Sizing Guide
Version 2.0

Sizing and Performance Considerations for Intel® Architecture-Based SAS® Solutions

Table of Contents

Introduction

As Business Intelligence software becomes core to company profitability, the necessary hardware investment is often an overlooked expenditure in leveraging such powerful software for maximum gain. Today, IT professionals have ever-decreasing budgets, but they are asked to sprovide greater return on investment with less hardware. The key is to use the best tool for the job or, in this case, the best CPU to achieve the sought-after results.

The sizing and performance considerations presented here will assist IT professionals in choosing the right Intel® architecture-based platform to get optimum performance of SAS software in a specific computing environment. In many cases, SAS System 9 and an Itanium® 2-based platform solves business problems most efficiently and economically. Updates to this sizing guide are available on the Intel and SAS Web sites when new platforms are tested and new data becomes available.

Executive Summary

Several system configurations were tested against a variety of SAS applications in order to evaluate the configurations that might best fill a customer's needs. This guide describes the tests that were run, the platforms and configurations used, the overall results (shown in Figure 1), and individual key findings that explain proven strategies for optimally sizing and configuring your Intel architecture-based SAS systems.

There are many nuances in the test results and key findings that illustrate price and performance advantages for one particular platform over another. This guide is best used when taken in its entirety. The general outlines of performance findings and system recommendations are summarized in the following matrix.

Using this matrix, you can evaluate your computing needs along four axes:

• The load placed by your application on system CPUs (left-side axis). For example, as a simple analysis of a data set becomes more CPU-intensive, the calculations per data point increase.

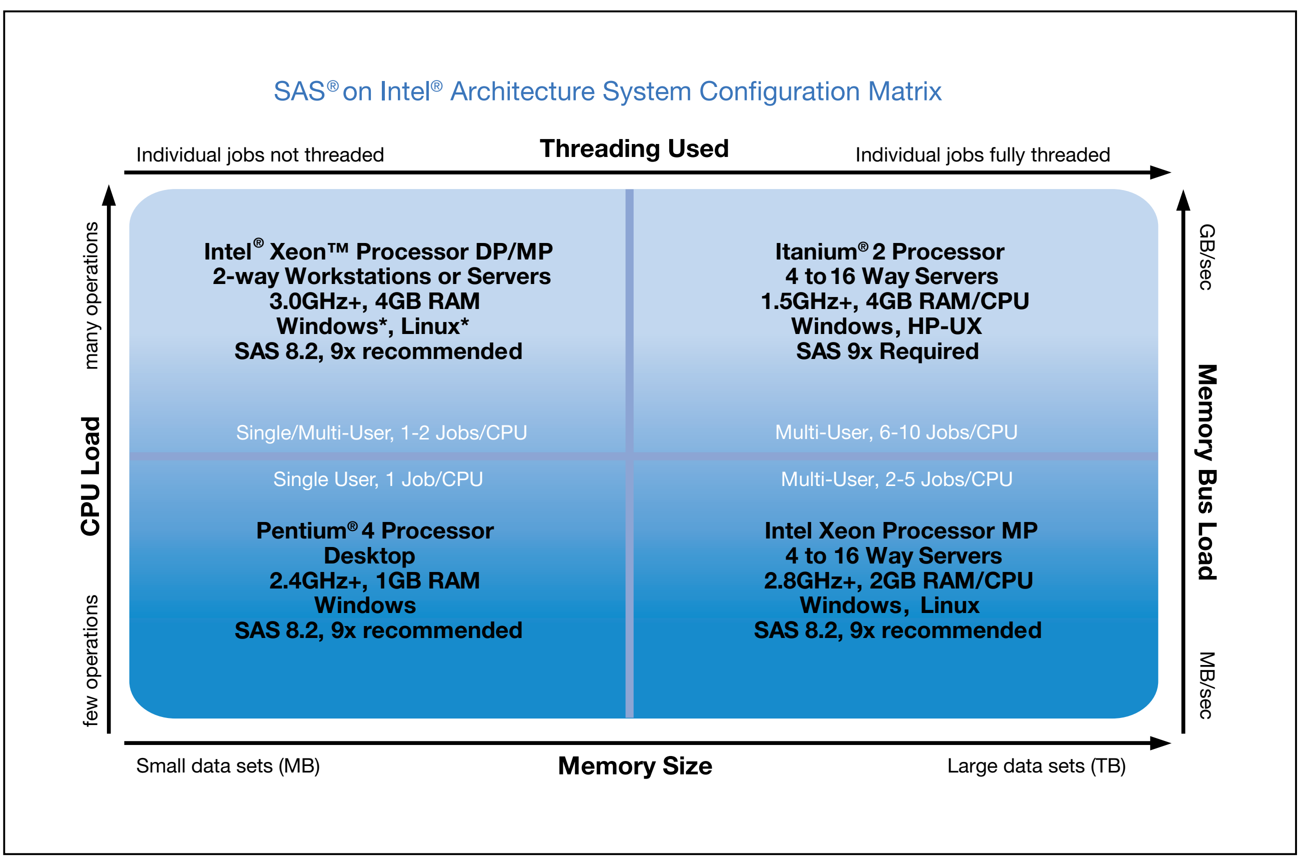

Figure 1: Quadrant Chart

- **The load placed by your application's data set size on the system (bottom-axis).** For example, as the size of the data set increases from MBs to GBs and more, you'll benefit from larger systems that can physically hold larger amounts of RAM.
- **The load placed on the memory bus (right-side axis).** For example, some statistical analysis of a large data set can be especially memory bandwidth-intensive.
- **The degree to which your applications depend on threaded routines (top axis).** Highly threaded applications such as general linear model (GLM) analysis benefit from higher-performing, multi-CPU platforms.

Another consideration when evaluating your computing needs is the number of simultaneous users or jobs. As more jobs are processed simultaneously, a higher-performing system configured with more CPUs can keep up with the total workload to provide optimum performance. The desktop and dual processor systems handle single users or single jobs. The larger systems in the right quadrant of the matrix handle more users or multiple jobs per CPU.

Dual- and multi-processing Intel Xeon™ processor-based systems complete the picture, providing excellent performance for a mix of computing needs. Dual-processor systems handle a small number of users and somewhat CPU-intensive small memory problems. Multi-processor systems add value in handling a large number of users and workloads that require more memory and are more CPU-intensive.

