

# ACTIVE WAREHOUSING

Stephen A. Brobst and AVR Venkatesa

Data warehouses are beginning to take on mission-critical roles supporting CRM, one-to-one marketing, and minute-to-minute decision making. This new generation of “active” warehouses demands new technology and performance requirements

**T**he scope and impact of data warehouse implementations are rapidly expanding and transforming to address today's business requirements. Data warehousing requirements have evolved to demand a decision-support capability that is not just oriented toward corporate staff in the ivory tower, but actionable on a day-to-day (minute-to-minute) basis.

Decisions such as when to replenish Barbie dolls at a particular retail outlet may not be strategic at the level of customer segmentation or long-term pricing strategies, but when executed properly, they certainly make a big difference to the bottom line. We refer to this capability as “tactical” decision support. Tactical decisions are the drivers for day-to-day management of the business. Businesses today want more than just strategic insight from their data warehouse implementations — they want better execution in running the business through more effective use of information for the decisions that get made thousands of times per day.

The bottom-line impact of a well-placed tactical decision-support capability often rivals that of strategic decision support. In the end, the business wants and needs both.



## FIRST-GENERATION WAREHOUSES

The first round of decision-support capability provided in the evolution of the data warehouse marketplace primarily targeted batch-oriented decision-support capability. These implementations focus on delivering cleansed and integrated data to information-starved marketing, finance, actuarial, and other corporate decision-making bodies. Information from across the organizational silos created by operational systems (usually along product boundaries) is made available for query purposes to knowledge workers. Value to the organization is huge: A customer-centric view of the business or, perhaps, integrated sales and inventory data is available to decision makers for the first time.

The refresh cycle for a first-generation data warehouse implementation is typically monthly or weekly. Queries are often launched in a batch-oriented environment with somewhat long turnaround times. A typical implementation allows ad hoc queries during the day and batch scoring algorithms for customer segmentation, profitability calculations, predictive response models, and so on at night. Advanced organizations may even take the results from the batch scoring jobs and feed them back into the operational

sue. If large numbers of users do exist, they are usually limited to prebuilt summary tables (or data marts) for satisfying queries in a highly parameterized reporting environment.

Despite their limitations, first-generation data warehouse designs deliver substantial value by bringing detailed data and a total organizational view of customer relationships together for the first time. Of course, a successful data warehouse implementation always begets demands for even further enhancement to the value proposition for the business. As a result, the current generation of decision-support implementations has begun moving toward “active” data warehouses — warehouses that bring new capabilities to the mainstream of the business.

## CURRENT-GENERATION WAREHOUSES

The key to maximizing the value derived from a data warehouse is to make the decision making actionable and relevant to the business. First-generation data warehouse implementations are often consigned to purely strategic endeavors such as studying long-term trends in the business or performing customer segmentation activities. The value of these activities, while certainly critical, is difficult to quantify.

aging them into a more proactive role in the management and execution of the enterprise.

One of the most effective approaches that has been used in the trend toward active data warehousing is “event-based triggering.” The basic idea is to apply quantitative decision-making techniques to the analysis of business events with the goal of proactively generating triggers for initiating operational activities. Customer retention is a particularly good application of event-based triggering. By detecting events that may lead to customer attrition, you can take proactive steps to retain your most valuable customers.

### Example 1: Transportation

A recent experience with a major airline provides an excellent example of the technique. After a long day at a client site, I made my way to the airport for a flight to a conference at which I was scheduled to speak early the next morning. At the airport, I found out that my flight — the last one that night — was delayed indefinitely due to a mechanical problem. I had encountered a similar delay (with the same airline) only a few weeks earlier. It is also important to note that I fly over 100,000 miles per year on this airline.

