



# **SAS Enterprise Miner Performance on IBM System p 570**

*Jan, 2008*

**Hsian-Fen Tsao  
Brian Porter  
Harry Seifert**

**IBM Corporation**

# TABLE OF CONTENTS

<b>ABSTRACT .....</b>	<b>3</b>
<b>EXECUTIVE SUMMARY .....</b>	<b>3</b>
<b>OVERVIEW OF IBM SYSTEM P 570 .....</b>	<b>4</b>
<b>SAS EM TESTING-ENVIRONMENT DESCRIPTION .....</b>	<b>5</b>
SAS EM Test Suite Description .....	5
Testing Configuration .....	6
<b>OVERVIEW OF RESULTS .....</b>	<b>6</b>
<b>CONCLUSIONS .....</b>	<b>9</b>
<b>REFERENCES .....</b>	<b>9</b>

# ABSTRACT

*The IBM System p™ 570 server with IBM AIX® Version 5.3 operating system demonstrated impressive performance, scalability and reliability when running the SAS Enterprise Miner (EM) 5.2 application in a test environment performed by IBM and SAS in November 2007.*

## EXECUTIVE SUMMARY

Turning increasing amounts of raw data into useful information have become increasingly important in today's highly competitive business environment, increasing the demand for predictive, analytics data mining solutions. SAS Enterprise Miner (EM), a powerful and comprehensive data mining software, includes services and training to help organizations get started right away exploring large quantities of data to discover relationships and patterns that lead to proactive decision making. IBM's unmatched expertise in hardware and software technology, along with services, enables the SAS Enterprise Miner solution for AIX 5L™ on IBM POWER™ processor-based systems offers significant benefits. It can be deployed on an infrastructure that is designed to improve reliability, performance and scalability across diverse data mining projects commonly used in real-life production environments.

An IBM p 570 server configured with 4.7 GHz POWER6™ processors achieved outstanding results for the SAS Enterprise Miner application:

- 500k observations and 500 variables were processed within 2-10 minutes per modeling node with nodes "decision tree", "neural network", and "regression".
- 1-Million observations and 200 variables were processed within 5-25 minutes per modeling node with nodes "decision tree", "neural network", and "regression".

These results were achieved in configurations ranging from 4-core to 8-core by using both small data (430 MB) and medium data size (1.6GB). These tests execute Enterprise Miner modeling flow diagrams (Figure 1 at page 5) which include "decision tree" modeling, predictive "neural network" modeling or traditional "regression" modeling. The results show how organizations can deploy the IBM System p 570 server to deliver good performance and scaling capabilities across diverse configurations to meet the demands of small and medium enterprise-mining requirements. With these powerful, yet affordable entry servers, companies are now able to perform business mining functions more frequently and on more mining projects concurrently while enjoying fast turn-around times. Table 1 is a brief summary of the recommended starter configurations that cover a range of user requirements.

**Note:** PO stands for performance-optimized and CO for cost-optimized.

Workload with AIX 5L TL6	SMALL				MEDIUM				
	PO	CO	PO	CO	PO	CO	PO	CO	
No. of Observations	500k	500k	500k	500k	1 Million	1 Million	1 Million	1 Million	
No. of Variables	100	100	200	200	200	200	200	200	
No. of parallel user sessions	2	4	4	8	2	4	4	8	
Max No. concurrent nodes	4	4	8	8	4	4	8	8	
IBM POWER6 Model	P570	P570	P570	P570	P570	P570	P570	P570	
No. of cores (4.7 GHz)	4	4	8	8	4	4	8	8	
RAM (GB)	16	16	32	32	16	16	32	32	
No. of Internal Drive (15KB 73.4GB)	4	4	4	4	4	4	4	4	
No. of 2 Gbps FC Adapter	4	4	4	4	4	4	4	4	
No. of IBM DS4700 FC HDD	16 x 73 GB 15KB	16 x 73 GB 15KB	16 x 73 GB 15KB	16 x 73 GB 15KB	24 x 73 GB 15KB	24 x 73 GB 15KB	24 x 73 GB 15KB	24 x 73 GB 15KB	
Node timing (in seconds) per modeling node	Neural	570	571	571	600	739	762	744	856
	Regression	98	101	101	109	317	351	326	408
	Tree	332	336	336	370	1,520	1,660	1,530	2,065

**Table 1. Recommended hardware/reference architecture deployment for SAS EM on IBM POWER6**

This document contains conclusions and technical insights produced from a comprehensive benchmarking and sizing project using small and medium data volumes. It is also intended to be used as a sample reference architecture guide for sizing and capacity planning of SAS EM deployments on IBM POWER6 servers.