Key Findings

The results of our tests support several key findings and offer additional insight into optimal sizing and configuration practices:

- **System scaling has a different effect on performance in single-user versus multi-user environments, as well as non-threaded versus threaded applications.**
- **Threads executing simultaneously can contend for processor time, which affects system performance.**
- **Memory bandwidth can have a significant effect on performance, depending on the application and data set.**
- **The number of simultaneous users provides a good gauge for determining the optimum number and type of processors. As the number of simultaneous users increases, the number of CPUs increases and the recommended type of processor should change.**
- **Larger systems, such as 16-processor Itanium 2-based systems, provide the best performance for many instances of SAS software offerings, large user bases, threaded applications such as general linear model analysis, and solution stack implementations.**
- **Different workloads run better on various platform architectures.**
- **On systems equipped with Intel Xeon processors MP, CPU-bound workloads can decrease in execution time linearly proportional to the increase in CPU speed. This is known as frequency scaling. Other workloads respond to different factors such as processor cache size.**
- **Beginning with version 9, SAS takes advantage of newer processor architecture features and enhancements.**

This guide discusses each of these key findings in more detail later in this paper and provides supporting data and advice about how to use the knowledge to best advantage. The key findings are an excellent starting point for researching the performance of SAS applications on Intel architecture-based systems.

Due to a large number of possible test scenarios and system configurations, only a finite set of workloads can be run. Because computing environments and solution sets are unique to each company, it is recommended that companies work with a specific hardware vendor and a SAS sales team to choose a system configuration that addresses their specific needs. In addition, Intel and SAS have created the SAS Customer Knowledge Practice, which can help in choosing and optimizing systems. For links to the SAS Customer Knowledge Practice from SAS and Intel, Intel® Solution Services, and other organizations that can provide professional sizing and configuration services, see "For More Information" at the end of this guide. Of particular interest is the I/O subsystem. While this paper provides a few high-level recommendations for subsystems, it is a good idea to size the subsystem carefully with the assistance of your company's specific hardware vendor.

Sizing Considerations

To simplify this discussion, let's begin with some basic definitions of computer platforms.

- **A desktop system** has one CPU with basic memory and I/O subsystem.
- **A workstation** usually has 1 or 2 processors, with more memory and multiple hard drives than a desktop system.
- **A server**, the largest and most robust system, has at least 2 to 4 processors, robust memory, and more hard drives than a workstation with independent paths to increase performance.

The key to correctly sizing platforms for your SAS applications is to carefully characterize the SAS job. By answering the following questions, you'll get a better understanding of the hardware and software requirements that are needed.

- **Which SAS procedures will be used?** Depending on the procedures that are used, a job that runs well on a single-CPU system might also benefit from using multiple processors. Certain job types run best on the Intel Xeon processor architecture, while others run best on the Itanium processor architecture. For more discussion, see "Explanation of Key Findings," numbers 1, 2, and 6, later in this guide.
- **Will the machine host more than one software application or user?** If there is only a single user and the workload is not too large, optimal performance can be achieved by using a Pentium® 4 processor-based desktop. However, if multiple users need to be supported concurrently, a workstation or server might deliver better performance. For more information about the optimum number of users per CPU, see "Explanation of Key Findings," numbers 2, 4, and 5, later in this guide.
- **Does data need to be shared among users and programs?** If users store their own small or medium-sized data sets, they might be well-served by individual copies of SAS on their own desktop systems. If data needs to be shared among users or programs, an external disk array attached to a server is the more manageable and effective solution.
- **Does the job have large data throughput requirements?** Many SAS 9 workloads run faster on a larger system with more CPUs. However, these systems also require a larger initial investment. For a discussion of workload scalability and headroom, see "Explanation of Key Findings," numbers 1 and 5, later in this guide. If multiple users typically run threaded workloads, the number of processors accessed by each user should be limited. For more information, see "Explanation of Key Findings," numbers 1 and 2, later in this guide.
- **Which version of SAS should be used?** Some of the analytical operations, along with sorting and aggregation procedures (PROCS), are threaded in SAS 9. If these PROCs are used heavily, then upgrading to SAS 9 should decrease the time it takes to run a job to completion.
- **What are the memory requirements of the application?** Depending on your workload, considerations of memory size and bandwidth can be important. Furthermore, if you support multiple users simultaneously, you should plan for enough memory to support each user. For further discussion of the effect of memory usage on performance, see "Explanation of Key Findings," numbers 3 and 6, later in this guide.
- **Do you perform many calculations per data point?** Extract Transform Load (ETL), Query-Reporting (Q&R) and data mining workloads can put a significant load on the processors in the system. These job types respond to the higher processor speeds offered by the Intel Xeon processor architecture. For more information, see "Explanation of Key Findings," numbers 3 and 6, later in this guide.
- **What operating systems are supported?** For 32-bit Intel architecture-based systems, SAS runs on your choice of Microsoft Windows* or Linux*. For systems based on the Intel Itanium processor family, SAS is supported on the 64-bit editions of Windows Server 2003 or HP-UX11* operating systems.
- **How much and what type of disk storage is needed?** For better performance, the SAS application, I/O files, and the temporary work space should be placed in separate file systems and folders should be placed on separate disks. Because SAS jobs use a lot of I/O, speeding up disk I/O is an effective way to improve overall system performance. In performance characterization tests, SAS workloads ran very well with a mid- to high-end disk subsystem that included at least a dozen disks with multiple independent data paths. I/O is a very important consideration in overall system performance.
- **Will SAS be used on a network?** Because SAS workloads often move a large volume of data, network bandwidth is an important consideration. Reducing network delays for the temporary work area is one effective way to optimize performance. Your hardware vendor might have additional advice about optimizing network performance and segmentation.

Note: For additional tuning guidelines, contact your hardware vendor. Several system manufacturers have published Tuning Guides, which are referenced on the Web at http://www.sas.com/partners/directory/intel.

The Tests

Individual User Test Suite (IUTS)

The Individual User Test Suite addresses a variety of usage models, including data manipulation, analytical operations and statistics.

