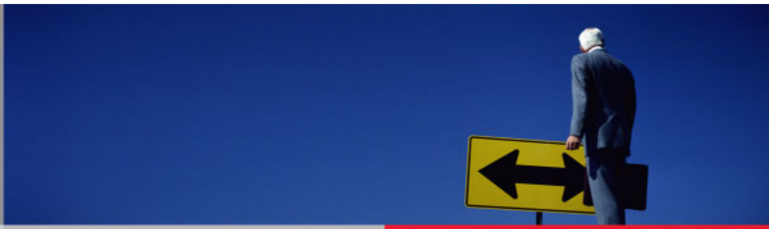
The Netezza Data Appliance Architecture: A Platform for High Performance Data Warehousing and Analytics



Redguides
for Business Leaders

Phil Francisco

- Exploit the power and simplicity of a purpose-built appliance for high speed metrics
- Improve the quality and timeliness of business intelligence
- Query data at lightening speed efficiently and economically

Executive overview

Success in any enterprise depends on having the best available information in time to make sound decisions. Anything less wastes opportunities, costs time and resources, and can even put the organization at risk. But finding crucial information to guide the best possible actions can mean analyzing billions of data points and petabytes of data, whether to predict an outcome, identify a trend, or chart the best course through a sea of ambiguity. Companies with this type of intelligence on demand react faster and make better decisions than their competitors.

Continuing innovations in analytics provides companies with an intelligence windfall benefiting all areas of the business. When you need critical information urgently, the platform that delivers it should be the last thing on your mind. It should be as simple, reliable, and immediate as a light switch, able to handle almost incomprehensible workloads without complexity getting in the way. It must be built for longevity, with a technology foundation able to sustain performance as more users run increasingly complex workloads and as data volumes continue to grow. Furthermore, to maximize returns to the business it should have the lowest total cost of ownership.

Extreme performance with appliance simplicity

Netezza, an IBM® company, transforms the data warehouse and analytics landscape with a platform built to deliver extreme, industry-leading price-performance with appliance simplicity. It is a new frontier in advanced analytics, with the ability to carry out monumental processing challenges with blazing speed, without barriers or compromises. For users and their organizations, it means the best intelligence for all who need it, even as demands for information escalate.

The Netezza data warehouse and analytics appliance's revolutionary design provides exceptional price-performance. As a purpose-built appliance for high speed analytics, its strength comes not from the most powerful and expensive components but from having the right components assembled and working together to maximize performance. Massively parallel processing (MPP) streams combine multi-core CPUs with Netezza's unique Field Programmable Gate Arrays (FPGA) Accelerated Streaming Technology (FAST) engines to deliver performance that in many cases exceeds expectations. And as an easy-to-use appliance, the system delivers its phenomenal results out of the box, with no indexing or tuning required. Appliance simplicity extends to application development, enabling organizations to innovate rapidly and bring high performance analytics to the widest range of users and processes.

This IBM Redguide™ publication introduces the Netezza Asymmetric Massively Parallel Processing (AMPP) architecture, and describes how the system orchestrates queries and analytics to achieve its unprecedented speed. You will see how Netezza software and hardware come together to extract the maximum utilization from every critical component, and how a system optimized for tens of thousands of users querying huge data volumes really works. It is a unique data warehouse and analytics platform with unparalleled price-performance, ready for today's needs and tomorrow's challenges.

Architectural principles

The Netezza appliances integrate database, processing, and storage in a compact system optimized for analytical processing and designed for flexible growth. The system architecture is based on the following core tenets that have been a hallmark of Netezza leadership in the industry:

- Processing close to the data source
- Balanced massively parallel architecture
- Platform for advanced analytics
- Appliance simplicity
- Accelerated innovation and performance improvements
- Flexible configurations and extreme scalability

Processing close to the data source

The Netezza architecture is based on a fundamental computer science principle: when operating on large data sets, do not move data unless absolutely necessary. The Netezza fully exploits this principle by utilizing commodity components called Field Programmable Gate Arrays (FPGAs) to filter out extraneous data as early in the data stream as possible and as fast as data streams off the disk. This process of data elimination close to the data source removes I/O bottlenecks and frees up downstream components such as the CPU, memory, and network from processing superfluous data, thus having a significant multiplier effect on system performance.