# OVERVIEW OF IBM SYSTEM P 570

The IBM System p 570 is the first system designed with POWER6 processors, resulting in performance and price/performance advantages that potentially benefit data- and compute-intensive software (e.g. SAS Enterprise Miner).

IBM System p 570 is powered by the 64-bit, 4.7 GHz, IBM POWER6™ processors. The POWER6 processor is the next generation of the IBM POWER5™ family of microprocessors. The POWER6 processor-based System p 570 incorporates performance, power efficiency, reliability and virtualization feature improvements as compared to its POWER5 predecessor (Table 2).

Summary of POWER6 enhancements	Its impact
Enhanced system architecture, for example: <ul style="list-style-type: none"> <li>• Smaller processor (CMOS 65 nm, dual-stress line) with faster clock speed (3.5-4.7 GHz) and advanced power management</li> <li>• Highly efficient simultaneous multithreading (SMT) technology</li> <li>• Enhanced caches and translation resources (for example, larger L1 and D-ERAT)</li> <li>• Predictive subspace snooping for vastly increased SMP coherence throughput</li> <li>• High-speed I/O-hub interface; multiplexed addressing and data; increased I/O frequency from 33 to 50 percent of the processor frequency</li> <li>• Advanced RAS features</li> <li>• All buses scale with processor frequency</li> </ul>	<ul style="list-style-type: none"> <li>• Increases performance and reduces costs, consumes less power with no performance impact</li> <li>• Increases scalability and throughput</li> <li>• Helps reduce cache misses</li> <li>• Helps reduce memory-access latency and improves efficiency of branch handling</li> <li>• Improves I/O throughput</li> <li>• Provides robust error detection, hardware checkpoints and recovery</li> <li>• Increases overall scalability, stability, and optimal throughput</li> </ul>
Improved high performance computing (HPC) performance; HPC capability depends on two critical performance parameters: <ul style="list-style-type: none"> <li>• Sustained bandwidth               <ul style="list-style-type: none"> <li>○ POWER6 memory bandwidth is about 250 percent greater than POWER5 bandwidth</li> <li>○ Five SMP interfaces on-chip for excellent SMP scalability                   <ul style="list-style-type: none"> <li>▪ Total SMP interface bandwidth is about 150 percent more than POWER5 bandwidth</li> </ul> </li> <li>○ I/O bandwidth is about 200 percent more than POWER5 bandwidth</li> <li>○ Advanced prefetch mechanism                   <ul style="list-style-type: none"> <li>▪ 200 percent more concurrent streams than on POWER5 systems</li> <li>▪ Store prefetching added in POWER6 processor-based systems</li> </ul> </li> <li>○ Load and store pair instructions for high bandwidth between L2 cache and microprocessor core</li> </ul> </li> <li>• Peak performance               <ul style="list-style-type: none"> <li>○ Matrix multiply is about 200 percent more than on POWER5 systems</li> </ul> </li> </ul>	Help overall performance, throughput and scalability of scientific or commercial workloads (including compute-intensive and memory-intensive HPC applications);  Many business and analytics intelligence applications are data-intensive and also carry the same characteristics of HPC workloads; thus, business-and analytics-intelligence workload should benefit from these features
Improved on-storage hierarchy, for example: <ul style="list-style-type: none"> <li>• Larger L2 cache; more than 400 percent greater than POWER5 systems</li> <li>• Greater associativity</li> </ul>	<ul style="list-style-type: none"> <li>• Helps reduce cache misses</li> <li>• Helps multithreaded execution and gets greater speedup for SMT in POWER6 systems</li> </ul>
Enhanced memory subsystem, for example: <ul style="list-style-type: none"> <li>• Two memory controllers, larger L3 cache, bus with higher frequency and bandwidth</li> </ul>	<ul style="list-style-type: none"> <li>• Helps reduce memory access latency</li> </ul>
Provides a series of innovative hardware features: AltiVec™ extensions, Hardware Decimal Floating-point, multiple page sizes, Live Partition Mobility.	Improve performance, increase system resource utilization, and lower total costs