- **Analytical** tests perform stepwise linear regression, general linear model, and other analytical operations on the data set. To observe the effect of data set size on performance, data sets of various sizes were tested. Tests of stepwise linear regression included 297 variables, ranging from about 10,000 to 200,000 observations. The general linear model tests also included various size data sets with approximately 10 variables and 1000–3500 observations.
- **Data mining** tests used SAS Enterprise Miner to find data trends on a data set. This test begins with a moderately large data store of approximately 200,000 observations, finds the data of interest, and eventually runs various models on a subset of approximately 40,000 rows.
- **Data manipulation** tests use a moderately large volume of input data, which is imported from approximately 50 flat ASCII files into two SAS data stores of 1.1GB and 2.8GB.
- **The Extract Transform and Load (ETL)** tests use the two SAS data sets that are generated in the data manipulation test and search, sort and manipulate the data in various ways. One test focuses on generating Multi-Dimensional Database (MDDB) cubes, mostly in memory.
- **Query and Reporting (Q&R)** tests use EBCDIC input files and binary files from a mainframe in addition to self-generated data. Various statistical tests are also used to evaluate the data.

Multi-User Test Suite (MUTS)

In an effort to simulate an enterprise work environment that has multiple users developing various usage models, the test environment included a Multi-User Test Suite. This suite includes approximately 10% analytical, 60% query-reporting, and 30% data manipulation tasks. The tasks were chosen from the ITS workload described above, plus some jobs that included SQL-style query and reporting operations. One example is a job that creates a view to a table and performs a count. The multi-user tests were especially valuable in learning what number of users can be supported on a given system, while maintaining reasonable performance. To establish a basis for our recommendations, a series of suites were run while varying the number of users per CPU, for example, 2 users per CPU, 4 users per CPU, 6 users per CPU and so on. In general, each suite was created by adding more jobs while maintaining the same distribution among data manipulation, analytical and query-reporting workloads.

Multi-user tests were initiated by starting all the selected tests at the same time. Because the individual workloads varied, some jobs finished before the others. The run time for the suite as a whole (monitoring real (clock) time, user CPU time, and system CPU time for each) was recorded in order to have many checkpoints to compare the results to. In addition, the Performance Monitor in the Windows operating system was used to track system information such as the Current Disk Queue Length and Processor Queue Length to gain insight about the use of system resources.

Test Hardware

For this study, both commercially available hardware and pre-production reference platforms provided by Intel were tested. The systems tested include:

- **Reference platform.** Four Itanium 2 processors at 1.0 GHz, with 1.5-MB L3 cache. 16-GB DDR memory, Windows Server 2003 64-bit Enterprise Edition RTM build 3790.
- **Large Itanium 2-based server.** Sixteen Itanium 2 processors at 1.0 GHz, with 1.5-MB L3 cache, 64-GB DDR memory, Windows 2003 Datacenter Server RC1 build 3663.
- **Mid-range Itanium 2-based server.** Four Itanium 2 processors at 1.5 GHz, with 6-MB L3 cache, 16-GB DDR memory, Windows Server 2003 Enterprise Edition for 64-bit Itanium-based Systems RTM build 3790.
- **Mid-range Itanium 2-based server, system #2.** Four Itanium 2 processors at 1.2 GHz, with 6-MB L3 cache, 16-GB DDR memory, Windows Server 2003 Enterprise Edition for 64-bit Itanium-based Systems RTM build 3790.

Test Hardware

- **Mid-range Intel Xeon MP processor-based server.** Eight Intel Xeon processors at 1.6 GHz, 4-GB SDRAM memory, Windows 2000 Datacenter Server, build 2195 Service Pack 2. This system was tested with Hyper-Threading Technology (HT Technology) disabled. In future releases of this guide, we will explore the effects of HT Technology.
- **Mid-range Intel Xeon MP processor-based server.** Four Intel Xeon processors MP at 2.0, 2.8, and 3.0 GHz, 12-GB DDR memory, Windows Server 2003 Enterprise Edition RTM build 3790.
- **Large Intel Xeon MP processor-based server.** 32 Intel Xeon processors MP at 1.4 GHz, 28-GB ECC memory, Windows Server 2003 Datacenter Edition RC1 build 3613.
- **Small Intel Xeon DP processor-based server.** 2 Intel Xeon processors DP at 3.06 GHz, 6-GB ECC memory, Windows Server 2003 Enterprise Edition RTM build 3790.

All platforms were connected to the same disk subsystem, and tests were run consecutively on each platform. The disk subsystem comprised two EMC CLARiiON* RAID arrays: a model FC4500 for input and output files, and a faster model FC4700-2 for the I/O-intensive SASWORK directory.

To gain insight into the effect of multiple-CPU configurations on the performance of various workload types and sizes, the systems were run with their full complement of processors as well as with a subset of CPUs. For example, the 8-CPU Intel Xeon processor-based system was run in 8-, 4-, and 1-processor configurations. These configurations might not always represent commercially available systems, but suggest the range of results one might see when using system partitions or when purchasing smaller but similarly configured systems.

A more complete picture of the overall system performance requires more tests with varied characteristics and differing workloads. This document will continue to be updated with the latest findings in order to provide customers with a planning tool for optimizing your SAS environment — today and tomorrow.

SAS Software Environment

Our testing included both SAS Release 8.2 and SAS 9.

- **SAS Release 8.2** is single-threaded and supports 32-bit Intel architecture-based platforms.
- **SAS 9** is the version that incorporates added support for systems based on the Itanium processor family. In addition, some of the analytical procedures, as well as sorting and aggregation PROCs, are threaded in SAS 9. Workloads using these procedures can take advantage of multiple CPUs in a single job. See key finding number 1 for a discussion of the performance benefits.
- **SAS 9.1** is the latest release and contains additional optimizations for the Itanium 2 processor platform.

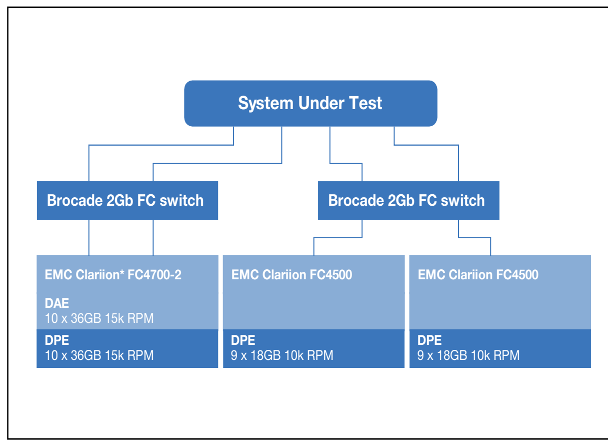

Figure 2: Hardware Diagram

Due to the publication time frame for this paper, the testing was done using an Early-Adopter version of SAS 9.1. In the future, this paper will be updated with results from the production release of SAS 9.1.

