

# ***Terascale Data Warehouse Appliances Overcome the Technology Bottleneck***

---

## **Whitepaper**

1 866 NETEZZA  
WWW.NETEZZA.COM

NOVEMBER 2006

## The Technology Bottleneck

Effective use of Business Intelligence (BI) can make or break a company. Data-intensive organizations across many industries are collecting terabytes of dynamic, detailed data on a daily basis. The timely and comprehensive analysis of this vast amount of data is vital to:

- Telecommunications carriers who must identify profitable customers at risk for churning and take preventative actions
- Retailers running CRM analyses to determine which on-the-spot deals will boost revenues
- Financial services institutions running credit risk analyses for loan approvals
- E-businesses running clickstream analyses to determine the shortest clicks-to-purchase path
- Life scientists analyzing massive databases of human genomic sequences to identify new drugs and diseases

With the ever-increasing size of the average data warehouse, and the need for more complex analyses and ad hoc reporting, achieving these goals with today's general-purpose systems has become increasingly difficult.

For decades, companies have been deploying general-purpose servers, storage and databases to store growing amounts of data and handle large volumes of transactions. They are now trying to apply this older, patchwork infrastructure to BI. Unfortunately, the older architectures, developed when data volumes and reporting needs were far less intense, weren't designed to handle terabytes of constantly growing and changing data—and the complex types of analyses business users now need to fully understand their customers and their business operations. Organizations have spent billions of dollars to implement, upgrade and maintain these solutions, but they are unable to effectively handle the demands of in-depth analyses on large amounts of enterprise data. Add to this the increasing numbers of users accessing the data and the strain has become crippling for many organizations.

Several factors contribute to the creation of a technology bottleneck that results in significant information latency for business decision makers:

- Disk transfer I/O rates (~60MB/sec) can't read terabytes of data quickly enough
- Network transfer rates (~125MB/sec) can't move terabytes of data quickly enough
- Memory density growth (~100%/year) can't keep up with data growth making traditional caching less effective over time

And increasing CPU performance becomes irrelevant since the gating factor in most implementations is moving the data off disk to the CPU.

Building upon traditional BI infrastructure comes at a high cost both in terms of initial investments and on-going maintenance cost. Companies attempt to address the information latency challenge with efforts to increase machine power including:

- Buying more CPUs and memory
- Upgrading servers
- Buying bigger disks
- Upgrading DBMS to the latest version
- Replicating/distributing databases

In addition to equipment upgrades, organizations must also have their costly database administrators (DBAs) spend a significant amount of their time fine-tuning for incremental query speed improvements. And it can take several DBAs to manage each terabyte of data. So, when the recommendations to purchase more equipment are combined with the costs of database administration staff, today's BI solutions soon become overwhelmingly expensive. But these expenditures provide only incremental performance gains and do nothing to address the underlying bottleneck. A new approach is required to see significant performance improvements.

Today's challenge is to find an affordable, integrated solution that is suited to the task of handling detailed, dynamic, terascale data, yet be compatible with existing BI applications and infrastructure. Furthermore, the solution should be simple to install and maintain, versus the highly complex and expensive systems currently available. Such solutions are data warehouse appliances—expandable, affordable and uniquely suited to the ever-growing demand for fast and comprehensive analyses.

## What is a Data Warehouse Appliance?

Appliances in the computer world are so common that we often forget about them. At its most basic level an appliance is an integrated box that can retrieve information at the request of external applications and keeps its inner workings hidden in order to maintain simplicity and ease of use. Take, for example, the network router. The majority of routers can be put in place with almost no configuration (other than setting the router's IP address) and will start storing and forwarding packets. Hubs and switches are even simpler. The point is that these devices aid network transport greatly, yet are essentially transparent to the user.

Simplicity is the name of the game. Why should DBAs be required to fine-tune the database system as the requirements increase? Appliances make our lives simpler. Why can't this analogy be carried into the database world?

It can. A terascale data warehouse appliance integrates the hardware, DBMS and storage into one opaque device. It combines the best elements of Symmetric Multiprocessing (SMP) and Massively Parallel Processing (MPP) architectures into a new architecture to allow a query to be processed in the most optimized manner without extensive DBA tuning and management.

## An Environment for Data Warehouse Appliances

Given the long history of database development and the existence of previous attempts at database appliances/machines, why is now the time for a terascale data warehouse appliance in BI?