**Table 2. IBM POWER6 systems improvements highlights**

A common computer design issue is the gap between a microprocessor core that stalls because of memory access. This gap is wide, and is getting wider by the year. The IBM p570 has been designed with scaling of processor performance and system design (cache sizes and bandwidth) in a balanced way. The POWER6 processor has a total cache size of 8 MB per processor (about 400 percent more than the POWER5 processor) to keep pace with the high processor bandwidth and enables operating at a higher data rate. This benefits

algorithm- and data-intensive business intelligence and analytics applications like SAS Enterprise Miner. As the calculations increase per data point, the processor-memory performance gap becomes one of the more challenging bottlenecks in achieving high performance. The balanced p570 system design has lower and more consistent latencies by using improved on-chip storage hierarchy, coupled with sustained and higher I/O bandwidth. Although low latency is essential for performance, consistent latency is substantially easier for operating systems to manage. Compared to IBM POWER5 system, IBM system p 570 showed a 59 percent performance improvement in a test using predictive modeling techniques and procedures commonly used in Enterprise Miner (EM) software.

The IBM System p 570 servers with advanced microprocessor technology, enhanced system design and operating system technologies are designed to be fast and secure business intelligence and analytics data mining servers.

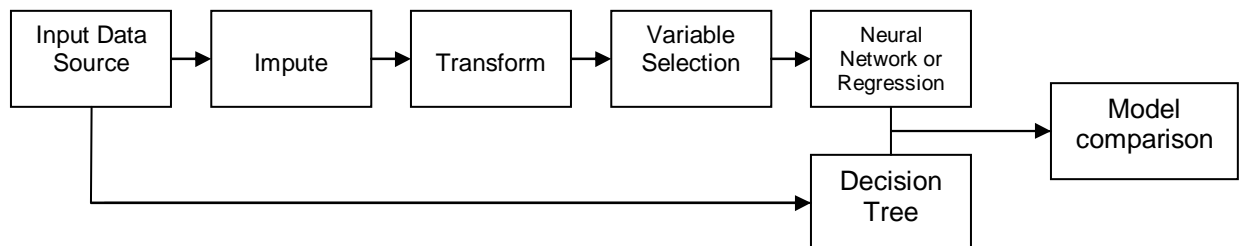
## SAS EM TESTING-ENVIRONMENT DESCRIPTION

This section describes the SAS Enterprise Miner testing environment that was used to gather the information presented in this white paper.

### SAS EM Test Suite Description

The SAS Enterprise miner test suite is designed to provide SAS clients with an understanding of the performance they can expect and system resources required to support their data mining activity. This test suite was defined based on the knowledge of common usage patterns for data mining and can be a useful tool to help clients size the deployment of SAS Enterprise Miner with POWER6 processor-based System p servers. It is well understood that individual client-resource requirements can vary and that tests specific to your environment might be necessary to optimize the configuration.

This test suite simulates users simultaneously submitting their modeling flow (Figure 1) via the SAS Enterprise Miner java-based graphical user interface. These tests use real-world data volumes and structures. Input tables for data mining projects typically contain many rows (observations) and many columns of data (explanatory variables). It is increasingly common for data miners to use several hundred independent variables to explain the behavior of a single dependent variable. For this paper, the test scenarios reflect this trend by using input data tables with 100-200 independent variables, and up to 1 million observations. In addition to other nodes depicted in Figure 1, each modeling flow executes "Decision Tree", "Neural Network" or "Regression" nodes to find the model which best fits their data to predict an event. The value of the models are evaluated in the Model Comparison node, which follows the modeling nodes in the EM project flow diagram.



**Figure 1. SAS Enterprise Miner modeling flow diagram**

## Testing Configuration

Small and medium configurations, using SAS Enterprise Miner were tested using a real-life environment. It consisted of:

- One System p 570 server with 32 GB of RAM. The server under test was configured with four to eight 4.7 GHz processor cores with 16 GB and 32 GB of RAM, respectively.
- 24 data-storage drives configured in an IBM System Storage DS4500 storage subsystem with two 2 Gbps fiber-channel adapters
- Eight 73.4 GB 15K SCSI internal drives
- All database files were created on the JFS2 file system and used a striped logical device with RAID5 for high availability and protection requirements
- The software stack contained 64-bit AIX V5.3 TL6, 32-bit/64-bit IBM JRE (v1.4.1 & v1.4.2), SAS Base Server 9.1.3 SP4, SAS Metadata Server, SAS Stored Process Server, SAS Object Spawner, SAS Analytics Platform, SAS Enterprise Miner 5.2 and jython 2.1.
- Based on the number of the concurrent EM user sessions running on 4-core and 8-core p570, a series of cost- and performance-optimized configurations were tested for SAS EM small and medium data volumes (See Table 1). While the performance-optimized configuration focused on improving performance, the cost-optimized configuration offered an affordable alternative with good performance.