Operating Systems

- **Intel Xeon processor-based systems:** SAS Release 8.2 and SAS 9 support Linux and the various forms of Windows. In this testing, Windows 2000 Datacenter Server and Windows Server 2003 Enterprise Edition were used.
- **Itanium 2-based systems:** For the 64-bit Itanium architecture, either Windows Server 2003 or Windows XP 64-bit Service Pack 1 or later are required. In this testing, Windows Server 2003 Enterprise Edition and Windows Server 2003 Datacenter Edition were used.

Multi-User Test Suite (MUTS)	Intel® Xeon™ processor-based systems vs. Itanium® 2 processor-based systems	Itanium 2 processor-based systems vs. Intel Xeon processor-based systems
2 Jobs/CPU	~10-50% benefit	
4 Jobs/CPU	Within 10%	Within 10%
6 Jobs/CPU		~10-50% benefit
8 Jobs/CPU		>50% benefit
16 Jobs/CPU		16-way: >50% benefit

Table 1: Multi-User (MUTS) Checkerboard

Individual Test Suite (ITS)	Intel® Xeon™ processor-based systems vs. Itanium® 2 processor-based systems	Itanium 2 processor-based systems vs. Intel Xeon processor-based systems
Analytical – Small	within ~10%	within ~10%
Analytical – Medium		~10-50% benefit
Analytical – Large		~10-50% benefit
Analytical – Small GLM		~10-50% benefit
Analytical – Medium Small GLM		~10-50% benefit
Analytical – Large GLM		50% benefit
Analytical – Mixed #2		50% benefit
Analytical – Mixed #1	~10%-50% benefit	
Analytical – Mixed #3	within ~10%	within ~10%
Data Manipulation		~10-50% benefit
ETL – Mixed #1	~10%-50% benefit	
ETL – Mixed #2	>50% benefit	
ETL – Mixed #3	~10%-50% benefit	
ETL – MDDB	within ~10%	within ~10%
Data Mining	~10%-50% benefit	
Q&R – Reporting	~10%-50% benefit	
Q&R – General	~10%-50% benefit	
Q&R – Finer Grid	~10%-50% benefit	

Table 2: Individual (ITS) Checkerboard

- Dark blue indicates a performance advantage of more than 50 percent for the highlighted platform.
- Pale blue indicates a 10 to 50 percent advantage.
- Tests with no highlighting ran with comparable performance on both platform architectures.

The Checkerboard Chart

The multi-user checkerboard chart in Table 1 gives a picture of the test results for varied workloads, user bases and system configurations. The chart shows which platforms are most likely to provide optimal performance for specific SAS applications, data sets, performance needs, and number of users.

The color coding in the chart indicates the differences in performance between an Intel Xeon processor-based platform and Itanium 2-based platforms. Dark blue signfies a performance advantage of more than 50% for the highlighted platform. Pale blue signifies a 10% to 50% advantage. Tests with no highlighting ran with comparable performance on both platform architectures. This chart has been updated since the December 2002 publication to include the Intel Xeon processor-based system results.

Due to its unusual scalability, the largest GLM test ran significantly better on the largest Itanium 2 processor-based system, so that platform is highlighted alone.

Explanation of Key Findings

In the initial round of testing, several performance trends were discovered that can assist in selecting the right system configuration. Below are the key findings based on number of users, types of applications, number of CPUs, and other common variables.

1. System scalability has a different effect on performance in single-user versus multi-user environments and in non-threaded versus threaded applications.

In a single-user environment, system scalability has the greatest performance benefit for applications that rely heavily on threaded routines. For example, a single user running only non-threaded procedures will not see scaling past 2 processors. However, a user will see some benefit using 2 processors rather than 1 processor because the operating system and SAS application can run on different processors.

Using SAS 9, a single user can take advantage of multiple CPUs to complete a job. This benefit is best achieved when using procedures that are threaded internally.

Scaling to 4 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 because they depend on other system resources that might not increase with the number of CPUs.

In the complete Individual User Test suite (IUTS) on SAS 9, the greatest performance gains occurred in scaling from 1 CPU to 4 CPUs. Performance continued to improve up to 8 CPUs. When using more than 8 CPUs, the less-threaded jobs began to affect results negatively.

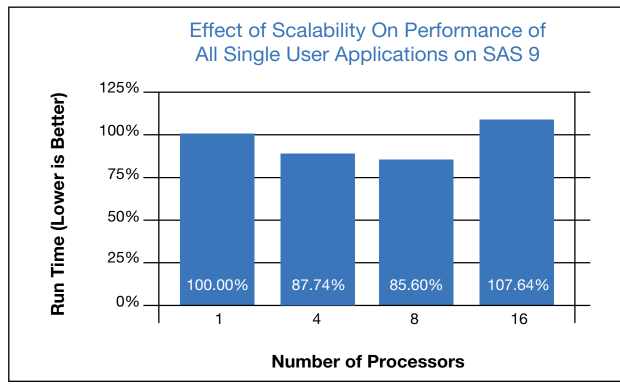

Figure 3: Individual Test Suite (ITS) Scaling

However, individual jobs that thread well can continue to improve in performance up to 16 CPUs. The primary example from the test suite is the general linear model (GLM) using a large data set. This job ran the fastest on a 16-CPU system, which indicates that large-scale multiprocessing can greatly improve performance for businesses that depend heavily on this type of application.

Note: This finding applies only to single-user environments. As you add more users (with each user submitting jobs simultaneously) 4-, 8-, and 16-CPU systems can process more jobs concurrently, optimizing throughput across the user base and application workload. See the following key findings for examples and configuration advice for high-volume environments.

2. Threads executing simultaneously can contend for processor time, affecting system performance.

In the tests on SAS 9 simulating a multi-user environment, the system spent significant time managing multiple threads. This was evident from observing the context switches as measured in the performance monitor. Similarly, in the Individual User Test Suite (ITS), the overall performance was best, not at the highest CPU configuration, but at 4 to 8 CPUs. Better overall system performance can be achieved if each job is restricted to use only a limited number of processors. For example, in a 16-CPU system, limiting individual jobs to 4 processors each increases performance compared to an unlimited thread count.