It was well past midnight by the time I arrived at my destination. When dragging myself out of bed at 6 A.M. for my keynote, my loyalty to the airline was definitely not at its peak. Allegiance to an airline goes only so far when on-time performance is not up to par. The interesting part of the story occurred when I arrived home that very next Friday night. A letter of apology from the airline was waiting for me in the mailbox, complete with some statistics about on-time performance (demonstrating that my recent two bad experiences were an unfortunate statistical anomaly) and two first-class upgrade certificates along with a free round-trip flight voucher. With an apology and certificates in hand, the inconvenience of a couple of late night flights didn't seem quite so bad anymore.

I happen to know that this particular airline is relatively sophisticated in its use of data warehousing. Clearly, after the second major flight delay in as many

## Measurable value comes when the results of the strategic analysis get translated into actionable decisions

systems to assist in customer care (for example, providing customer product affinity scores to customer service representatives to assist in cross-selling and up-selling).

Extensive support for multiuser query workloads is a rarity in first-generation data warehousing. Decision-support capability is rolled out to only a small number of power users and analytically oriented decision makers, so supporting widespread deployment is not yet an is-

Measurable value comes when the results of the strategic analysis get translated into actionable decisions. Many such decisions are possible with first-generation data warehouse implementations: pricing, market positioning, business strategy development, and so on. Getting to the next level of value, however, requires wider use of decision support in the business. Rather than relegating data warehousing to the narrow domain of think tank deployment, organizations are lever-

weeks, they identified me as an at-risk customer based on a scoring initiated by the delayed flight event. Interestingly, the airline didn't react to my first delayed flight that month (nor was I particularly annoyed with this isolated incident, given that these things are tolerable as long as they are infrequent). Even more interesting is that not all passengers on the plane received compensation for the inconvenience. The airline sent nothing at all to my companion traveler, who flies much less frequently than I and who had not experienced the first delay I encountered. Selective inclusion in the retention program is an important aspect of its effectiveness. The combination of my risk of attrition and high value to the airline triggered inclusion in an aggressive retention program designed to rebuild my confidence in on-time performance and reinforce my loyalty through flight awards.

This customer retention program can be translated into a variety of other industries and scenarios. Providers of wireless service often use dropped call events to trigger customer retention activities. Missed or late deliveries may be the events to look for in the package shipping business, whereas a botched wire or transfer is a likely candidate in the financial services arena. Sophisticated event-based triggering uses combinations of a number of different events to implement an optimized customer retention program. The key is to have integrated detailed data (and events) from across all customer relationships and touch points within the enterprise. This requires a single view of the customer across product boundaries within an organization, which an enterprise-oriented data warehouse (as opposed to product-oriented data marts) can provide.

### Example 2: Health Care

Event-based triggering is certainly not confined to customer retention programs. In healthcare data warehouse implementations, for example, it is used in proactive intervention programs where high-risk patients are identified and steered into case-management programs. By scoring patients as early as possible when ac-

quiring preadmission certification for illnesses, healthcare costs and quality can often be managed more effectively by case-managing the individual back to health rather than relying on traditional healthcare processes. Scoring will usually be initiated based on an event such as a preadmission certification request or online

on the customer's individual needs. Assessment of these needs should be based on an understanding of the investment portfolio, risk tolerance, and so forth of the individual and should be scored considering the multiple account relationships within the bank.



## The goal is to provide an interaction that is uniquely customized to each individual based on specific needs

nursing call and will use both patient characteristics (age, gender, heredity, and so forth) and past medical history to make individual decisions regarding placement into a case-management scenario. Medical claims data may also be used to initiate scoring, but lags in acquisition of this data often make its use outside of chronic illnesses less beneficial than data that can be acquired earlier in the medical process.

Identification of candidates for case management using quantitative scoring techniques with significant breadth and depth of patient data enables earlier and more effective allocation of case-management resources. Using these quantitative decision-making techniques augments the traditional physician referral process for case management, benefiting the health of individual patients as well as the overall management of healthcare costs. Areas of particular success include case management for high-risk pregnancies, chronic diabetes, chronic asthma, congestive heart failure, and other illnesses where close monitoring by skilled nursing professionals can significantly impact an individual's health.