## OVERVIEW OF RESULTS

The results of the performance tests provide clients with an understanding of the synergetic impact of the SAS software and IBM hardware platform combination from a system-wide perspective. It can also correlate with the anticipated usage profiles of possible deployments.

### Performance Using Small Data Volumes

Figure 2 shows the average time it took each Enterprise Miner node to run, using different number of users and CPUs with “neural network” and “regression” modeling techniques. Node timings remain consistent across these diverse configurations as shown, provided there is at least one processor available for each Enterprise Miner node that is executing.

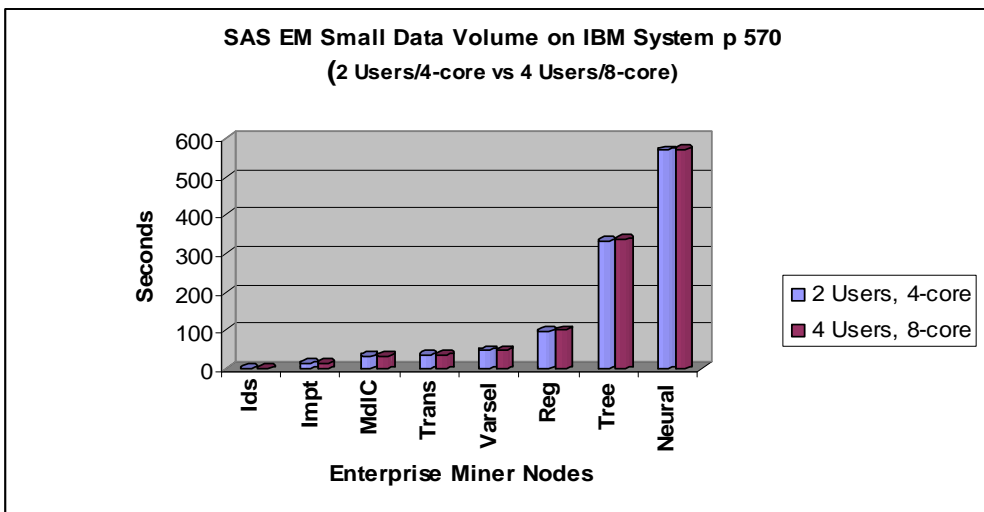


Figure 2. Small data volumes on p570 - Node timing with different number of user sessions and CPUs

SAS Enterprise Miner is designed to efficiently utilize hardware resources. The p570 server enables high reliability and stability in scale-up environments. Figure 3 demonstrates that the SAS Enterprise Miner small configuration can successfully scale upward to support a heavy, yet realistic, user workload on p570 server with diverse configurations (from 2 Users, 4-core to 4 Users, 4-core and from 4 Users, 8-core to 8 Users, 8-core) and still delivers almost linear scaling of performance.