You can limit the number of processors allocated to each individual job by doing one of the following:

- Setting the SAS thread count (CPUCOUNT). Establish a default setting for CPUCOUNT and allow individual users to change the setting if needed. For example, findings indicate that a workload using a general linear model with a large data set continued to demonstrate performance gains past 4 CPUs. Therefore, users running this type of job can set CPUCOUNT to 8 while other users retain the default setting of 4.
- Dividing the system into partitions. Partitioning effectively breaks down the system into several smaller ones. If supported by the system manufacturer, this option enables the administrator to allocate system resources to optimize the performance of applications running in each partition. For example, partitions can be individually configured for the number of processors, memory size, hard disk space and processor type.

Both these system configuration strategies modularize the server's processing power, so that no single job attempts to distribute operations across all the available processors at the same time. By minimizing conflicts between jobs, this strategy ensures maximum throughput for all applications and users.

Note: Assuming similar resources, such as memory and I/O, a 4-CPU server offers excellent performance for most single instances of non-threaded SAS applications. However, it's usually best to deploy extra computing capacity to handle future needs. If you anticipate supporting multiple SAS users in the future, it probably makes sense to choose a system with more processors.

3. Memory bandwidth can have a significant effect on performance, depending on the application and data set.

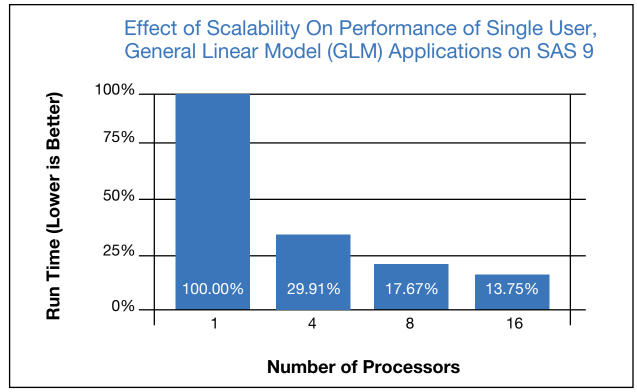

Figure 4: General Linear Model with Large Dataset

When your application 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 data sets can benefit greatly from a system that offers the widest data bus and the highest memory bandwidth. This type of application often achieves maximum performance by using the wide-open data paths of an Itanium 2-based system.

Extremely complex processing tasks, such as data mining and statistical analysis on large datasets, benefit from the high memory bandwidth of the Itanium 2 processor.

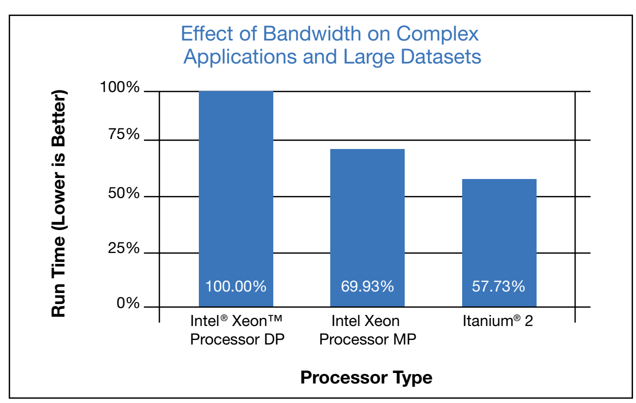

Figure 5: MUTS Throughput on CPU Types

CPU-intensive applications that place less demand on memory bandwidth might provide better price and performance on multi-CPU servers that are based on the Intel Xeon processor. For example, data analysis using relatively smaller data sets can benefit from the extremely quick ramp-up time of this processor, without incurring a significant penalty in memory throughput.

Note: The checkerboard chart in Table 1 illustrates the point at which a larger data set size makes a difference in I/O performance under various system configurations. For example, the larger analytical tests performed best on the Itanium 2 processor, which has a faster memory architecture than the Intel Xeon processor-based system. The “Analytical-Large” test included 183K observations and 297 variables. Similarly, the largest GLM did best on the Itanium 2 microarchitecture. This effect became evident even earlier, at approximately 2100 observations and 11 variables, with the medium-small GLM.

4. The number of simultaneous users provides a good gauge for determining the optimum number and type of processors needed. As the number of simultaneous users increases, the number of CPUs increases, and the recommended type of processor should change.

The testing on the Itanium 2-based systems reflected optimum price and performance at an average of 8 to10 users per CPU, while the Intel Xeon MP processor-based platform performed better with 2 to 6 users per CPU. These conclusions were reached by examining the lengths of the processor queue and the disk queue with various user loads. Workloads that produced processor and disk queue lengths at or near 1.0 were used to determine the ideal number of simultaneous users per CPU. (These queues can be observed by using the Performance Monitor under Windows.)

The findings suggest that, ideally, there should be no more than one process at a time in each queue, showing that the processor and disk I/O are keeping up with the tasks to be completed. Queues grow longer when the processors or disks are unable to keep up with the workload, indicating a drop in system efficiency.

At a load of 8 simultaneous jobs per CPU on an Itanium 2-based system, the average processor queue and disk queue lengths remain at 1.0, signifying minimal competition for bandwidth.

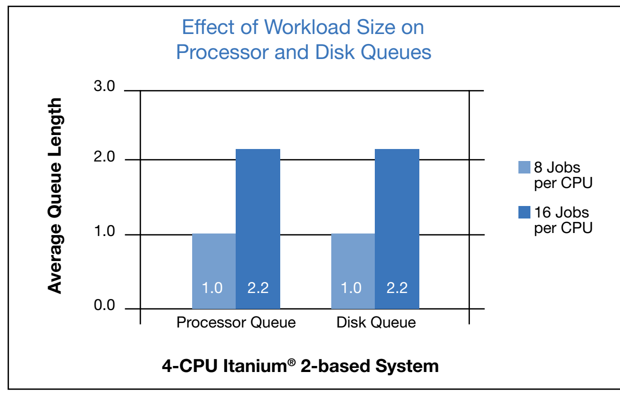

Figure 6: Workload Size

The recommended number of users per CPU is only an estimate. The ideal number of users per CPU is influenced by multiple factors, including whether the application is especially processor-intensive and keeping each CPU busier as well as the other resources available on the system. For example, a system with insufficient memory is unlikely to benefit from more CPUs. Also, remember that the price and performance of the I/O subsystem can be an important consideration in your decision. When the system is consumed with multiple long jobs, the results illustrate that short jobs can be submitted and completed quickly, with relatively little impact on overall performance. For more tuning techniques and advice, see the “SAS Tuning Checklist” on the SAS Web site.

