

Intel® Architecture for Flexible, Reliable SAS* Computing: Sizing and Performance Considerations for the IBM® @server™ xSeries™ 450

WHITE PAPER

The Intel® Itanium™ 2 processor is superbly suited to the demands of SAS* computing. The IBM® @server™ xSeries™ 450 puts the Itanium 2 processor to work in a highly reliable, flexible and scalable platform that delivers exceptional performance and value for I/O- and memory-intensive SAS solutions. This paper discusses these two powerhouse technologies and provides SAS sizing and performance test results that can help you apply the Itanium 2 processor and the xSeries 450 server to your SAS computing needs.

intel®

IBM®

Contents

When Business Intelligence is Mission-Critical	2
Power Where It Matters for SAS Computing	2
Determining Your Performance Requirements	3
Sizing and Performance Considerations	4
A Powerful Architecture for SAS Computing	5
Summary	6
Learn More	6

When Business Intelligence is Mission-Critical

Companies around the world rely on software solutions from SAS to help them turn raw data into practical, actionable knowledge. This knowledge provides the basis for decisions that shape the direction of the business, its relationships with customers and collaborators, and the efficiency of its supply chain. Companies that extract that knowledge rapidly, efficiently and cost effectively can gain a significant competitive edge.

The Intel Itanium 2 processor and the IBM xSeries 450 help companies gain that edge, providing a powerful, reliable, flexible and cost-efficient platform for mission-critical SAS solutions.

Power Where It Matters for SAS Computing

The Intel Itanium 2 processor extends Intel volume economics to the most data-intensive, business critical and technical applications, typified by SAS solutions. Intel and SAS have collaborated to ensure outstanding performance for SAS System 9 on Itanium 2 processor based platforms.

The Itanium 2 processor is uniquely architected for outstanding transaction processing performance on demanding enterprise and technical applications. The processor offers massive execution resources, including 6.4 GB/sec system bus bandwidth, a 3 MB integrated L3 cache and 64-bit memory addressability as well as outstanding floating performance and scalability. With its outstanding memory bandwidth and memory size, it delivers exceptional performance and throughput for large SAS datasets. Windows Server 2003 users can take advantage of the operating system's larger memory buffers to improve the performance on large datasets.

Based on IBM's Enterprise X-Architecture™ technology, the xSeries 450 provides high performance and exceptional value in a reliable, flexible platform. X-Architecture technology is drawn from IBM's vast enterprise server heritage, applying technologies that have already revolutionized larger IBM systems to the Intel® processor-based platform. The results: industry-standard servers designed to provide enterprise-inspired power, scalability, control and service at very attractive prices.

Available with up to four Itanium 2 processors per server, the xSeries 450 offers industry-leading density, with a compact 4U package that saves data center floor space. The IBM xSeries 450 features the IBM OnForever™ technology to help protect valuable data and ensure high system availability. The server provides out-of-the-box flexibility via IBM XpandOnDemand™ scalability, with outstanding PCI-X I/O capacity and award-winning IBM system management tools. Storage options include the IBM TotalStorage™ FAST700 Storage Server, a midrange storage solution that scales up to 32.8 terabytes, providing data protection with dual redundant components, multiple RAID levels, LUN masking and enhanced management options. The IBM TotalStorage™ SAN Switch F08, an entry-fabric switch for server clustering, LAN-free backup, storage consolidation and remote disk monitoring is also included.

Determining Your Performance Requirements

In choosing platforms for SAS computing, your capacity requirements are determined by a variety of factors. Among the most relevant ones:

- **Number of simultaneous users or jobs.** As more jobs are processed simultaneously, a more powerful system configured with more CPUs can provide optimum performance. With up to four Itanium 2 processors, the xSeries 450 is well equipped to handle multiple jobs.
- **Number of calculations per data point.** As the calculations per data point increase, the demands on the processor(s) rise. Extract-Transform-Load (ETL), Query-Reporting (Q&R) and data mining workloads in particular may put a significant load on the processors and benefit from the performance of the Itanium 2 processor.
- **Dataset size.** As dataset sizes increase from megabytes to gigabytes and terabytes, the Itanium 2 processor's virtually unlimited addressability becomes increasingly valuable.
- **Memory bandwidth requirements.** Tasks such as the statistical analysis of a large dataset can be particularly memory bandwidth-intensive and therefore likely to benefit greatly from the xSeries 450's fast memory bus and the Itanium 2 processor's memory bandwidth.