There are several reasons why the data warehouse appliance is now possible, but the most important of these is the maturity of database technology. Database standards have been set, and this allows the system to be built completely around the desires and needs of the end user. Furthermore, the concept of a relational database is well defined and users are eager to run increasingly complex queries. A wide variety of sophisticated applications and tools with standard interfaces including ODBC, JDBC and SQL allow widespread access to the database. And, as noted earlier, terabyte-sized databases, an influx of users and a demand for complex queries have placed unprecedented strain on the existing general-purpose patchwork infrastructure. Simultaneously, high-powered, commercially available off-the-shelf components, which are dropping in price, are readily available to construct a terascale appliance.

These technological advances pave the way for a whole new paradigm. The way of the future is a terascale appliance for BI that delivers robust performance and linear scalability at a reasonable cost.

## Characteristics of a Performance-Driven Data Warehouse Appliance

As business and technical demands continue to grow and change, the new enterprise-class data warehouse appliance is designed from the outset to scale with data size, scope and performance needs. And it must do all of this at an affordable price. When evaluating special-purpose data warehouse appliance technology there are several important criteria to consider:

**Performance** — The goal of any data warehouse appliance is to provide dramatic performance improvements over today's traditional solutions. As a result, complex and ad hoc queries on terabytes of information go from hours or days to minutes or seconds.

**Affordability** — Perhaps the most pressing need for data warehouse appliances is the need for a reasonable total cost of ownership (TCO). Appliances built for analytics must have lower upfront costs but must also reduce on-going maintenance costs, including expensive DBA time. Quality appliances meet this need by automating the optimization of the integration of storage, server and database functions.

**Ease of Use** — terascale data warehouse appliances leverage existing skills within the organization by providing appliance packaging, utilizing well-understood SQL language commands and limiting the need for pre-tuning and on-going database administration.

**Effortless Scalability** — From hundreds of gigabytes to hundreds of terabytes of user data, data warehouse appliances must scale with little response-time degradation.

**High User Concurrency** — The growth of business users within today's computing environment demands that BI appliances must be able to handle many requests simultaneously, including short, tactical queries and long, complex ones. BI appliances must be able to serve the needs of hundreds of geographically dispersed users and provide information when and where they need it.

**Compatibility** — It's critical that any enterprise-class data warehouse appliance fit seamlessly within the organization's current environment, leveraging the significant investments they've already made in technology. By taking advantage of mature industry standards, such as ODBC, JDBC and SQL, special-purpose data warehouse appliances not only exist within the current infrastructure, but maximize the performance of existing ETL tools, BI and CRM applications and transaction processing. All of this is done using current data sources, whether from the web, legacy databases or mainframes.

**Flexibility** — As computing history has shown, environments change, business requirements evolve and individuals simply can't accurately predict what type of analytics will be needed in the future. Organizations must therefore build a flexible system that's designed to evolve and grow with the easy addition of modular processing units for scalability and reconfiguration.

**Reliability** — Leveraging advanced mirroring techniques, a quality data warehouse appliance will be fully ACID- and RAID- compliant to secure maximum structural reliability.

## Architected for Success

The Netezza Performance Server® (NPS®) appliance architecturally integrates servers, storage and database into a single powerful, enterprise-class appliance created from the outset to process queries faster. The NPS appliance enables complex analyses that were not previously possible because of large and fast-growing data sizes, and is transforming the way organizations can leverage their BI.

Netezza has designed the NPS system to ensure maximum performance that exceeds expectations and fully scales for future data volumes and analytic requirements. Delivering 10-100x the performance of existing data warehouse solutions, the NPS system's unique integration of hardware, DBMS and storage into one opaque device accomplishes that objective through its Asymmetric Massively Parallel Processing™ (AMPP™) architecture, built to harness the processing power of SMP and combine it with the scalability of MPP. At the front end, an SMP-based Host compiles queries into parallel execution plans and provides the right amount of processing power to sort and aggregate large sets of query results. On the back end, data is distributed across many nodes (or "spindles") to minimize I/O latency and increase scalability. Query functions are implemented in silicon to optimize throughput. Simply put, AMPP applies SMP and MPP approaches to the areas where they can provide the largest benefits in performance and scalability at a fraction of the cost of existing solutions.

The technology innovations described above provide a unique framework to drive unparalleled performance and scalability for terascale BI. This revolutionary approach not only provides unprecedented performance, but does so for half the cost of traditional solutions by reducing the time spent by costly DBAs maintaining the system and by leveraging off-the-shelf components for low acquisition costs.