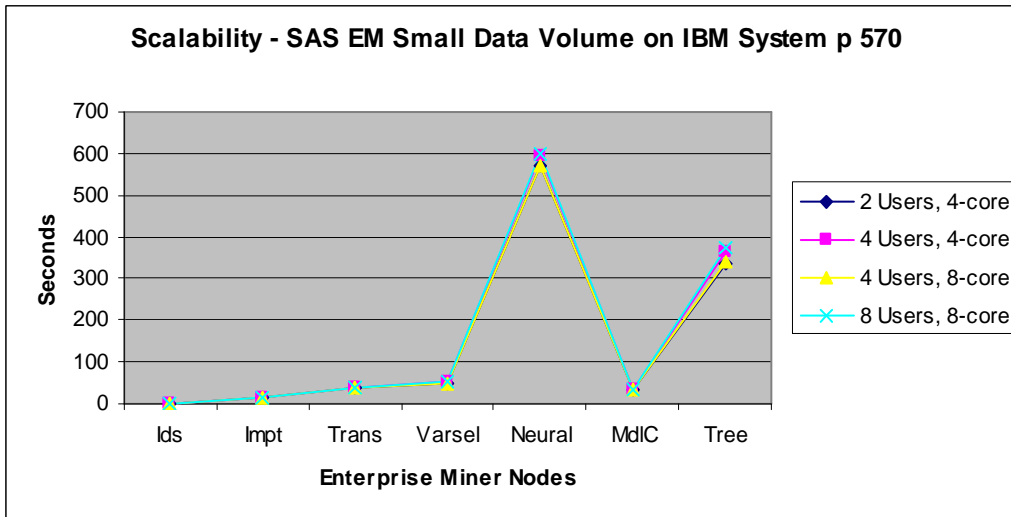


Figure 3. SAS EM small data-size scalability across diverse configurations

### Performance Using Medium Data Volumes

Similar to the small data volumes conclusion, the test results with medium data volume also shows that the SAS Enterprise Miner Server scaled linearly from the 4-core to 8-core configurations on System p 570, as displayed in Figure 4.

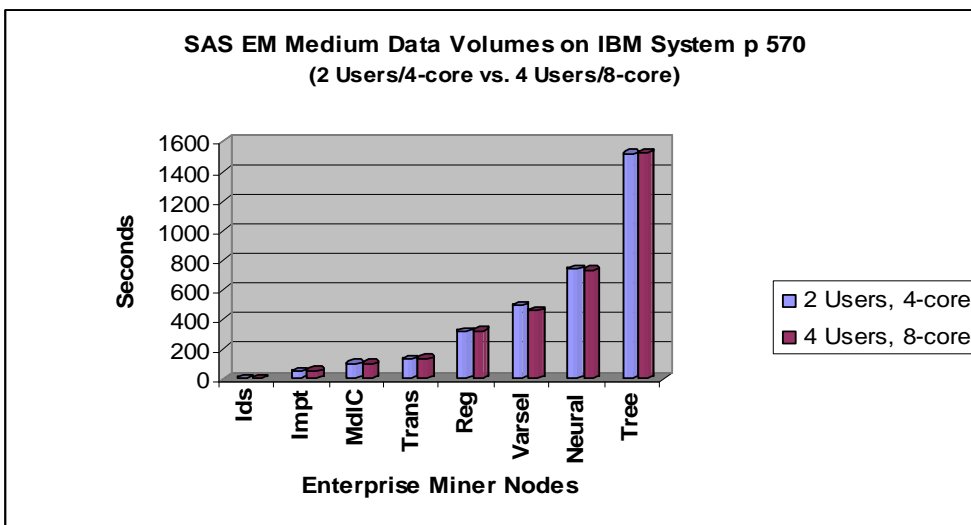


Figure 4. Medium data volumes on p570 – Node timing with different number of user sessions and CPUs

Although the 8-core IBM p570 server is usually considered an entry system, it is fully capable of handling larger business-mining workloads, as shown in the medium data volume test. This provides clients an affordable solution for performing mining functions. By supporting multiple users and data mining projects, p570 reduces the number of servers that are needed and delivers good performance to increase productivity.

As displayed in Figure 5, using the same number of processors (8 cores), P570 server demonstrates its capability of driving twice the workload, which is targeted for a 16-core server, and achieves comparable performance to the 4-user, 8-core configuration.

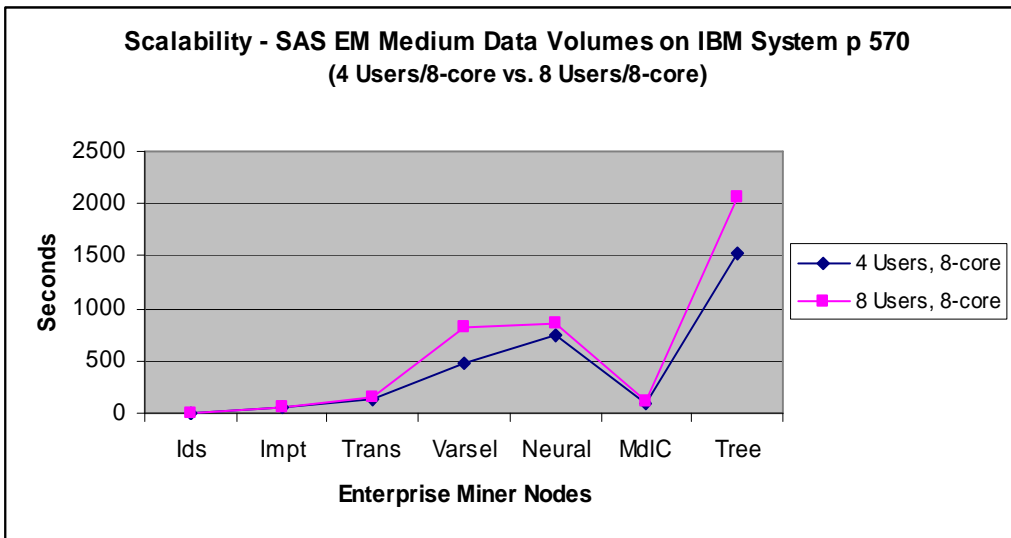


Figure 5. SAS EM medium data-size scalability across diverse combination

These results demonstrate the great price/performance advantage and sustained throughput performance of the System p 570 platform, which is critical in today's enterprise environment.

