

VENDOR NEEDS AND STRATEGIES

Netezza: A Business Analytics Revolution in the Making

Dan Vesset

IDC OPINION

As the competitive environment has intensified, business analytics have taken on an increasingly strategic role as tools that enable analysis and optimization of both revenue-enhancing and cost-control processes. Analysis of sales trends, customer behavior, supplier performance, inventory levels, financial trends, fraud, and other decision-making processes is based on the successful implementation and utilization of business analytics software. Traditionally, data analysis has been handled by IT architectures that combine numerous software, server, and storage technologies. However, the time may have arrived for the purpose-built, business analytics–dedicated, open standards–based data appliance. In such a solution, software, server and storage components are integrated by one IT vendor to optimize analytic processing, decrease integration and maintenance costs, and improve query processing performance. A vendor offering one such data appliance solution is Netezza. With its Netezza Performance Server (NPS) product, the company is likely to have a significant impact on the business analytics market by:

- ☒ Setting new standards for performance of data warehousing platforms based on Asymmetric Massively Parallel Processing (AMPP) architecture
- ☒ Enabling compression of decision cycles, thus providing competitive advantage to organizations implementing the new platform
- ☒ Disrupting the competitive landscape by establishing new price/performance benchmarks made possible by the use of database standards, open source software, and standard hardware components

IN THIS STUDY

In this study, IDC looks at the need of organizations for just-in-time business analytics as the basis for business process optimization and competitive advantage. The trends of rapidly increasing data volumes and user populations in the data warehousing and business intelligence markets are examined. The study highlights the challenges faced by organizations in serving the needs of end users while controlling IT costs.

A solution based on a dedicated, purpose-built appliance for data warehousing is proposed as an alternative to traditional multicomponent, multivendor solutions. The study highlights such an appliance from Netezza, an IT vendor offering its Netezza Performance Server as a solution proven to increase performance 10 to 20 times over competing solutions at a substantially lower price.

SITUATION OVERVIEW

INTRODUCTION

In 2001, the data warehousing tools (DWT) market reached \$6.5 billion and is expected to grow at a compound annual growth rate (CAGR) of about 10% through 2006. The DWT software market is segmented into three primary segments, including data warehouse generation, management, and access. The access market is composed of a portion of business intelligence (BI) tools, itself a \$3.7 billion market in 2002 that is expected to grow at a CAGR of about 10% over the next five years. However, the largest segment of DWT is data warehouse management, which is dominated by the RDBMS vendors such as IBM, Oracle, Microsoft, Teradata, and Sybase.

While some of the earliest products in DW and BI can be traced to early 1970s, the first major growth spurt occurred in the mid-1980s. Since then, the market has expanded rapidly and continues to be one of the bright spots in the otherwise lackluster IT market of the past two years.

As the competitive environment has intensified, business analytics have taken on an increasingly strategic role as tools that enable analysis and optimization of both revenue-enhancing and cost-control processes. Analysis of sales trends, customer behavior, supplier performance, inventory levels, financial trends, fraud, and other decision-making processes is based on the successful implementation and utilization of business analytics software.

Over the years, data analysis has been handled by IT architectures combining numerous software, server, and storage technologies. Most of today's business analytics environments are based on standard RDBMS and ETL software used for generating and managing data warehouses, with BI tools and analytic applications as end user-facing software that is implemented on top of such data warehouses.

These solutions often fulfill their intended goal of providing reporting, multidimensional analysis, and data mining support. However, they come with their own challenges:

- The requirement to perform considerable integration services to make the various software and hardware components work together.
- Constant monitoring, tuning, and maintenance by database and systems administrators to optimize the integrated solution.