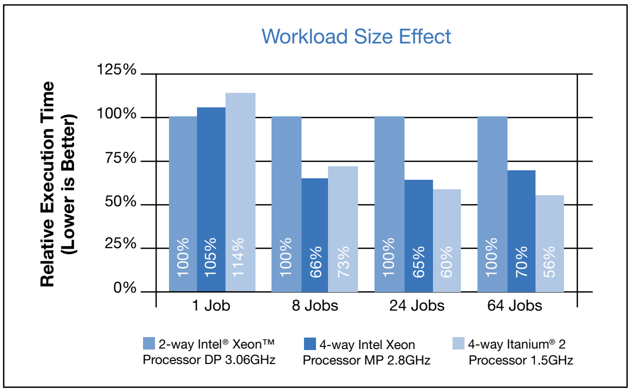

Figure 7: MUTS Throughput on CPU Types

This finding also illustrates the effect of the various factors influencing the Quadrant Chart (see Figure 1). When comparing overall system throughput for MUTS workloads, the ability of the higher "class" CPU types to better handle the workload becomes evident.

While 2-way Intel Xeon processor MP-based systems perform well for a maximum of 2 jobs/CPU, 4-way Intel Xeon processor MP-based systems exhibit greater throughput after you exceed that workload level. When the workload level reaches 6 to 8 jobs/CPU, 4-way Itanium 2-based systems become more effective. For single jobs, other factors influence the performance on Itanium 2 or Intel Xeon processor-based systems. For more information about single jobs, see Table 2.

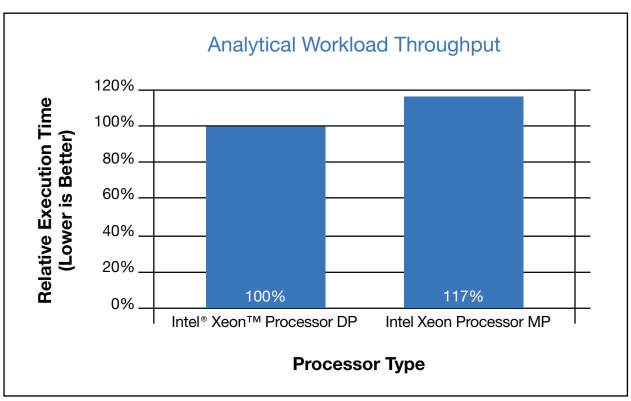

Figure 8: ITS Throughput on CPU Types

For the Individual User Test Suite, performance is significantly influenced by architectural features such as the processor speed and System Bus speed. For our testing, the two-way Intel Xeon processor system included a 533-MHz System Bus, vs. 400-MHz System Bus for the Intel Xeon processor MP-based system. In addition, the Intel Xeon-based system featured a 3.06-GHz processor speed vs. a 2.8 GHz-processor speed for the Intel Xeon processor MP-based system. The Analytical workload jobs in the Individual User Test Suite results show the impact of these differences, with the Intel Xeon DP-based system outperforming the MP system.

For those workloads whose data fits into cache, the Intel Xeon processor MP-based system provided superior performance, largely driven by the L3 cache, which is a differentiator between DP and MP systems. In general, for any given system architecture, increasing the cache or memory will improve performance.

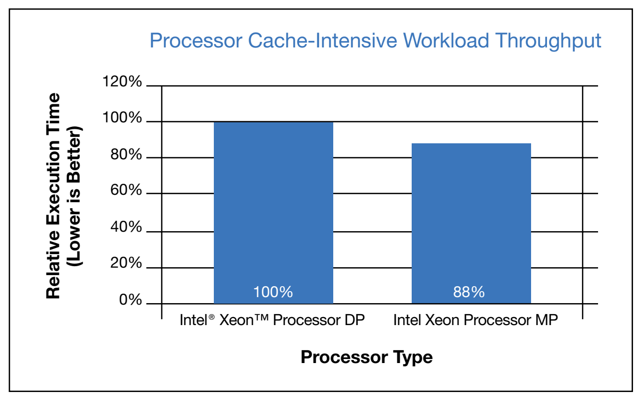

Figure 9: Cache Benefit of Intel™ Processor MP-based System

The decision to use a DP or an MP system should include factors other than the performance of the IUTS or MUTS. For example, system memory and I/O slot capacities vary among manufacturers and models. In particular, the number of PCI slots might be a factor if your workload includes a large amount of data that resides on a Storage Area Network (SAN). In these situations, better performance can be achieved by using multiple Host Bus Adapters (HBAs), each of which consumes a PCI slot. A dual Intel Xeon processor-based system configured as a high-end workstation might have fewer available PCI slots than an Intel Xeon processor MP-based system because these slots can be completely consumed by the SAN, making them unavailable for other uses.

5. Larger systems such as the 16-CPU Itanium 2-based systems provide the best performance for:

- Large numbers of users, following the general guidelines given in key finding number 4. Large systems might also be the best solution for hybrid installations of SAS, for example, where one system supports multiple SAS solutions that are product suites targeted to a particular purpose.
- Highly threaded applications such as general linear model analysis of very large data sets, for example, an application that analyzes consumer preferences for a given product category across a large retail chain. Itanium 2-based systems scale up to increase throughput in large multi-user environments.

Note: Throughput of a multi-user workload increases as you increase the number of CPUs. If throughput is very important, you can use a large-scale system to reduce run times in multi-user and some single-user workloads.

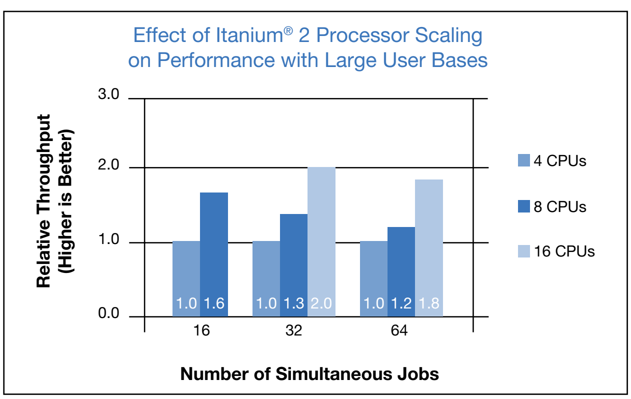

Figure 10: Effects of Large User Base

6. Different workloads run better on various platform architectures.

Analytical tests generally ran best on the Itanium 2-based systems. Within that group, the most threaded test (Analytical-Large GLM) ran best on the largest system, with 16 processors (this result is discussed further in key finding number 5).

