September, 2004

SAS®9 Software Performance on IBM® eServer® p5 Systems

IBM Systems and Technology Group

Introduction

When customers deploy SAS software on powerful, resilient IBM servers and storage systems, the result is an outstanding combination of performance and flexibility for advanced BI and decision making. SAS and IBM offer joint solutions that are scalable, reliable and supported by trusted business partners that stand behind their products and services.

This paper details the SAS9 software performance on the IBM POWER5™ based eServer p5 systems in the areas of workload, single user, simultaneous jobs, scaling, and PROC SORT and GLM. The effect on performance of POWER5 Simultaneous Multi-Threading (SMT) is also documented. For comparison purposes we used the previous generation, the POWER4™ based pSeries™ servers. All tests demonstrate that the SAS9 performance on POWER5 based p5 servers is significantly better than the already very high performing SAS 8.2 and 9.1 on POWER4 based pSeries servers.

Cornerstones of SAS9

SAS 9.1 represents a revolutionary new generation of SAS technology. SAS 9.1 is based on four design cornerstones:

Usability
SAS 9.1 provides a new suite of interfaces that are developed for various types of users and their functional needs.

- MIGRATE procedure
- New SAS/GRAPH styles.
- Output Delivery System writes to more destinations
- User-created formats and informats with names longer than eight characters
- Enhanced SAS Data Quality Server™

Scalability
SAS 9.1 delivers enhanced support for the increasingly intensive processing tasks that many organizations require.

- Threaded kernel
- Parallel I/O – optimizes use of available processors
- SAS Scalable Performance Data Server™
- Parallel processing with SAS/CONNECT™
- Threaded PROCs, SAS Storage Engine, SAS Servers, DBMS SAS/ACCESS™ engines, and access to DB2
- Improved 64-bit performance

Manageability
SAS 9.1 delivers one central point of control. It provides standardized management for:

- Disseminating application updates
- Managing user access
- Administering third-party applications

Interoperability
SAS 9.1 extends support for open standards, ensuring that SAS continues to effectively operate within any existing IT environment. In addition to the Open SAS Platform provided by SAS 8.2 it provides:

- Certified as J2EE-compatible
- Full suite of Java SDKs
- IBM WebSphere MQ

SAS 9.1 Single User Performance

When compared to SAS 8.2, SAS 9.1 offers benefits of

- more concurrent users, and
- higher computational density

SAS 9.1 is optimized to take advantage of the design and features of POWER architecture. A threaded kernel, along with threaded procedures such as GLM and SORT allow SAS 9.1 to handle increased computational density and increased concurrent number of users without negatively impacting performance.

The following systems were used to generate the test data[1]:

IBM eServer pSeries 630	**IBM eServer pSeries 655**	**IBM eServer p5 570** (partition)
4 POWER4 1.0 GHz	8 POWER4 1.1 GHz	4 POWER5 1.9 GHz
8-GB DDR memory	32-GB DDR memory	15-GB DDR memory
AIX 5L Version 5.2	AIX 5L Version 5.2	AIX 5L Version 5.3

SAS® 9.1 provides benefits for single-user jobs. Figure 1shows the performance boost[2] measured when running a single user instance of Data Manipulation[3], Query & Reporting[4], and Data Mining workloads[5].

[1] The report is based on tests performed by SAS and IBM in August 2004.
[2] Except where noted, all single user tests were executed on both a 4-way IBM pSeries p630 with 1.2 GHz POWER4 processors and 8 GB of DDR memory and a 4-way SMT enabled partition of an IBM p570 with 1.9 GHz POWER5 processors and 15 GB of DDR memory. Memory requirements in single user tests never exceeded more than 1.52 GB on main memory. No swapping was observed.
[3] Single User Data Manipulation workload tests data step merges. Test comprised DATA and SORT procedures. Maximum memory requirement for the test was 505 MB.
[4] Single User Query & Reporting workload summarizes and collects frequency information on two SAS tables. The test uses the SUMMARY, FREQ and SORT procedures. Maximum memory requirement for the test was 506 MB.
[5] Single User Data Mining workload uses the DMREG, NEURAL, DMDB, FREQ, DATA step, PRINT, CIMPORT, MEANS, SORT, TRANSPOSE procedures. Maximum memory requirement for the test was 55 MB.

