ActiveAnswers

einfrastructure group

january 2003

hp invent

performance brief

scalability and performance of hp ProLiant BL20p Generation 2 server blades (with 2.8GHz processors)

contents

abstract: HP ProLiant p-Class server blades are designed for a multi-tiered data center environment, offering density, high performance and high availability. These server blades are ideal for use as high performance front-end servers or for single-function applications such as Thin Client Server Computing (TCSC).

ProLiant BL20p Generation 2 (G2) server blades feature up to two of the latest Intel® 2.8GHz Xeon processors, 512KB level-two cache, up to 8GB of PC2100 DDR memory, two hot-plug SCSI hard drive bays, and HP Integrated Lights-Out (iLO) Advanced management capability.

This Performance Brief outlines scalability and performance metrics for a ProLiant BL20p G2 server blade in a TCSC environment. HP projects the optimal number of users that can be supported by one-way and two-way servers running Microsoft® Windows® 2000 Server, Terminal Services and Citrix MetaFrame XP.

Overall, test results indicate that the ProLiant BL20p G2 server blade is a performance leader.

notice

Technical information in this document is subject to change without notice.

Intel and Pentium are US registered trademarks of Intel Corporation.

Itanium is a registered trademark of Intel Corporation in the US and other countries and is used under license.

Microsoft, Windows and Windows NT are registered trademarks of Microsoft Corporation.

january 2003
P/N 5981-5627ENA1

introduction

This performance brief discusses the ProLiant BL20p G2 ultra-dense server blade, which is optimized for rapid deployment and provisioning. This server blade is ideal for space-constrained enterprises or service providers needing a high performance front-end server, or for single-function applications such as Thin Client Server Computing.

The combination of increasing costs and the scarcity of skilled IT resources is leading more and more organizations to implement enterprise-wide TCSC solutions based on Citrix MetaFrame XP. For these solutions to be cost-effective, the MetaFrame XP server farm must be implemented using highly-available servers that can be rapidly scaled and easily managed.

server blades in a TCSC environment

Using ProLiant BL20p G2 server blades to deploy a server farm fulfills the following TCSC needs:

- **scalability**
  The server farm can be scaled out to distribute risk and load while minimizing the impact on users.

- **performance**
  ProLiant BL20p G2 server blades provide the performance necessary to support the most demanding line-of-business applications.

- **rapid server deployment and redeployment**
  Rapid deployment capability allows the organization to save valuable time by quickly deploying the TCSC solution, then quickly and dynamically responding to changing business needs.

  – It takes only seconds to install server blades and power supplies once the rack infrastructure is in place. This allows dynamic scaling – without powering down the system.

  – There is single-sided access to most pluggable components.

  – ProLiant Essentials Rapid Deployment Pack (RDP) software allows the administrator to pre-configure each server bay before installing server blades. After installation, server blades configure automatically, assuming the role assigned by the administrator.

  – RDP also provides rip-and-replace capability. When a server blade is replaced, the new server blade automatically assumes the role assigned to the original server blade.

- **anytime, anywhere access**
  Integrated Lights-Out (iLO) Advanced management allows anytime, anywhere access to the server farm for remote management.

- **reduced total cost of ownership**
  ProLiant BL20p server blades can further reduce Total Cost of Ownership (TCO) in the following ways:

  – consuming 25% less power per user, which reduces utility and air conditioning costs

  – supporting 10% more users per rack, which reduces licensing and real estate costs

  – reducing cabling complexity

audience

This performance brief is intended for the use of field systems engineers, resellers and other professionals.

ProLiant server blade overview

Businesses are deploying more and more servers for TCSC or edge-of-the-network applications such as web serving, media streaming, load balancing, caching, and firewall protection. However, adding servers increases operating costs, consumes more power and space, and increases the complexity of system administration; existing resources may become inadequate.

HP has developed the ProLiant BL line of modular server blades specifically to address the needs of space-constrained enterprises and service providers for increased server density, rapid deployment and provisioning, and remote manageability. The ProLiant BL line is optimized for use with ProLiant Essentials Rapid Deployment Pack (RDP) for automatic, high-volume server deployments or redeployments.

This Performance Brief focuses on the ProLiant BL p-Class system.

p-Class system components

The modular ProLiant BL p-Class system is comprised of the following components:

- **server blades**

  The ProLiant BL20p G2, dual processor server blade is designed for performance front-end and mid-tier computing with enterprise availability.

- **server blade enclosure**

  Up to eight ProLiant BL20p G2 server blades can be housed in each 6U server blade enclosure. Server blades blind-mate into the enclosure's backplane for power and data connections.

- **power distribution**

    Redundant –48VDC power (from a power enclosure or externally-generated) is distributed to server blade enclosures through bus bars or a power bus box.

    The power enclosure houses hot-plug power supplies for the ProLiant BL p-Class system infrastructure. The power supplies convert 200 – 240VAC to the –48VDC required by the server blade enclosures.

- **rack**

    ProLiant BL p-Class system infrastructure is optimized for use with HP 10000 Series racks but fits into other ProLiant racks, telco racks or third-party racks.

more information

More information on ProLiant BL p-Class system infrastructure is available on the HP website.

benefits

Table 1 lists the key benefits of ProLiant BL20p G2 server blades.

Table 1. ProLiant BL20p G2 server blade benefits

Benefit	Description
Processor power	The ProLiant BL20P G2 server blade deploys up to two 2.8GHz Intel Xeon processors with 512KB level-2 cache and a 533MHz system bus.
Reliability	The ProLiant BL20P G2 server blade has the proven reliability that is essential for business-critical applications. The following features are supported: • two hot-plug SCSI hard drive bays with integrated RAID (HP Smart Array 5i Plus) and optional battery-backed write cache • redundant NIC connections • redundant ROM • redundant AC and DC power feeds • optional redundant interconnect switch

