

***Answers at the Speed of Thought:  
Breaking the Latency Barrier with the  
Netezza Performance Server® System***

---

**Whitepaper**

1 866 NETEZZA  
WWW.NETEZZA.COM

NOVEMBER 2006

## Introduction

Latency is an ongoing threat to data warehouses of multi-terabyte size. Speed is vital, and delays in finding nuggets of information are an albatross when you're up against aggressive, fast-moving competitors.

Latency plagues the entire data warehousing experience and often results in the failure of large BI/DW initiatives. Latency issues force companies to work with stale data, wait hours or days for queries to complete, and settle for coarse summary results because detailed analysis takes too long. As a result, many enterprises are not getting the business intelligence they need to succeed.

Today's data volumes are so huge, and business timeframes so compressed, that latency is a challenge for any data-driven enterprise where profitability depends on detecting and responding to emerging trends:

- For the telecommunications carrier analyzing billions of rows of data to reduce churn
- For the online retailer querying millions of daily transactions to develop targeted promotions
- For the service provider looking to maximize yield on advertising clickstream traffic

Business intelligence has become increasingly strategic in most organizations, and its use has spread across the enterprise. However, traditional data warehouse architectures have not kept up to speed with current demands.

The Netezza Performance Server (NPS®) family of data warehouse appliances eliminates the latency problems that affect other data warehouse systems. The system's performance comes from a unique architecture that reduces latency by orders of magnitude compared to other systems, allowing users to promptly get the information they need, based on current data. In addition to greatly accelerating loading and querying, minimal latency enables Netezza customers to adopt innovative new processes that fully leverage the intelligence available within the enterprise.

## Latency in the Data Warehouse Environment

Most people think of latency in terms of delays affecting specific electronic processes. In large data warehousing environments, latency must be considered far more comprehensively – including anything that delays access to data. With traditional systems, latency affects the entire data warehouse experience from the time equipment arrives on the loading dock:

### Installation

A data warehouse consists of three main elements – server, storage and database software – typically patched together from multiple vendors. Set-up often takes weeks, involving testing, debugging and fine-tuning of parameters to make sure all elements work together. While this may not seem like latency in the traditional sense, the time lost in gaining access to data has real business value and should not be overlooked.

### Loading and updating data

Many data warehouse users wish they could load data more frequently – sometimes hourly or even continuously throughout the day. Immediate access to data is vital in many industries, including telcos, click-stream companies and enterprises using RFID data for business intelligence. However, ETL

(extract, transformation and load) is traditionally the slowest, most cumbersome and most resource-intensive part of the data warehousing process.

Many companies are forced to delay loading data because it puts too much strain on the enterprise data warehouse, disrupting users who are trying to perform queries. Rather than loading data as it becomes available, IT departments often resort to updating the database as a single batch job in the evening. However, batch windows are shrinking, even for these nightly loads. There's more data to be loaded with each passing month, and because of the 24x7 nature of global businesses, the time available for batch loading is increasingly squeezed.

### **Running queries**

When a complex analysis takes hours to complete, it places a company at a business disadvantage. By the time results are returned, conditions will have changed, and users are not making decisions with the best information. In many cases, users decide not to bother running the query in the first place because they expect latency to cause unacceptable delays. In turn, this can cause users to limit their creative thinking on innovative and more iterative analyses that could provide some of the most valuable information for business decisions.

However, delays are a fact of life with traditional data warehouse systems based on architectures that were not designed for high-performance, terascale business intelligence. These systems operate by reading data from storage, bringing it across an I/O interconnection and loading it into memory for processing. The approach is fundamentally inefficient for shuttling billions of rows of data back and forth through the system in order to process a query – and business intelligence users pay the price.

### **Performing backups**

As with data loading, the bulk data movement required by backups can disrupt user operations. Accordingly, backups are often performed after hours. However, many IT departments have limited opportunities during the week and on weekends for performing backups, and available windows are shrinking. As a result, more data is being left at risk – often for timeframes considerably longer than the company is comfortable with.

