

Teradata

MAGAZINE
ONLINE

SOLUTIONS THAT POWER GROWTH

2Q
2001[Articles](#)[Archives](#)[Tech2Tech](#)[Media Kit](#)[Subscribe](#)[Contact us](#)[HOME](#)

INSIDER'S WAREHOUSE

Articles Spring 2001



The five stages of an Active Data Warehouse evolution

By Stephen Brobst and Joe Rarey

Data warehousing is a journey. The most successful data warehouse implementations deliver business value on an iterative and continuous basis. Each iteration builds upon its predecessor to increase the business value proposition for information delivery. In recent years, the evolution of data warehousing has reached a new pinnacle with the deployment of decision support capability throughout an organization and even beyond its conventional boundaries to partners and customers.

In the early days, data warehousing focused almost entirely on providing strategic decision-making capability to knowledge workers in the corporate ivory tower. End users for the data warehouse were traditionally in areas such as marketing, strategic planning and finance. Access to information dramatically increased the quality of their decision-making. However, developing a superior corporate strategy is only part of what it takes to succeed in today's intensely competitive business environment. A great strategy is nothing without great execution.

The emerging generation of data warehouse deployments improves the execution of a business strategy in addition to its development. This evolution imposes an ever-increasing set of service levels upon the data warehouse architect. In this article we discuss the evolution of data warehousing through the five stages that are most common in the maturation of decision support within an organization.

Stage 1: Reporting



The initial stage of data warehouse deployment typically focuses on reporting from a single source of truth within an organization. The data warehouse brings huge value simply by integrating disparate sources of information within an organization into a single repository to drive decision-making across functional and/or product boundaries. For the most part, the questions in a reporting environment are known in advance. Thus, database structures can be optimized to deliver good performance even when queries require access to huge amounts of information.

The biggest challenge in Stage 1 data warehouse deployment is data integration. The challenges in constructing a repository with consistent, cleansed data cannot be overstated. There can easily be hundreds of data sources in a legacy computing environment—each with a unique domain value standard and underlying implementation technology. The hard work that goes into providing well-integrated information for decision-makers becomes the foundation for all subsequent stages of data warehouse deployment.

Stage 2: Analysis



In a Stage 2 data warehouse deployment, decision-makers focus less on what happened and more on why it happened. Analysis activities are concerned with drilling down beneath the numbers on a report to slice and dice data at a detailed level. Ad hoc analysis plays a big role in Stage 2 data warehouse implementations. Questions against the database cannot be known in advance. Performance management relies a lot more on advanced optimizer capability in the RDBMS because query structures are not as predictable as they are in a pure reporting environment.

Performance is also a lot more important in a Stage 2 data warehouse implementation because the information repository is used much more interactively. Whereas reports are typically scheduled to run on a regular basis with business calendars as a driver for timing, ad hoc analysis is fundamentally a hands-on activity with iterative refinement of questions in an interactive environment. Business users require direct access to the data warehouse via GUI tools without the need for programmer intermediaries. Support for concurrent query execution and large numbers of users against the warehouse is typical of a Stage 2 implementation.

Business users, however, are a very impatient bunch. Performance must provide response times measured in seconds or a small number of minutes for drill-downs in an OLAP (online analytical processing) environment. The database optimizer's ability to determine efficient access paths, using indexing and sophisticated join techniques, plays a critical role in allowing flexible access to information within acceptable response times.

Stage 3: Prediction



As an organization becomes well-entrenched in quantitative decision-making techniques and experiences the value proposition for understanding the “whats” and “whys” of its business dynamics, the next step is to leverage information for predictive purposes. Understanding what will happen next in the business has

huge implications for proactively managing the strategy for an organization. Stage 3 data warehousing requires data mining tools for building predictive models using historical detail.

The number of end users who will apply the advanced analytics involved in predictive modeling is relatively small. However, the workloads associated with model construction and scoring are intense. Model construction typically involves derivation of hundreds of complex metrics for hundreds of thousands (or more) of observations as the basis for training the predictive algorithms for a specific set of business objectives. Scoring is frequently applied against a larger set (millions) of observations because the full population is scored rather than the smaller training sets used in model construction.

The power of an active data warehouse allows decisions to be made in an optimal fashion on an item - by - item, store - by - store and second - by - second basis.

Advanced data mining methods often employ complex mathematical functions such as logarithms, exponentiation, trigonometric functions and sophisticated statistical functions to obtain the predictive characteristics desired. Access to detailed data is essential to the predictive power of the algorithms. Tools from vendors such as SAS and Quadstone provide a framework for development of complex models and require direct access to information stored in the relational structures of the data warehouse. The business end users in the data mining space tend to be a relatively small group of very sophisticated analysts with market research or statistical backgrounds. However, beware in your capacity planning! This small quantity of end users can easily consume 50% or more of the machine cycles on the data warehouse platform during peak periods. This heavy resource utilization is due to the complexity of data access and volume of data handled in a typical data mining environment.

Stage 4: Operationalize



Operationalization in Stage 4 of the evolution starts to bring us into the realm of active data warehousing. Whereas stages 1 to 3 focus on strategic decision-making within an organization, Stage 4 focuses on tactical decision support.

Think of strategic decision support as providing the information necessary to make long-term decisions in the business. Applications of strategic decision support include market segmentation, product (category) management strategies, profitability analysis, forecasting and many others. Tactical decision support is not focused on developing corporate strategy in the ivory tower, but rather on supporting the people in the field who execute it.