Balanced, massively parallel architecture

The Netezza architecture combines the best elements of Symmetric Multiprocessing (SMP) and Massively Parallel Processing (MPP) to create an appliance purpose-built for analyzing petabytes of data quickly. Every component of the architecture, including the processor, FPGA, memory, and network, is carefully selected and optimized to service data as fast as the physics of the disk allows, while minimizing cost and power consumption. The Netezza software orchestrates these components to operate concurrently on the data stream in a pipeline fashion, thus maximizing utilization and extracting the utmost throughput from each MPP node. In addition to raw performance, this balanced architecture delivers linear scalability to more than a thousand processing streams executing in parallel, while offering a very economical total cost of ownership.

Platform for advanced analytics

The principles of MPP and processing data close to the source are equally applicable to advanced analytics on large data sets. Netezza appliances simply process on a massively parallel scale complex algorithms expressed in languages other than SQL, with none of the intricacies typical of parallel and grid programming. Running analytics of any complexity *on stream* against huge data volumes eliminates the delays and costs incurred moving data to separate hardware. It accelerates performance by orders of magnitude, making Netezza the ideal platform to converge data warehousing with advanced analytics.

Appliance simplicity

By automating and streamlining day-to-day operations, the Netezza architecture shields users from the underlying complexity of the platform. Simplicity rules whenever there is a design tradeoff with any other aspect of the appliance. Unlike other solutions, it just runs, handling demanding queries and mixed workloads with blistering speed, without the tuning required by other systems. Even normally time-consuming tasks such as installation, upgrades, and ensuring high availability and business continuity are vastly simplified, saving precious time and resources.

Accelerated innovation and performance improvements

One of the key goals of the Netezza architecture is to deliver price-performance improvements and innovative functionality faster than competing technologies over the long run. While the use of open, blade-based components allows the Netezza architecture to incorporate technology enhancements very quickly, the turbocharger effect of the FPGA, a balanced hardware configuration, and tightly coupled intelligent software combine to deliver overall performance gains far greater than those of individual elements. In fact, the Netezza platform has delivered more than 4 times performance improvement every two years (double that of Moore's Law) since its introduction.

Moore’s law: Gordon Moore, Intel® co-founder, predicted in 1965 that the number of transistors on a chip will double about every two years. Software applications generally rely on these processor improvements to accelerate performance over time.[a]

[a] “Cramming more components onto integrated circuits”, Gordon Moore, Electronics, Volume 38, Number 8, April 19, 1965

Flexible configurations and extreme scalability

The Netezza platform scales modularly from a few hundred gigabytes to tens of petabytes of queryable user data. The system architecture serves the needs of different segments of the data warehouse and analytics market. The use of open blade-based components allows the disk-processor-memory ratio to be easily modified in configurations that cater to performance- or storage-centric requirements. The same architecture also supports memory-based systems that provide extremely fast, real-time analytics for mission-critical applications.

The following sections examine how the Netezza solution puts these principles into practice.

System building blocks

A major part of the Netezza solution's performance advantage comes from its unique AMPP architecture (shown in Figure 1), which combines an SMP front end with a shared nothing MPP back end for query processing. Each component of the architecture is carefully chosen and integrated to yield a balanced overall system. Every processing element operates on multiple data streams, filtering out extraneous data as early as possible. More than a thousand of these customized MPP streams work together to *divide and conquer* the workload.