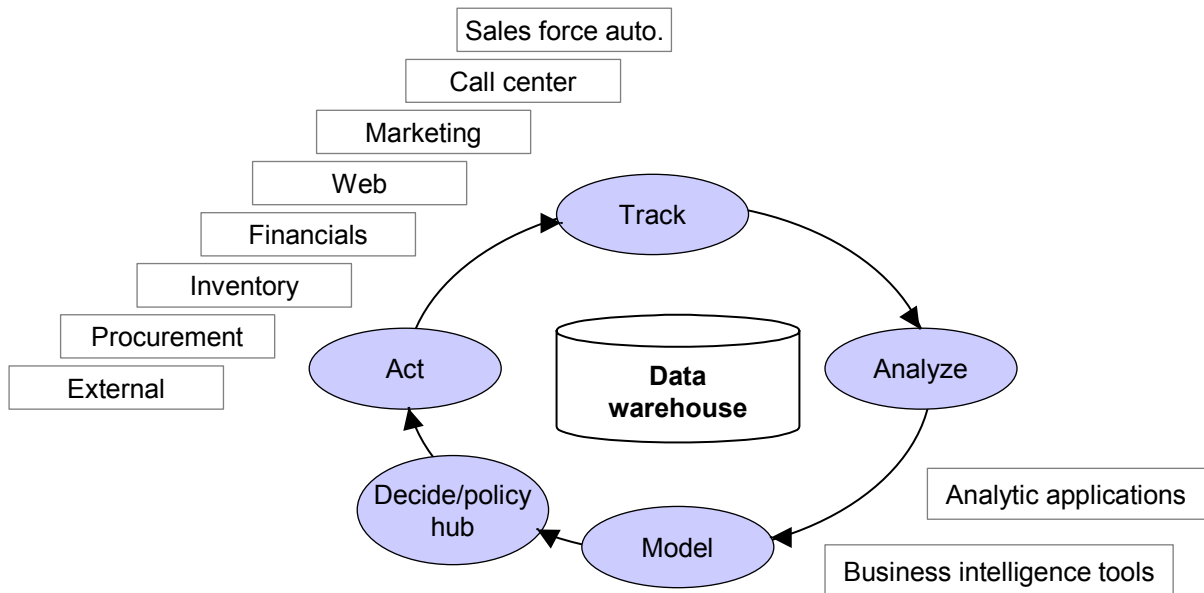
- ☒ The need for continuous upgrades to keep up with growth in user populations and growth in the volume of data that organizations are interested in tracking, storing, and analyzing:
 - ☐ A research study conducted in 2002 by DM Review and IDC showed that 65% of all organizations expect their data warehouses to grow by at least 100% over a three-year period.
 - ☐ Market trends indicate that user populations are growing both within and outside organizations, with suppliers, partners, and customers increasingly getting access to information held in data warehouses.
- ☒ New user requirements based on the need for faster (real-time or near-real-time) access to data for analysis dictated by either competitive forces or legislative mandates such as the USA Patriot Act, Sarbanes-Oxley, or Basel II.

These trends present organizations with the challenge of placing increasing demands and stresses on their business analytics infrastructures while facing IT budget constraints.

To illustrate the role of business analytics software in organizations, IDC developed the model for closed-loop applications shown in Figure 1.

FIGURE 1

CLOSED LOOP APPLICATION MODEL



Source: IDC, 2003

The five major steps of the closed-loop application model are:

- ☒ **Track.** The process of extraction, transformation, loading, and integration of data into a data warehouse.
- ☒ **Analyze.** The process of analyzing the data using business intelligence tools such as query and reporting, multidimensional analysis, and data mining.
- ☒ **Model.** The process of formulating models based on the analysis, using various descriptive and predictive statistical methods, and resulting in scoring or other models used in decision making.
- ☒ **Decide/policy hub.** The process of arriving at a decision based on analysis and preexisting or newly developed models, which often combines individual and group input, facilitated by collaboration tools or personal interaction.
- ☒ **Act.** The process of acting on the decision based on the particular business process being addressed. Examples include:
 - ☐ Launching a new marketing campaign based on the analysis of previous campaign results, customer behavior, a new promotional plan, or inventory levels
 - ☐ Approving or denying a request for credit based on past financial activity
 - ☐ Renegotiating sourcing contracts based on supplier delivery trends, product quality, and warranty activity trends

While other closed-loop systems have been developed in academia and the military, they share the notion of a sequential process through the steps, with human and information technology processes interacting in various combinations. The goal of organizations in all cases is to traverse the steps of the closed loop as fast as possible.

Compression of tracking-analysis-decision time and acceleration of the decision to action time have been the basis of competitive advantage in the past, and are even more so in today's global economy.

However, speed of decision-making — whether we're talking about real-time or near-real-time cycles — is still only a goal for most organizations. There are those that have achieved a desired level of real-time decision-making capabilities — for example, the U.S. military (as shown in the current campaign in Iraq) and some fraud-detection applications in financial services or quality control applications in manufacturing.

For the vast majority of organizations, however, current DW platforms present a roadblock to achieving such a goal, in a time of increasing data volumes and user populations, and stable or decreasing IT budgets.

One approach to resolving the limitations of most current DW platforms is to look at the DW appliance as an alternative to the patchwork of software, server, and storage offerings from various vendors. A DW appliance provides a dedicated, purpose-built solution that eliminates the need for most of the integration work, while providing an optimized environment for performance gains that would satisfy real-time analytic needs.