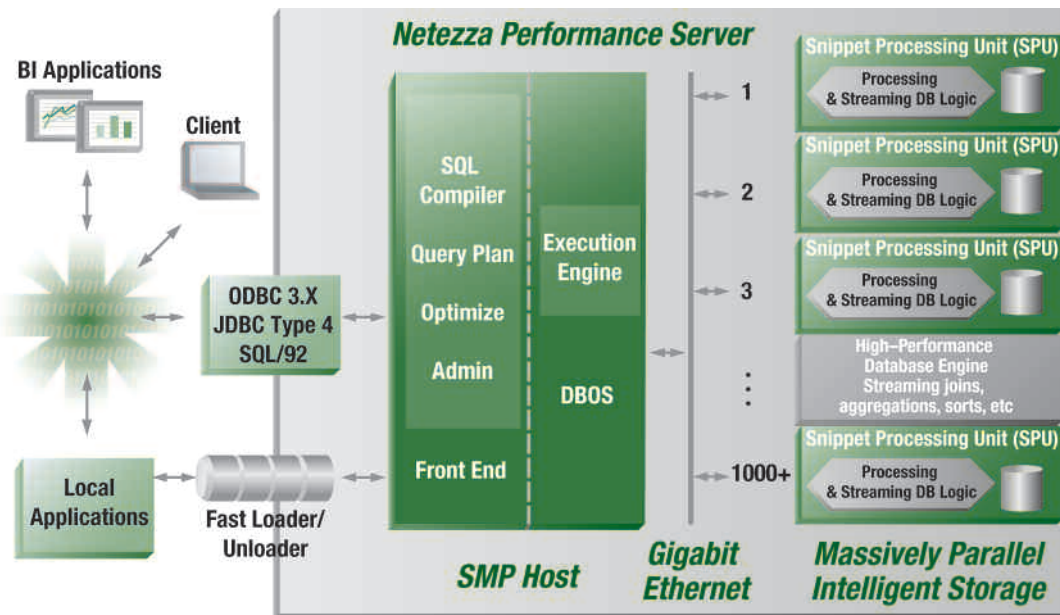


Figure 1.

Through standard interfaces including ODBC, JDBC and SQL, the NPS system is fully compatible with existing BI applications, tools and data. And it is extremely simple to use. The NPS system can be easily installed without changing analytic applications or tuning data sources. The system requires less than one DBA to manage because, with its intelligent design, traditional performance tuning is no longer required. That means no indexing, materialized views, temp space allocation and monitoring, logical volume creations of files or any of the other time-intensive manual optimization procedures required in the general-purpose environment.

Further, the NPS system is designed to remove all the bottlenecks to data flow so that the only remaining limit is the disk speed—a ‘data flow’ architecture where data moves at ‘streaming’ speeds. At the heart of the NPS appliance’s technology is a breakthrough, patent-pending Intelligent Query Streaming™ design that optimizes the flow of information by placing silicon processors in close proximity to the storage device. The NPS system intelligently filters and processes records immediately as they come off the storage disk, bringing over only the relevant information for each query. Processing that would usually overwhelm a general-purpose server now occurs right at the source of the data to deliver a tremendous boost in performance.

## Imagine the Possibilities

The benefits of an enterprise-class data warehouse appliance optimized for BI are far-reaching. Business users who have been forced to sample and summarize their data and wait hours or days for analysis on traditional systems will benefit from tremendous speeds with the new technology, which can deliver analyses of complete, current data in minutes or seconds. Companies can harness information located in tens, even hundreds, of terabytes of dynamic, detailed data in order to recover lost revenues, improve risk management, reduce fraud, compress decision cycles, strengthen customer relationships and even enable new, previously impossible business processes. Fast, comprehensive analyses of terabytes of data enable organizations to optimize their BI for a competitive advantage.

And it’s not only the business user who benefits from the appliance. By nature, appliances provide simplicity for the administrator, effortless scalability and the lowest TCO. Together, these benefits offer powerful value to data-intensive organizations. Imagine the possibilities for your business—orders of magnitude performance improvements, for half the cost. Everyone wins.

## About Netezza

Netezza, the global data warehouse appliance market leader, enables enterprises to make all of their data actionable - quickly, simply and affordably. The Netezza Performance Server family of products delivers breakthrough performance, unmatched ease of deployment and operation, and innovative flexibility and scalability at a fraction of the cost of traditional data warehouse solutions. By architecturally integrating database, server and storage within a single appliance, the NPS system delivers 10 to 100 times the performance at half the cost of existing systems. Based in Framingham, Mass., Netezza has offices in Washington, DC, the United Kingdom and Asia Pacific. The Company is backed by leading venture capital firms, including Matrix Partners, Charles River Ventures, Battery Ventures, Orange, Sequoia Capital and Meritech Capital Partners. **For more information about Netezza, please visit [www.netezza.com](http://www.netezza.com).**

Netezza Corporation : 200 Crossing Boulevard : Framingham, MA : 01702-4480  
+1 508 665 6800 tel : +1 508 665 6811 fax : [www.netezza.com](http://www.netezza.com)

**NETEZZA**  
*The Power to Question Everything™*