BUILDING A DATA WAREHOUSE
IS NO EASY TASK ... THE RIGHT
PEOPLE, METHODOLOGY, AND
EXPERIENCE ARE EXTREMELY CRITICAL

Eleven Steps to Success in Data Warehousing

by Sanjay Raizada
General Manager, Northwest Operations
Beaverton, Oregon

SYNTEL

Eleven Steps to Success in Data Warehousing

More and more companies are using data warehousing as a strategic tool to help them win new customers, develop new products, and lower costs. Searching through mountains of data generated by corporate transaction systems can provide insights and highlight critical facts that can significantly improve business performance. Until recently, data warehousing has been an option mostly for large companies, but the reduction in the cost of warehousing technology makes it practical, often even a competitive requirement, for smaller companies as well. Turnkey integrated analytical solutions are reducing the cost, time, and risk involved in implementation. While access to the warehouse was previously limited to highly trained analytical specialists, today corporate portals are making it possible to grant access to hundreds, even thousands of employees.

BY SANJAY RAIZADA • SYNTEL • BEAVERTON, OREGON

1.

RECOGNIZE THAT THE JOB IS PROBABLY HARDER THAN YOU EXPECT

Experts frequently report that 30 percent to 50 percent of the information in a typical database is missing or incorrect. This situation may not be noticeable or may even be acceptable in an operational system that focuses on swiftly and accurately processing current transactions. But it's totally unacceptable in a data warehousing system designed to sort through millions of historical records in order to identify trends or select potential customers for a new product. And, even when the data is correct, it may not be usable in a data warehouse environment. For example, legacy system programmers often use shortcuts to save disk space or CPU cycles, such as using numbers instead of names of cities, which makes the data meaningless in a generic environment. Another challenge is that database schema often change over the lifecycle of a project, yet few companies take the time to rebuild historical databases to account for these changes.

2.

UNDERSTAND THE DATA IN YOUR EXISTING SYSTEMS

The first step in any data warehousing project should be to perform a detailed analysis of the status of all databases that will potentially contribute to the data warehouse. An important part of understanding

"The reduction in the cost of warehousing technology makes it practical."

the existing data is determining interrelationships between various systems. The importance of this step lies in the fact that interrelationships must be maintained as the data is moved into the warehouse. In addition, the implementation of the data warehouse often involves making changes to database schema. A clear understanding of data relationships among heterogeneous systems is required to determine in advance the impact of any such change. Otherwise, it's possible for changes to create inconsistencies in other areas that ripple across the entire enterprise, creating enormous headaches.

3.

BE SURE TO RECOGNIZE EQUIVALENT ENTITIES

One of the most important aspects of preparing for a data warehousing project is identifying equivalent entities and heterogeneous systems. The problem arises from the fact that the same essential piece of information may appear under different field names in different parts of the organization. For example, two different divisions may be servicing the same customer yet have the name entered in a slightly different manner in their records (such as AIG and American International Group). A data transformation product capable of fuzzy matching can be used to

identify and correct this and similar problems. More complicated issues arise when corporate entities take a conceptually different approach in the way that they manage in-store data. This situation frequently occurs in cases of a merger. The database schema of the two organizations was established under the influence of two entirely different corporate cultures. Establishing a common database structure can be just as important as merging the corporate cultures and crucial to obtaining the full effects of synergy.

still been made to minimize the work involved in dealing with historical data. This emphasizes the importance of starting as soon as possible to create and capture metadata for interfaces, business processes, and database requirements. Several vendors have developed products that have the potential to integrate metadata from disparate sources and begin to establish a central repository that can be used to provide the information needed by both administrators and users.

4.
USE METADATA TO SUPPORT DATA QUALITY

Metadata is crucial to a successful data ware-housing implementation. By definition, metadata is data about data; for example, tags that indicate the subject of a World Wide Web document. There are many types of metadata that can be associated with a database: to characterize and index data, to facilitate or restrict access to data, to determine the source and currency of data, etc. One major challenge is trying to synchronize the metadata between different vendor products, different functions, and different metadata stores. A major data center might easily have a 500-page schema layout or COBOL copybooks with 200,000 lines and up. But the chances are that many undocumented changes have

5.
SELECT THE RIGHT DATA TRANSFORMATION TOOLS

Data transformation tools extract data from the operational sources, clean it, and load it into the data warehouse while capturing the history of that process. This transformation process may include creating and populating new fields from the operational data, summarizing data to an appropriate level for analysis, performing error checking operations to validate the integrity of the data, etc. Look for a tool that makes it possible to map data from source to target with a simple point-and-click interface. The ability to track and manage the relationships of interrelated data entities that are affected when changing database structures is also useful. Finally, try to find a tool that can capture and store metadata during the conversion process.

"The biggest technological improvements in recent years in the data warehousing field have come in the area of information delivery."

6.

TAKE ADVANTAGE OF EXTERNAL RESOURCES

External sources of information, such as data from a customer's transaction processing system or market research data provided by a third party, can often greatly increase the value of internal information. Rather than simply comparing sales against the same month last year, use of external data might make it possible, for example, to compare sales growth against the increase in the overall market. Or suppose that you want to estimate the income of each of your customers. You may be able to obtain a database containing the average income for every ZIP Code in the country. While this will not provide accurate answers for individual customers, it will enable you to obtain a good estimate of an aggregate grouping, such as all 30 to 45 year olds in Kansas, that might be used in a targeted marketing campaign. The integration challenge is even greater when external data sources are involved. In some cases, external data will differ so drastically from existing schema that data transformation algorithms will be required to make use of the external resource.