DATA APPLIANCES

Are data appliances new? If we look back far enough the answer is a resounding no. The following excerpt is from *Back to the Future: Trends in the Overall IT Market Portend Major Change for the Packaged Enterprise Applications Market* (IDC #25822, November 2001):

Beginning in the 1960s, lasting well into the 1990s, most users deployed an enterprise application dependent on their underlying infrastructure hardware and software. The "stacks" were characterized by the following examples:

Only a few real "packages" (e.g., the GE Model 58 or System/3 RPG-based accounting features) were available.

Most applications were heavily customized behind the glass doors (e.g., Honeywell's initial Bill of Material Processor) or the market experienced total application customization (e.g., AT&T communications switches).

In addition to the obvious hardware/OS dependence, the OS was highly dependent on a proprietary file system.

For a variety of technical and nontechnical reasons (e.g., U.S. government antitrust actions versus IBM and AT&T), the stacks began to come apart during the 1970s for the following reasons:

Separate file system concepts (DRDA from IBM and Multics from MIT/Bell Labs/GE/Honeywell) began to emerge from within application/infrastructure stacks and from within the operating system.

These two emerging "band"-like technical trends (separate file systems and operating software) in a sense became the separate database and the minicomputer markets respectively, and the two enjoyed a symbiotic relationship for 20 years.

Throughout the 1980s, the movement to bands accelerated. By the end of the 1990s, it looked like the stack was history, but it never really went away. As the 2001–2010 decade rolls in, the movement to bands seems stalled, and stacks are back.

While the above excerpt references enterprise applications, in data warehousing, a stacked solution has been provided by Teradata for years. However, several recent trends have enabled a new type of purpose-built, low-cost, high-performance data warehousing appliance. These trends include:

- Evolution of mature database standards such as SQL, ODBC, and JDBC
- Availability of standard hardware components at rapidly falling prices
- Emergence of better-understood parallel query optimization methods

An IT vendor currently providing such an appliance is Netezza.

COMPANY BACKGROUND

Founded in September 2000, Netezza develops, builds, and markets the Netezza Performance Server — a dedicated, purpose-built data warehousing appliance that combines software, server, and storage technologies to deliver unparalleled performance at significantly lower prices than other current solutions.

Headquartered in Framingham, Massachusetts, Netezza is led by CEO and cofounder Jit Saxena, a veteran of the BI industry and former founder and CEO of Applix. The company has attracted other management and staff from Compaq, Sun, Cisco, and Teradata, and counts among its board members Ed Zander, former president and COO of Sun Microsystems.

Netezza is backed by three major VC firms with over \$33 million in funding. The company has 75 employees, and its initial sales efforts were focused on direct sales through a team of 10 sales and presales technical experts. However, Netezza's strategy is to balance direct sales with indirect channel sales. The company has so far signed two direct customers, recently closed a deal through one of its channel partners, and is in the late proof-of-concept stage with another two companies in the United States and the United Kingdom.

PARTNERSHIPS AND ALLIANCES

Netezza's data appliance for business intelligence is based on industry-standard components and interfaces compatible with existing ETL and BI tools and analytic applications.

The company has announced partnerships with several leading BI software vendors and integrators. Microstrategy, SPSS, and InfoGain have joined Netezza's Advantage Program — a program that provides joint development, support, and sales initiatives. In addition, Netezza itself has joined Business Objects' technology alliance partner program, recognizing the compatibility of the offerings of the two companies.

However, because of its use of industry-standard interfaces, Netezza has already been proven to work at client sites with other BI and analytic applications vendors such as SAS, Unica, Crystal, Hyperion, and MS Excel.

NETEZZA PERFORMANCE SERVER

Netezza Performance Server is a data appliance optimized as a dedicated data warehousing platform for large-scale (multiterabyte) environments where real-time access to data and ad hoc, complex queries are common. Through its patent-pending solution that integrates RDBMS software, server, and storage components, Netezza has developed an appliance that has demonstrated lower costs and 10 to 20 times processing improvements over existing alternative solutions that combine offerings from established database, server, and storage vendors.

NPS 8100, Netezza's base product, is the smallest configuration appliance in a series of appliances with increasing technical characteristics and cost. At a starting price of just over \$600,000, the NPS 8100 is based on standard, off-the-shelf components and comes as one standard 19in. rack with two-way Pentium III Xeon 2.8GHz host processors, 114 processors dedicated to its patent-pending hardware-aware Intelligent Query Streaming, and 4.5TB of total data storage. NPS 8100 software includes Red Hat Linux OS, a database built on an open source PostgreSQL, and support for SQL, ODBC, and JDBC APIs. Currently, there are another two larger NPS appliances available with either two or four racks, 226 or 450 processors, and total data storage of 9TB or 18TB. Additional software modules for all NPS appliances

include a Windows-based administration and data loader and a Web-based monitoring module.