### Example 3: Banking

In the banking industry, events such as a large deposit into a checking account may trigger communications to a customer regarding alternative investment vehicles (with higher yields). The specifics of the communication should be based

An impending payment date on a life insurance policy may be used to trigger customer communications to minimize the probability of lapse. Life events also play a big role in triggering action within the business. For example, a "right-selling" approach in the insurance industry would certainly involve working with a customer to re-evaluate the appropriate coverages upon major life changes such as marriage, birth of a child, or purchase of a new home. Depending on the customer relationship, life events may be reported directly by the customer or acquired from external service bureaus.

Notice that in all of these examples the data warehouse is playing a *proactive* role in triggering action within the business based on event detection and scoring. First-generation data warehouse implementations often provide the ability to perform list selection and extracts for purposes of direct mail and telemarketing campaigns. However, these list pulls are largely initiated by human intervention or predefined scheduling criteria that are largely independent of individual customer events. The current generation of data warehouse implementations is much more focused on proactive management of customer events.

Part of making a decision actionable and relevant is ensuring that it is timely. Monthly refresh cycles are not nearly frequent enough for proactive management of events. Nightly (incremental) refresh is typically a minimum requirement for



proactive event-based triggering. For example, winback campaigns in the long-distance calling marketplace degrade almost exponentially in effectiveness for each day lost after a competitive disconnect.

Current-generation data warehouse implementations are differentiated from their predecessors by the need for more frequent updates and workload characteristics that are much more sophisticated. The need to score many different subsets

ing able to service these needs rather than produce the same response for every customer asking about a particular product. While cross-selling and up-selling based on a product inquiry can certainly be effective, this technique falls short of one-to-one relationship building because it does not seek to understand the customer as an individual.

Next-generation customer service interactions will need to use scoring techniques based on the product of inquiry

a large number of business-to-business and business-to-consumer enterprises. Reaction to this trend from an operational perspective has been very positive. Reduced costs in customer service resulting from the self-service model are very attractive. However, marketing organizations in astute corporations are already grappling with the impact of losing direct contact with the customer.

Brand image and quality of a customer relationship is largely based on the “experience” of doing business with an enterprise. Since human customer service representatives are becoming disintermediated in the world of e-commerce, the intelligence necessary for superior customer care must be built into the Web interactions. The intimacy of experience previously provided by human interaction now needs to be delivered using advanced CRM techniques in an automated fashion. This means that individual customer profiles and online scoring capabilities need to become accessible from within the transactional model for implementing best-of-breed e-commerce solutions.

So far we’ve seen e-commerce solutions adopt data warehousing capabilities for cross-selling, up-selling, yield management with customer-specific pricing, fraud detection, and so on. Each of these applications requires heavy-duty analytics and yet must be delivered in the transactionally oriented world of e-commerce. The goal is to provide an interaction that is uniquely customized to each individual based on specific needs that can be gleaned at the time of the e-commerce event combined with as much demographic, purchase, inquiry, and other information as can be leveraged to refine the interaction.

## NEW CHALLENGES

Next-generation data warehouse implementations will push the envelope of technology in a big way. Response-time requirements for online scoring of individual customer events will be near instantaneous. This requirement combines the large database size and query complexity of a traditional decision-support environment with the response-time re-

## Next-generation data warehouse implementations will push the envelope of technology in a big way

of customers based on event classifications (using different quantitative models appropriate to each scenario) leads to more sophisticated use of database optimizer capabilities in query execution.

### NEXT-GENERATION WAREHOUSES

The next generation of data warehouse implementations will need to support real-time analytics to assist in managing the customer relationship at any and all touch points. The focus on customer relationship management (CRM) in today’s competitive business environments will drive businesses to leverage warehousing capabilities to direct customer interactions for one-to-one relationship building. The active data warehousing concept comes to maturity as analytical decision-support capabilities are integrated directly with customer interactions.