## **Driving Out Latency: The Netezza Difference**

The NPS systems minimize data latency and transforms data warehousing operations:

### **“Loading dock to loading data” in two days or less**

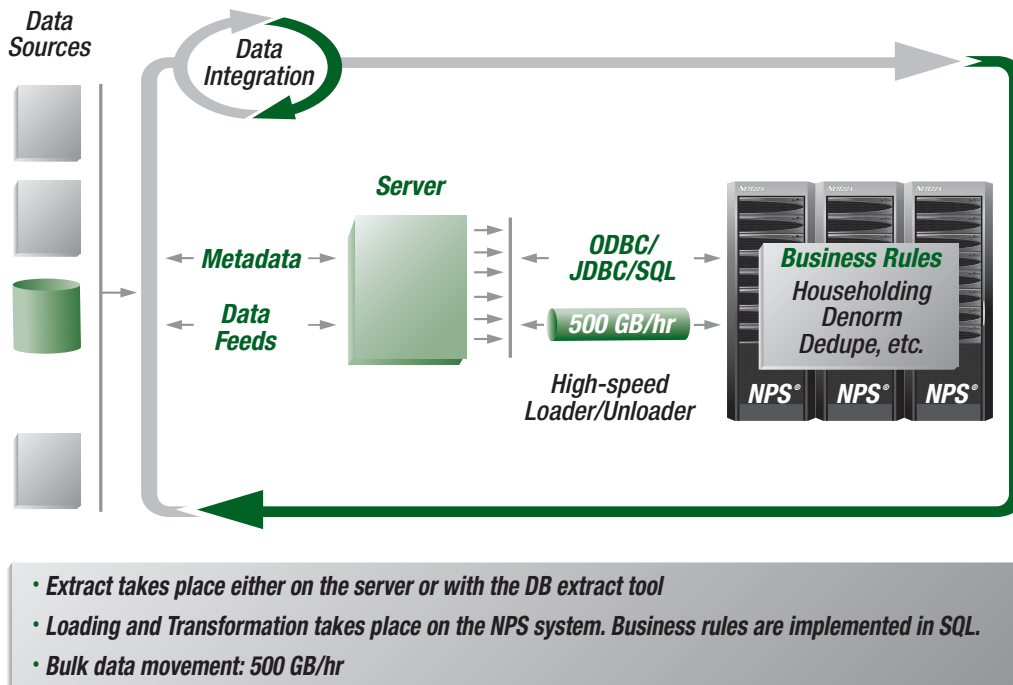
The NPS system is a data warehouse appliance – a fully integrated system built specifically to provide near real time business intelligence and analytics on terabytes of data. Its “load and go” implementation process takes hours, not weeks or even months like traditional data warehouse systems.

### **Fast ETL and ELT**

The NPS system integrates easily with all major ETL tools from vendors including Ab Initio, Ascential, Informatica and Sunopsis. With a data transfer speed of up to 500 GB/hour, Netezza can load and unload data much faster than most competitors. The NPS system enables data to flow into the Netezza data warehouse as quickly as it can stream off the source system.

The raw loading speed also allows Netezza customers to adopt variations of the traditional ETL process. For example, as shown in Figure 1, customers sometimes decide that they're going to convert a portion of their ETL process to an ELT process, saving time by performing data transformation tasks within the NPS system. After initial "scrubbing" on an external server, data is loaded at high speed into the NPS system for more in-depth cleaning or aggregating using SQL queries, taking advantage of Netezza's unmatched query processing performance.

**Figure 1 E-L-T Process**



### Continuous loading while querying

Many large enterprises collect hundreds of gigabytes of detailed data every day through e-business, wireless communications, RFID and other technologies. With data flowing into organizations at unprecedented rates, the ability to access this data in near real time can provide vital, current intelligence about customers, products and operations.