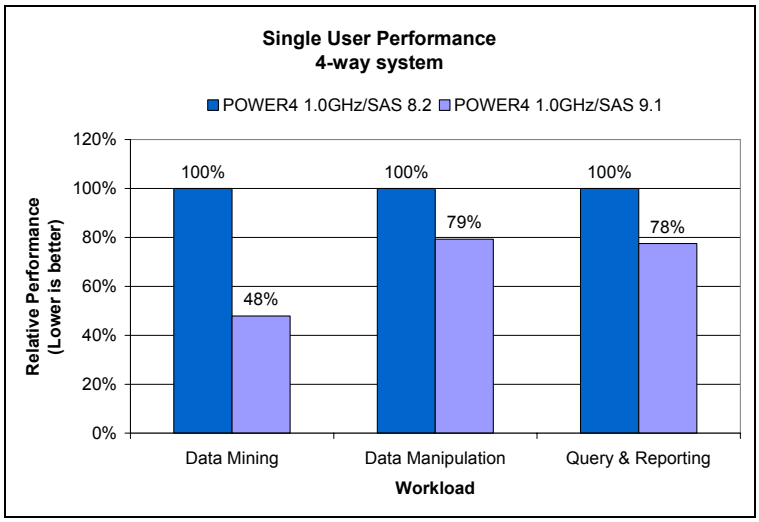

Figure 1 Single user performance of different SAS workloads

Using SAS® 9.1, a single user can take advantage of multiple CPUs to complete a job. This benefit is best realized when using procedures that are threaded internally. Scaling to more CPUs provides significantly better performance for threaded procedures, such as PROC GLM.

Figure 2 shows performance results for running the general linear model (GLM) procedure using a large dataset.

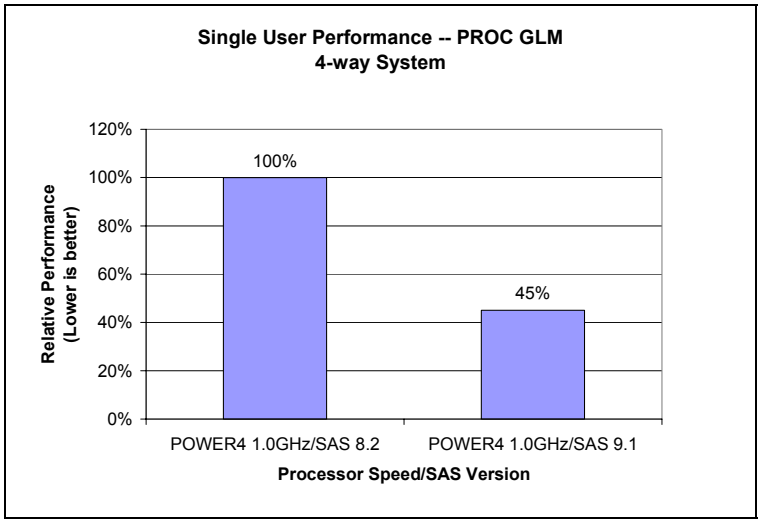

Figure 2 Single user performance of PROC GLM

To demonstrate the performance improvements, a single user test was executed. As illustrated in figure 3, the relative performance difference is dramatic: the same workload executed

on the 4-way POWER5 system in almost half the time (and half the memory) than on the 8-way POWER4 system.