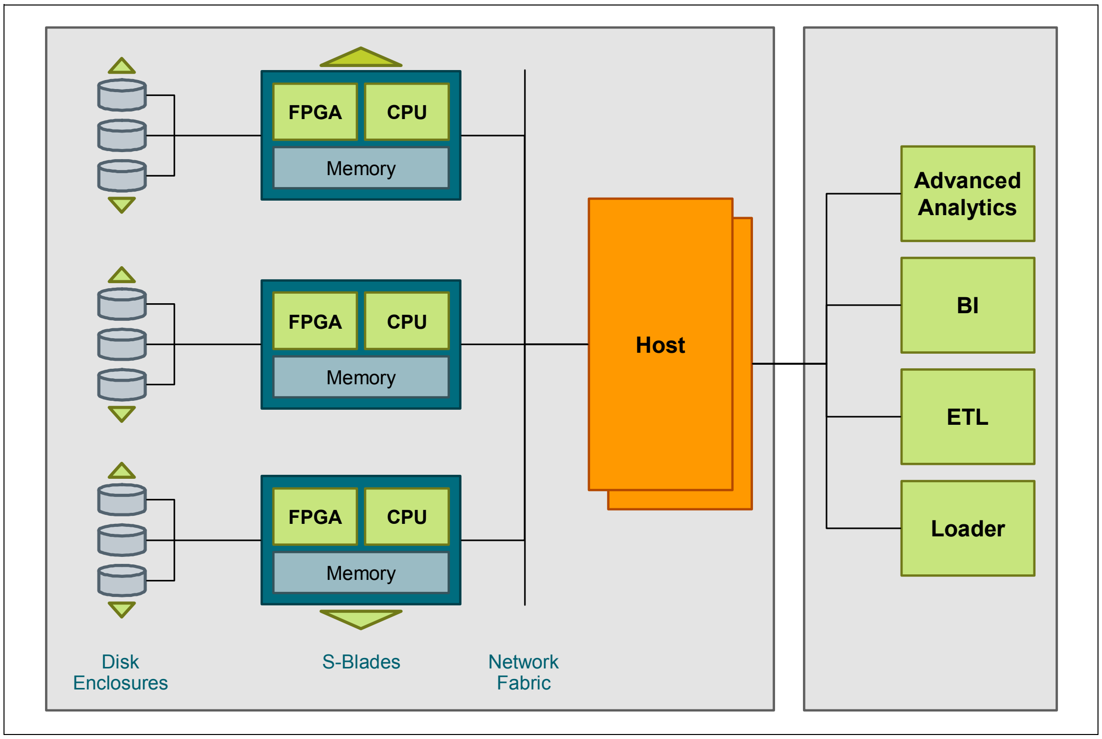


Figure 1 AMPP architecture

Let's examine the key building blocks of the appliance:

- Netezza hosts

  The SMP hosts are high-performance Linux® servers set up in an active-passive configuration for high availability. The active host presents a standardized interface to external tools and applications. It compiles SQL queries into executable code segments called snippets, creates optimized query plans, and distributes the snippets to the MPP nodes for execution.

- Snippet Blades (S-Blades)

  S-Blades are intelligent processing nodes that make up the turbocharged MPP engine of the appliance. Each S-Blade is an independent server containing powerful multi-core CPUs, multi-engine FPGAs, and gigabytes of RAM, all balanced and working concurrently to deliver peak performance. The CPU cores are designed with ample headroom to run complex algorithms against large data volumes for advanced analytics applications.

- Disk enclosures

  The disk enclosures' high-density, high-performance disks are RAID protected. Each disk contains a slice of every database table's data. A high-speed network connects disk enclosures to S-Blades, allowing all the disks in a Netezza to simultaneously stream data to the S-Blades at the maximum rate possible.

- Network fabric

  A high-speed network fabric connects all system components. The Netezza appliance runs a customized IP-based protocol that fully utilizes the total cross-sectional bandwidth of the fabric and eliminates congestion even under sustained, bursty network traffic. The network is optimized to scale to more than a thousand nodes, while allowing each node to initiate large data transfers to every other node simultaneously.

> **Note:** All system components are redundant. While the hosts are active-passive, all other components in the appliance are hot swappable. User data is fully mirrored, enabling better than 99.99% availability.