continued

Table 1. ProLiant BL20p G2 server blade benefits *(continued)*

Scalability	A p-Class system scales out to eight ProLiant BL20p G2 server blades per enclosure, 48 per 42U rack.
Manageability	ProLiant BL20P G2 server blades can improve productivity, uptime and deployment time. Capabilities include: • plugging in new servers or power supplies in seconds for dynamic scaling – without powering down the system • pre-configuring each enclosure bay so that, when plugged in, a new blade automatically deploys the appropriate operating system and applications • taking advantage of RDP to quickly deploy or redeploy blades – use multicasting to configure two – 100 server blades in 30 minutes • using rip-and-replace technology to replace a defective blade with a new blade that automatically assumes the role of the defective blade • relieving cable congestion – the GbE interconnect switch option can reduce network cabling requirements to a single cable for every eight blades • improving serviceability – the modular design makes upgrades and service easy • integrating with the existing environment – components fit into HP, telco and some third-party racks, and mix with traditional servers and storage
Total cost of ownership	ProLiant BL20P G2 server blades can significantly reduce IT costs and resource requirements. Benefits include: • reducing the IT headcount through remote management, innovative features, and ease of deployment • saving valuable floor space through increased density – up to 48 blades in a 42U rack – without sacrificing high-availability features • taking seconds to snap in new server blades and power supplies • offering headless management – no keyboard, video or mouse • requiring fewer network and power cables • protecting the customer's investment with support for next-generation ProLiant BL p-Class server blades and built-in power headroom

features

The ProLiant BL20p G2 server blade offers the features listed in Table 2.

Table 2. Key features of a ProLiant BL20p G2 server blade

Feature	Description
Processor	Up to two 2.8GHz Xeon processors with 512KB level-2 cache
Chipset	ServerWorks GC-LE
Memory	Four DIMM slots – up to 8GB of ECC PC2100 DDR, 2 x 1 interleaved
Disk	Two universal hot-plug SCSI disk bays, supporting up to 288GB; individual drives up to 144GB supported
Fault tolerance	Embedded Smart Array 5i Plus controller with RAID1 or RAID0, supporting Wide Ultra3 technology and offering an optional battery-backed write cache
Network connectivity	Three auto-sensing 10/100/1000 Ethernet connections; additional dedicated Ethernet connection for Integrated Lights-Out (iLO) management capability (described below)
LED display	LEDs on the front of the blade provide information on unit identification, health, NIC status, disk drive activity, and power.
Headless	Virtual keyboard, video, mouse and media drive available through iLO (described below)
PXE capability	Embedded Pre-boot Execution Environment (PXE) capability, allowing the server blade to boot from the network – particularly useful in an RDP environment (described below)
Diagnostic station	The optional diagnostic station allows the administrator to work with a server blade or GbE interconnect switch that has been removed from the rack. Capabilities include: • diagnosing problems • building an initial server image • testing the server blade after an upgrade • configuring the switch
Integrated Lights-Out (iLO) Advanced	The iLO Advanced feature allows the administrator to access and control the server blade from any location on the network, regardless of the state of the blade's operating system or hardware. iLO provides a virtual power button, keyboard, video, mouse and media drive.

continued

Table 2. Key features of a ProLiant BL20p G2 server blade *(continued)*

Related HP software tools	
Insight Manager 7	HP Insight Manager 7 brings together in one location all performance, management and fault information relating to the IT environment, allowing the administrator to monitor and manage groups of server blades from a standard web browser.
	In addition, Insight Manager 7 includes a Blade Server Visual Locator tool that offers a graphical representation of server blades, server blade enclosures and shared infrastructure.
ProLiant Essentials Rapid Deployment Pack (RDP)	Optional RDP software allows the administrator to deploy and redeploy blade servers, rapidly and automatically. High-volume server deployments are now possible within minutes rather than hours or days.
	RDP also allows the administrator to assign a role to each bay in a server blade enclosure. When a new server blade is plugged in to a particular bay, the pre-assigned software is automatically deployed to the blade; if the blade is replaced, the new blade automatically receives the pre-assigned software, taking over the role of the replaced blade.

new processor technologies

The [Xeon processor](#) used in the ProLiant BL20p G2 server blade features the key innovations and improvements described in Table 3.

Table 3. New processor technologies

New Technology	Description
Hyper-threading technology	Processor chips often use parallelism to improve performance; for example, Intel's Itanium™ architecture is based on Explicitly Parallel Instruction Computing (EPIC) technology, which features instruction-level parallelism. However, Intel's new [hyper-threading technology](#) delivers two logical processors that can simultaneously execute different tasks using shared hardware resources, improving performance by up to 40%.
	With instruction level-parallelism, some hardware resources are mutually-exclusive; however, with hyper-threading, two processors share the resources of a single physical processor so that applications or threads can execute simultaneously. A single physical processor appears as two logical processors to the operating system or application.
	Hyper-threading can boost the performance of multi-threading and multi-tasking operations, and will enable the world's first Simultaneous Multi-threaded (SMT) processor.
	Hyper-threading delivers the following benefits to the customer: • improved performance • improved response times • more users supported • more transactions processed
	Hyper-threading is largely invisible to the platform. Today's multi-processor-aware software is compatible with a hyper-threading-enabled platform; only small changes are needed to optimize many multi-threaded applications for hyper-threading. However, further performance gains can be realized by specifically tuning for hyper-threading.
Rapid execution engine	Two Arithmetic Logic Units (ALUs) on the Xeon processor run at twice the core processor frequency, allowing basic integer instructions (such as Add, Subtract, Logical AND, and Logical OR) to execute in one-half of a clock cycle.

continued

Table 3. New processor technologies *(continued)*

533MHz system bus	The Xeon processor features a 533MHz system bus with capabilities up to 3.2GB per second (as opposed to 1.06GB per second delivered by the Pentium® III's 133MHz system bus).
Hyper-pipelined technology	Hyper-pipelined technology can significantly increase the performance, frequency, and scalability of the processor.
	The pipeline depth of the Xeon processor is double that of the Pentium III; for example, the branch prediction/recovery pipeline is now implemented in 20 rather than 10 stages.

what's new

Key design and technology improvements implemented in the ProLiant BL20p G2 server blade include:

- 2.8GHz Xeon processors
- Gigabit NICs
- PC2100 DDR memory
- up to 8GB memory
- Storage Array Network (SAN) support

performance testing overview

HP tested a variety of ProLiant BL20p G2 server blade configurations in a TCSC environment to determine the following critical performance characteristics:

- **performance**
  HP has made a distinction between the optimal number[1] of users that can be supported by a particular server configuration and the maximum number. The optimal number is a more practical metric that reflects the number of users that can be supported **without impacting response times**.
  The performance testing used various metrics to establish the optimal number of users supported by ProLiant BL20p G2 2.8GHz server blades.

