

Unisys ES7000 Speeds Through Large U.S. Census Bureau Data Sets

Tests Show Benefits of Unisys ES7000 Servers with Intel® Itanium® 2 Processors Running Windows*-based SAS® 9.1 for ETL and Data Warehousing Applications*

Overview

The most data-intensive computing requirements — even those of the U.S. Census Bureau — can achieve outstanding performance, reduce risk, gain flexibility and lower your total cost of ownership (TCO) for SAS®9 on the Unisys ES7000 family of Enterprise Servers*. Designed to support both Intel® Xeon™ processors and Intel® Itanium® 2 processors, the Unisys ES7000 platform delivers the power and versatility to meet diverse and growing business intelligence (BI) requirements and delivers the performance of UNIX* or RISC-based platforms.

Sixty-four-bit Itanium® 2-based servers clearly demonstrated their value by supporting much larger workloads and data sets than possible on 32-bit servers. For example, due to the inherent limitations of the 32-bit Microsoft Windows* operating system, a 16 processor server could compute up to only 6 million data records, but the 8 processor Itanium 2-based system could easily handle 120 million records. (See Figure 1). The 64-bit Itanium 2-based Unisys ES7000 system easily calculated 120 million records of U.S. Census Bureau data, yielding the outstanding scalability and performance required for running SAS 9.1 applications.

CENSUS BENCHMARK BACKGROUND

The U.S. Census Bureau runs more than 100 tests annually on a wide variety of computing platforms, including mainframes, UNIX servers and Windows clients and servers. U.S. Census Bureau software developers created their own benchmark, "bench.sas," to provide vendors with benchmark tests and scripts to independently run hardware performance tests running SAS 9.1.

The benchmark program is a resource-intensive SAS program that consists of 12 individual SAS processing steps to provide a broad spectrum of operations. It creates its own records and measures the time to complete each processing run. For these tests, the benchmark program took advantage of the new features in SAS 9.1.

UNISYS BENCHMARK PROCESS

The Unisys Center of Excellence (CoE) performed the tests in their lab at SAS.

The Unisys Test Environment

Unisys performed its tests on a 16-way ES7000 built on Intel Xeon processors and on an 8-way ES7000 built on Itanium 2 processors. The objective was to compare how the U.S. Census Bureau benchmark tests ran SAS on the two different processing environments.

As Figure 1 shows, the complete range of tests could compute up to only 6 million records on an Intel Xeon processor-based 32-bit system and required a 64-bit ES7000, Windows Server 2003 system to manage an increasing number of records.

THE HARDWARE

The two servers used for the Unisys test bench comprised:

- Unisys ES7000, Model 520 (32-bit system)
 - 16x 2.8 Ghz Intel Xeon processor with 2MB integrated Level 3 cache
 - 32GB Memory
 - Two EMC Clariion 4700FC disk subsystems with 80 x 18GB 10K RPM Disk
 - Eight Emulex* LP9802 HBAs at 1 Gbps
- Unisys ES7000, Model 420 (64-bit system)
 - 64GB Memory
 - 8 x 1.5 GHz Intel Itanium 2 processors with 6MB integrated level 3 cache
 - Two EMC* Clariion CX600 Disk Subsystems with 120 x 36GB 15K RPM Disks
 - Four Emulex* LP9802 HBAs at 2 Gbps

Figure 6 on page 5 shows the ES7000 system and disk layout for the U.S. Census Bureau data tests.

THE SOFTWARE

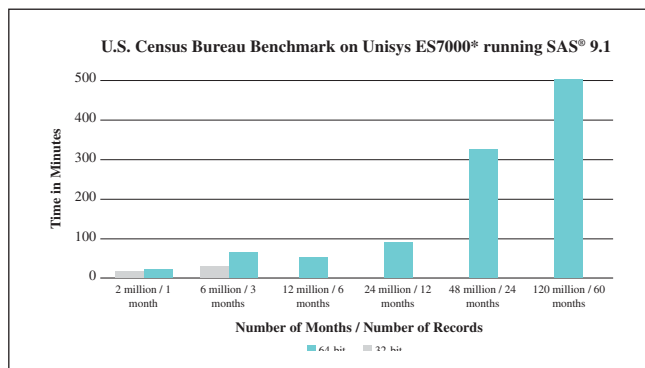
- For 32-bit systems
 - Windows Server 2003 for 32-bit Xeon-based systems, Datacenter Edition
 - Emulex Drivers for LP9802
 - SAS 9.1 for Windows [Early Adopter Edition]
 - Classic SAS components and SAS OLAP Server
- For 64-bit systems
 - Windows Server 2003 for 64-bit Itanium-based systems, Datacenter Edition
 - Emulex drivers for LP9802 for 64-bit
 - SAS 9.1 for Windows 64-bit [Early Adopter Edition]
 - Classic SAS components and SAS OLAP Server

The ES7000 Results

Unisys conducted the benchmark testing during December 2003 through January 2004 and measured single-user processing time for different record sizes and numbers of records. Testing for each system was conducted separately and results were compiled. The Itanium 2-based Unisys ES7000 server ran a 24 million record workload in just over 1.5 hours, and it processed a 120 million record workload in under 35 hours. Figure 1 provides details.

