

***Data Warehousing without the Complexity:  
How the Simplicity of the NPS® Data  
Warehouse Appliance Lowers TCO***

Whitepaper

## Introduction

With its Netezza Performance Server® (NPS) family of data warehouse appliances, Netezza has created a new paradigm in data warehousing. As the data warehouse appliance market leader, the NPS system dramatically changes data center operations, improving query performance by orders of magnitude. Customers worldwide have recognized the market-leading performance of the NPS system and its ability to power deep analyses of customer and operational data.

As they deploy and run their Netezza appliance, customers quickly realize that its simplicity and ease of use also bring tremendous value. In addition to saving significant time and money, this simplicity enables customers to do more with their data warehouse: to expand into new applications, generate new sources of revenue and solve pressing business challenges.

This paper compares the complexity of traditional data warehouse systems with the unprecedented simplicity of the Netezza appliance. It begins by examining the design limitations of other systems, and the effort required to squeeze performance from architectures with inherent shortcomings. This complexity comes at a steep cost: in lengthy deployments, in training fees, in over-burdened staff and in costly professional services.

The paper then examines how the simplicity of the Netezza appliance transforms the data warehousing experience, making life easier for users, administrators and their companies. Netezza's focus on simplicity goes beyond easy deployment and administration – it also means simplicity and transparency in the relationship between Netezza and its customers.

Finally, the paper shows how the simplicity of the Netezza approach provides a significant reduction in total cost of ownership (TCO) compared to alternative systems. In addition to delivering a quantum leap in performance, the NPS system has a dramatic impact on the data center's bottom line.

## Traditional Systems: Layers of Complexity

Traditional data warehouse systems are based on architectures that were not designed for high-performance, terascale business intelligence. They consist of general-purpose components – servers, storage and database software from multiple vendors – patched together as a data warehouse. Some systems such as Oracle 10g RAC consist of the underlying database together with the customer's choice of server, storage and operating system. Others such as IBM's BCU (Balanced Configuration Unit) or Oracle's Database Packs pre-package some or all of the core components into a bundled solution or "pseudo appliance".

The fact remains that these systems are collections of generic piece parts with a fundamental design flaw for a data warehouse. They all read data from storage, bring it across an I/O interface and load it into memory for processing. The approach is extremely inefficient for shuttling millions or even billions of rows of data back and forth in order to process a query. The result – in addition to poor performance and excessive cost – is a tremendous amount of care and feeding required to keep the system running. This complexity occurs at every step of the way, from procurement and deployment through system administration and ongoing vendor relationships.

## **Procurement**

The complexity of a traditional data warehouse system is apparent even before you buy it. There are multiple product and service contracts from different suppliers, with varying terms and conditions. A price quote will list page after page of line items to wade through and evaluate. There will be quotes for servers, storage, cabling and other hardware. There will be still more quotes for the database, the operating system, the management tools and other assorted software.

Traditional solutions also open the door to all kinds of complicated pricing arrangements. For example, customers often have to pay separately for database software for each hardware node. Then they have to pay an additional subscription fee, also per node, to receive software updates. Many vendors rely heavily on revenues from consulting/professional services engagements, which also adds to the complexity of the buying process – not to mention the cost of planning, training and upkeep.

Some vendors offer pricing based on various “performance” configurations, further complicating procurement. These offers assume that all performance-related variables are known at the time of purchase, and often mislead the customer into purchasing a cheaper, lower-performance option upfront, only to find after deployment that they must purchase high-cost upgrades to get to the expected performance levels.

## **Installation**

A traditional data warehouse system is a complex environment that must be assembled and configured on site. Installation can take weeks, requiring assembly, testing, debugging and fine tuning of system parameters. While the recent introduction of “pseudo appliances” gives the impression of greater simplicity, the fact remains that these are still systems based on piece parts that must be custom-created for each deployment.

Simply figuring out where to fit all the hardware can be a challenge when dealing with large equipment footprints such as banks of external storage. When SCSI-based disk storage is used, cabling length constraints may dictate specific rack lineups and give customers less flexibility in their data center space.

Software installation has its own issues. How much time and effort is required? Will this version of the database work with this particular operating system and server? On some systems, installation must be performed manually on every node in a cluster or bladed rack. Once installed, it becomes a separate managed operating system on each of nodes – not exactly elegant or convenient.