- **System Paged Table Entry/Paged Address Pool limitation**
  The performance testing was also intended to determine if lack of System Paged Table Entry (PTE) or Paged Address Pool space could limit performance when there were no other system bottlenecks.

performance metrics

Performance metrics monitored by HP included:

- **processor utilization** – to establish the number of users at various levels of processor utilization

- **processor queue length** – to establish the lengths of the processor queues at various levels of user activity

- **canary time** – to establish the number of users that are active when response times exceed the baseline by over 10%

Note: HP used the Windows NT® Performance Monitor analysis tool to monitor processor utilization (%CPU Utilization) and processor queue length (Processor Queue Length) metrics.

Because users share the Terminal Server's processors to run applications, the speed and number of processors are critical for maximizing server performance. Though Terminal Services divides processing type equally between each user, processor-intensive applications may degrade overall performance.

---

[1] Historically, HP has defined the optimal number of users as the number of users that are active when processor utilization reaches 80%. Additional users are supported but response times may become unacceptable.

tested configuration

This section provides information on the following topics:

- Server configurations – An overview of the servers used in the tested configuration

- Recommendations – Recommendations for configuring servers

- System summary – A system summary of the server under test

- Client machines – An overview of the clients used in the tested configuration

- Test scripts – An overview of the test scripts used by HP to simulate the activities of Heavy, Medium and Light Users
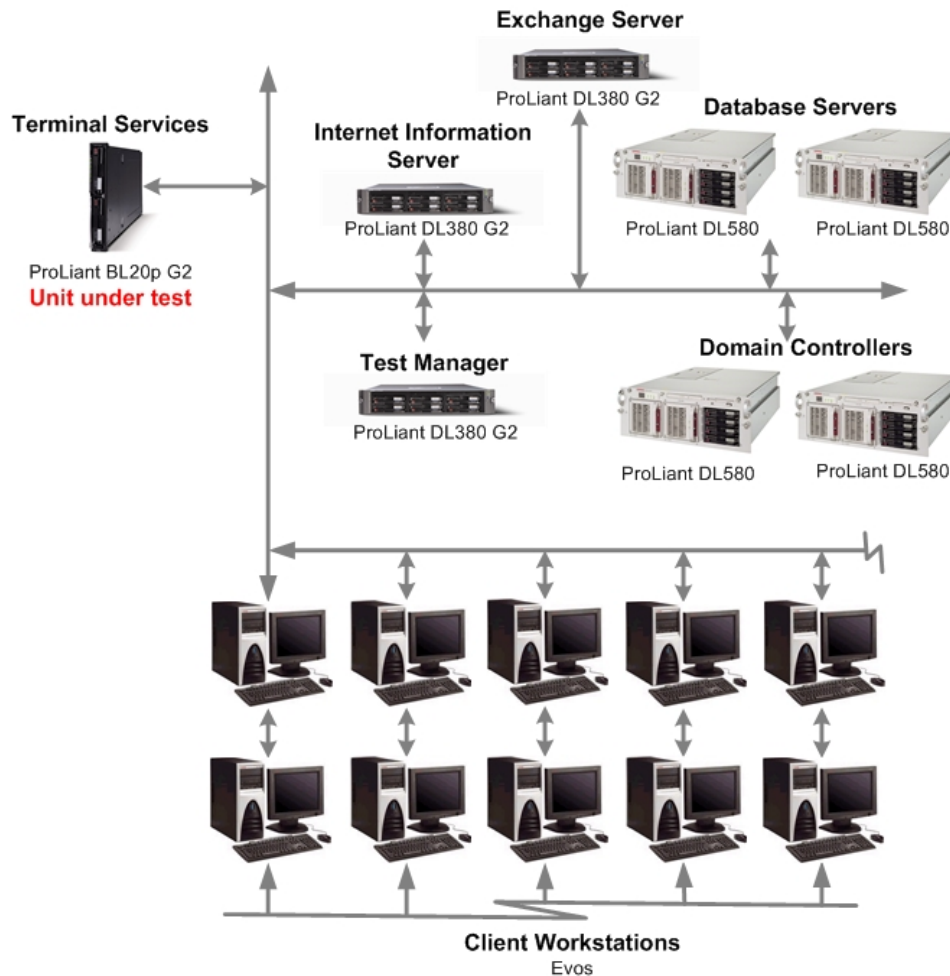
Figure 1 illustrates the test environment.

Figure 1: The tested configuration

Users accessed applications running on the ProLiant BL20p G2 server blade from HP Evo client machines with various configurations. HP varied the numbers of users and the levels of complexity of the work being carried out.

server configurations

Table 4 summarizes the configurations of servers deployed in the test environment.

Table 4. Server Configurations

Server	Software	Hardware
Terminal Services	• Windows 2000 Advanced Server, Terminal Services enabled (see **IMPORTANT** below) • Citrix MetaFrame XP	One ProLiant BL20p G2 server blade with: • One-way or two-way 2.8GHz Xeon processor configuration • 4GB RAM • Two hot-pluggable 18.2GB 10,000RPM Ultra SCSI-2 hard drives • Integrated Smart Array 5i+ controller with RAID 1 • Three available NC7770 PCI-X Gigabit NIC ports
Exchange Server	Microsoft Exchange	Two ProLiant DL380 G2 servers, each with: • Two-way 1.266GHz Pentium III processor configuration • 512KB level-two cache • 2GB RAM • Two hot-pluggable 9.1GB 10,000RPM Ultra SCSI-2 hard drives • Integrated Smart Array 5i controller with RAID 1 • Netelligent 10/100 TX PCI UTP controller
Internet Information Server	Microsoft Internet Information Server 4.0	Two ProLiant DL380 G2 servers, each with: • Two-way 1.266GHz Pentium III processor configuration • 512KB level-two cache • 2GB RAM • Two hot-pluggable 9.1GB 10,000RPM Ultra SCSI-2 hard drives • Integrated Smart Array 5i controller with RAID 1 • Netelligent 10/100 TX PCI UTP controller

continued

Table 4. Server Configurations *(continued)*

Database Server	Microsoft SQL Server 6.5	Two ProLiant DL580 Servers, each with: • Four-way 700MHz Pentium III processor configuration • 512KB level-two cache • 512MB RAM • Four internal hot-pluggable 9.1GB Ultra3 hard drives • ProLiant storage system with six 9.1GB hard drives • Integrated Smart Array controller with RAID 1 • Netelligent 10/100 TX PCI UTP controller
Domain Controller		Two ProLiant DL580 Servers, each with: • Two-way 700MHz Pentium III processor configuration • 512KB level-two cache • 512MB RAM • Integrated Smart Array controller with RAID 1 • Four internal hot-pluggable 9.1GB Ultra3 hard drives; RAID 1 • Netelligent 10/100 TX PCI UTP controller