The NPS system allows continuous high-speed data loading while querying is taking place, with minimal impact on query performance. During concurrent loading and querying, the system maintains atomic, temporally consistent processing in its ACID (Atomicity, Consistency, Isolation and Durability) compliant database. While new or updated data is loaded into tables, ongoing queries complete using data as it was when the transaction began, preventing incorrect answers and preserving data integrity.

### Fast aggregations

Because the NPS system is so fast, many customers find that they no longer need to use data aggregates, and can query all their detailed data. In some cases, however, aggregations are used for basic reporting when full data granularity is not required. With an architecture that allows query processing at “physics speed”, the NPS system can build aggregate tables from the full database in a fraction of the time required by traditional systems. Then, when accessing the aggregate data, the Netezza system speeds through those queries too.

Aggregation is one of many areas where minimal latency allows the NPS system to be used in new ways to improve business performance. For example, some Netezza customers use the NPS system to build aggregate tables for other systems that are used as “marts” in the data center. This approach tends to be ten to fifty times faster than using incumbent systems (e.g., Teradata, IBM, Oracle, etc.) and helps boost data center productivity.

### Streaming query processing = fast results

Unparalleled query processing speed is a core strength of the Netezza system. In contrast to traditional architectures based on “moving data to the query”, the NPS system takes a different approach: “moving the query to the data” using Netezza’s patent-pending Intelligent Query Streaming technology. By processing data as records stream off the disk, the NPS system improves performance by at least an order of magnitude for BI applications analyzing terabytes of data.

With fast results, business users are in a much stronger position to make effective decisions. But the implications go further: Netezza customers are able to do in-depth analysis and complex business intelligence operations that previously would have been out of reach. The “Customer Examples” below highlight some of the ways Netezza customers are using fast query processing together with fast data loading to improve profitability and market share.

*Note: More details about how the NPS appliance processes queries and how it differs from traditional data warehouse systems can be found in another Netezza White Paper, **An Architectural Comparison**.*

### Fast backups

The NPS system also offers outstanding performance for the bulk data movement required by backups and restores. The 500 GB/hour (2 TB/hour compressed) data transfer speed minimizes the time required to copy new data to a backup site or to rebuild a data warehouse onto a new NPS system. Compression factors of up to 4:1 or better enable extremely fast backup of data, at rates of up to 2 TB/hour. Backups finish faster, providing IT departments with opportunities to perform backups more often and reduce data vulnerability.

### **Datamarts on Demand**

*The NPS system integrates readily into existing data center environments, where its minimal latency and exceptional query throughput allow significant improvements in productivity. One example is the ability to build specialized “data marts” from existing data marts in the data center.*

*Because the NPS system is data schema agnostic, it can collect data tables from multiple existing data marts (or an overburdened data warehouse) without having to reset individual schemas on the tables. For example, there may be a business need to analyze flat files from an external source using star schema tables from a DBMS such as Oracle or IBM, together with Third Normal Form schema tables from Teradata, and perhaps data from a mainframe system as well. Because the NPS system can work with all of these, there’s no need to rework the schema of the tables: just load and go.*

*With its fast load speed, full tables can be loaded into the NPS system as rapidly as the data can be extracted from the other systems. While the initial business requirements may be only to examine a subset of the source tables, the complete tables can be saved in the Netezza data mart for more detailed (and extremely fast) follow-on analysis.*

## Customer Examples: Minimal Latency and Business Results

This section examines how four Netezza customers from different business sectors are taking advantage of the NPS system's low latency combined with unmatched processing power. While these customers have different objectives and business models, they all use the NPS system in innovative ways that provide a competitive edge.

### Major Retail Chain: From Weekly to Daily Loads

A leading retailer was struggling with data latency that resulted from the limited load performance characteristics of its DB2-based data warehousing system. With data volumes growing quickly, the daily load processes were taking more than seven hours, preventing the company from loading sales data within the nightly batch processing windows. Instead, the company was forced to update the warehouse on a weekly basis. This meant that business users had to wait several days before they could analyze and identify critical sales trends for marketing purposes. Using the Netezza system, the retailer has realized a dramatic increase in load performance, and is now transitioning to daily updates. Business users can now identify and react to evolving sales trends at greater speeds.