Where extreme performance happens: inside an S-Blade

Commodity components and Netezza software combine to extract the utmost throughput from each MPP node. A dedicated high-speed interconnect from the storage array delivers data to memory as quickly as each disk can stream. Compressed data is cached in memory using a smart algorithm, which ensures that the most commonly accessed data is served right out of memory instead of requiring a disk access. FAST Engines (shown in Figure 2) running in parallel inside the FPGAs uncompress and filter out 95–98% of table data at physics speed, keeping only data needed to answer the query. The remaining data in the stream is processed concurrently by CPU cores, also running in parallel. The process is repeated on more than a thousand of these parallel Snippet Processors running in the Netezza appliance.

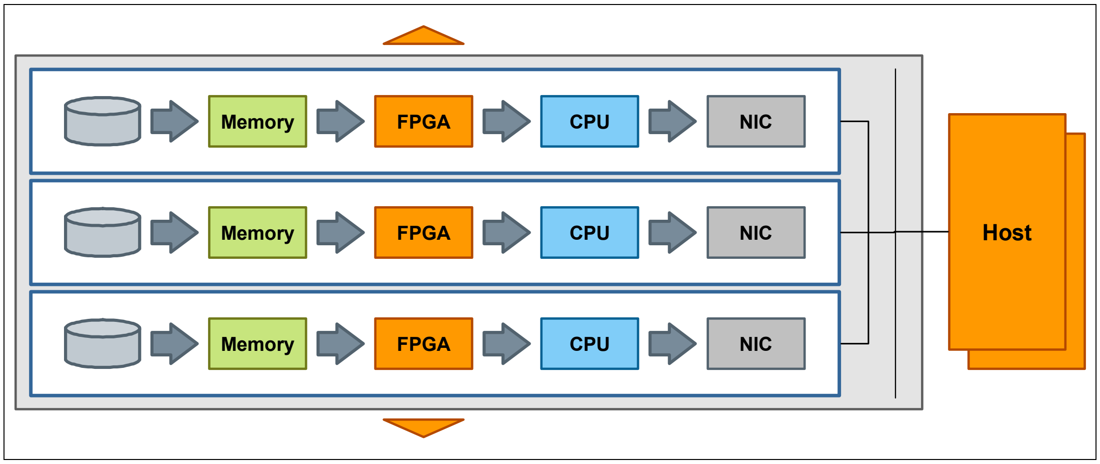


Figure 2 Inside S-Blade

Turbocharging the S-Blades: the power of Netezza FAST engines

The FPGA is a critical enabler of the price-performance advantages of the Netezza platform. Each FPGA contains embedded engines that perform filtering and transformation functions on the data stream. These FAST engines (shown in Figure 3) are dynamically reconfigurable, allowing them to be modified or extended through software. They are customized for every snippet through parameters provided during query execution and act on the data stream delivered by a Direct Memory Access (DMA) module at extremely high speed.

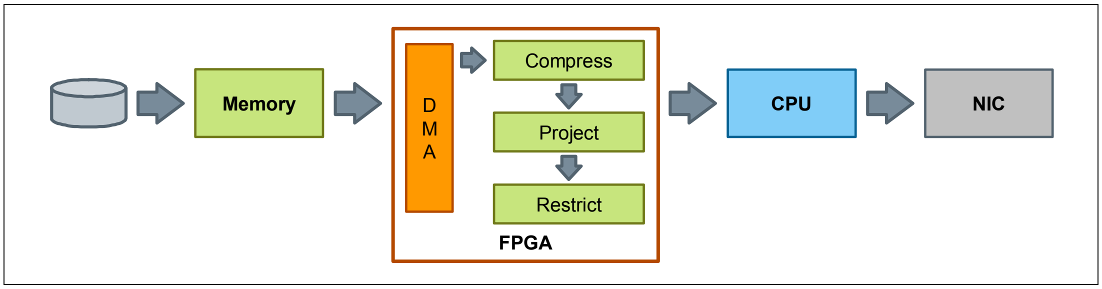


Figure 3   Netezza FAST engines

FAST engines include:

- The Compress engine, a Netezza innovation boosting system performance by a factor of 4 to 8 times. The engine uncompresses data at wire speed, instantly transforming each block on disk into 4 to 8 blocks in memory. The result is a significant speedup of the slowest component in any data warehouse, the disk.
- The Project and Restrict engines, which further increase performance by filtering out columns and rows respectively, based on the parameters in the `SELECT` and `WHERE` clauses in a SQL query.
- The Visibility engine, which plays a critical role in maintaining Atomicity, Consistency, Isolation, and Durability (ACID) compliance at streaming speeds in the Netezza platform. It filters out rows that should not be *seen* by a query; for example, rows belonging to a transaction that is not yet committed.

The Netezza FAST engines provide an extensible framework for innovative future functions to be added through updates to the Netezza software. These new functions promise further improvement in system performance, security, and reliability.

Orchestrating queries on the Netezza platform

The Netezza hardware components and intelligent system software are closely intertwined. The software (shown in Figure 4) is designed to fully exploit the hardware capabilities of the appliance and incorporates numerous innovations to offer exponential performance gains, whether for simple inquiries, complex ad-hoc queries, or deep analytics. In this section, we examine the intelligence built into the system every step of the way.

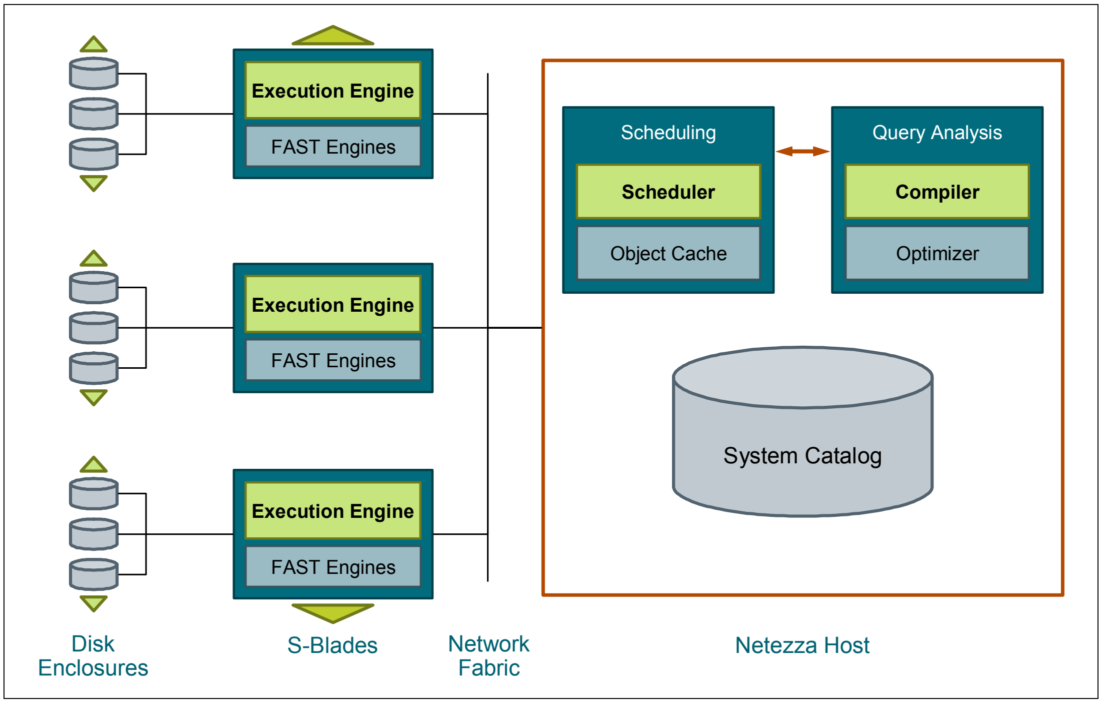


Figure 4 Software architecture

Netezza software components include:

- A sophisticated parallel optimizer that transforms queries to run more efficiently and ensures that each component in every processing node is fully utilized
- An intelligent scheduler that keeps the system running at its peak throughput, regardless of workload
- Turbocharged Snippet Processors that efficiently execute multiple queries and complex analytics functions concurrently
- A smart network that makes moving large amounts of data through the Netezza system a breeze

Let's see how these elements work together, starting when a user submits a query. Technology-savvy readers will see that the Netezza processes queries very differently than other data warehouse systems.