IMPORTANT: With 50 – 100 users logged on, a Terminal Services server may intermittently stop responding for up to 30 seconds. HP has limited experience with this issue but there have been numerous reports from other end-users.

Microsoft has responded to this issue in Windows Service Pack 3 (SP3) but an additional hotfix is still required. For more information, refer to the Microsoft website.

The customer may not achieve the test results detailed in this Performance Brief unless the Terminal Services server is running SP3 with the hotfix.

recommendations

HP offers the following recommendations for configuring servers in the test environment:

- **Memory**
  All applications reside and are executed on the server blade running Terminal Services. Adequate memory is essential to optimize system performance. HP makes the following recommendations based on user type:

  – 5 – 6MB for each active Light User

  – 8 – 10MB for each active Medium User

  – 12MB for each active Heavy User

  For more information on these user types, refer to the section on Test Scripts.

- **NIC speed**
  Since network traffic is light (only keystrokes, mouse clicks and display updates), HP set all server NICs to 100Mbps.

system summary

Table 5 provides a system summary of the ProLiant BL20p G2 server blade under test.

Table 5. System summary

System Component	Description
Operating system	Microsoft Windows 2000 Advanced Server 5.0.2195 Build 2195
Version	5.0.2195 Service Pack 3 Build 2195
System name	Administrator4
System model	ProLiant BL20p G2
System type	X86-based PC
Processors	X86 Family 15 Model 1 Stepping 1 GenuineIntel~2999MHz
BIOS version	07/30/2002
Windows directory	C:\WINNT
System directory	C:\WINNT\System32
Boot device	\Device\Harddisk0\Partition1
Locale	United States
User name	\Administrator
Time zone	Central Daylight Time

continued

Table 5. System summary *(continued)*

Total physical memory	3,767,828KB
Available physical memory	3,537,016KB
Total virtual memory	11,158,716KB
Available virtual memory	10,850,820KB
Page file space	7,390,888KB
Page file	c:\pagefile.sys

client machines

HP simulated a TCSC environment using a variety of Pentium-based 800MHz – 1GHz Evo workstations as clients. Each Evo was equipped with at least 256MB of memory and operated at 640×480/256-color resolution.

HP set all client NICs to 10Mbps.

test scripts

HP used test scripts that simulated the activities of three types of users (Heavy, Medium and Light). Table 6 describes these users and their simulated activities.

Table 6. Simulated User Activities

User Type	Activities
Heavy user	Heavy Users (also known as Structured Task Workers) tend to open multiple applications simultaneously and remain active for long periods of time. Heavy Users often leave applications open when not in use.
	Heavy users have an average typing speed of 40 – 50 words per minute.
	The Heavy User script included the following activities: • creating, saving and printing documents using Microsoft Word • browsing Web pages, and sending and receiving e-mail using Microsoft Outlook • performing extensive Microsoft Excel activity (using PivotTable dynamic views, graphing and charting data, and so on)

continued

Table 6. Simulated User Activities *(continued)*

Medium user	Medium Users (also known as Knowledge Workers) tend to open and close applications more quickly than Heavy Users. Medium Users typically do not leave applications open when not in use.
	Medium Users have an average typing speed of 35 – 40 words per minute.
	The Medium User script included the following activities: • creating small documents with Microsoft Word • performing simple data entry in Microsoft Excel • sending and receiving e-mail with Microsoft Outlook (seven e-mail messages per hour) • engaging in minimal browsing of an intranet site using Microsoft Internet Explorer 5.5
Light user	Light Users are also known as Task Workers.
	The Light User script simulated a single line-of-business application – a call-center application that started a Visual Basic program to access an SQL database.
	The script queried the database, read, wrote, and printed information.

performance test results

This section provides information on the following topics:

*   Individual performance metrics – uses individual performance metrics to determine the optimal numbers of Heavy Users supported by one-way and two-way server blades

*   System PTE/Paged Address Pool limitation – shows how server performance was limited by a lack of Paged Address Pool space

*   Results – displays the optimal numbers of users supported by one-way and two-way server blades with hyper-threading capability turned on

*   Performance summary – tabulates the performance of one-way and two-way server blades

*   Performance comparison – compares the optimal numbers of users supported by one-way and two-way server blades

*   Canary execution time – offers an alternative performance metric to identify the optimal number of Heavy Users supported by one-way and two-way server blades

Note: HP determined that there were no disk, memory or network bottlenecks in the test environment.

individual performance metrics

HP used individual performance metrics (such as processor utilization and processor queue length) to quantify server performance.

HP varied the following characteristics to create a series of test scenarios:

- server configuration – one-way or two-way

- user type – Heavy, Medium, or Light

- hyper-threading – on

For each test scenario, HP began by running the appropriate script (simulating Heavy, Medium, or Light Users, as appropriate) for a group of ten users. Start times were staggered to eliminate authentication overhead. After the sessions finished, HP added ten more users, then repeated the testing.

This methodology allowed HP to obtain the following performance metrics:

- **optimal number of users**
  Historically, HP has defined the optimal number of users as the number of users that are active when processor utilization reaches 80%. Additional users are supported but response times may become unacceptable.

- **queue length**
  ProLiant G1 servers became congested with a queue length of 10 – 15. However, higher clocking rates, larger caches, and new processor technologies such as Hyper-threading allow G2 servers to process the queue more quickly while still maintaining optimal response times.