The basis for NPS is the DBMS, server, and storage disk integration based on its Asymmetric Massively Parallel Processing (AMPP) architecture and Intelligent Query Streaming. With this architecture, Netezza is able to capitalize on the best features of both symmetric multiprocessing (processing power) and massively parallel processing (scalability) environments. Symmetric multiprocessing (SMP) systems are usually considered best for OLTP systems, while massively parallel processing (MPP) systems are often found in analytic or data warehousing environments.

The queries are managed through Netezza's Intelligent Query Streaming (IQS) method that optimizes flow of data throughout the system by limiting data movement operations. Silicon processors are placed near storage to enable streaming database operations right at the source of the data, bringing over only the relevant data for queries, thus minimizing I/O and network transfer resource needs.

Current performance is not limited by CPU speeds; instead, bottlenecks are due to the limitations of disk transfer I/O and network transfer rates. Therefore, attempts at gaining scalability by just increasing CPU and memory provide only limited benefits.

AMPP with IQS enables extremely fast queries, such as large multidimensional aggregations, ad hoc aggregations, 10- to 12-way joins, and building of summary tables and refreshing them when data is loaded. The combined solution spreads data across multiple nodes or snippet processing units (SPUs), each with its own dedicated CPU, memory, controller, and disk, eliminating the bottleneck of slow disk I/O. With each additional disk spindle, performance can be further increased linearly.

Initial data processing, such as sorting and aggregation, is performed in parallel, close to the disk within the SPUs. Early filtering of data is accomplished through SPUs, each of which can evaluate simple "where" clauses as data streams off the disk. The result is less data flowing through the system, as irrelevant data is excluded before it gets into memory. The CPU and memory of each SPU is therefore optimized for processing only relevant data.

The SMP host coordinates communication among SPUs, including the aggregation of preprocessed data from each SPU. It optimizes the query execution plan by breaking the query into snippets that are executed in parallel by SPUs. This optimized approach is based on automated evaluation techniques that use systems performance statistics gathered periodically.

NPS also offers advanced mirroring techniques that ensure reliability.

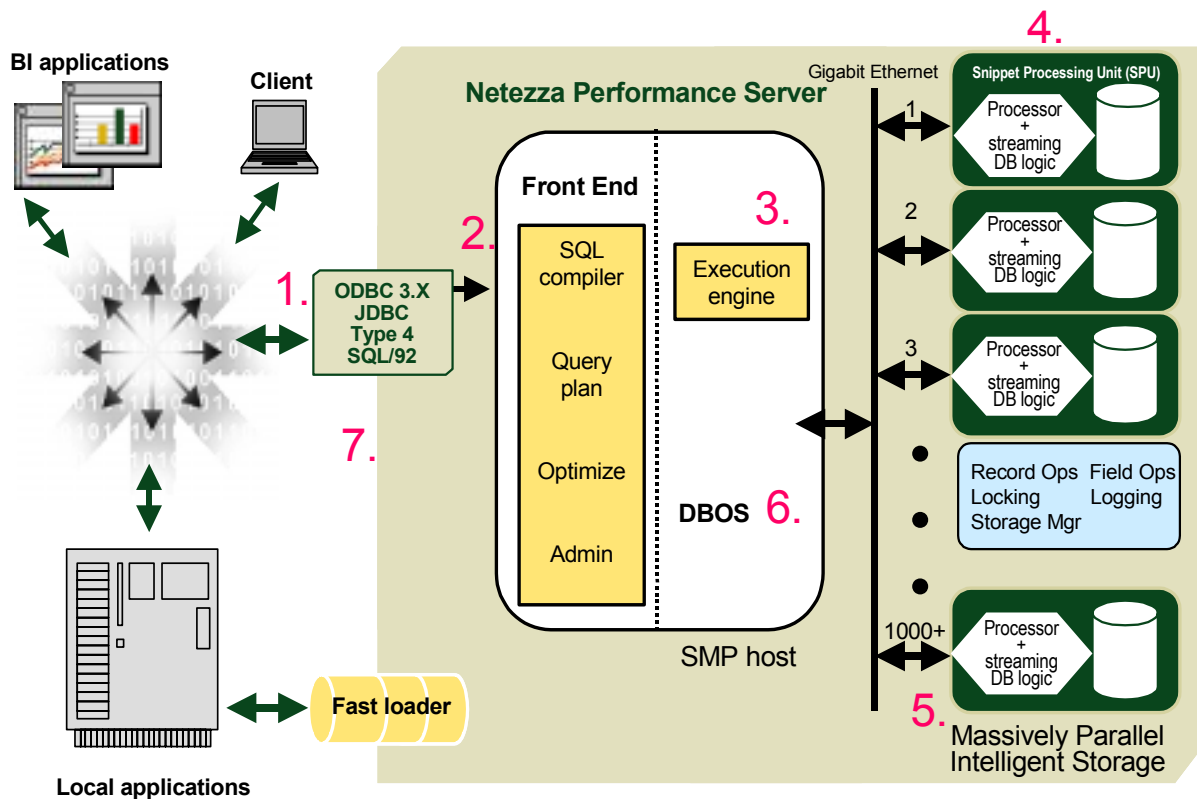
A query is handled by Netezza in the following manner (see Figure 2):