Make an optimized query plan

The host compiles the query and creates a query execution plan optimized for the Netezza AMPP architecture. The intelligence of the Netezza optimizer is one of the system's greatest strengths. The optimizer makes use of all the MPP nodes in the system to gather detailed, up-to-date statistics on every database table referenced in a query. A majority of these metrics are captured during query execution with very low overhead, yielding just-in-time statistics that are individualized per query. The appliance nature of the Netezza system, with integrated components able to communicate with each other, allows the cost-based optimizer to more accurately measure disk, processing, and network costs associated with an operation. By relying on accurate data rather than heuristics alone, the optimizer is able to generate query plans that utilize all components with extreme efficiency.

Intelligence in the optimizer (calculating join order): One example of optimizer intelligence is the ability to determine the best join order in a complex join. For example, when joining multiple small tables to a large fact table, the optimizer can choose to broadcast the small tables in their entirety to each of the S-Blades, while keeping the large table distributed across all Snippet Processors. This approach minimizes data movement while taking advantage of the AMPP architecture to parallelize the join.

By utilizing these statistics to transform queries before processing begins, the optimizer minimizes disk I/O and data movement, the two factors slowing performance in a data warehouse system. Transforming operations performed by the optimizer include:

- Determining correct join order
- Rewriting expressions
- Removing redundancy from SQL operations

Convert it to snippets

The compiler converts the query plan into executable code segments, called snippets, which are query segments executed by Snippet Processors in parallel across all the data streams in the appliance. Each snippet has two elements: compiled code executed by individual CPU cores and a set of FPGA parameters to customize the FAST engines' filtering for that particular snippet. This snippet-by-snippet customization allows the Netezza platform to provide, in effect, a hardware configuration optimized on the fly for individual queries.

Intelligence in the compiler (the object cache): The host uses a feature called the object cache to further accelerate query performance. This is a large cache of previously compiled snippet code that supports parameter variations. For example, a snippet with the clause, `where name = ‘bob’` might use the same compiled code as a snippet with the clause, `where name = ‘jim’` but with settings that reflect the different name. This approach eliminates the compilation step for over 99% of snippets.

Schedule them to run at just the right moment

The Netezza scheduler (shown in Figure 5) balances execution across complex workloads to meet the objectives of different users, while maintaining maximum utilization and throughput. It considers a number of factors, including query priority, size, and resource availability, in determining when to execute snippets on the S-Blades. The scheduler uses the appliance architecture to gather up-to-date and accurate metrics about resource availability from each component of the system. Using sophisticated algorithms, the scheduler maximizes system throughput by utilizing close to 100% of the disk bandwidth and ensuring that memory and network resources are not overloaded, a common cause of thrashing for other, less efficient systems. This is an important characteristic of the Netezza platform, ensuring the system keeps performing at peak throughput even under very heavy loads.

When the scheduler gives the green light, the snippet is broadcast to all Snippet Processors through the intelligent network fabric.