For example, Figure 2 shows performance metrics for a two-way ProLiant BL20p G2 server blade with Heavy Users and hyper-threading on.
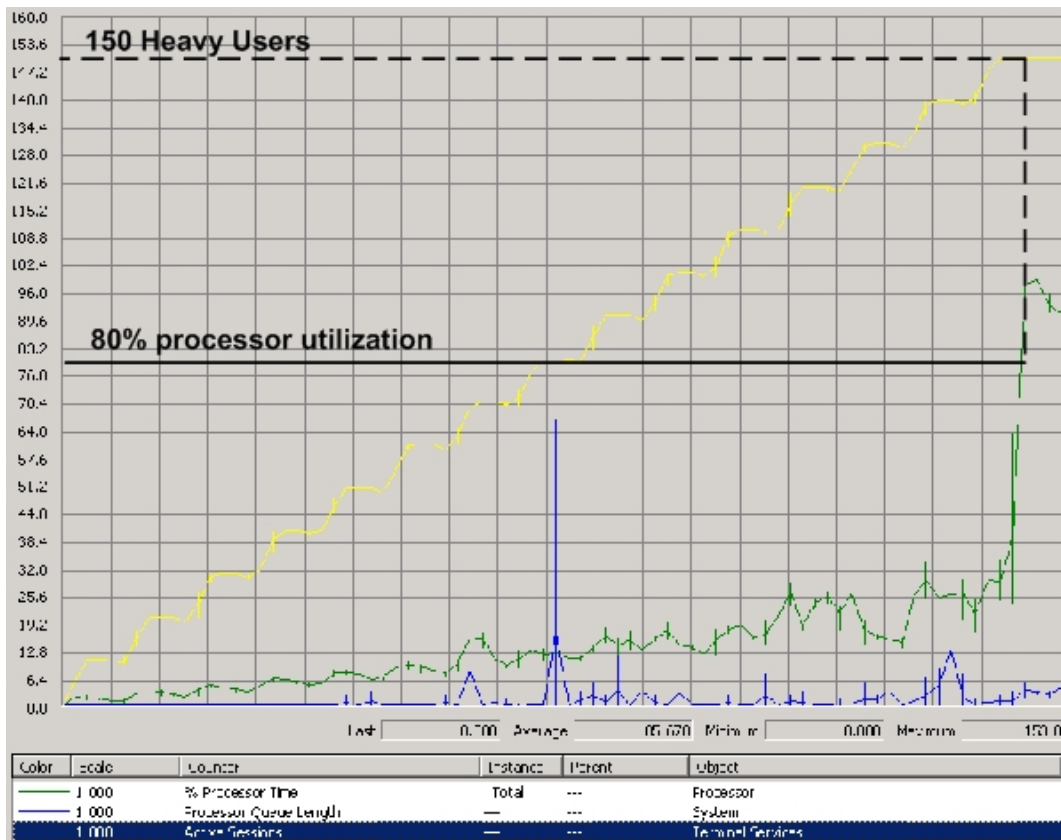

Figure 2: Individual performance metrics for a two-way server blade with Heavy Users and hyper-threading on

Figure 2 shows that 80% processor utilization first occurred with more than 150 Heavy Users logged on.

In addition, Figure 2 shows that, with 150 Heavy Users, the processor queue length was 8, 4 per processor. Note that HP is placing less importance on processor queue length as a server performance metric.
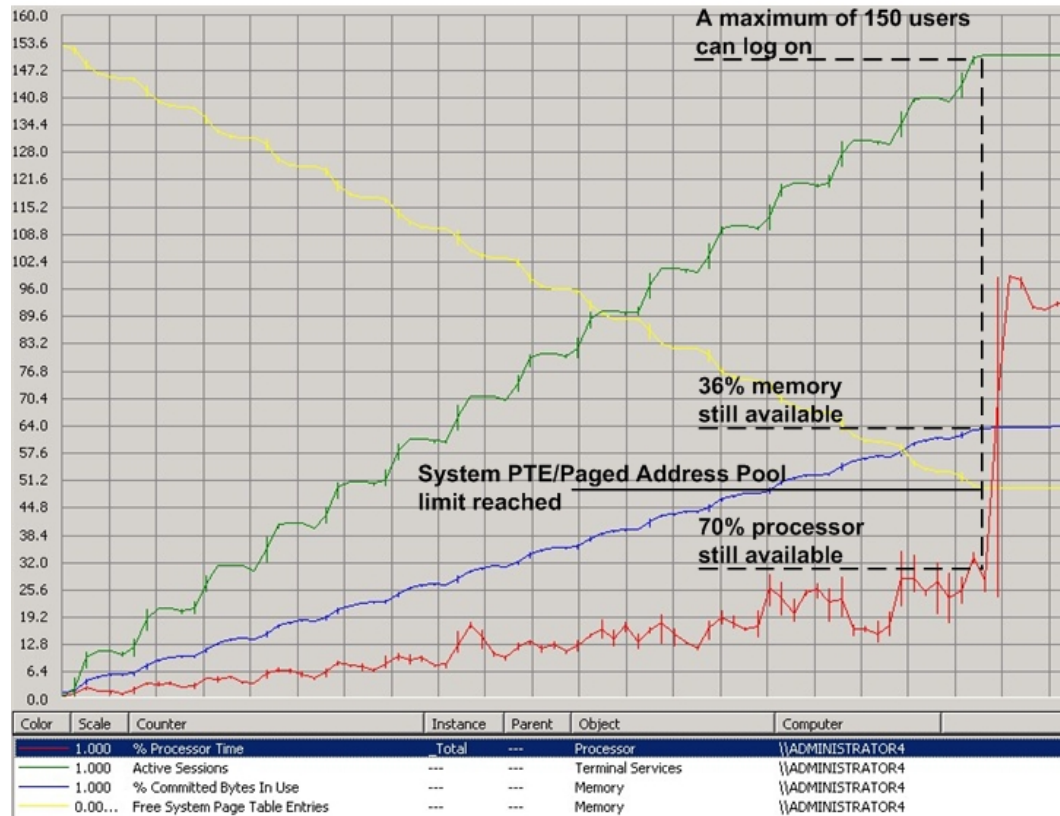
System PTE/Paged Address Pool limitation

The 32-bit Windows 2000 operating system can directly address up to 4GB of address space. By default, 2GB of this space is allocated to processes, 2GB to the kernel. The kernel area includes the following areas:

- Paged Address Pool – memory allocations for users

- System Paged Table Entry (PTE) area – kernel stack allocations

If available memory space becomes low in one of the areas, the system may exhibit unexpected errors or the inability to accept new logins, effectively limiting system performance even though there are no other resource limitations.