Consider an interaction with a customer service representative while in pursuit of a retail catalog purchase. Companies often provide customer service representatives with standard cross-selling and up-selling scripts for interacting with customers who have inquired about a specific product. This is not what it means to deliver one-to-one customer service. One-to-one means understanding the individual needs of each customer and be-

ing able to service these needs rather than produce the same response for every customer asking about a particular product. More-over, “Customer Specific Marketing” techniques (à la Brian Woolf) will allow special offers and pricing to individual customers based on the loyalty and profitability of each customer. Notice that real-time scoring is essential because a predefined scoring (even if it is based on customer-specific attributes) ignores the most influential attributes of all — those that describe the content of the customer interaction at this very moment. Providing online scoring to help representatives interact with customers is only the beginning of real-time customer relationship management from the data warehouse. The trend toward integrating the data warehouse with customer interactions will be adopted even more aggressively and evolve toward more sophisticated use of information as corporations explore ways of developing a customer relationship management strategy for conducting business in the face of e-commerce.

Conducting business over the Web will eliminate traditional channels of relationship building. Bank tellers, travel agents, retail salespersons, brokers, and so on are quickly being outmoded by self-service models using the Internet for



quirements of a traditional transaction-processing system. Moreover, data acquisition needs to occur closer to real time than in current- or previous-generation data warehouse implementations. We're not talking two phase commit here, but certainly trickle feed architectures and advanced replication mechanisms will be required.

The bar will also be raised for RASR (Reliability, Availability, Serviceability, and Recoverability) characteristics in next-generation data warehouse implementations. Active data warehousing is all about

integrating advanced decision support with day-to-day decision making. Success in this endeavor inevitably escalates the RASR requirements for the data warehouse, because enabling tactical decision making becomes mission critical. Moreover, in the world of e-commerce there is no down time — the channel for delivering products and services is open 24 hours a day, 365 days a year.

Another requirement that first- and current-generation data warehouse implementations are not prepared for is the orders of magnitude increase in concur-

rent query workload. No longer are we talking about dozens or even hundreds of simultaneously executing queries — concurrency will reach into the thousands when customer service representatives and Web sites begin accessing the data warehouse for sophisticated scoring capabilities. Middleware software, such as BEA Systems' Tuxedo and Top End, will become a critical component of next-generation data warehouse architectures. The ability to assist in managing service levels

*(continued on page 41)*

## PROMPT: Active Data Warehousing at Work

**I**n September 1998, NCR bought 50 percent of Stirling Douglas Group, Inc. (SDG), a privately held Canadian software manufacturer, to expand its active data warehouse solutions for retailers. Six months into the relationship, codevelopment efforts are bearing fruit — namely the first customer implementation of the SDG product PROMPT (Product Replenishment on Merchandise Policy Techniques) in a Teradata environment. PROMPT lets Teradata users leverage detailed sales and inventory data for operational decision making, an essential capability for retail implementations of an active data warehouse.

A retail assortment management system, PROMPT is supported by three modules: contribution, sales profiling, and automated forecasting and replenishment.

The contribution module creates weekly summary rankings for products and locations that are used by the other two modules. The sales profiling module runs as needed, with daily or weekly tuning, to build sales profiles at any of a wide variety of location/product-level combinations. The final module — forecasting and replenishment — has many components, but the primary two are the forecasting and replenishment engines. The forecasting engine runs weekly, and the replenishment engine produces recommended stock levels daily or weekly. This module uses the information produced by the other two modules, and its engines are reused by other SDG applications.

PROMPT uses these modules to support product mix management applications via contribution and profitability metrics, plan vs. actual reporting with multilevel objectives, exception monitoring based on predefined policies, and what-if scenarios. It uses profile clusters and store groups in addition to the usual granular store/product combinations.