### Leading Healthcare Service Provider: Rapid ELT and Aggregation

This service provider for U.S. hospitals and healthcare systems helps researchers and care providers analyze patient outcomes on a massive scale. Its 3 TB clinical database maintains more than six billion records, and is growing rapidly.

Coping with greater data volume was a top priority. As part of the ELT and data loading process, three billion rows of aggregate data have to be reloaded with each release of data, a process that involves building dozens of aggregate tables from more than 25 base level fact tables. The customer streamlined a portion of the process and now uses the unparalleled speed of the NPS system to build aggregation tables nearly 30 times faster than previously with its dedicated enterprise server.

Another priority was faster execution of queries to support 4,000 users running more than 70,000 reports per month. Complex multi-step analyses of patient outcomes are now ready in minutes where previously they took days – if they ever ran at all. Data mining for combinations of diagnosis, treatment and demographics completes in about three minutes – down from over a day.

In addition to system performance, staff productivity has improved dramatically. The help desk is no longer swamped with calls from frustrated users, analysts can perform more in-depth analysis and DBA staff are free to work on new project development.

### Global E-Commerce Company: “Teramarts” on Demand

This e-commerce customer coined the term “Teramart” in reference to the massive amount of data that can be assembled around a set of issues, such as which ads are clicked on the most and where they generate the most revenue for the customer's advertising clients. In this variation of the datamart concept introduced earlier, teramarts are built within the enterprise data warehouse on the NPS system.

Because the NPS system allows these teramarts to be quickly built and executed – even in multi-terabyte sizes – the customer can deploy them as needed to examine specific problems. With an architecture designed specifically for querying massive databases, the Netezza system can create very large scale marts at blinding speed. For example, this customer was able to create a 2 TB mart on demand in less than one hour and a 5 TB mart within two hours. When the customer is finished with the analysis, they may decide to keep the mart for future use, or simply delete it to free up system space.

In addition to pure speed, the data mart concept on the NPS system is highly flexible. For example, multiple database instances within the data warehouse can act as marts for individual departments and/or business needs. Data within the data warehouse (or in departmental marts) can also be combined with externally-sourced data for one-off analysis (e.g., end-of-year planning, or rapid response to an abnormal market condition such as a merger/acquisition).

### **Tier-1 Telecommunications Company: Continuous Loading while Querying**

For this telco customer, the Netezza data warehouse appliance handles hundreds of millions of call detail records (CDRs) streaming in daily. The telco uses the data warehouse to address issues ranging from determining the most efficient way to route calls through the network to catching mistakes before a bill is cut.

Rather than performing a nightly or weekly batch load, the telco customer loads data at least once per hour to operate with more current data. Analysis of the CDRs is used to determine least-cost routing and load factors, and depends upon the NPS system's ability to provide fully ACID-compliant querying while data loading is ongoing. Because the NPS system does not lock tables while loading to them, and because the system guarantees atomic, temporally consistent transaction processing per ACID requirements, analytic operations can continue without interruption or degradation while data is streaming into the data warehouse. For the customer, it means in-depth understanding of operations combined with minimal delay in accessing the most current, near real time data.

## **Conclusion: Why Compromise on Business Performance?**

Users of traditional data warehouse systems are often encouraged to accept a certain amount of latency as a fact of life, especially as data volumes continue to soar. But with today's competitive pressures, organizations cannot afford to wait – for fresh data to load or for queries to return the intelligence that users need to make vital decisions.

With its unique appliance approach and streaming architecture, the NPS family of systems breaks the data latency barrier across the data warehousing environment. Now enterprises can understand what their customers are thinking today – not last week or last month. They're able to recognize new opportunities – and take advantage of them before it's too late. In short, business intelligence becomes a whole lot smarter.

With the NPS appliance, knowledge workers get the answers they need, when they need them. For Netezza customers, that's a new and welcome fact of life.

## 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™*