Figure 3 shows performance test results for Medium Users on a two-way ProLiant BL20p G2 server blade, demonstrating a performance limitation due to System PTE or Paged Address Pool space becoming low.

Figure 3: Test results showing the impact of low System PTE or Paged Address Pool resources

Figure 3 shows that, in this test scenario, the server blade was able to support a maximum of 150 users; after this time, no additional users were able to log in.

Note: The number of users supported when the System PTE/Paged Address Pool limitation occurs is system-specific – the ProLiant DL380 G3 2.8GHz and ProLiant DL580 G2 1.6GHz servers are limited to 175 Medium Users. In addition, HP has determined that the limitation varies based on user-type.

This performance limitation was not caused by processor or physical memory bottlenecks. Figure 3 shows that the System PTE/Paged Address Pool space was almost consumed (as demonstrated by the Free System Page Table Entries value leveling out at 50KB), which imposed a limit on server performance. At this time, approximately 70% of processor power and 36% of the 4GB memory array were still available for use.

HP was able to determine that performance was limited by a lack of Paged Address Pool space (rather than System PTE space); however, this conclusion could not be supported explicitly using Performance Monitor.

The current version of the operating system, by default, sizes System PTE to be as large as possible[2], maximizing the space available to the kernel. While it is possible to reconfigure System PTE space to increase or decrease the space available to the Paged Address Pool, HP does not recommend this. Reducing the space available to the kernel may make the system unstable, outweighing the benefits of additional user support. More information on tuning System PTE and Paged Address Pool resources is available on the Microsoft website.

results

HP tested one-way and two-way ProLiant BL20p G2 server blades with hyper-threading on.

one-way server

HP determined the optimal numbers of users (as shown in Figure 4) supported by a one-way server blade with hyper-threading on.
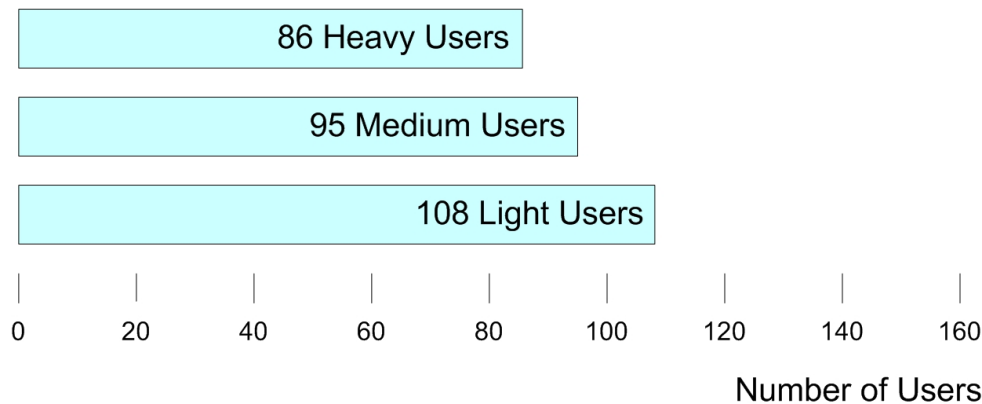
Figure 4: Test results for a one-way server with hyper-threading ON

As shown in Figure 4, a one-way ProLiant BL20p G2 server blade with hyper-threading on can support up to 108 Light Users, 95 Medium Users, or 86 Heavy Users before the users begin to experience unacceptable response times.

---

[2] On a server with Terminal Services enabled

two-way server blade

HP determined the optimal numbers of users (as shown in Figure 5) supported by a two-way server blade with hyper-threading on.

150 Heavy Users	
150 Medium Users*	*Maximum number of users supported due to the System PTE/Paged Address Pool limitation
150 Light Users*	

```
|     |     |     |     |     |     |     |     |
0    20    40    60    80   100   120   140   160
```
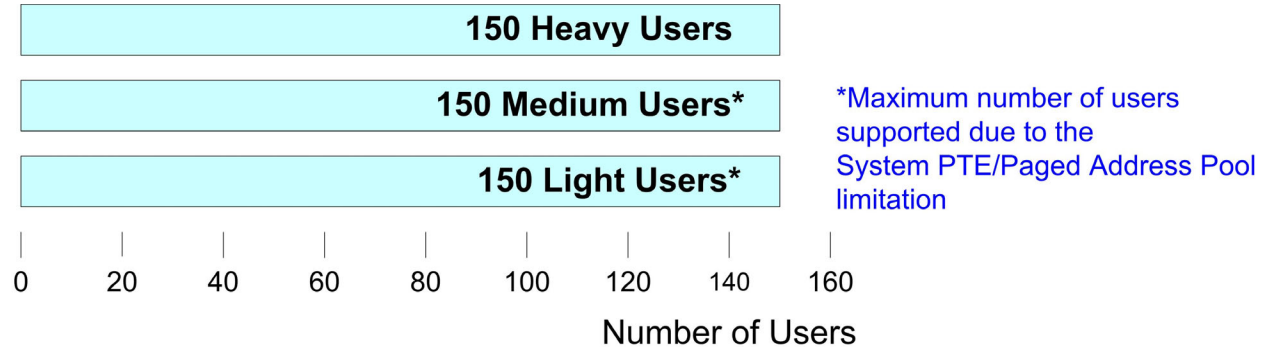
Number of Users

Figure 5: Test results for a two-way server with hyper-threading ON

As shown in Figure 5, a two-way ProLiant BL20p G2 server blade with hyper-threading ON can support up to 150 Heavy Users before the users begin to experience unacceptable response times. Alternatively, the server blade can support 150 Medium Users or 150 Light Users before lack of System PTE/Paged Address Pool space limits performance.

performance summary

Table 7 summarizes performance test results with hyper-threading ON.