## **Tuning the System**

Traditional data warehouse systems depend on elaborate tuning to squeeze performance out of their older architectures. Loads have to be balanced, indexes created, disk partitions and logical volumes defined – and one setting almost invariably interferes with another. This multi-step process takes weeks or months, and almost always requires an expensive professional services engagement.

For many vendors, tuning has become so complex that it resembles a scientific discipline. For example, Oracle published a 23-page paper: “The Top Five Performance Tips for Tuning” – reflecting the extensive training required for even a subset of the actual tuning work required to manage the data warehouse. With IBM DB2, tuning is exacerbated by the need to manage separate operating systems on each blade or server node, which all must be kept in sync as the system evolves. As might be expected, all this tuning greatly increases setup and maintenance complexity and cost.

## **Loading Data**

Loading the huge data volumes associated with an enterprise data warehouse is a tricky, cumbersome exercise on traditional systems. With some vendors, data loading appears to be a problem that has not been solved. For example, some vendors offer a choice of loaders, all with basic limitations. They may allow querying while loading data, but run relatively slowly. Or they may be extremely fast, but users are locked out from accessing tables while data is being loaded – defeating the purpose of the high-speed tool.

Some data warehouse systems have restrictions about what database schema they accept, requiring time-consuming conversion before data can be loaded. For example, Teradata strongly recommends using the Third Normal Form (3NF) database schema, requiring conversion of source files from their original format (flat files, star schema, snowflake structure, etc). Other systems like Oracle can require conversion from a 3NF model to a dimensional model. Mapping the database structure to the new format and converting the data can keep a team of software designers, developers, testers and data modelers busy for months - an additional cost above the purchase price of the system. In addition, as business requirements change, traditional systems often require extensive redesign of their physical and logical data models, adding further cost and delay.

## **Ongoing Administration**

Managing a traditional data warehouse is a delicate balancing act. Systems are very sensitive to minor changes in partitioning and indexing, draining productivity from IT staff. DBAs are forced to spend an inordinate amount of time monitoring and tweaking the system rather than developing new projects or mining critical information from the database. Ambiguous errors, poorly understood performance characteristics, compatibility problems after upgrades, and finger pointing between vendors are just some of the issues that IT departments have to deal with.

In short, looking after a data warehouse is anything but simple. To give an idea of how much work is required, implementations are often staffed by teams: a system administration team, a storage administration team and a network administration team. Even the latest pre-configured models such as the IBM BCU require at least two full time administrators.

Management tools are another reminder that these are really just a collection of piece parts. Most traditional systems come with separate management programs for servers, storage and the database, but not the system as a whole. Sometimes solutions include elaborate system management tools such as IBM's Tivoli that have to be learned (and paid for). Customers are saddled with a great deal of extraneous functionality for solving problems that have nothing to do with data warehousing.

## **Training**

Learning how to operate and maintain a traditional data warehouse system involves weeks of training – an expensive proposition given typical training fees. For example, by browsing through Teradata's 31-page training catalog (<http://www.tensupport.com/catalog/CurrentCatalogAmericas.pdf>), one can get a sense of their system's complexity, and the extensive curriculum to become a proficient user or administrator. One may wonder whether a four-day class devoted to indexing – or IBM's three-day class on database partitioning – is really a good use of an administrator's time. (Netezza users know better.)

### **Expansion and Upgrades**

Scaling a traditional data warehouse also means scaling up complexity. Because of the customized nature of a traditional data warehouse system, capacity upgrades require resetting the indexing and partitioning parameters originally set during the installation. This is not only time-consuming; it can – and does – affect performance if not done correctly.

Delays and risk are the norm when upgrading an architecture based on piece parts. For example, can the vendor guarantee that an OS upgrade will not adversely impact performance of the finely-tuned DBMS? Given this uncertainty, elaborate testing is needed to verify that all system elements continue to interoperate. Upgrades also usually mean an expensive engagement with professional services organizations, and a “circular firing squad” where vendors blame each other when things go wrong.

### **Business Relationships**

In addition to system complexity, the business strategy of traditional data warehouse vendors is often complex and at odds with the customer’s best interests. For example, some vendors are increasingly focused on packaged solutions – with the data warehouse embedded within a stack of vendor offerings. Consequently, the sales force incentive is to sell the company’s solutions rather than the data warehouse – denying customers the opportunity to select the best of breed tools they want.