Figure 1 also shows that the Intel® Xeon™ processor-based Unisys ES7000 server was unable to successfully complete tests requiring more than 6 million records (three month workload). After 6 million records, SAS PROC MDDDB operations exceeded the 4 gigabyte memory address space limitation in 32-bit computing environments. Itanium 2-based Unisys ES7000 servers deliver enterprise-class 64-bit computing power and are able to successfully process the larger tests. Figure 2 summarizes each process step by workload.

Completion times for the Unisys ES7000 servers* configured with Intel® Itanium® 2 processors compared to an Intel® Xeon™ processor-based 32-bit system.



Months	Number of Records	Elapsed Times (hh:mm:ss)	
		32-bit	64-bit
1	2,000,000	00:13:01.1	00:06:56.1
3	6,000,000	01:00:07.4	00:21:30.0
6	12,000,000	-----	00:44:35.6
12	24,000,000	-----	01:35:11.3
24	48,000,000	-----	05:26:17.1

Figure 1. The 32-bit system exceeded capacity after three months of data (6 million observations) but the Unisys ES7000 performance on SAS 9.1 ranges from less than seven minutes for one month of data (2 million observations) to 60 months of data (120 million observations) running in less than 35 hours.

PERFORMANCE METRIC BY PROCESS STEP UNDER DIFFERENT DATA WORKLOADS