IMPORTANT: In a different TCSC environment, the performance of a ProLiant BL20p G2 server blade may not match the results summarized in Table 6. To identify server configurations for alternate applications, please use the <u>online sizer tool</u>.

Table 7. Performance Summary with Hyper-threading ON

Processor			Memory	Optimal Number of Users		
Type	Number	Speed		Heavy	Medium	Light
Xeon	1	2.8GHz	4GB	86	95	108
Xeon	2	2.8GHz	4GB	150	150*	150*

* This is the maximum number of users that can be accommodated before lack of System PTE/Paged Address Pool space limits system performance. When this limitation occurs, processor utilization is still below the 80% threshold.

performance comparison

Figure 6 compares the optimal numbers of users supported by one-way and two-way ProLiant BL20p G2 server blades with hyper-threading on.
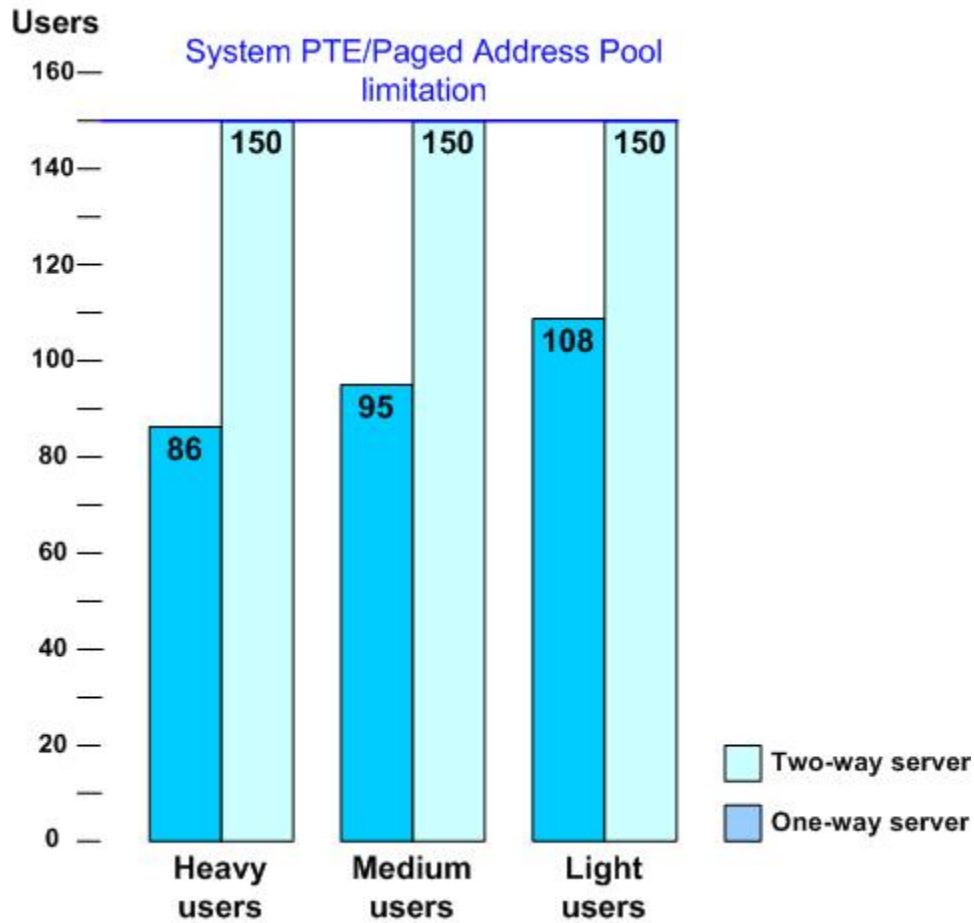

Figure 6: Performance comparison

canary execution time

HP often uses canary execution time as a performance metric. With no load on the system, the canary script is executed and the time taken to execute the canary is recorded to establish a baseline. As each group of users is added and a steady state achieved, the canary is executed again. Canary scripts are executed until execution time has increased by more than 10% over the baseline.

Table 8 lists canary execution times for a two-way server blade with Heavy Users.

Table 8. Canary Execution Times

Elapsed Time	Number of Heavy Users	Execution Time
0:00	1 – 10	10:16
0:19	11– 20	10:09
0:39	21– 30	10:10
0:58	31– 40	10:12
1:18	41– 50	10:17
1:37	51– 60	10:11
1:57	61– 70	10:08
2:16	71– 80	10:15
2:36	81– 90	10:08
2:55	91– 100	10:11
3:15	101– 110	10:08
3:34	111– 120	10:24
3:54	121– 130	10:15
4:13	131– 140	10:09
4:33	141– 150	10:21
4:52	151– 160	136:15:00

Table 8 indicates that the two-way server blade can support 141 – 150 Heavy Users before canary execution time exceeds the baseline by 10%[3]. This validates the figure of 150 Heavy Users obtained using 80% processor utilization as the performance threshold.

---

[3] At 11:18

sizing recommendations

As with any laboratory benchmark, the performance metrics quoted in this Performance Brief are idealized. In a production environment, these metrics may be impacted by the following factors:

- **overhead**
  Agents and services (virus scanning, backup and restore, provisioning, security, management and more) automatically consume overhead. Rogue applications can consume additional overhead.
  The system architect may wish to provide a 25% – 30% buffer to accommodate this overhead.

- **future growth**
  To accommodate future growth, the system architect may wish to provide an additional buffer. Alternatively, servers can be added as needed, taking advantage of the MetaFrame XP server farm's inherent ability to scale out.

- **user profiles**
  For accurate sizing, the system architect must closely match the profiles of the users in the production environment with those described in Table 6. If the profiles do not match, more are available from the [TCSC online sizer](); alternatively, the system architect can consult [HP Services]() for more information.

System PTE/Paged Address Pool considerations

There is a new sizing consideration with ProLiant BL20p G2 server blades, which are susceptible to performance limitations caused by lack of System PTE or Paged Address Pool space.

