” Randy expressed caution about benchmarks that are based only on information technology-based criteria, such as: “Query A used to take 10 hours to execute and now it only requires 5 hours to execute.” This caution is due to the lack of connection between the allocation of data warehousing resources and corporate objectives.

Next, Randy said, the need for a data warehouse blueprint is essential. If the shop doesn't have such a blueprint, they can be built in-house, or outsourced to data warehouse architects. It is also essential that the team involved in data warehouse construction know how to read these blueprints. In Randy's experience, SAS programmers are traditionally good at knowing “how to build a wall,” but when it comes to building the room, do they know where to place the wall and tie it in w/the other walls? When considering the scope of planning involved in developing a well-architected system (as opposed to a one-off report or analysis), a new mind-set may be needed to work in this environment.

A third guideline is that administration is a big part of the work, falling into three areas.

1. Collection of metadata, a tedious yet critical task
2. Maintaining the mechanical aspect of the data flows from source to target
3. Developing feedback mechanisms to handle new requirements when data in the warehouse does not satisfy end-user needs

Randy advised members of the audience to seek the underlying needs of the people in the business unit generating the demand for data warehousing. To develop this understanding, you need involvement from members of the business unit, database administrators and network administrators, and executives.

Managing user expectations is a tricky process. Randy's advice is to be very clear about what the data warehouse is going to deliver, by stating specifically which problem the warehouse is going to solve. Recall that this problem is usually not an issue of query speed or performance, but more typically the needs of business analysts in

solving operational, marketing, or financial issues.

Randy noted how finance departments often have the best data quality and a single general ledger. This simplifies the mechanics of building a financial data warehouse for common financial analyst needs, such as consolidation and roll-up. In contrast, marketing departments often demand a customer-centric data warehouse that enables them to perform data mining tasks.

Finally, Randy addressed the issue of data warehouse design considerations and how they are influenced by business metrics. The essential question by which designers need to be influenced is “Are we making better business decisions?” In his experience, believes 50% to 75% of the actual work in data warehousing is actually committed to data transformation, emphasizing the role of involvement by analysts familiar with the business issues.

Discussion from the floor

What is truly new about data warehousing? Is it a new name for a traditional line of work, or are there some fundamentally new issues involved?

The panel believed there are a blend of traditional and new elements to the field of data warehousing. What have been known for some time, the panel agreed, are two fundamental drivers for data warehousing: the usefulness of decision support tools for organizational management, and the use of subject areas supported by metadata. What is new about data warehousing falls under four categories:

1. There are better tools available for the standardization process.
2. Distributed servers and interactive operating systems are delivering greater computing speed and flexibility, enabling in turn the ability to conduct more insightful analyses.
3. There has been a continued emphasis across the industry to develop clear specifications and nomenclature for data warehousing. A trade group literature has evolved very

quickly in data warehousing, which facilitates enhanced communication and learning among data warehousing constituencies.

4. Resulting in part from these developments, data warehousing appears to be delivering greater value to organizations than more traditional programs of decision support information systems.

How do you convince decision-makers to pursue data warehousing projects despite their incremental costs to the organization?

There may be no silver bullet that simplifies the corporate justification process specific to data warehousing. However, Randy Betancourt wanted to clear up a few myths about the role played by SAS software in corporate data warehousing situations, in an effort to focus the discussion about data warehousing investments.

The first myth is that "you have to store all your raw data in SAS format" in order to use the SAS system for data warehousing. Randy reported several instances of companies storing their raw and transformed data stores in any number of relational database management systems, where SAS software was used to load data into a data warehouse via SAS access methods.

The second myth is that the results of SAS-facilitated transformations and loadings have to be stored in a SAS data format. Again, fully loaded and transformed data need not be stored in SAS format if the corporate desktop systems prefer to work with information in a different format. Randy advised the audience to draw on the strengths of the SAS warehousing solution for extracting, transforming and loading processes. He said many organizations are using SAS to solve problems of multiple passes on the data between systems and across network architectures. Another strength of the SAS solution is in comprehensive maintenance of metadata across platforms.

How should you control expectations and time lines for data warehousing projects?

The panel agreed that data warehousing is less a project and more an ongoing, continual process. As Randy Betancourt advised, "Let the genie out of the bottle." Panelists and audience members nodded sympathetically at the notion that once a data warehouse has helped analysts answer one set of questions, another set of questions arises from the first set of answers, begetting another round of data warehousing demands. There is a growing collection of literature in the industry that data warehouse designers should study in order to leverage the lessons of other engagements. Finally, the panel agreed that data warehousing processes should start by picking a small, well-defined objective that requires a relatively short time frame and can demonstrate rapid return on investment. This will build credibility for a budding data warehousing effort.

What is the SAS Data Warehouse Administrator?

Randy Betancourt summarized the features of the new SAS Data Warehouse Administrator (DWA) product this way:

- DWA provides a visual interface on top of existing tools for extracting, transforming, and summarizing data
- DWA serves as an intelligent agent that collects all the parts of a data warehouse under a common framework
- DWA helps manage and automate an increasingly mechanical process
- DWA is driven by metadata

What is the difference between OLAP and Data warehousing?

On-line analytic processing (OLAP) is essentially dimensional analysis enhanced by drill-down and roll-up functions built into a visually driven, interactive environment. It allows the analyst to ask questions like, "How is this product line performing in this region this month as opposed to the same period last year?"

Data warehousing is the process of getting the data organized for the subsequent process of OLAP and other decision-support techniques. In contrast to OLAP's interactive nature, data warehousing is typically a batch-driven process.

Predictions

The panel concluded with each panelist delivering their short-term predictions for the field of data warehousing.

Paul Oldenkamp predicts vendors will develop a standardized set of application programming interfaces (APIs) as a link between data stores and data warehousing tools.

Mark Thompson believes the field will begin to deliver an increasing variety of industry-specific data warehousing products and packages.

Randy Betancourt had three predictions:

1. Microsoft will drive down the cost of database licensing, primarily through increased volume sales of their SQL Server product.
2. There will be increased problems with the metadata control, as the explosion of end-user tools will make it easier to develop multiple metadata sources within the same organization.
3. New products, such as Hewlett-Packard's Intelligent Warehouse, will examine queries with feedback loops and permit administrators to tune databases and queries more dynamically.