7.

UTILIZE NEW INFORMATION DISTRIBUTION METHODS

The biggest technological improvements in recent years in the data warehousing field have come in the area of information delivery. In the past, highly skilled analysts were needed to prepare information in the format requested by users. Today, information can be delivered in several different ways directly to the people who need it. Users can subscribe to regular reports and have them delivered via e-mail. If, when run, the report contains data that meets the subscription criteria, a customer-filtered view of the data will be delivered to the user by means of an e-mail attachment. The user only has to click on the e-mail attachment to bring up the report. The report data stays securely and economically on the server and only the pages requested by the authorized user are sent on demand. Another option is to let the user log in, search for data, and open the reports that provide the needed information. Only the pages requested by the user are sent over the network, to minimize unnecessary traffic.

8.
FOCUS ON HIGHER PAYBACK MARKETING APPLICATIONS

Most of the really hot applications in data warehousing involve marketing because of the potential for an immediate payback in terms of increased revenues. For example, catalog manufacturers are using data warehousing to match personal characteristics to purchases of specific items. This empirical data is then used to produce special demographic editions that generate higher sales since they are targeted more closely to their customers' needs. Companies selling more complicated and higher value products, such as brokerages, insurance companies, credit card firms, etc., are weaving data warehousing into their relationship-based businesses in order to dramatically increase the information flow between their firm and their customer. Innovative banks are even using similar methods to sort out customers whose trading characteristics match the profiles of known money launderers.

9.
EMPHASIZE EARLY WINS TO BUILD SUPPORT THROUGHOUT THE ORGANIZATION

The availability of a wide range of off-the-shelf solutions has made it possible to drastically reduce cost and lead-time requirements for data warehousing applications. Off-the-shelf solutions won't usually complete project objectives, but they often can be used to provide point solutions in a short time that serve as a training and demonstration platform and most importantly, build momentum for full-scale implementation. Even for the largest scale applications, technology surveys should be performed in order to maximize the use of pre-built technology.

10.
DON'T UNDERESTIMATE HARDWARE REQUIREMENTS

The hardware requirements for data warehousing database servers are surprisingly high. The primary reason is the large number of CPU cycles required to slice and dice data over and over again to meet the varying needs of users throughout the organization. Database size also plays a part in the server performance requirements — requirements range up to terabytes and higher. For these reasons, be sure to select a scalable platform regardless of how much headroom you have provided in your server specification. The typical data warehouse implementation starts out at the departmental level and grows over time to an enterprise wide solution. Purchasing servers that can be expanded with additional processors is one possible approach. A more ambitious idea is to combine loosely coupled systems that enable the database to be spread out over multiple servers, although it still appears as a single entity to users.

11.

CONSIDER OUTSOURCING YOUR DATA WAREHOUSE DEVELOPMENT AND MAINTENANCE

A large percentage of medium and large companies use outsourcing to avoid the difficulty of locating and the high cost of retaining skilled IT staff members. Most data warehousing applications fit the main criteria for a good outsourcing project — a large project that has been defined to the point that it does not require day-to-day interaction between business and development teams. There have been many cases where an outsourcing team is able to make dramatic improvements in a new or existing data warehouse. Typically, these improvements do not necessarily stem from an increased level of skill on the part of the outsourcing team, but rather flow from the nature of outsourcing. The outsourcing team brings fresh ideas and a new perspective to their assignment and is often able to bring to bear methods and solutions that they have developed on previous assignments. The outsourcing team also does not have to deal with manpower shortages and conflicting priorities faced by the previous internal team.

12.

CONCLUSION

By now, you've realized that building a data warehouse is no easy task. With the average cost of a system valued at $1.8 million, the right people, methodology, and experience are critical. The reliance on technology is only a small part in realizing the true business value buried within the multitude of data collected within an organization's business systems. Data warehouses touch the organization at all levels, and the people that design and build the data warehouse must be capable of working across the organization as well. The industry and product experience of a diverse team coupled with a business focus and proven methodology may be the difference between a functional system and true success.

SYNTEL

about SYNTEL:

about the author

Sanjay Raizada

   With Syntel's acquisition of IMG in September 1999, Sanjay was appointed General Manager of Northwest Operations based in Beaverton, Oregon. As President of IMG, Sanjay led its Consulting Services Division and helped build a Fortune 500 clientele. His responsibilities included business and methodology development and quality control for all projects in the region. Prior to joining IMG and Syntel, Sanjay was a Senior Manager at a Big 5 professional services firm leading their Western Region Data Warehousing Practice for North America. From 1991 to 1994, Sanjay served as Principle Consultant at an industry-leading provider of global e-business solutions. Sanjay's 15 years IT experience includes all aspects of systems architecture design, infrastructure planning, and systems development. He specializes in data warehousing architecture and decision support solutions using automated process methodologies involving CASE, ETT, and other emerging technologies.

**Sanjay Raizada
General Manager,
Northwest Operations**

Syntel

Syntel (www.syntelinc.com) is a global e-Business and applications outsourcing company that delivers real-world technology solutions to Global 2000 corporations. Syntel's portfolio of services includes e-Business development and integration, wireless solutions, data warehousing, eCRM, as well as complex application development and enterprise integration services. The company was the first U.S.-based firm to launch a Global Delivery Service to drive speed-to-market and quality advantages for its customers. Syntel's 2,600 employees operate from centers in the U.S., Europe, and Asia.

SYNTEL
HEADQUARTERS
2800 Livernois Road
Troy, MI 48083
phone **248/619-3503**
fax **248/619-2889**

SYNTEL