Some vendors have separate business units with different business practices and conflicting sales agendas – even to the point of different groups selling against each other. Sometimes storage, server and software teams cannot even agree on how to configure the system. This sense of dealing with different companies is also reflected when things go wrong, and there is no single point of help for the customer.

## **The NPS System: Simple is Beautiful**

The NPS system is a true data warehouse appliance – a fully integrated device built specifically to enable real-time business intelligence and analytics on terabytes of data. Netezza’s Intelligent Query Streaming™ technology combines innovations in system architecture, hardware and software to enable query processing at full disk speed, or “physics speed.” By moving processing intelligence to a record stream adjacent to storage, the NPS system eliminates traditional bottlenecks – and tuning headaches – that occur when moving blocks of data to a processor.

Netezza customers notice the difference immediately: in phenomenal query performance combined with operational simplicity that complex patchwork offerings cannot match. From procurement and deployment through the entire data warehousing experience, all the previous layers of complexity are stripped away.

### ***Simple Procurement...***

The simplicity of the Netezza data warehouse appliance begins with system acquisition. There are two line items on the Netezza contract: a license for the NPS system and a license for maintenance. There is one part number – what could be simpler than that?

### ***Simple Installation...***

“Load-and-go” deployment means that customers are performing terascale analytics within two days, rather than weeks to months. The Netezza appliance is factory-configured and tested – just wheel it in from the loading dock and plug it in. The compact footprint with processors and storage in the same cabinet greatly simplifies installation. Customers or technicians don’t have to know anything about

SCSI cabling or disk arrangement. There are no complicated wires – simply power, network and modem connections.

### **Simple Data Loading...**

Seamless data loading is another dimension of the Netezza data warehousing experience. There are two simple load methods in the NPS system: using a load utility (nzload) or via SQL (interactive or batch). The NPS system integrates with all major ETL vendors including Ab Initio, Ascential, Informatica and Sunopsis through ODBC or JDBC.

Full ACID (Atomicity, Consistency, Isolation and Durability) compliance allows analytic operations without interruption or degradation while data is streaming into the data warehouse. There is no need to lock tables or otherwise restrict users in any way.

The NPS system is also completely schema agnostic. Whether data is in third normal form, star schema, snowflake or any other database structure – just “load and go.” The high-speed loader captures data as quickly as it can flow off the source system. And, as business requirements change, there’s no need to make changes to physical or logical data models.

### **Simple Tuning...**

As a self-regulated data warehouse appliance, the NPS system delivers 10 to 100 times better performance "out of the box" than competing RDBMS architectures designed decades ago. There is no tuning in the conventional sense, eliminating maintenance tasks that have been the bane of administrators for years:

- No indexing required
- No dbspace/tablespace sizing and configuration
- No redo/physical log sizing and configuration
- No journaling/logical log sizing and configuration
- No page/block sizing and configuration for tables
- No extent sizing and configuration for tables
- No temp space allocation and monitoring
- No RAID level decisions for dbspaces
- No logical volume creations of files
- No integration of OS kernel recommendations
- No maintenance of OS recommended patch levels
- No JAD sessions to configure host/network/storage
- No software to install
- Simple partitioning strategies: HASH or ROUND ROBIN

### **Simplicity and Your Electric Bill**

*As a purpose-built appliance, the NPS system not only provides much higher query performance than traditional systems, it also consumes far less power. Each Snippet Processing Unit (SPU) node consumes less than 30 Watts, with the embedded processor consuming only about 5 Watts. Thus, the power consumption and cooling specifications for a full-rack system (host plus 112 SPUs) are only 3,500 Watts and 12,000 BTU/hour respectively. These figures include all server and data warehouse storage capacity required for up to 12.5 TB of raw data per equipment rack.*

*Compare this with two widely used configurations:*

- *Teradata 5500H with 2X dual Intel Xeon processors and NCR 6843 storage arrays will require 5 racks of equipment, 18,500 watts to power and will produce 63,000 BTU/Hr that need to be ventilated and cooled.*
- *IBM BCU AIX on P5-575 servers will require 4 racks of equipment, 29,700 Watts to power and will produce 101,400 BTU/Hr that needs to be ventilated and cooled.*

*As density of racks and components continues to rise, more powerful cooling equipment is required to keep temperatures down, resulting in more power consumption, money and occupied space in the data center.*

### Simple Administration...

The NPS system typically requires less than a single administrator's time. System administrators and DBAs appreciate the system's simple graphical user interface and CLI tools. There are no obscure "Netezza-isms" that administrators have to learn – just standard ANSI SQL. The system comes with ODBC and JDBC interfaces to support a wide range of standards-based tools and applications, including all the major BI applications.