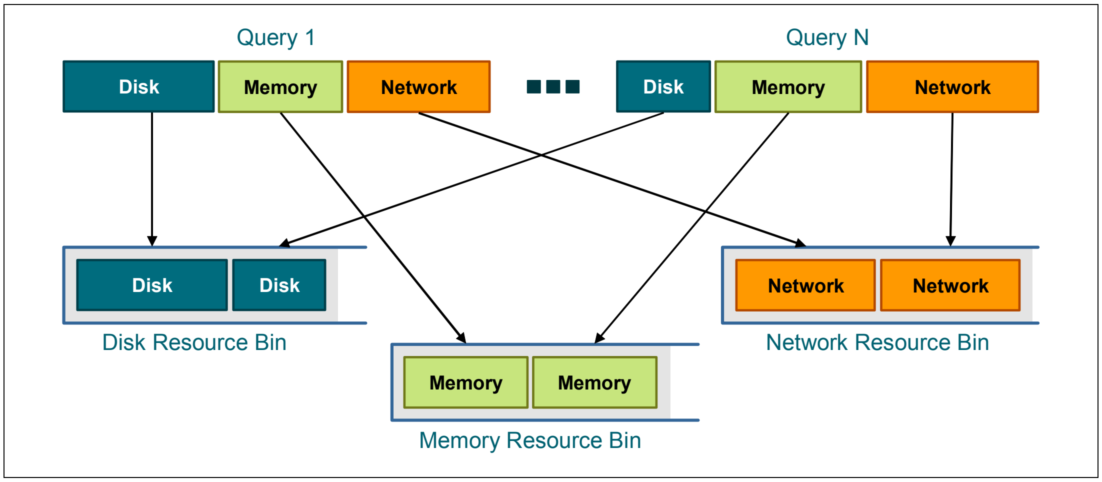


Figure 5 Intelligence in the Scheduler: no resource overloading

Execute them in parallel

Each Snippet Processor on every S-Blade now has the instructions it needs to execute its portion of the snippet. In addition to the host scheduler, the Snippet Processors have their own smart preemptive scheduler that allows snippets from multiple queries to execute simultaneously. The scheduler takes into account the priority of the query and the resources set aside for the user or group that issued it to decide when and for how long to schedule a particular snippet for execution. When that instant arrives, it's show time:

1. The processor core on each Snippet Processor configures the FAST engines with parameters contained in the query snippet and sets up a data stream.
2. The Snippet Processor reads table data from the disk array into memory, utilizing a Netezza innovation called ZoneMap^{TM} acceleration to reduce disk scans. The Snippet Processor also interrogates the cache before accessing the disk for a data block, avoiding a scan if the data is already in memory.
3. The FPGA then acts on the data stream. It first accelerates the data stream by a factor of up to 4 to 8 times by uncompressing the data stream at wire speed.
4. The FAST engines then filter out any data not relevant to the query. The remaining data streams back to memory for concurrent processing by the CPU core. This data is typically a tiny fraction (2–5%) of the original stream, greatly reducing the execution time required by the processor core.
5. The processor core picks up the data stream and performs core database operations such as sorts, joins, and aggregations. It also applies complex algorithms embedded in the Snippet Processor for advanced analytics processing.
6. Results from each Snippet Processor are assembled in memory to produce a sub-result for the entire snippet. This process is repeated simultaneously across more than a thousand Snippet Processors, with hundreds or thousands of query snippets executing in parallel.

ZoneMap acceleration (the Netezza anti-index): ZoneMap acceleration exploits the natural ordering of rows in a data warehouse to accelerate performance by orders of magnitude. The technique avoids scanning rows with column values outside the start and end range of a query. For example, if a table contains two years of weekly records (~100 weeks) and a query is looking for data for only one week, ZoneMap acceleration can improve performance up to 100 times. Unlike indexes, ZoneMaps are automatically created and updated for each database table, without incurring any administrative overhead.

And return the results!

All Snippet Processors now have snippet results that must be assembled. The Snippet Processors use the intelligent network fabric to communicate flexibly with the host and with each other to perform intermediate calculations and aggregations.

Intelligence in the network (predictable performance and scalability): The Netezza custom network protocol is designed specifically for the data volumes and traffic patterns associated with high volume data warehousing. The Netezza protocol ensures maximum utilization of the network bandwidth without overloading it, allowing predictable performance close to the line rate.

Traffic flows smoothly in three distinct directions:

- From the host to the Snippet Processors (1 to 1000+) in broadcast mode
- From Snippet Processors to the host (1000+ to 1), with aggregation in the S-Blades and at the system rack level
- Between Snippet Processors (1000+ to 1000+), with data flowing freely on a massive scale for intermediate processing

The host assembles the intermediate results received from the Snippet Processors, compiles the final result set and returns it to the user's application. Meanwhile, other queries are streaming through the system at various stages of completion.