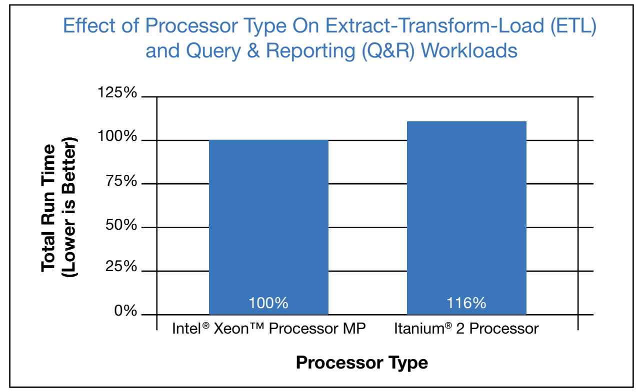

Figure 11: CPU Speed-Sensitive Tasks

The Analytical-Small GLM test ran slightly better on Intel Xeon processor MP-based systems; the Analytical-Large GLM test ran slightly better on Itanium architecture. The medium-sized test showed little difference between these platforms. Overall, the tests that required more memory or memory bandwidth ran better on Itanium architecture. In keeping with this trend, the largest GLM attained the best performance on multi-CPU Itanium 2-based systems.

The data manipulation tests show little performance difference between these two platforms. It is likely that the I/O subsystem primarily impacts this type of workload and overshadows any differences between processors.

ETL, Data Mining, and Query-Reporting (Q&R) tests generally performed best on the Intel Xeon processor-based system. These tests benefited from the high speed and low overhead of the Intel Xeon processor. The primary exception to this trend was ETL-MDDB, which scaled fairly well and used a significant amount of memory. These characteristics best matched the Itanium 2-based system with 16 CPUs.

CPU speed-sensitive tasks such as ETL workloads might benefit from the higher speed and quick ramp-up time of Intel Xeon processor MP-based systems.

In order to provide the best performance for those customers using single-CPU systems, some SAS procedures have optimized versions for 1-way operation. An example of this behavior is the ETL test generating MDDB cubes, which performed better in a 1-way configuration than in a 4-way configuration. Multi-CPU performance improves beyond 4-way configurations and quickly supersedes the 1-way result.

This improvement is more apparent on systems that are equipped with faster Intel Xeon processors MP. When comparing 1.6-GHz against 2.0-GHz Intel Xeon processor MP-based systems, the test results suggest that there are cases where more than a 25% speedup (decrease) in elapsed time was observed during IUTS testing. This is explained by the better memory bandwidth (see key finding number 3) on the 2.0-GHz Intel Xeon processor MP and increased L3 cache size compared to the previous generation processors. The 2.0-GHz system was better able to handle the higher workloads of multi-users. For example, when running the MUTS with 32 users per CPU, a 30% speedup was observed on the 2.0-GHz system versus the 1.6-GHz system.

7. On systems equipped with Intel Xeon MP processors, CPU-bound workloads decrease in execution time linearly proportional to the increase in CPU speed. This is known as frequency scaling. Other workloads respond to different factors such as processor cache size.

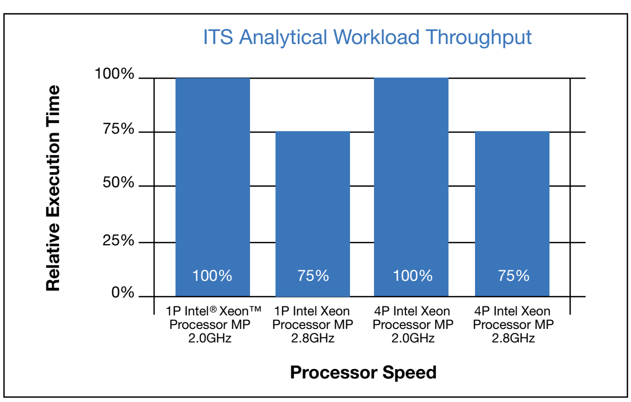

Figure 12: CPU Clock Speed Effect on ITS Workload

When observing systems equipped with CPUs from the Intel Xeon processor MP family, CPU clock speed, or frequency rating, has a nearly 1:1 correlation to execution time for the IUTS analytical workload as measured with the “Real Time” metric provided in SAS software.

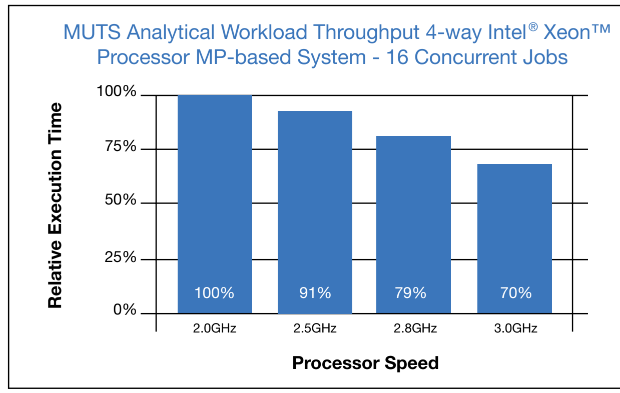

Figure 13: CPU Clock Speed Effect on MUTS Workload

Because the analytical components of the IUTS workload are CPU-intensive, these results indicate that the architecture of the Intel Xeon processor MP platform demonstrates frequency scaling benefits when moving to CPUs that have higher clock speeds. The MUTS workload on Intel Xeon processor MP-based systems also showed similar execution time benefits for the analytical components.

Because the MUTS workload places additional stress on a system with the overhead task of balancing system resources among multiple processes and threads, the above results indicate that the architecture of the Intel Xeon processor MP platform is capable of adequately handling the additional load with minimal degradation in execution time for the analytical components of the workload.

Workloads that are bound by other architectural factors such as I/O or memory will probably experience less benefit from higher processor frequencies. When observing the results for non-CPU-intensive workloads such as Data Manipulation and Query & Reporting, frequency-scaling effects are not fully realized.