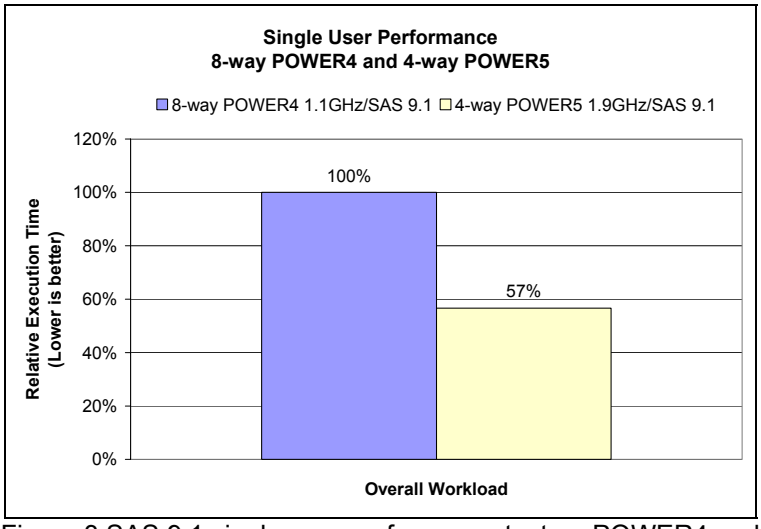

Figure 3 SAS 9.1 single user performance test on POWER4 and POWER5

SAS 9.1 Concurrent Users Performance

SAS 9.1 support for multiprocessor threading, 64-bit architecture, and large memory addressing takes advantage of the POWER based eServer large cache size and high floating point and memory bandwidth.

While important in demonstrating performance improvements, single user tests are not typically representative of customer environments. More commonly, customers run SAS jobs simultaneously. More often than not, these simultaneous jobs will show different workload characteristics. Figures 4 and 5 show some preliminary results for the POWER5 processor.

Figure 4 shows that a 4-way POWER5 server running 8 simultaneous jobs will complete the workload in approximately 65% of the time required by a 4-way POWER4 server running the same workload.

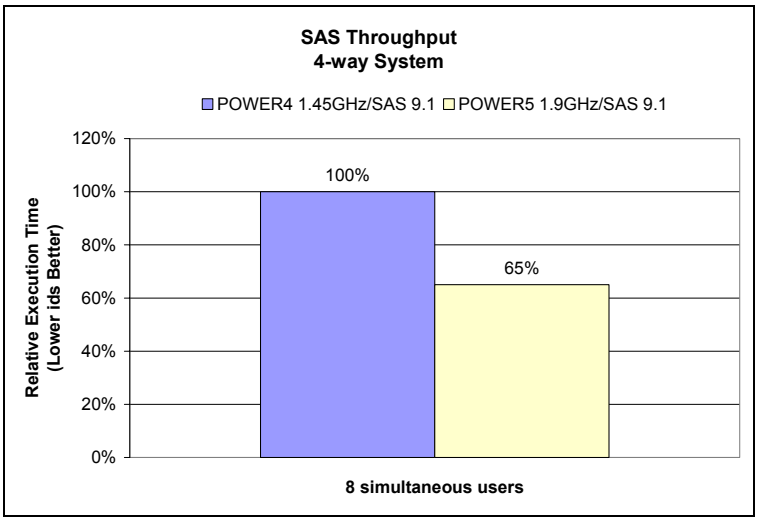

Figure 4 Relative performance of 8 simultaneous jobs on POWER4 and POWER5

Below, Figure 5 illustrates the scalability of the POWER architecture as the number of simultaneous jobs executed increases.

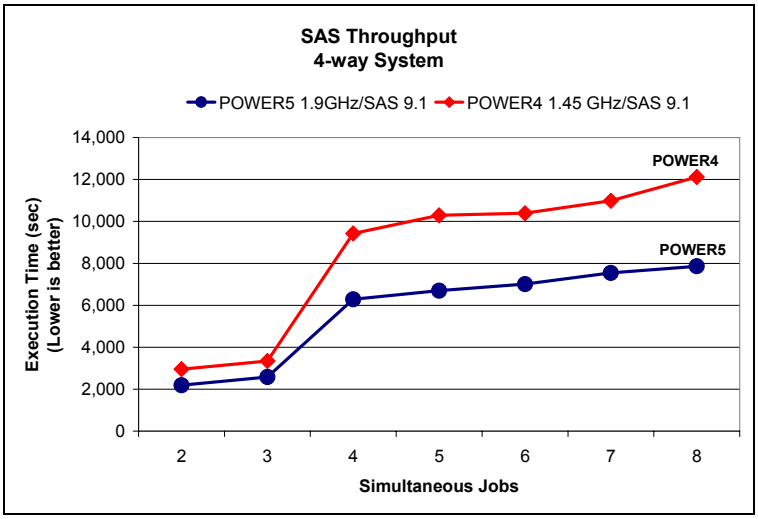

Figure 5 Scaling of SAS 9.1 on IBM POWER processors

As illustrated in Figure 6 below, single user workloads benefited greatly from the POWER5 processor's powerful memory and data throughput capabilities.

