**Data Warehouse Optimization**

Server virtualization helps analytic database take accelerated business insight to a new level.

The ability to transform data into information and knowledge is at the center of business success. Pertinent information is required by decision-makers at every level of the organization. What’s more, an uncertain economic environment makes effective business intelligence (BI) even more important.

Effectively managing data and extracting value from it isn’t lost on most business and IT executives. Yet, nowhere is the task more challenging than in today’s massive data warehouses. Relentlessly accumulating real-time data, the sheer size and complexity of today’s repositories stretch many organizations to their limit — and sometimes beyond.

As a result, many companies are now searching for a more effective way to leverage their data. At the center of this trend is a move toward much higher-performance analytic database systems and data warehouse appliances. These solutions typically consist of an integrated set of servers, storage and an operating system, along with database software, that’s specifically designed for complex analysis, data mining and reporting.

A key player in the analytic data warehouse space is Cupertino, Calif.-based ParAccel Inc. Leveraging the benefits of high-performance and existing enterprise ecosystems, the firm’s flagship ParAccel Analytic Database (PADB) software has proven to take data analysis to record levels. And taking note are large and small firms, like OfficeMax, Provisio Inc. and others, looking to better address sophisticated and ad hoc analytic requirements.

**GOING WITH THE DATA FLOW**

The idea of storing data in a huge repository is nothing new. The first data warehouses were created in the 1980s. They provided architecture capable of moving data from operational systems to decision support environments — all while minimizing redundancy and costs. Before this, many organizations were forced to utilize multiple decision support environments.

Data warehousing has evolved over the years. However, along the way, storage requirements have swelled from gigabytes to terabytes and petabytes. Despite complex and robust architectures designed to offer high scalability, the ability to execute tasks in these environments has begun to bog people down.

In fact, the tech analyst and research firm Gartner notes that clients increasingly report performance-constrained data warehouses. The firm goes on to estimate that nearly 70 percent of data warehouses experience performance-constraining issues of various types.

Current economic conditions, including the financial crisis and subsequent recession, have also contributed to changes in the world of database management. Further, there is much more interest from upper-level corporate management to achieve greater data quality, understanding and transparency all across the enterprise.

“Many organizations are now pushing performance to an extreme,” says Dipesh Patel, senior manager of product marketing at ParAccel. “The length of time to complete a query or answer particular questions can take hours or days using traditional database processes.”

**DATA WAREHOUSE SOLUTIONS**

Enter high-performance analytic database solutions and data warehouse appliances. These solutions — servers, storage, operating system plus database software — can be modular, making them easy to install and scale up as needed.

In some cases, they can even be configured as a cluster with the analytical database software preinstalled on the appropriate servers and storage. They rely on a massively parallel processing (MPP) environment to divide tasks across independent servers and processors, an approach that offers superior performance along with high scalability.
Founded in 2005, ParAccel is a thriving participant in this emerging market segment. The 100-person software company provides what can be said to be one of the highest-performing analytic databases on the market today. (See below, “World Record for Analytic Performance in Virtualized Environment.”)

ParAccel's PADB, a high-performance database system, aims to take data analysis to a higher level and, in the process, reduce processing time and human involvement, Patel says. “The goal is to bring greater performance to the data warehouse.”

PADB is designed to spot patterns, relationships and correlations through complex and iterative analysis. It also offers operational agility, including an ability to explore any hypothesis regardless of the business questions.

In the end, it allows the enterprise to respond to fast-changing business conditions while there still is a chance to shape the future. These speed and efficiency gains provide a foundation for better and faster performance. Yet, as Michael Weir, senior director of marketing at ParAccel notes, boosting performance also allows an organization to ask questions and receive answers so quickly that it's possible to venture into new territory.

In some cases, Patel says, PADB reduces the time it takes to complete a task by 95 percent or more. But the benefit extends beyond time and expense. “The ability to drill down to an unexpected answer that nobody else has previously tapped into is extraordinarily valuable,” he says. “Organizations have an opportunity to gain a competitive advantage by leveraging data more efficiently.”

BUSINESS TAKES NOTICE

The analytic database warehouse is designed for a wide range of industries and situations. These include government, educational institutions, financial services, manufacturing, online marketing, pharmaceuticals, retail, cloud-based analytics and information technology.

OfficeMax, for example, has turned to a ParAccel solution to ratchet up its analytics and drive more efficient buying decisions. Improved analysis, data mining and reporting allow the electronics and office products retailer to conduct much more frequent market analysis to better understand customer behavior. The firm can even do this on an intra-day basis, if needed.

OfficeMax decision-makers come away with a deeper understanding of the types of products that customers are interested in. And it provides valuable insight into what products they will add to their real and/or virtual shopping carts.

“The answers they get allow OfficeMax executives to figure out how to position merchandise, bundle items and take advantage of product combinations,” Patel says. “The strategy is designed to boost the number of products in shopping carts and, in the end, increase the retailer's share-of-wallet.”

Another firm embracing the analytic database concept is Provisio Inc. The Nashville, Tenn.-based healthcare company develops and markets technologies and services that enhance clinical trial processes. The company focuses on developing iTrials, which automate and streamline the way drug trial candidates are identified, notified and enrolled.

The process includes leveraging an iTrials data warehouse — with the comprehensive longitudinal health histories on over 70 million patients. Here, patient level data is combined to provide valuable insight into patient populations and identify success factors which aid in protocol design, site selection and patient recruitment.

In the past, Provisio managed its database across several SQL Server clusters — totaling approximately 230,000 database tables, Patel notes. Pinpointing the relevant patient information was difficult and time consuming. In some cases, queries could take weeks to complete.