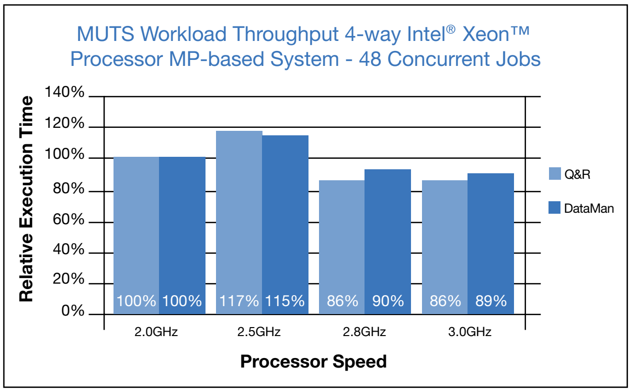

Figure 14: CPU Clock Speed Effect on MUTS Non-Analytical Workload

For these workloads, CPU frequency did not appear to have a corresponding effect on the relative execution time. Also notice that the 2.5-GHz processor produced longer execution times. This behavior is probably due to the smaller (1-MB) L3 cache on the 2.5-GHz processor. For the discussion of how different L3 cache sizes impact system performance, see key finding number 4. The data that supports this finding will be further developed and presented in upcoming revisions of this paper as various processors with higher clock speeds become available in the different platform architectures (that is, the two-way Intel Xeon processor-based, Intel Xeon processor MP, and the Itanium 2-based systems).

8. Beginning with version 9, SAS takes advantage of newer processor architecture features and enhancements.

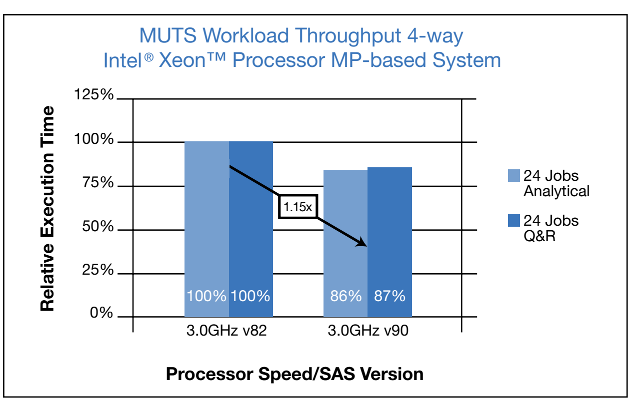

Figure 15: SAS 8.2 vs. SAS 9.0

SAS, with assistance from Intel, will continue to improve all its product offerings in many ways that are important to its customer base.

In this example, the results of the processor-sensitive workloads (Analytical and Q&R) show a 1.15x performance gain is realized when comparing SAS 9 to SAS 8.2. These results were observed on systems that have the same hardware and OS platforms. The only difference in the tests was the version of SAS that was used.

Working with the 64-bit enabled versions of SAS, the following results were observed:

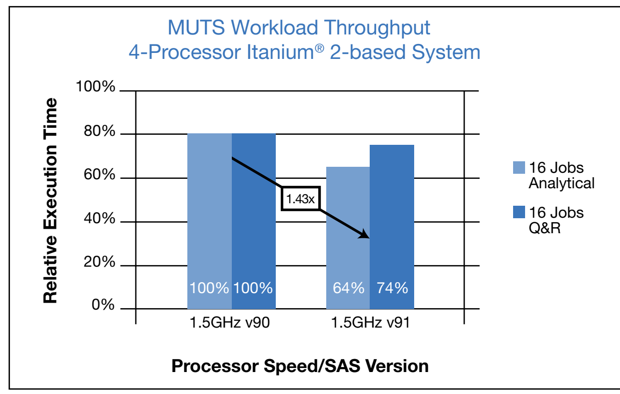

Figure 16: SAS 9.1 Early Adopter vs. SAS 9.0

The Early Adopter version of SAS 9.1 exhibited a 1.43x performance gain when compared to SAS 9.0 results.

The data that supports this finding will be further developed and presented in upcoming revisions of this paper as newer versions of SAS become publicly available.

Looking Ahead

This report contains several guidelines for evaluating, selecting, and configuring the best system for a specific application set, user base size, throughput needs, and budget by carefully characterizing the SAS job and computing environment. In general, when selecting a hardware configuration, it is recommended that you take into account future growth for new applications, increased volumes of data, and larger user bases.

The SAS Customer Knowledge Practice from SAS and Intel and Intel Solution Services is provided by organizations that are uniquely positioned to apply the best available knowledge for achieving the goal of optimizing SAS applications on Intel architecture. Please contact your Intel representative or SAS Account Executive anytime with your questions and comments.

For More Information

To learn more about SAS and Intel and our continuing efforts to improve and quantify the performance of SAS solutions on Intel architecture, visit: http://www.sas.com/partners/directory/intel/index.html or http://www.intel.com/go/SAS.

SAS Customer Knowledge Practice from SAS and Intel

If you need help to install or tune SAS on Intel processor-based systems, or you need answers to questions about SAS, Intel and SAS have built a practice to help. Please contact your local SAS or Intel representative for additional information or contact:

Rob Hamm, SAS Institute Inc. (919) 531-4970 or Robert.Hamm@sas.com

Jerome Esteban, Intel Corporation (770) 368-7634 or jerome.w.esteban@intel.com

The SAS Customer Knowledge Practice is provided by Intel Solution Services in conjunction with SAS.

Staffed by Intel engineers who are trained in SAS by SAS, the practice provides a knowledge base and support for SAS installations and offers a single point of technical support, helping remove barriers to Intel architecture-based deployments of SAS.

Intel Solution Services

Intel Solution Services provides a worldwide consulting and services R&D organization, specializing in distributed solutions and enterprise e-Business infrastructure. Intel Solution Services works with key industry leaders such as SAS Institute and early adopters, to deliver cutting-edge solutions that can be offered to customers. You can accelerate the adoption and application of the latest technology advances on Intel architecture-based platforms by drawing on the expertise of Intel Solution Services. The mutually beneficial relationship between Intel and SAS enables all parties to reap the benefits of leading technology and precious time-to-market advantages. To learn more, visit www.intel.com/internetservices/intelsolutionservices or e-mail us at solution-services-questions@intel.com

1. Hyper-Threading Technology requires a computer system with an Intel Pentium® 4 processor at 3.06 GHz or higher, a chipset and BIOS that utilize this technology, and an operating system that includes optimizations for this technology. Look for systems with the Intel Pentium 4 Processor with HT Technology logo which your system vendor has verified utilize Hyper-Threading Technology. Performance will vary depending on the specific hardware and software you use. See http://www.intel.com/info/hyperthreading for information.

Performance tests and ratings are measured using specific computer systems and/or components and reflect the approximate performance of Intel products as measured by those tests. Any difference in system hardware or software design or configuration may affect actual performance.

Version 2.0 First Published March 2004
0304/BTB/ITF/XX/PDF