Summary

The best solutions are not necessarily the biggest or most expensive, they are the ones that have the smartest design. The Netezza team recognized and exploited the inherent advantage that streaming processing provides over the traditional computing architectures used by other analytic and data warehousing systems. The result is a compact appliance with performance dwarfing that of much larger systems, with blinding speed for running complex algorithms against huge data volumes and the mixed workloads created by thousands of concurrent users. Processing performance is complemented by other capabilities that make the Netezza solution a unique platform to help businesses succeed, including:

- Simplicity of use

  The Netezza platform is self-managed, as an appliance should be, and is always running at its peak throughput. The system software ensures that without human intervention.

- Better decisions across the enterprise

  Embedded functions bring a new generation of analytics into the database with minimum development effort. There is no need for separate server hardware or time lost in massive data transfers – just lightning-fast results and the ability to bring crucial business intelligence to everyone who could benefit, in all sectors of an organization.

- Agility for the future

  The system is built not just for today's challenges, but for years to come, scaling linearly to tens of petabytes of user data and with performance acceleration far beyond the conventional speed-up governed by Moore's Law.

The Netezza platform allows you and your company to make decisions with maximum clarity while taking performance for granted. But do not just take our word for it. The best way to appreciate the Netezza solution is to see it in action. We think you will agree there is simply nothing else like it for making the most of your data.

Other resources for more information

For additional information, refer to the Netezza website:

http://www.netezza.com/

The author who wrote this guide

This guide was produced by a specialist working with the International Technical Support Organization (ITSO).

Phil Francisco is the Vice President, Product Management and Product Marketing in the USA for Netezza, an IBM company. He has over 20 years of experience in development and global technology marketing. Phil holds BS degrees in electrical engineering and computer science from Moore School of Electrical Engineering at the University of Pennsylvania, a Master's degree in electrical engineering from Stanford University, and he completed the Advanced Management Program at the Fuqua School of Business at Duke University.

Thanks to the following people for their contributions to this project:

Stephanie Caputo
IBM Software Group, Information Management

David Carter
IBM Software Group, Information Management

LindaMay Patterson
International Technical Support Organization, Rochester Center

Now you can become a published author, too!

Here's an opportunity to spotlight your skills, grow your career, and become a published author—all at the same time! Join an ITSO residency project and help write a book in your area of expertise, while honing your experience using leading edge technologies. Your efforts will help to increase product acceptance and customer satisfaction, as you expand your network of technical contacts and relationships. Residencies run from two to six weeks in length, and you can participate either in person or as a remote resident working from your home base.

Find out more about the residency program, browse the residency index, and apply online at:

ibm.com/redbooks/residencies.html

Stay connected to IBM Redbooks

- Find us on Facebook:

  http://www.facebook.com/IBMRedbooks

- Follow us on Twitter:

  http://twitter.com/ibmredbooks

- Look for us on LinkedIn:

  http://www.linkedin.com/groups?home=&gid=2130806

- Explore new Redbooks® publications, residencies, and workshops with the IBM Redbooks weekly newsletter:

  https://www.redbooks.ibm.com/Redbooks.nsf/subscribe?OpenForm

- Stay current on recent Redbooks publications with RSS Feeds:

  http://www.redbooks.ibm.com/rss.html

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs.

This document, REDP-4725-00, was created or updated on January 14, 2011.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. These and other IBM trademarked terms are marked on their first occurrence in this information with the appropriate symbol (® or ™), indicating US registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at http://www.ibm.com/legal/copytrade.shtml

The following terms are trademarks of the International Business Machines Corporation in the United States, other countries, or both:

The following terms are trademarks of other companies:

IBM®
Redbooks®
Redguide™
Redbooks (logo) ®

Intel, Intel logo, Intel Inside logo, and Intel Centrino logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Linux is a trademark of Linus Torvalds in the United States, other countries, or both.

Other company, product, or service names may be trademarks or service marks of others.