### Enhanced SMT Performance

The new POWER6 architecture includes SMT technology enhancements that leverage and use hardware resources more effectively and efficiently than previous POWER5 SMT technology. This has been proven with a variety of SAS workloads. Use predictive and analytic SAS Forecast Server as an example, POWER6 SMT enhancements improved the performance of SAS Forecast Server by 10-77 percent across 2- to 8-user configurations compared to single thread (ST) performance using the same configuration. In comparison, SMT results on POWER5+ processor-based systems showed 0-43 percent improvements, respectively, when compared to ST performance results on the same systems.

POWER6 SMT enhancements improved the performance of SAS Enterprise Miner by 2-54 percent across 2- to 8-user configurations (compared to ST performance using the same configuration). These results validate the advantages of SMT, which primarily benefits application throughput; thus, greater SMT performance is expected by running more threads/jobs of data- and compute-intensive tasks such as SAS Enterprise Miner.

Enabling SMT is a permanent requirement for performance improvements in the SAS EM environment.

## CONCLUSIONS

The results of this SAS Enterprise Miner performance test suite successfully demonstrate the performance, reliability and scalability on the IBM System p 570 server. IBM breakthrough hardware and operating system technology, together with the robust scalability of SAS software, delivers an ultimate solution to address analytical mining demands for our clients.

## REFERENCES

1. For more information about SAS  
[www.sas.com](http://www.sas.com)
2. For more information about IBM System p servers  
[ibm.com/systems/p](http://ibm.com/systems/p)
3. For more information about AIX  
[ibm.com/servers/aix](http://ibm.com/servers/aix)
4. For more information about IBM storage  
<http://www.ibm.com/servers/storage>



© IBM Corporation 2008  
IBM Corporation  
Systems and Technology Group  
Route 100  
Somers, New York 10589

Produced in the United States of America  
December 2007  
All Rights Reserved

This document was developed for products and/or services offered in the United States. IBM may not offer the products, features, or services discussed in this document in other countries.

The information may be subject to change without notice. Consult your local IBM business contact for information on the products, features and services available in your area.

All statements regarding IBM future directions and intent are subject to change or withdrawal without notice and represent goals and objectives only.

IBM, the IBM logo, AIX, AIX 5L, Micro-Partitioning, POWER, POWER4, POWER5, POWER5+, System p, System p5, System Storage, are trademarks or registered trademarks of International Business Machines Corporation in the United States or other countries or both.

Java and all Java-based trademarks and logos are trademarks of Sun Microsystems, Inc. In the United States and/or other countries.

UNIX is a registered trademark of The Open Group in the United States, other countries or both.

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

IBM hardware products are manufactured from new parts, or new and used parts. In some cases, the hardware product may not be new and may have been previously installed. Regardless, our warranty terms apply.

Copying or downloading the images contained in this document is expressly prohibited without the written consent of IBM.

Information concerning non-IBM products was obtained from the suppliers of these products or other public sources. Questions on the capabilities of the non-IBM products should be addressed with those suppliers.

All performance information was determined in a controlled environment. Actual results may vary. Performance information is provided "AS IS" and no warranties or guarantees are expressed or implied by IBM. Buyers should consult other sources of information, including system benchmarks, to evaluate the performance of a system they are considering buying.

When referring to storage capacity, 1TB equals total GB divided by 1000; accessible capacity may be less.

Many of the features described in this document are operating system dependent and may not be available on Linux. For more information, please check: [ibm.com/systems/linux/power/whitepapers/linux\\_overview.pdf](http://ibm.com/systems/linux/power/whitepapers/linux_overview.pdf).

The IBM home page on the Internet can be found at: [ibm.com](http://ibm.com).

The IBM System p home page on the Internet can be found at: [ibm.com/systems/p](http://ibm.com/systems/p).

[PSW03028-USEN-00](#)