Administration of a Netezza data warehouse is constant, regardless of database size. This means that even a 20-terabyte database system can be supported by one DBA, with plenty of time left over for other critical activities. Rather than constantly tweaking and monitoring the system, DBAs can use their skills to provide real business value:

- Bringing on new applications and groups
- Quickly building out new data marts
- Providing more functionality to end users

#### Simpler to Manage

The following table illustrates the simplicity of the NPS system compared to Oracle. It was prepared for a leading telco, and shows the relative amount of work required to allow querying of just one large table of call detail records. The table shows the various elements that require interactive management and tuning in an Oracle environment, such as indices and data partitions, versus management of just the primary base table in the NPS data warehouse appliance. Even before accounting for the additional work to configure storage files for Oracle, this represents an excellent example of the power of Netezza simplicity.

<b>Telecom Call Detail Record (6 billion rows)</b>	<b>Oracle Object Count*</b>	<b>Netezza Object Count</b>
Tables	1	1
Indexes	12	
Table Partitions	47	
Index Partitions	564	
Table Partitions Tablespaces	47	
Index Partitions Tablespaces	47	
Table Data Files	170	
Index Data Files	122	
<b>Total</b>	<b>1,010</b>	<b>1</b>

\* Oracle figures do not account for additional effort required to configure the file system to accommodate this index management scheme.

### Simple Training...

NPS system training takes a maximum of three days, since virtually all tuning and system management tasks are handled automatically. As with any good appliance, there's simply less extraneous material that a user or administrator needs to learn. Netezza customers save many thousands of dollars in training fees, and their administrators and users are able to get started quickly on meaningful work.

### ***Simple Expansion and Upgrades...***

Unlike other data warehouse systems, scaling the NPS family of systems takes hours – not days or weeks. Simply plug in additional SPU blades or connect a new rack – there’s no lengthy on-site assembly, configuring or testing. The NPS system is also extremely compact, packing tremendous processing power on up to 100 terabytes of data into a maximum of eight cabinets (as opposed to dozens of cabinets with traditional vendors).

Software upgrades on the NPS system are also fast and easy, as they should be on a data warehouse appliance. By running a simple setup procedure in a manner familiar to Windows PC users, upgrades are loaded automatically across the system. Unlike with traditional systems, there’s a single, simple upgrade process, managed by Netezza, for all software components in the warehouse.

### ***Simple – and Rewarding – Customer Relationships***

The simplicity of the Netezza approach extends to the company’s relationship with its customers. Netezza has only one mission: to help customers succeed by delivering the best possible data warehouse solution. Clear pricing guidelines, appliance simplicity and the singular focus of Netezza’s sales force makes the buying process easy. Netezza’s sales force is dedicated to delivering a terascale data warehouse that meets customer needs, without being compromised by solution sales goals, cross-channel selling or any of the other pressures that other sales organizations have to deal with.

Netezza has a large and diverse partner program with world-class software firms to deliver packaged and customized solutions in vertical markets. Netezza’s list of partners continues to grow as the company builds relationships with major software, technology and system integration providers. (See <http://www.netezza.com/partners/partners.cfm> for a list of Netezza’s partners.) Customers get the solution they need – not one imposed on them by the vendor.

The simplicity of the appliance approach also means that there is no pressure or agenda driving service revenues. There’s one simple maintenance contract, without any pricey “extras.” Customers interact with one vendor, minimizing the time, complexity and cost involved in acquiring, deploying and supporting the data warehouse. There’s one part number, and one number to call.

## **The Impact on TCO**

This paper has examined how the NPS system requires only minimal administration and maintenance, providing comprehensive ease of use that traditional systems cannot match. This unparalleled simplicity translates to significantly lower cost of ownership, from upfront purchase through ongoing operations over time.

### **Cost of Ownership in the Data Center**

Today, a CIO is faced with an ever-growing challenge in the data center: how to maintain the existing legacy systems already in place, invest in innovation, yet at the same time lower costs and do more with less. According to BusinessWeek, over 70% of IT spending is consumed by maintenance and administration of existing systems<sup>1</sup>, so reducing these costs and simplifying IT management efforts have a clear bottom-line impact. Within the data center, spending on data warehouse and business intelligence initiatives has grown into the tens of millions for many enterprises, with the data warehouse platform (including DBMS and hardware), IT staff and support/maintenance accounting for



over 80% of total investment according to Gartner<sup>2</sup>. A data warehouse appliance that directly reduces all these key cost components provides dramatic cost and staffing benefits for the CIO, opening the door for innovative projects and greater productivity.