The customer should size the server blade to accommodate the desired workload while ensuring that there will always be adequate System PTE or Paged Address Pool space available. HP recommends allowing a 25% buffer between the desired number of users and the number of users supported when the performance limitation occurs. This buffer should ensure that there is always sufficient System PTE or Paged Address Pool space available to accommodate the workload.

In addition, HP recommends considering the impact of workload on available System PTE or Paged Address Pool space:

- **heavier workload**
  Performance on a server blade with a heavy workload tends to be limited by processor power rather than lack of System PTE or Paged Address Pool space – particularly since customers tend to deploy additional memory for heavier users (or reduce the number of users)

- **light-to-medium workload**
  Performance on a server blade with a light-to-medium workload is more likely to be limited by a lack of System PTE or Paged Address Pool space since operations are less processor-intensive and there tends to be more users.

summary

To quantify server performance, HP selected a metric that relates directly to the user experience – the number of Heavy Users that can be supported before response times start to become unacceptable. Test results show that a two-way ProLiant BL20p G2 2.8GHz server blade can support 150 Heavy Users, an improvement of 56% over the first generation ProLiant BL20p server blade.

Overall, test results show that the ProLiant BL20p G2 2.8GHz server blade is a performance leader.

appendix a – Windows 2000 Terminal Services

Terminal Services is an integral part of Windows 2000, allowing desktops to access Windows-based applications running on the server. Windows 2000 Terminal Services expands the capabilities available with Windows NT, using Remote Display Protocol (RDP) to connect multiple, diverse clients to the server.

Terminal Services is an optional feature that can be enabled on any Windows 2000 server (standard, Advanced or Datacenter).

Terminal Services provides the following benefits:

- **Lower cost of ownership and improved manageability**:
  Deploying applications on the server rather than on the client reduces administrative costs by 35% to 50% over the life of the system because management of back-ups, updates, troubleshooting, and many help-desk requests can be centralized on the server instead of being performed on individual clients. In addition, task-based users can utilize thin clients with no performance degradation since the server handles storage, memory, and processor demands.

- **Strong limited-bandwidth application performance**:
  Mouse clicks, keystrokes, or screen updates consume minimal bandwidth. Terminal Services reduces an application's network usage by as much as 90% and allows users to enjoy excellent performance even with limited bandwidth or during periods of heavy network traffic.

- **Windows applications compatibility**:
  Terminal Services can run 16-bit or 32-bit Windows-based applications that operate on a Windows 2000 system. Typically, clients require no further modifications to use the software, regardless of the hardware configuration.

- **Deployment in new environments**:
  Terminal Services allows the deployment of Windows 2000 applications in new environments. For example, task-based users such as bank tellers, insurance claim adjusters or fast food operators can use a stationary client with a single business application in an office setting. At the same time, mobile task-based users can use Terminal Services to take restaurant orders, register car rental returns or monitor bedside terminals in hospitals.

- **Network load balancing**:
  Terminal Services allows the deployment of applications across server farms, providing higher availability and easy expansion to create a NonStop solution for applications' availability.

more information

- A [summary](#) of the features of Terminal Services

- [More](#) about Terminal Services, including system architecture, client licensing, remote administration, and optimizing applications for use with Terminal Services

- [Links](#) to additional technical information on Windows 2000

appendix b – Citrix MetaFrame XP

Citrix MetaFrame XP for Microsoft Windows offers a complete solution for organizations that want more from their application serving environments. IT administrators need more reach to connect with users in any location, on any device, over any network, and more power to manage applications from a single, centralized location. Based on proven Citrix technology, Citrix MetaFrame XP extends the reach of the Windows 2000 Server family as well as advanced platforms. It securely deploys Windows-based applications with the highest levels of manageability so administrators can easily scale servers as organizations grow. Citrix MetaFrame XP offers a more predictable, cost-effective computing experience for all users.

Single-point control over applications, servers and server farms, licenses and resources means better utilization of limited IT departments.

MetaFrame XP family

Citrix specifically customized the MetaFrame XP family into three tailored solutions that accommodate a wide spectrum of customer requirements for functionality and scale:

- **MetaFrame XPs**:
  Delivers the level of control that departmental workgroups require to extend the reach of Windows 2000 servers to multiple devices

- **MetaFrame XPa**:
  Maximizes the availability of applications across the Internet for growing organizations

- **MetaFrame XPe**:
  Offers extensive scalability and manageability, as well as rapid application deployment, for global enterprises in a Windows 2000 server-based environment.

The MetaFrame XP family features a highly scalable communications and management foundation that provides the capability to manage and deploy applications across the enterprise. MetaFrame XP delivers much more than application support. It gives IT managers the ability and confidence to maintain a fast, predictable, and efficient application-serving network.

key benefits

- **Unparalleled manageability and scale – systems, applications and users**
  MetaFrame XP offers the power to manage servers and server farms anywhere across an enterprise, with robust system management capabilities that can be integrated with third-party network management solutions. At last IT managers can centrally configure and manage software distribution to multiple servers in numerous locations in just minutes, as well as monitor and control application usage, resource utilization and user activity. With these capabilities, IT managers can proactively predict system requirements and maximize availability and performance across the network.

- **Total Net leverage – integration, personalization and control**
  The adoption of the Internet is the driving force behind the concept of a universal Net – the seamless integration of LANs, WANs, the Internet, intranets and extranets – that enables organizations to optimize their communications. Citrix NFuse empowers businesses with the ability to deliver any application to a single portal view without rewriting code. This is the fastest way to provide single-point access to applications and content information through any device that supports a standard Web browser – improving productivity and delivering true user mobility.

- **Ultimate flexibility – any application, any device, over any Net connection**
  MetaFrame XP provides the power to deploy any application to any device via any Net connection – quickly and with a reduced total cost of ownership. By shifting application processing to the server, administrators can ensure the rapid, user-friendly deployment of business-critical applications with a higher level of performance predictability. The flexibility of seamless application access from PCs, Macintosh computers, UNIX or Linux workstations, laptops, wireless devices and other network appliances means fewer hardware roadblocks and far less technology churn. It affords a consistent user experience across the enterprise, complete freedom and mobility, unparalleled speed, and simplified management. MetaFrame XP is the powerful, behind-the-scenes framework that makes it all happen.

more information

For more information on MetaFrame XP, please refer to the [Citrix website](#).