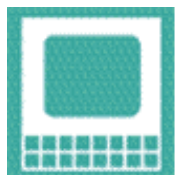
Operationalization typically means providing access to information for immediate decision-making in the field. Two examples are (1) inventory management with just-in-time replenishment and (2) scheduling and routing for package delivery. Many retailers are moving toward vendor managed inventory, with a retail chain and the manufacturers that supply it working as partners. The goal is to reduce inventory costs through more efficient supply chain management. In order for the partnership to be successful, access to information regarding sales, promotions, inventory-on-hand, etc. must be provided to the vendor at a detailed level. Manufacturing, delivery and so on can then be executed efficiently based on inventory requirements on a per-store and per-SKU level. To be useful, the information must be extremely up-to-date and query response times must be very fast.

In the example of package shipping with less than full load trucking there are very complex decisions involved in how to schedule trucks and route packages. Trucks generally converge at break bulks wherein packages get moved from one truck to another so that they ultimately arrive at their desired destination (in a way very analogous to how humans are shuffled around between connecting flights at an airline hub). When packages are on a late-arriving truck, tough decisions need to get made in regard to whether the connecting truck that the late package is scheduled for will wait for the package or leave on time. If it leaves without the package, the service level on that package may be compromised. On the other hand, waiting for the delayed package may cause other packages that are ready to go to miss their service levels.

How long the truck should wait will depend on the service levels of all delayed packages destined for the truck as well as service levels for those packages already on the truck. A package due the next day is obviously going to have more difficulty in meeting its service levels under conditions of delay than one that is not due until many days later. Moreover, the sending and receiving parties associated with the package shipment should also be considered. Higher priority on making service levels should be given to packages associated with profitable customers where the relationship may be at risk if a package is late. Alternative routing options for the late packages, weather conditions and many other factors may also come into play. Making good decisions in this environment amounts to a highly complex optimization problem.

It is clear that a break bulk manager will dramatically increase the quality of his or her scheduling and routing decisions with the assistance of advanced decision support capabilities. However, for these capabilities to be useful, the information to drive decision-making must be extremely up-to-date. This means continuous data acquisition into the data warehouse in order for the decision-making capabilities to be relevant to day-to-day operations. Whereas a strategic decision support environment can use data that is loaded once per month or once per week, this lack of data freshness is unacceptable for tactical decision support. Furthermore, the response time for queries must be measured in a small number of seconds in order to accommodate the realities of decision-making in an operational, field environment.

Stage 5: Activate



The larger the role an active data warehouse plays in the operational aspects of decision support, the more incentive the business has to automate the decision processes. Both for efficiency reasons and for consistency in decision-making, the business will want to automate decisions when humans do not add significant value. In e-commerce business models there is no choice but to automate decision-making when a customer interacts with a Web site. Interactive customer relationship management (CRM) on a Web site or at an ATM is all about making decisions to optimize the customer relationship through individualized product offers, pricing, content delivery and so on. The very complex decision-making associated with interactive CRM takes place without humans in a completely automated fashion and must be executed with response times measured in seconds or milliseconds.

As technology evolves, more and more decisions become executed with event-driven triggers to initiate fully automated decision processes. For example, the retail industry is on the verge of a technology breakthrough in the form of electronic shelf labels. This technology obsoletes the old-style Mylar labels, which require manual labor to update prices by swapping small plastic numbers on a shelf label. The new electronic labels can implement price changes remotely via computer controls without any manual labor. Integration of the electronic shelf label technology with an active data warehouse facilitates sophisticated price management with as much automation as a business cares to deploy. For seasonal items in stores where inventories are higher than they ought to be, it will be possible to automatically initiate sophisticated mark-down strategies to drive maximum sell-through with minimum margin erosion. Whereas a sophisticated mark-down strategy is prohibitively costly in the world of manual pricing, the use of electronic shelf labels with promotional messaging and dynamic pricing opens a whole new world of possibilities for price management. Moreover, the power of an active data warehouse allows these decisions to be made in an optimal fashion on an item-by-item, store-by-store and second-by-second basis using event triggering and sophisticated decision support capability. In a CRM context, even customer-by-customer decisions are possible with an active data warehouse.

Intense competition and technology innovations are motivating these advances in decision support deployment. An active data warehouse delivers information and enables decision support throughout an organization rather than being confined to strategic decision-making processes. However, tactical decision support does not replace strategic decision support. Rather, an active data warehouse supports the coexistence of both types of workloads. Notice in Figure 2 that a significant amount of workload in a Stage 5 data warehouse is still focused on strategic thinking. The operationalized and event triggered decision support of stages 4 and 5 provide the execution capability for strategies developed from traditional data warehouse analysis characterized in stages 1 to 3.

Conclusions

The successful deployment of an active data warehouse is evolutionary. It is not advisable to jump straight to Stage 5 in a first iteration implementation. A risk managed implementation builds upon the integrated, single source of truth in a traditional data warehouse. Use of the data warehouse for strategic decision support creates demand for tactical decision support to execute the business strategy. The business value of an active data warehouse is amplified by its deployment throughout an organization. Providing information to (tens of) thousands of decision-makers throughout an organization (and even to customers indirectly through CRM applications) has huge leverage for the business. It also requires a new approach to architecting a data warehouse solution. Design of the data warehouse to support extreme service levels in terms of performance, availability and data freshness in a scalable solutions environment cannot be achieved as an afterthought. **T**

Stephen Brobst is the chief technology officer for Teradata. He specializes in very large database implementations for data warehouse and customer relationship management solutions. You can e-mail Stephen at stephen.brobst@ncr.com.

Joe Rarey is a lead consultant at Strategic Technologies & Systems. He specializes in

high-end systems integration for data

warehouse solutions at Fortune 500 companies worldwide. You can e-mail Joe at

jrarey@strattech.com.



Copyright © 2001 NCR CORPORATION. All rights reserved. Revised: July 03, 2001 .