### The Netezza Advantage

The following table compares the major factors in cost of ownership of the NPS system with offerings from IBM, Oracle and Teradata. The competitor configurations selected represent “mid-tier” offerings that are reasonable to compare with the NPS system – there are clearly many other competitor configurations with lower performance at a lower price, as well as those with higher performance at a still higher price. Additionally, in most cases, the other vendors will have additional cost factors for professional and/or consulting services.

Not only is the Netezza appliance significantly cheaper in terms of initial cost, but also in cost of installation, maintenance and ongoing administration. The Netezza appliance is also far more compact, as indicated by the difference in the number of racks.

### NPS System vs. IBM, Oracle and Teradata

	<b>Netezza</b>	<b>IBM BCU on AIX</b>	<b>Oracle on HP SuperDome</b>	<b>Teradata</b>
<i>System Model</i>	<i>NPS 10100</i>	<i>P5 575 6 nodes, 48-way</i>	<i>32 CPU</i>	<i>5500H - 12 nodes</i>
<i>Discount Level</i>	<i>0%</i>	<i>40%</i>	<i>40%</i>	<i>40%</i>
<i>Number of Racks</i>	<i>1</i>	<i>4</i>	<i>6</i>	<i>4</i>
<i>SysAdm/DBA FTE/year</i>	<i>0.5</i>	<i>3.5</i>	<i>3.5</i>	<i>3.0</i>
<i>Initial System Cost</i>	<i>\$1,110,000</i>	<i>\$3,647,000</i>	<i>\$3,295,000</i>	<i>\$4,702,000</i>
<i>3-Year Maint. &amp; S/W Subscr.</i>	<i>\$599,000</i>	<i>\$1,160,000</i>	<i>\$1,618,000</i>	<i>\$2,538,000</i>
<i>3-Year SysAdm/ DBA Cost</i>	<i>\$187,000</i>	<i>\$1,313,000</i>	<i>\$1,313,000</i>	<i>\$1,125,000</i>
<b>Full 3-Year TCO</b>	<b>\$1,896,000</b>	<b>\$6,120,000</b>	<b>\$6,226,000</b>	<b>\$8,365,000</b>
<b>\$ Savings w/Netezza</b>		<b>\$4,224,000</b>	<b>\$4,330,000</b>	<b>\$6,469,000</b>
<b>Savings % w/Netezza</b>		<b>69.0%</b>	<b>69.5%</b>	<b>77.0%</b>

*Typical system configurations in support of up to 12.5 TB of user data.*

*Annual loaded FTE rate for System and Database Administrators assumed to be \$125,000.*

*Estimates for System and Database Administration staffing levels are conservative, based on general industry knowledge and interactions with Netezza customers.*

*List pricing for IBM, Oracle and Teradata systems is composed from recent publicly available sources and market experiences.*

*To learn more about this TCO analysis or to conduct one for your company please contact Netezza at [salesinfo@netezza.com](mailto:salesinfo@netezza.com).*

## Summary: Simply the Best

In an economy that increasingly operates in real time, businesses have to be extremely responsive or get left behind. Companies with traditional data warehouse systems are realizing that they cannot afford to spend weeks or months getting a system running, dedicate a team of experts to care for it or spend a fortune in outside services. The lost time and resources are simply not economically viable.

In contrast, the Netezza data warehouse appliance is designed to be as simple and intuitive as a PC. Analysts are able to identify emerging trends and respond before windows of opportunity close – without extraneous maintenance tasks getting in the way. DBAs can use their skills to solve business problems rather than simply keeping the system in working order. Seamless integration and upgrades allow users to take advantage of new functionality as it becomes available. Lengthy, expensive professional service engagements are history.

Welcome to the new data warehousing experience. It's amazingly fast, highly affordable and low, low maintenance. It's as simple as that.

---

<sup>1</sup> "Next Generation Computers", Peter Burrows, BusinessWeek, December 12, 2005.

<sup>2</sup> Gartner Symposium ITxpo 2002, "Data Warehousing & Business Intelligence: The Challenges of TCO and ROI", presented by Kevin Strange.

## 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™*