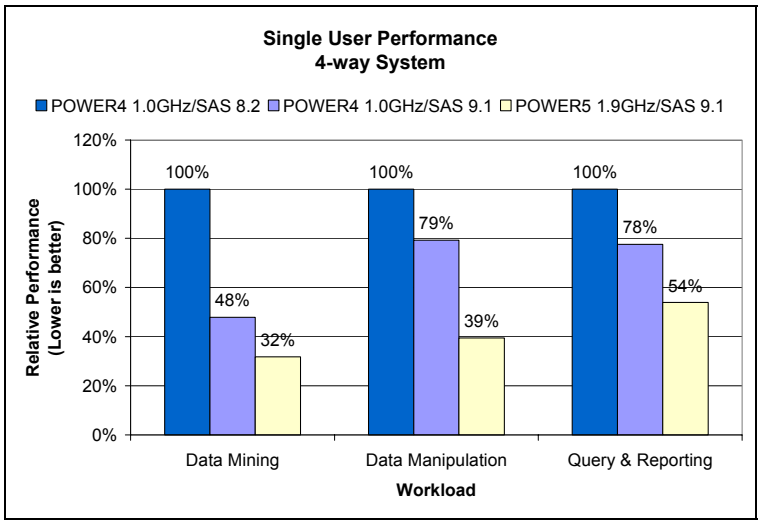

Figure 6 Workload performance of POWER5 compared to POWER4

The Data Mining workload is characterized by a test that uses three procedures found in the SAS Enterprise Miner product: NEURAL, DMREG and DMDB. The test is computationally intensive test acting on a 40,000 row table containing 297 variables. The Data Manipulation workload is characterized by a memory test that uses the SORT procedure. The Query & Reporting workload is characterized by an I/O intensive test that utilizes SUMMARY, FREQ and SORT procedures on two files of 1.1 GBytes and 2.8 GBytes.

A number of these procedures have been threaded in SAS9. The numerically intensive data mining workload shows the greatest improvement in performance when moving from SAS 8.2 on IBM POWER4 to SAS 9.1 on IBM POWER5. The I/O intensive query and reporting workload shows the smallest performance improvement. It should be noted that in an effort to keep the I/O characteristics as common as possible across all the test platforms, the localized SCSI and IBM SSA disks where used rather than a more efficient I/O subsystem.

Figure 6 clearly shows the benefits of the IBM POWER5 architecture for numerically intensive workloads. Given the appropriate I/O bandwidth, the POWER5 processor can significantly improve I/O intensive workload performance.

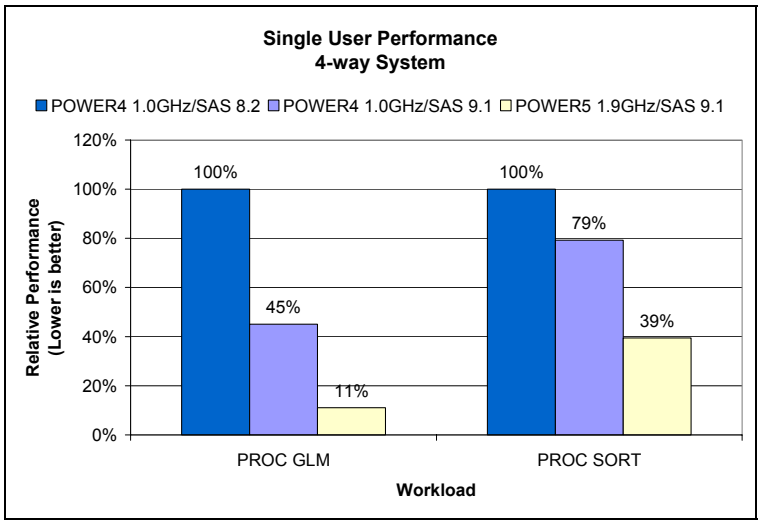

Figure 7 PROC SORT and GLM performance on POWER5 and POWER4

With a definite performance improvement observed in different workload types, it is important to understand the source of the increase in performance. Figure 7 shows that SAS 9.1 running on POWER5 processors boosts the performance of internally threaded SAS procedures. SAS 9.1 PROC GLM ran 89% faster on POWER5 than SAS 8.2 PROC GLM on POWER4 processors. SAS 9.1 PROC SORT ran 61% faster on POWER5 than SAS 8.2 PROC SORT on POWER4 processors.

