



# **IBM Netfinity Web Server Accelerator**

## ***Executive Summary***

As businesses grow and their use of the Internet in e-commerce expands, systems administrators are faced with the complexities associated with heavier system loads and demand for higher bandwidth. And with the increasing number of users accessing Internet and intranet sites and applications, minimizing response time is a critical factor in optimizing users' productivity while maintaining the most efficient use of hardware.

IBM Netfinity® servers running Microsoft® Windows NT® and using the IBM Netfinity Web Server Accelerator—standard caching reverse proxy software—technology developed by IBM, can cut server response time in half while substantially lowering the time your server's processor spends responding to requests for static Web page content. This frees the processors to spend more time on other tasks, such as processing dynamic content.

With its support for heterogeneous applications, the Netfinity Web Server Accelerator gives you an advantage you can't readily get from other Intel processor-based platforms. And with the purchase of any Netfinity server, you can download the software from the Netfinity Web site at no additional charge and add the application to your Internet Web server running Windows NT.

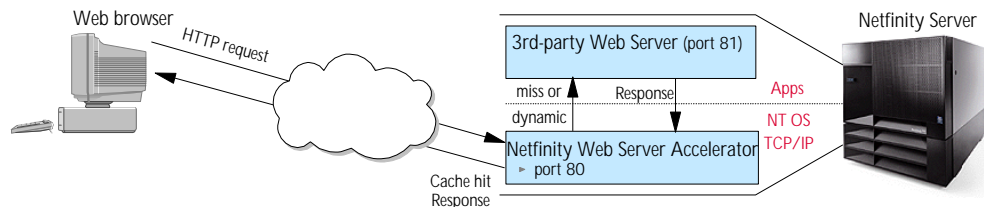
Internet Service Providers (ISPs), who deliver Web solutions, can also benefit from the Web Server Accelerator. Facing a competitive marketplace, ISPs need an infrastructure that can help them provide additional applications and host more sites at an affordable price. They often can't afford to invest the time and expense required to develop complex applications on UNIX® platforms. Yet, customers are demanding that Web sites be brought on-line in weeks, not months. And current levels of service must be maintained if not improved to retain customers and attract more.

The Web Server Accelerator for Windows NT helps ISPs deliver Web solutions with enhanced price and performance value. End users are able to access static Web pages fast. This means increased productivity for end users and a way for ISPs to improve service levels on each individual server. Increased capacity and customer satisfaction pave the way for an increased customer base for ISPs.

This paper provides a brief overview of the IBM Netfinity Web Server Accelerator, its flexibility and scalability, and how they can help deliver performance and reliability where they are needed—in your business. Of course, results will vary based on your specific implementation and Web strategy.

## **Leading-edge Technology from IBM Research**

The IBM Netfinity Web Server Accelerator caches static file content on behalf of a Web server. Requests for this content are intercepted at the lowest levels of the protocol stack, avoiding overhead and reducing access time. Web-based files that are cached stay in sync with the file system, allowing the latest Web content to be available. If one of the Web site's files is modified (by a database or an editor), the Web Server Accelerator immediately detects this change and updates the cache accordingly. By operating at the system level, cache coherency problems are eliminated. The Web Server Accelerator quickly detects whether it can service the request made to the Web server. When the request is for uncacheable static content or for dynamic content, the request is passed to the Web server for servicing. Requests for dynamic content such as Java™ servlets and CGI are sent directly to the Web server.



The Web Server Accelerator is administered using a customized Microsoft Management Console (MMC) snap-in. Administration is straightforward: The Web administrator first points the Accelerator to the Web server's new HTTP port and configures the Accelerator to take over the Web site's original HTTP port. Next, the home directory for the Web site is specified in the Accelerator's cache dialog. The last step is to specify which file types and directories the Accelerator will include in its cache. This step determines which requests will be handled by the Accelerator and which by the Web server. The Web administrator has complete control over what requests the Accelerator will respond to.

The Web Server Accelerator processes requests quickly by acting as a kernel driver. The Accelerator driver extends the Windows NT TCP stack using published interfaces collectively called Transport Driver Interface (TDI). Using TDI, the Accelerator is notified when connection requests occur and data packets arrive. The incoming data is accumulated until a well-formed HTTP request is obtained. The response for this request is looked up and, if possible, returned quickly to the remote browser.