1. A BI tool and analytic application launches an SQL request.
2. The query is sent to the SMP-based host, which compiles queries into parallel execution plans for hundreds of SPUs and processes eventual aggregation of query results. Note that the executable is compiled into machine instructions, further enhancing processing speed.
3. The host compiles the plan and ships the plan to SPUs through a Gb switch.
4. Each SPU initially performs row and column filtering processes (before the data is brought into memory) to isolate only the required data from storage disks, thus minimizing subsequent I/O needs.

5. Preaggregation is then performed in memory, and results are returned to the host.
6. The host aggregates the aggregates (i.e., combines the results from all SPUs that were processing in parallel).
7. Results are returned to the client applications through ODBC/JDBC.

FIGURE 2

NETEZZA PERFORMANCE SERVER — ASYMMETRIC MASSIVELY PARALLEL PROCESSING



Source: IDC, 2003

CUSTOMERS

At present, Netezza has three clients, with several others testing the NPS appliance. While the company's first two clients were acquired through direct sales, there has already been the first indirect channel sale to a large telecom, which purchased NPS with one of its partner applications. In addition, the company is in the final stages of negotiating its first deal in Europe, again through one of its partners.

VIBRANT SOLUTIONS

Vibrant Solutions is a cost and revenue management software and services company dedicated to the telecommunications industry. Headquartered in Fairfax, Virginia, the company was founded in 1991. One of its main business lines is cost management,

where the company processes and audits carrier-to-carrier charges. This process allows its telco clients to recapture millions of dollars in extra bottom-line savings. In recent years, as the industry has fallen on hard times, telcos have placed greater emphasis on scrutinizing such bill errors, which represent 2–5% of their total cost. Based on information provided by Vibrant, a study by the Deloitte & Touche estimates revenue leakage to be a \$100-billion industry problem.

With all usage-based services, like long distance and wireless services, for telcos to generate a bill, they need access to call records, which are collected from switches. However, given the reality of switch errors, incorrect call details often lead to errors in billing and subsequent direct top-line impact, even before fraud or credit issues, which further cut into revenue.

Vibrant's solution to the industry's problem is its Network Usage Assurance (NUA) application. The application helps increase billable revenue by monitoring the entire network to detect losses and allow for the speedy recovery of otherwise unbilled events.

Until now, standard industry practice in managing revenue leakage was to either verify integrity by using test calls and checking their (billing) accuracy as the calls go through switches, or to use data sampling techniques and base the analysis of the entire network on a fairly small sample. Both methods are inherently limited in their ability to provide a true picture of call-detail records — the basis for recovering lost revenue. The ability to monitor the entire network all the time addresses these limitations.

The bottleneck to implementing a system that constantly monitors all network activity and provides complete visibility was the lack of scalable technology at affordable prices. Given the daily volume of records — ranging from hundreds of millions to billions of records — Vibrant was unable to process the necessary queries in a reasonable time to make the analysis useful in lost-revenue recovery.

For Vibrant, this bottleneck was removed when it selected Netezza's NPS in October 2002, after a period of extensive performance testing. As Andrew Hurrell, director of product marketing, who is responsible for launching and planning new software and services offerings, put it: "Vibrant is now revolutionizing the process of revenue leakage management. Through constant monitoring, Vibrant's NUA application can detect problems sooner and more accurately." The add-hoc analysis that is needed on the massive amounts of data is now possible with the NPS appliance. For example, Vibrant has been able to perform unstructured queries on Netezza's tera-scale, unindexed database with 1–2 minute response times.