The forecasting engine utilizes advanced, self-monitoring forecasting techniques in its weekly runs. This engine continuously monitors the current effectiveness of each of several forecasting models (including regular, promotional, and seasonal variations) for each store/product combination, auto-

matically selecting the model that produces the best forecast of recent activity to produce the current forecast. For new stores and products, the system permits the assignment of a "like" store or product to be used to bootstrap the forecasting calculations until enough real history is available.

The replenishment engine calculates recommended stock levels on a daily basis. The frequent runs and large data volumes processed mean that operational performance is critical to this component. These recommendations take into account a safety stock level based on vendor lead time and vendor quantity-shipped errors (determined from historical vendor performance data), estimated forecasting inaccuracies, and consumer service-level targets.

PROMPT's intense use of large volumes of detailed data from a wide variety of operational sources makes excellent use of Teradata's strengths as a high-capacity/high-performance data warehousing platform — the effectiveness of PROMPT's profiling and forecasting systems improves dramatically as more data is made available. The full breadth of the product requires detailed and frequently updated cleansed and integrated data for the entire supply chain, including orders and shipments, warehouse and selling-site capacities, inventory levels and transactions, and promotions. PROMPT is integrated with Teradata directly on top of existing data warehouse structures, thereby avoiding the need to change the existing system or create a special-purpose data mart.

Additional SDG components that are built on the same three engines as PROMPT are due to be available for Teradata in 1999. AIMM (Active Integrated Merchandise Management) integrates PROMPT forecasts into the retailer's purchase order system, closing the loop of the decision-making and implementation cycle for a powerful active data warehousing solution. SDG's MAPS (Merchandise Activity Planning System) supports merchandise plan development through forecasting, planning models, and contribution metrics.

— *Gregory N. Purdy and Stephen A. Brobst*

## Active Warehousing

(continued from page 33)

for performance and availability in the face of thousands of concurrently launched queries will be a key role of the middle-ware layer in these implementations.

The Teradata database is positioned exceptionally well for stepping up to the challenges related to high availability and large multiuser workloads required for an active data warehouse implementation. And, of course, handling complex queries has always been a strength of the Teradata optimizer. The main challenges in implementing active data warehousing with Teradata will be the transaction-like response times required for online scoring. Teradata's inclusion of efficient and scalable triggers in 1999 is designed to support the kind of event-based triggering discussed in this article. In addition, the planned deployment of user defined functions (UDFs) in Teradata will also allow sophisticated scoring logic to be performed directly in the database engine so as to avoid the overhead of data move-

ment across address spaces to and from a traditional program. Bringing down the overhead of transactional latencies, which is ongoing in Teradata performance work, will also play a particularly important role in next-generation data warehouse implementations. The join indexing capabilities introduced in the most recent release of Teradata are also marked for future augmentation with additional advanced indexing techniques and built-in aggregate tables to enhance performance in mixed-workload environments.

The marketplace is coming of age as we progress from "passive" decision-support systems to current- and next-generation "active" data warehouse implementations. Convergence of requirements from traditional decision-support and transactional systems presents challenges previously not dealt with in high-end database implementations.

Next-generation implementations will be possible within the next two to three years with the introduction of advanced indexing techniques and a focus on mixed-workload support. The key is to design

today with an eye toward tomorrow. This approach is particularly important in light of the impending tidal wave of e-commerce capability that will be required in our customer care solutions. ■



**Stephen A. Brobst** is a cofounder and senior consultant at Tanning Technology Corp. in Denver. He is also a founder and managing partner at Strategic Technologies & Systems in Boston. He specializes in very large database implementations and recently coauthored *Building a Data Warehouse for Decision Support* (Prentice Hall, 1998). You can reach him by email at [sabrobst@jj.lcs.mit.edu](mailto:sabrobst@jj.lcs.mit.edu).

**AVR Venkatesa** is director of east coast delivery services at Tanning Technology Corp. and has over 15 years experience in designing, building, planning, and implementing advanced customer relationship management solutions. You can reach him by email at [avr@tanning.com](mailto:avr@tanning.com).