For best performance, the Netfinity Web Server Accelerator ensures that popular HTTP responses and headers are "ready to go" when requests arrive from remote browsers. The Accelerator caches responses and headers in a two-level memory hierarchy, where very popular pages use additional memory manager resources to achieve better performance.

The Netfinity Web Server Accelerator is designed to scale on Netfinity SMP servers. This means that adding more processors to your Netfinity system will improve the effectiveness of the Accelerator. A number of kernel techniques are used to achieve scalability, including hierarchically locking of cached responses and per-processor thread worker pools.

## ***Flexibility and Scalability***

You need not sacrifice performance for application flexibility. The IBM Netfinity Web Server Accelerator supports Windows NT based Web servers, such as Apache, Netscape Fastrac and Microsoft Internet Information Server 4.0, to deliver fast Web response for end users.

The Web Server Accelerator scales from one or two processors (Netfinity 5500) to four processors (Netfinity 7000 M10) to deliver outstanding performance as demonstrated in the SPECweb96 benchmark reported in July 1998. This type of performance can support the needs of heavily trafficked e-commerce Web sites, with performance depending on response file sizes and access patterns. The Web Server Accelerator frees the server to deliver more cycles for processing dynamic Web content in e-commerce applications such as on-line shopping, secure electronic transactions and database searches.

And Netfinity servers do not force you to change your Windows NT Web server software, because IBM designed the Web Server Accelerator for compatibility. Businesses can have substantial investments in programs written for specific Web server software. IBM protects those investments by allowing customers to continue using their existing, supported Windows NT based Web server software—no need to sacrifice these programs to get the Web Server Accelerator's performance benefits on Netfinity servers.

Designed by IBM Research for maximum performance, installation and administration ease, the Web Server Accelerator is integrated into the MMC so that administrators can configure the Web Server Accelerator through a Windows NT standard interface. No additional, specialized management applications are required—no hoops to jump through to set up and operate.

## ***Supercharged Web Server Performance***

Performance isn't about just handling the greatest number of requests ("hits"); it's also about the most efficient use of hardware. Netfinity servers using Web Server Accelerator technology can cut server response time in half, based on SPECweb96 results reported later in this paper, while substantially lowering the time your server's processor spends responding to requests for static Web page content. This is performance value. Netfinity servers using this new technology can simply do more with less. In similar processor-memory configurations (same number of processors, same amount of memory), no other server matches our 1-, 2- and 4-processor SPECweb96 performance.

### ***About SPECweb96***

SPECweb96, with its standardized workload and implementation, measures a system's ability to perform as a World Wide Web server for static pages. The workload simulates the accesses to a Web service provider, where the server supports multiple pages for a number of different organizations. This benchmark is useful in evaluating systems that handle millions of hits per day and multiple hits per second. SPECweb96 provides one of the most objective, most representative benchmarks for measuring Web server performance.

## *Enhancing Web server performance*

SPECweb96 reports for results cited in this paper, as well as all other results, are available on the World Wide Web at [www.specbench.org/osg/web96](http://www.specbench.org/osg/web96).

### **Netfinity Web Server Accelerator Performance**

The SPECweb96 benchmark results, reported in July 1998, demonstrate the robust capabilities of Netfinity servers for handling Web page delivery and e-commerce at heavily trafficked Web sites. The **Netfinity 7000 M10** enterprise server, configured with four 400MHz<sup>1</sup> Intel® Pentium® II Xeon™ processors and 2GB of memory, running Windows NT 4.0 Server, Apache HTTP Web Server Version 1.3.3 and IBM's Web Server Accelerator technology, achieved peak results of 5,277 Web page requests per second, with an average response time of 8.3 msec per request. This is the highest SPECweb96 performance result recorded for a four-way, Windows NT based Web server.

The **Netfinity 5500** server achieved the best one- and two-way SPECweb96 performance on any server/platform. Configured with two 400MHz Pentium II processors and 1GB of memory, and, alternately, with one Pentium II processor and 1GB of memory, the Netfinity 5500 achieved peak results of 3,111 Web page requests per second with an average response time of 16.7 msec per request, and 2,575 Web page requests per second with an average response time of 17.2 msec per request, respectively. The Netfinity 5500 server's throughput is nearly two times higher than that of the vendor whose Web server had the second-best performance (in the similar processor-memory configurations mentioned previously).