Vibrant's NUA application based on its ViewLogic software platform includes visual displays such as dashboards that, in addition to ad hoc querying, provide early warning detection and reporting functionality. In the current process, Vibrant acquires the call detail data from existing mediation and collection systems that poll switches. (It is common practice to use at least two of such collection systems to verify data quality.) Data is then loaded into NPS through hourly feeds. On average, 200 million call records are added per day, and 60 days of history are kept prior to data archival. The size of Vibrant's Netezza-based data warehouse is currently approximately 6TB. With 24 hours worth of data, a query that previously took six hours to process is now completed in 1.5 minutes, and a query on 120 days of history takes only 20 minutes.

In Vibrant's case, the experience with NPS has been truly revolutionary. The new architecture for NUA doesn't simply improve previous processes, it enables previously impossible ones. For example, prior to installing Netezza, it was impossible to run an ad hoc, unstructured query on 1 billion records overnight (without prior indexing). The benefits of the new NUA include support for increased accuracy in usage billing and timely detection of billing errors that supports immediate corrective action, significantly reducing the potential for back-billing.

While in the past Vibrant has realized more revenue from its cost management business line, given the new capabilities, Andrew Hurrell expects the revenue management business to provide faster growth.

Because Vibrant's business model is based on tight service level agreements (SLAs) and contingency (win split) or fixed fee-based revenue, delivered either on a licensed or outsourced basis, the new capabilities in analyzing call detail records for revenue leakage management provide it with a new competitive advantage.

Vibrant has been using NPS for about a year. Following the initial installation, which took a few hours, the NPS appliance has not had any service-affecting failures.

EPSILON

Epsilon, headquartered in Burlington, Massachusetts, has a 30-year history in marketing services. Epsilon is a relationship marketing provider specializing in the strategic development, deployment, and management of online and offline solutions.

With leading technology and functional experience, Epsilon delivers solutions with bottom-line impact and gives clients an edge in improving their direct dialogue with their most valuable prospects and customers. In November 2001, Epsilon was acquired by Relizon a document imaging, billing, and fulfillment company. Epsilon lists among its clients many leading travel, financial services, and telecommunications companies.

Like many organizations, Epsilon continues to have various server (mostly Unix from Hewlett-Packard [HP], IBM, and SGI) and database software (Informix, DB2, SQL Server) installations to support its data warehousing efforts. However, more recently, the company has mostly standardized on an Oracle/SUN platform. In addition, the company uses both Business Objects and Crystal Decision's BI software for accessing the data warehouse.

Epsilon's business model consists of building IT systems for marketing departments of their client companies or hosting such systems for the clients. Today, over 40 hosted systems exist that provide either full service or just access to data.

Cost, performance, and scalability are all key aspects of Epsilon's success, especially in hosted engagements, where lower costs of delivering agreed-upon service levels directly improve Epsilon's profitability. In addition, managing the hardware and software maintenance and support costs (especially for its larger clients, where such costs can run over \$1 million per year) has been a challenge for Epsilon in the past.

Currently, Epsilon is in production with NPS 8100 as the data warehousing platform for data analysis for one of Epsilon's business units. In this environment, NPS is supporting the company's campaign management application provided by Unica and BI tools from Business Objects. Both software packages have worked well with Netezza through ODBC with ANSI standard SQL. In fact, Bob Doyle, vice president of Database Architecture at Epsilon, noted that he was pleasantly surprised at the ease of application and the BI tools' interoperability with NPS.

Mike Coakley, vice president of Marketing Technology Products at Epsilon, noted: "In selecting NPS we couldn't rely on industry experience with the product, given that it's fairly new to the market. Epsilon therefore went through rigorous testing of NPS prior to acquisition. We were impressed not only with the performance but also with the ease of maintenance of NPS."

Epsilon analysts are also running complex SAS queries on NPS. Some of these queries, which took over two hours on an Oracle/SUN platform, have been reduced to less than 15 minutes on NPS. The data warehouse load times have also improved: For example, data load times have been reduced from 11 hours to 3 hours.

Epsilon's experience has shown that maintaining NPS does not require any special training beyond common DBA skills and, in general, requires only minimal intervention. With its service-based business model, Epsilon expects significant cost reductions from NPS, which will in turn provide the company with competitive advantage manifested in the ability to submit more competitive bids for new business and reduce costs in servicing existing clients.

WELLS FARGO

Wells Fargo is 150 year-old diversified financial services company providing banking, insurance, wealth management and estate planning, investments, mortgage, and consumer finance services. Headquartered in San Francisco, Wells Fargo has \$349 billion in assets and 134,000 employees. The company was ranked fourth in assets and third in market capitalization as of December 31, 2002, among its peers.