The SAS 9.1 PROC SORT job on POWER5 processors completed in approximately one-third of the time it took the job to run in SAS 8.2 on POWER4 processors. This demonstrates that SAS multi-threaded procedures can clearly benefit from the POWER5 features, such as Simultaneous Multi-Threading.

IBM SMT is designed to integrate into existing SAS environments. SMT is a kernel based technology that allows two threads to execute on each POWER5 processor. SMT can be disabled if desired and no specialized coding techniques are required in the application running on AIX, so development teams can take advantage of their existing resources, tools, and skill sets to utilize this technology.

The improvements that SMT technology provides on the PROC GLM procedure are illustrated in Figure 8 below.

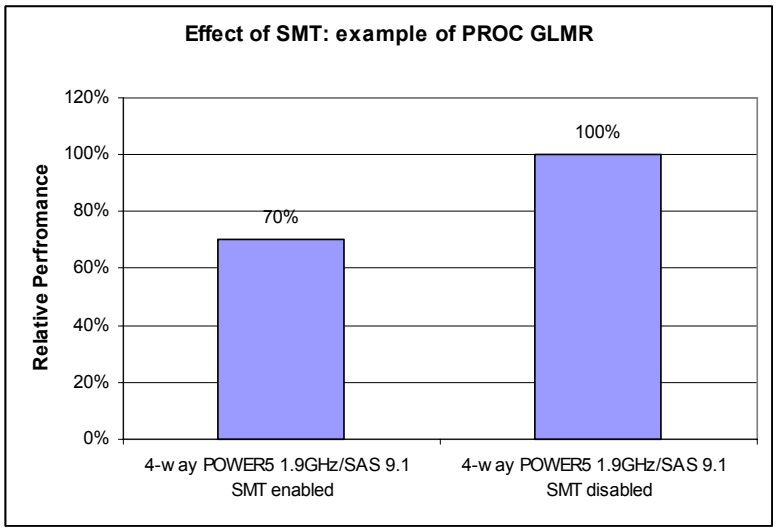

Figure 8 Effect of POWER5 SMT on the performance of PROC GLM in SAS 9.1

The POWER5 processor is designed to support both simultaneous multi-threading (SMT) and single threaded (ST) operation modes.

The SMT mode maximizes the usage of the execution units. To improve SMT performance for various workload mixes and provide robust quality of service, POWER5 provides two features:

- Dynamic resource balancing ensures that the two threads executing on the same processor flow smoothly through the system. Depending on the situation, the POWER5 processor resource balancing logic has different thread throttling mechanisms.
- Adjustable thread priority lets software determine when one thread should have a greater (or lesser) share of execution resources. POWER5 supports eight software-controlled priority levels for each thread.

However, not all applications benefit from SMT. Having threads executing on the same processor will not increase the performance of applications that consume the processor's entire memory bandwidth. For this reason, the POWER5 supports the ST execution mode. In this mode, the POWER5 processor gives all the physical resources to the active thread, allowing it to achieve higher performance than a POWER4 processor-based system at equivalent frequencies.

Conclusion

SAS 9.1 Software running on IBM AIX 5L Version 5.3 and the IBM POWER5 processor based eServer p5 systems bring together the world's leading business intelligence solution with the most affordable and fastest growing 64-bit computing environment.

For more information

To learn more about joint solutions offered by IBM and SAS, please visit www.ibm.com/servers or www.sas.com/partners/directory/ibm/index.html

IBM Corporation
Marketing Communications
Systems and Technology Group
Route 100
Somers, New York 10589

Produced in the United States

September 2004