<sup>1</sup> MHz only measures internal clock speed, not application performance. Many factors affect application performance.

## **Conclusion**

IBM's Netfinity Web Server Accelerator clearly offers the technological edge for customers who need fast, flexible, scalable and reliable access to static Web page content. With the Web Server Accelerator, you can handle heavier systems loads and achieve higher bandwidth on your Netfinity server simply by adding another processor and maximizing your server investment.

The Web Server Accelerator, from installation onward, helps you achieve:

- Accelerated performance of Internet and intranet sites
- Improved response time for end users
- Reduction of load on your Web server
- Lower CPU utilization for the same number of Web page requests
- More Web page requests per second
- More CPU utilization for other e-commerce applications
- Increased number of Web sites that can be hosted on each server
- Enhanced price/performance value

## *Enhancing Web server performance*

- Increased responsiveness and capacity
- Increased productivity for end users

When you combine these benefits with those of IBM's limited, three-year on-site warranty<sup>2</sup> for Netfinity servers and Options by IBM that are installed on our servers, you have the solution you need to keep your business up and running 7 days a week, 24 hours a day and 365 days a year. Our customers would expect nothing less.

<sup>2</sup> For terms and conditions or copies of IBM's limited warranty, call 1 800 772-2227 in the U.S. Limited warranty includes International Warranty Service in those countries where this product is sold by IBM or IBM Business Partners (registration required). Telephone support may be subject to additional charges.

### ***Additional Information***

For more information on IBM Netfinity direction, products and services, refer to the following white papers, available from our Web site at [www.pc.ibm.com/netfinity](http://www.pc.ibm.com/netfinity).

*Lotus Domino Clusters Overview*

*Lotus Domino Clusters Installation Primer*

*Integrating IBM Netfinity Manager with Intel LANDesk Server Manager*

*IBM Netfinity System Management Processor*

*IBM Netfinity Hot-Plug Solutions*

*IBM Netfinity Storage Management Solutions Using Tape Subsystems*

*IBM Netfinity X-architecture*

*IBM Netfinity Servers and Intel Architecture*

*IBM Netfinity 8-Way SMP Directions*

*IBM Netfinity Cluster Directions*

*IBM Netfinity Fibre Channel Directions*

*IBM Netfinity Ultra2 SCSI Directions*

*IBM Netfinity Server Quality*

*IBM Netfinity ServerGuide for Netfinity and PC Server Systems*

*IBM Netfinity Predictive Failure Analysis*

*Integrating IBM Netfinity Manager with Microsoft Systems Management Server*

*Examples Implementing IBM Netfinity Server Management: Air Conditioning Failure*

*At your service...Difference beyond technology*

*IBM Netfinity System Management for Servers*

## *Enhancing Web server performance*



© International Business Machines Corporation 1998

IBM Personal Computer Company  
Department LO6A  
3039 Cornwallis Road  
Research Triangle Park NC 27709

Printed in the United States of America  
9-98  
All rights reserved

References in this publication to IBM products or services do not imply that IBM intends to make them available in all countries in which IBM operates. IBM reserves the right to change specifications or other product information without notice.

IBM Netfinity servers and PC servers are assembled in the U.S., Great Britain, Japan, Australia and Brazil and are comprised of U.S. and non-U.S. parts.

IBM and Netfinity are trademarks of International Business Machines Corporation in the United States and/or other countries.

Intel, Pentium and Pentium II Xeon are trademarks or registered trademarks of Intel Corporation.

Java is a trademark of Sun Microsystems, Inc.

Microsoft, Windows, Windows NT and the Windows 95 logo are trademarks or registered trademarks of Microsoft Corporation.

UNIX is a registered trademark licensed exclusively through X/Open Company Limited.

Other company, product and service names may be trademarks or service marks of other companies.

THIS PUBLICATION MAY INCLUDE TYPOGRAPHICAL ERRORS AND TECHNICAL INACCURACIES. THE CONTENT IS PROVIDED AS IS, WITHOUT EXPRESS OR IMPLIED WARRANTIES OF ANY KIND.