One of the company's business units is Wells Fargo Home Mortgage (WFHM), whose analytics data management group is based in Maryland. Some of the analytic business processes within WFHM include customer behavior analysis, delinquency analysis, and several other risk management-related processes performed by analysts using modeling, data mining, and high-performance statistical methods.

WFHM has a central data warehouse and several data marts. The central data warehouse is loaded with operational, servicing, and credit history data. In the past, the analysts in WFHM have used an analytic environment from SAS Institute for doing forecasting and analysis. The data marts have been based entirely on SAS tools, where flat files are loaded into relational-like schema for sophisticated statistical analysis. However, this architecture resulted in WFHM running into storage and performance limitations. For example, it took too much time to process and prepare data for final analysis. To solve these issues, WFHM set out to find a more scalable data warehousing platform for its applications.

WFHM's analytic data management team, led by Frank Eichorn, director of Data Management, was first introduced to Netezza in mid-2002, when the company decided to test the new NPS appliance by providing a subset of data to be processed at Netezza headquarters. After seeing dramatic improvements in processing speed, WFHM took the second step in its evaluation process, which involved setting up a NPS appliance at its site in Maryland. WFHM ran benchmarks of NPS 8100 against a comparable Teradata configuration (four-node 5350, 1.8TB storage), which is also used at Wells Fargo, and found 5 to 10 times processing speed improvements at a quarter of the price.

From a business standpoint, WFHM is now able to get faster answers to its clients and credit and risk analysts. In the past, analysts have been limited in the number of iterations they have been able to run during a fixed time period to develop decision models. With NPS, analysts are able to perform 10 to 20 times more iterations (or "what if" analyses), resulting in more robust models for:

- Marketing campaigns
- Loss forecasting
- Prepayment analysis
- Credit forecasting
- Delinquency analysis

The new data management environment is based on an NPS data mart that sources its data from the central data warehouse and is currently accessed by 200 users. These include both ad hoc users, such as analysts and managers, and model builders. The former group does so using Business Objects, while the latter group continues to use SAS analytic tools for model-building by pointing modeling and statistical tools directly at SAS tables running on NPS, without the need to extract and prepare data.

New data is loaded into the data mart monthly, with some loads taking place weekly. Today, the data mart size is 750GB, and it is expected to grow 15–25% per year. While NPS also improves the load speeds, this aspect of the appliance was not a primary consideration for Wells Fargo, since data loads have not posed a problem historically. It is the processing speed improvement of NPS that is making the difference for WFHM. For example, query processing time on 300 million rows of servicing or payment history data has decreased from five hours with SAS running on Sun to 47 seconds with SAS running on Netezza.

Frank Eichorn has also been impressed with NPS's ease of maintenance. His team of seven staff members (including two DBAs), who provide all analytic data management support for WFHM, did not need any specialized training for maintaining NPS. WFHM has been using NPS for one of its data marts for about four months, and as Frank Eichorn put it, "DBAs don't pay more than one hour per week attention to it. There are no indexes, no triggers, or materialized view to manage, so ongoing tuning and support are unnecessary."

The only point of failure so far has been an instance of storage disk crash. However, Netezza's mirroring with automatic failover functionality worked as intended, without any actual downtime for the NPS. During the recovery time, there was some processing time slowdown; however, as soon as primary storage was restored, processing speeds increased to previous levels. It is interesting to note that even the slowdown in processing speed compared favorably with historical results seen at WFHM prior to the NPS implementation.

While NPS has been installed for one of the data marts at WFHM, its scalability has not been tested yet beyond this mart. Since WFHM keeps history "forever," its needs include management of 30TB+ data warehouses, which the company continues to run on Teradata — a solution well known for hosting some of the world's largest data warehouses. Thus, the coexistence of the two best-known purpose-built data warehousing platforms in one organization is likely to continue for the foreseeable future.

FUTURE OUTLOOK

ANALYSIS OF CHALLENGES AND OPPORTUNITIES

The time for the purpose-built, business analytics–dedicated, open standards–based data appliance may indeed have arrived. NPS is certainly a revolutionary offering. While it may be possible today to replicate NPS's performance using general-purpose database, server, and storage products, it would be impossible to do so at comparable prices.

The opportunities for Netezza abound, especially in environments of complex, ad hoc queries. Such environments are often the case in the telecommunications and financial services industries — specifically in the credit and insurance segments of the latter, where teams of analysts run sophisticated analytic routines on a regular basis.

Other opportunities will continue to come from service providers whose business models rely on providing a certain level of service at the lowest cost, including that of IT installation and maintenance.

Not only are the performance characteristics of NPS impressive, but the ongoing maintenance requirements are also attractive. For example, database indexing is optional, and much of the tuning by DBAs and network administrators is not needed.