- **Amount of threading.** Highly threaded applications like General Linear Model (GLM) analysis benefit from the xSeries 450's high performance and multiple CPUs.
- **Business urgency.** Business intelligence results must be timely to be useful. The xSeries 450 offers excellent throughput for complex jobs.
- **SAS version.** The xSeries 450 is an excellent choice for SAS System 9 computing, since Intel and SAS worked together to make SAS System 9 run superbly on the Itanium 2 processor. In addition, many SAS System 9 workloads run much faster on a system with more CPUs. Some analytical operations, along with sorting and aggregation routines, are threaded in SAS System 9; if you're using these routines, you may want to use SAS System 9 on a larger system for faster job completion.

Sizing and Performance Considerations

In the Fall of 2002, Intel evaluated sizing and performance considerations for SAS solutions on Intel processor-based servers. Several system configurations were tested against a range of SAS applications, and the results provide information that may be helpful to businesses in determining how to use the xSeries 450 platform to enhance the reliability and flexibility of SAS deployments. Both SAS System 8.2 and SAS System 9 were tested on a range of platform configurations in both single-user and multi-user suites. Single user workloads included:

- Analytical tests (stepwise linear regression, general linear model and others)
- Enterprise Miner* to perform data mining tests
- Data manipulation on a large volume of input data
- Extract Transform and Load tests
- Query and reporting tests

Multi-user tests encompassed these same tests plus a mix of usage models—10 percent analytical, 60 percent query and reporting, and 30 percent data manipulation.

These results do not eliminate the need to benchmark your own solutions, but they do provide some general guidance. Because your environment and solution set are unique, it is recommended that you work with your sales team to configure a system that suits your particular needs. In addition, Intel and SAS have created the SAS Customer Knowledge Practice to help you choose and optimize your system.

The test results are organized around four key findings:

- **Scaling.** System scaling has a different effect on performance in single-user versus multi-user environments, as well as on non-threaded versus threaded applications. Using SAS System 9, a single user can take advantage of multiple CPUs to complete a complex job. This benefit is best realized when using procedures that are threaded internally. Scaling to four or more CPUs can provide significantly better performance for threaded procedures such as REG (regression), GLM (general linear regression), DMINE (data mining), LOESS, DMREG, SORT, SUMMARY/MEANS, and SQL. The relative scalability of these procedures varies, since they depend on other system resources that may not increase with the number of CPUs. The four-way xSeries 450 is well suited to handle multiple concurrent users.
- **Threading.** Threads executing simultaneously can contend for processor time, affecting system performance. Thus, in the single-user test suite, overall performance was best at systems with four to eight CPUs. Limiting the number of processors allocated to any individual job helps minimize conflicts between jobs and ensures maximum throughput for all applications and users.
- **Memory bandwidth.** Memory bandwidth can have a significant effect on performance, depending on the application and dataset. When applications must process a large volume of data in a single pass, memory bandwidth can be the chief determinant of overall performance. For example, statistical analysis using large datasets can benefit greatly from the wide-open data paths of the Itanium 2-based xSeries 450. Extremely complex processing tasks, such as data mining and statistical analysis on large datasets, benefit from the high memory bandwidth of the Intel Itanium 2 processor. As the analytical tests increased in size, performance was best on the Itanium 2 processor, which has the faster memory architecture. The Itanium 2 processor performed well for large data manipulation script, and performed particularly well and scaled linearly on large GLM.
- **Simultaneous users/jobs.** The number of simultaneous users provides a good gauge of the optimum number of processors. Testing on the Itanium 2-based systems reflected optimum price/performance at an average of 8–10 users per CPU.

A Powerful Architecture for SAS Computing

The Itanium 2 processor offers five architectural capabilities that make the xSeries 450 superbly suited to mission-critical SAS computing:

- **Memory addressability and memory bandwidth.** Large physical memory and high memory bandwidth provide the ability to access more data more quickly from system memory. Greater amounts of data can be held near the processor for faster calculations and data analysis. This capability improves application performance by allowing larger file system caches for read-ahead and write-behind I/O operations. This enables SAS applications to retain large amounts of data in memory instead of repeatedly reading the data from disk. Ultimately, this feature helps companies extract business intelligence data from warehouses much more quickly. In SAS data mining applications, for example, companies can perform analyses on very large volumes of data that might previously have been impractical.
- **Scalability.** Many enterprise customers find value in using one server for multiple users. The primary challenge in this model is providing sufficient memory and I/O throughput to deliver excellent performance to all users. The xSeries 450's balanced, high-performance architecture meets that challenge, giving system administrators the flexibility to support large and small jobs on the same system simultaneously. A single job can utilize all the processors on a large server system. This capability is made possible by new threaded procedures in SAS System 9. Additional scalability can be achieved by splitting a single job into pieces and executing each piece on a different processor. This utilizes server resources effectively and provides high throughput for end users. The Itanium 2 processor's large addressable memory space and outstanding floating point performance allow SAS System 9 to scale effectively on multi-processor systems.
- **Multiple levels of parallelism.** Instruction level parallelism (ILP) is the ability to execute multiple bundles (three instructions in a bundle) at the same time. The Intel Itanium 2 microarchitecture can deliver faster performance by executing multiple bundles per clock cycle. Both at the instruction level and at the SMP system level, the microarchitecture enhances scalability by enabling more efficient use of virtually all system resources. Scalability is critical to data warehouse architects, who have found through experience that many data warehouses may start out modest in size and usage but then skyrocket as strategists and planners see the business intelligence benefits of analyzing clean, warehoused data. Thus, warehouses should be planned from the ground up with platforms that can scale to meet this demand, to deliver the performance and throughput business users require. As processing loads rise, the Itanium 2 processor's inherent support for parallelism allows it to keep pace.
- **Speculation and predication** reduce memory latency during data access and improve the efficiency of branch handling. These features are particularly vital in calculation-intensive applications that manipulate large volumes of data, such as SAS business intelligence applications that model customer behavior for inclusion in Customer Relationship Management (CRM) systems. Predication removes branches through parallel execution to increase performance. Speculation reduces the impact of memory latency, thus providing significant benefits to applications with many cache accesses. When used in a large-scale decision support system, these features promote higher throughput since the processor obtains data from memory more efficiently. It also runs program code in a more predictable and faster way.
- **Large number of registers and rotating registers** allow more data to be handled within the registers, a computer's fastest storage, producing a faster and more flexible applications environment. For a complex application like the SAS System, which processes enormous amounts of data, these additional registers translate into higher throughput rates. Meanwhile, the rotating registers further enhance loop performance with an innovation called software pipelining, which avoids code expansion and thus reduces cache misses. When software pipelining is used effectively in a large application like the SAS System, processor resources such as registers and logic units can be kept more uniformly busy, resulting in higher throughput.

Summary

The Intel Itanium 2 processor is designed for mission-critical enterprise computing applications such as data warehousing and business intelligence solutions from SAS. SAS and Intel have collaborated to ensure that SAS System 9 delivers outstanding performance on Itanium 2 processor-based platforms. Building on the power of the Itanium 2 processor, the IBM xSeries 450 is designed to scale, store, process and access large amounts of data. This platform is an outstanding choice to support large numbers of users and memory-, I/O and/or memory-intensive SAS solutions.

Learn More

To learn more about SAS computing on the Itanium 2 processor and the xSeries 450, contact your Intel or IBM representatives.

- **SAS and Intel Advanced Research Center (SIARC)** is jointly funded and staffed by Intel and SAS to enhance the performance of SAS solutions on new Intel platforms. SIARC identifies promising performance-enhancing opportunities for new Intel platforms and incorporates them into the next SAS release. The work performed at SIARC has led to significant optimizations in SAS System 9 for the Intel Itanium 2 microarchitecture.
- **SAS Customer Knowledge Practice from SAS and Intel** offers support to customers in installing, tuning or answering questions about SAS on Intel-based systems. The SAS Customer Knowledge Practice is provided by Intel® Solution Services in conjunction with SAS and is staffed by Intel engineers trained in SAS by SAS. For information, contact your local SAS or Intel representative, visit us on the Web at www.intel.com/internetservices/intelsolutionservices, or e-mail us at solution-services-questions@intel.com
- **Read the full sizing guide report.**
See *Sizing and Performance Considerations for Intel® Architecture-Based SAS* Solutions*, December, 2002. Visit www.intel.com/ids and click on Strategy & Solutions in the left-hand navigation bar.

Intel and Itanium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

eSeries, OnForever, X-Architecture, XpandOnDemand and xSeries are trademarks or registered trademarks of IBM Corporation or its subsidiaries in the United States and other countries.

* Other names and brands may be claimed as the property of others.

Copyright © 2003 Intel Corporation and IBM Corporation. All rights reserved.