Today, Provisio uses PADB on a Linux platform to deliver patient information in columnar form. “The firm has consolidated 230,000 tables down to just 12,” Patel says. “In many cases, answers are available within seconds or minutes.”

One thing that's particularly attractive to companies is the fact that ParAccel's PADB doesn't require specific hardware. It's able to run on a diverse array of server and storage systems from the likes of such vendors as HP, NetApp and Oracle (Sun Microsystems) among others.

“This makes it possible to install the software using existing server and storage standards in the data
“Companies are able to approach things in a less disruptive way and configure the environment to their specific price and performance requirements.”

**SERVER VIRTUALIZATION CONNECTION**

Deploying ParAccel’s PADB in a virtual environment could allow a business to achieve record levels of analytic performance for their private cloud. Such was the case when the firm established an industry benchmark for analytic processing performance in April 2010.

“VMware has revolutionized IT with virtualization solutions that enable efficient hardware utilization, reduced administrative costs and significant energy savings,” says Barry Zane, CTO and founder of ParAccel. “Organizations have been evaluating virtualization for one of the most hardware-intensive and vital corporate utilities — complex analytic processing.

“Working with VMware, we’ve been able to show the industry that the powerful compression and processing capabilities of the ParAccel Analytic Database can be virtualized to deliver exceptional performance and savings,” he adds. “We have demonstrated new levels of analytic performance and IT efficiency in standard deployments.”

Make no mistake, the data warehousing landscape is changing and ParAccel is at the vanguard of this revolution. As a growing array of organizations turn to analytics to slice and dice data in new and more efficient ways, they’re looking for faster and more efficient ways to get the job done.

Concludes Patel: “Managing data more effectively and gaining insight into patterns, trends and behavior is critically important. In today’s business environment, organizations that put data to maximum use gain a distinct competitive advantage.”

**World Record for Analytic Performance in Virtualized Environment**

In April 2010, ParAccel established an industry benchmark for analytic processing performance. This included achieving the fastest 1TB Transaction Processing Performance Council Benchmark H (TPC-H) on record, using VMware vSphere 4, with 7.7 times better price/performance than the prior performance record holder.

This combination beat the previous TPC-H record using 37 percent fewer servers and with a superior database load time that was 8.7 times faster than the previous performance record holder. This new benchmark result establishes a path for organizations to create and deploy on-demand, ultra-performance analytic processing in private clouds.

**Specifications**: An 80-node cluster of ParAccel Analytic Database (PADB) version 2.5 consolidated on 40 HP ProLiant DL380 G6 Servers running VMware vSphere 4 achieved 1,316,882 Composite Queries per Hour (QphH) @ 1000GB, a price/performance of U.S. $ .70/QphH. The benchmark summary and full disclosure reports are available online at www.paraccel.com or www.tpc.org. The TPC Benchmark H (TPC-H) specification is available at www.tpc.org.

**Six Tips to Better Server Virtualization**

1. **Conduct a comprehensive assessment of the server environment.**

   This process will help to identify which pieces of hardware you can and should virtualize. Use free tools from vendors such as VMware and Microsoft to help identify exactly which parts of the IT infrastructure are good candidates for consolidation and virtualization.

2. **Ensure that applications are compatible with virtualization software.**

   Most virtualization software vendors will help a business determine if its server environment and applications are suitable for virtualization.

3. **Choose a virtualization platform that best suits your hardware platforms.**
Make an informed decision about which hypervisor will operate most effectively with your hardware platforms and meet your specific business needs. The hypervisor acts like the train conductor on the host system to make sure that the virtual machines (VMs) don’t collide or overload the hardware. Technology advances quickly, and the virtualization platform selected may not be compatible with the server hardware in the environment. The most commonly used hypervisors — Citrix, Microsoft and VMware — are very distinct platforms.

4. Revisit the terms and conditions of the current backup contract.  

Many vendors offer enhanced backup products for a virtual infrastructure, some priced by the number of virtual servers on a machine (also called a “socket” or “host”). These can reduce costs considerably. But a business needs to consider whether making a change will break an existing contract it may already have if the business is tied to a long-term maintenance contract for its current data backup infrastructure. Breaking the contract could result in financial penalties.

5. Implement a plan that eliminates as many physical servers as possible.  

By eliminating as many physical servers as possible, a business will obviously realize the quickest cost savings and return on its investment. Turn off and recycle or sell unused servers — that’s where the most savings materialize up front.

6. Train your staff.  

Invest in training about maintaining a virtual infrastructure. Improperly maintained virtual environments can cost much more in repairs and lost functionality.

**Server Virtualization: Benefits to Business**

- Slash capital costs. This is accomplished through consolidating servers and containing additional hardware spends.

- Enhance staff productivity. Virtualization solutions provide IT professionals with the freedom to be released from menial tasks, enabling them to take a more strategic role in the enterprise.

- Improve business continuity. Take advantage of complete data protection, truly continuous application availability and automated disaster recovery across physical sites.

- Improve business responsiveness. Managing a virtual infrastructure allows IT professionals to quickly connect and better manage resources to meet business needs. Examples: faster provisioning of new applications and acceleration of change request response times.

- Improve application quality and deployment. With server virtualization, small businesses can test more applications by optimizing preproduction staging environments, resulting in less downtime for the applications that run and drive the business.

- Strengthen security. Server virtualization enables automated patch management of server hosts and virtual machines and an integrated firewall that maintains security policies across the mobile, flexible environment. It also decreases the operational costs of managing them.

Source: VMware

**LOOK TO CDW FOR SERVER VIRTUALIZATION SOLUTIONS THAT LOWER COSTS, OPTIMIZE PERFORMANCE AND SIMPLIFY MANAGEMENT. CALL 800.800.4CDW TO TALK TO A SPECIALIST TODAY.**

120110