Netezza's closest competitor is Teradata, a company that dwarfs Netezza in size and brought to market the first server/database stack dedicated to data warehousing. However, the current target markets seem to be slightly different for these two vendors — and (at least at present) Netezza NPS has a price/performance advantage over Teradata and other industry solutions (e.g., Oracle/Sun or Sybase/Sun), as seen in initial comparison tests.

Netezza is likely to thrive in environments of strong ad hoc query demand with sizable data volumes. It is likely to see data mart implementations focused on particular business processes. On the other hand, Teradata continues to be a strong data warehousing solution for very large data warehouse (VLDW) environments that sometimes reach into the 100TB+ range.

However, regardless of the nuances, the competition to existing data warehousing vendors from Netezza will only serve as a positive market force in fostering further technical innovation and price competition, ultimately benefiting end users.

Finally, Netezza will have to deal with the issue of being a newcomer, with the usual questions from prospects about long-term viability and survivability. One argument to counteract such concerns would point to the standard components and interfaces of NPS, which would make any potential porting of applications running on NPS to another standards-based system fairly painless.

ESSENTIAL GUIDANCE

As Netezza proceeds in executing its stated market strategy, its partners and potential clients should consider the guidance outlined below.

Although NPS is a horizontal offering, it is likely to continue to pursue industry verticals with "big data" issues. Initially, these will include financial services and telecommunications, where the volume of data growth is high, and where end users are comprised of large populations of analysts conducting sophisticated analysis requiring complex queries and immediate access to information.

Other likely initial applications will come from click-stream data analysis applications; these can span industries, but are more prevalent in retail. Finally, service providers to the abovementioned industries are attractive targets for Netezza because of the reliance on service level agreements. The performance improvements of NPS can provide strategic advantage to such service providers by enabling introduction of more competitive bids and improving customer service levels.

At the same time, Netezza's best customers are those that have tried to perform the type of analysis that NPS enables, but have failed due to the performance and scalability limitations of their existing solutions in an environment of tight IT budgets. Some prospects that have never even tried to run ad hoc queries on several hundred-million records may be in a position of not realizing the benefits or not knowing what they are missing. Such cases are likely to complicate Netezza's selling process.

Netezza should also continue to aggressively develop its channel sales with partners that should be encouraged to develop proof-of-concept applications running on Netezza NPS.

BI and analytic applications vendors should acknowledge as quickly as possible that their own software will get a better response and be used by more end users if the performance of the underlying data warehouse platform can deliver processing speeds 10 to 20 times faster than most competing platforms.

Organizations facing performance and scalability challenges in their efforts to capitalize on the growing availability of data should consider test runs of Netezza NPS. Such initial tests can be conducted by simply providing Netezza with data that will be processed at Netezza headquarters. Therefore, there is no direct cost to testing, and the cost of the time to perform the test run might pay for itself quickly if an organization takes advantage of the potential performance and cost benefits offered by Netezza: By compressing decision cycles, NPS can provide organizations with competitive advantage by allowing them to react faster to competitive pressures, customer and prospect demands, and supplier or partner initiatives.

LEARN MORE

RELATED RESEARCH

- ☒ *Worldwide Business Intelligence Software Forecast, 2003–2007* (IDC #29035, March 2003)
- ☒ *Business Analytics Implementations Challenges: Top 10 Considerations for 2003 and Beyond* (IDC #28728, January 2003)
- ☒ *Worldwide Data Warehousing Tools Competitive Analysis, 2001: Market Concentration* (IDC #28031, September 2002)
- ☒ *Worldwide Data Warehousing Tools Preliminary Forecast Summary, 2002–2006* (IDC #27401, June 2002)
- ☒ *Back to the Future: Trends in the Overall IT Market Portend Major Change for the Packaged Enterprise Applications Market* (IDC #25822, November 2001)

COPYRIGHT NOTICE

This IDC research document was published as part of an IDC continuous intelligence service, providing written research, analyst interactions, telebriefings, and conferences. Visit www.idc.com to learn more about IDC subscription and consulting services. To view a list of IDC offices worldwide, visit www.idc.com/offices. Please contact the IDC Hotline at 800.343.4952, ext. 7988 (or +1.508.988.7988) or sales@idc.com for information on applying the price of this document toward the purchase of an IDC service or for information on additional copies or Web rights.

Copyright 2003 IDC. Reproduction is forbidden unless authorized. All rights reserved.

Published Under Services: Business Intelligence and Data Warehousing Strategies;
Analytics and Data Warehousing Software