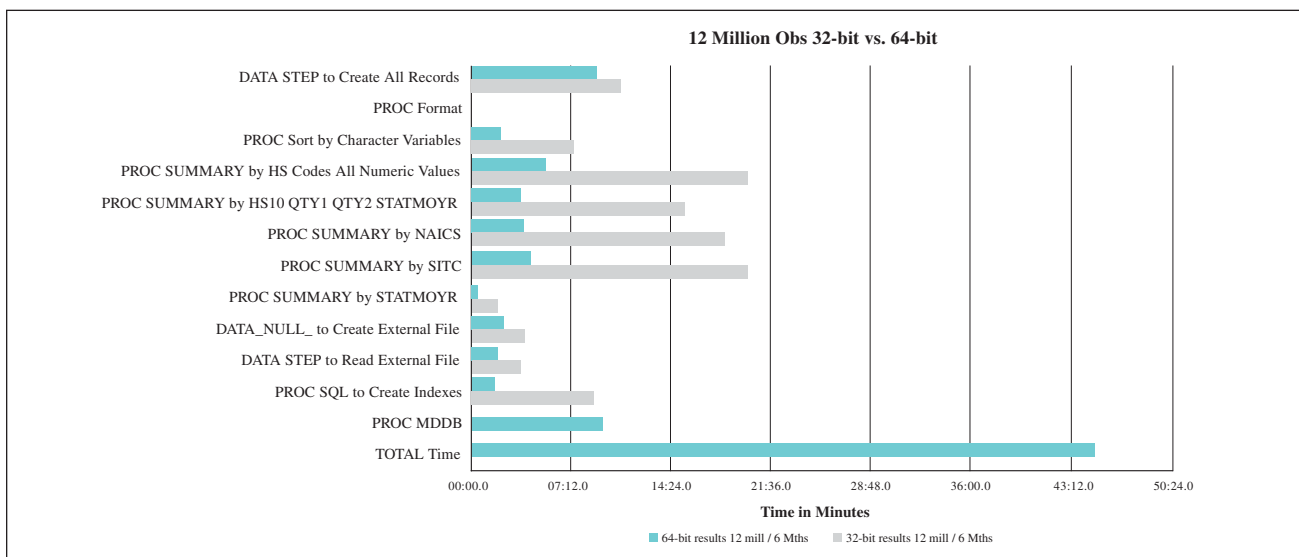
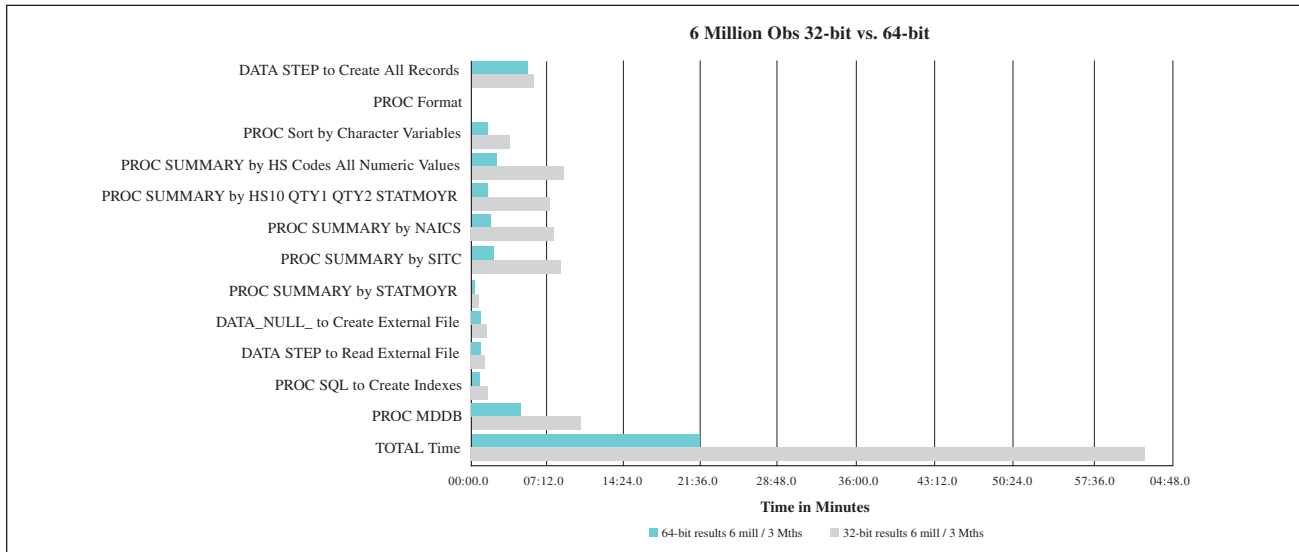
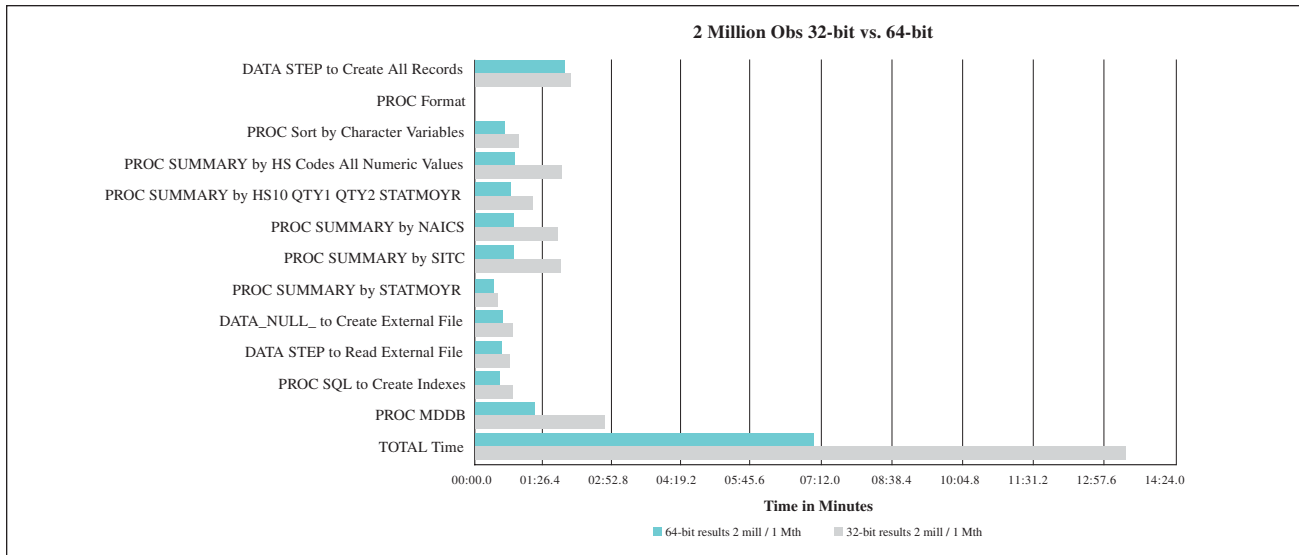


Figure 2. Comparing the results of a 32-bit and 64-bit system on U.S. Census Bureau data shows that the Unisys ES7000 64-bit system outperformed the 32-bit system in all test operations. The 32-bit system exceeded its memory limitation after 6 million records (three months of data). The final bar chart in this series shows the process step (PROC MDDB) at 12 million records (six months of data) that the 32-bit system was unable to complete.

The majority of processing time was consumed by the “DATA CREATE” and varied “PROC SUMMARY” test procedures. The performance of the 8-way Itanium 2-based ES7000 is summarized for all data loads (up to 120,000,000) for these specific functions in Figure 1 on page 2.

The “DATA CREATE,” “PROC SORT” and “PROC SUMMARY” procedures demonstrated particularly high memory usage requirements. These SAS “PROCS” have been optimized under 64-bit to take advantage of more than 4GB of memory. The 64-bit ES7000 environment enabled all operations to run in memory, resulting in superior performance of approximately one-third of the time of the 32-bit system to perform the benchmark tests on 6 million records.

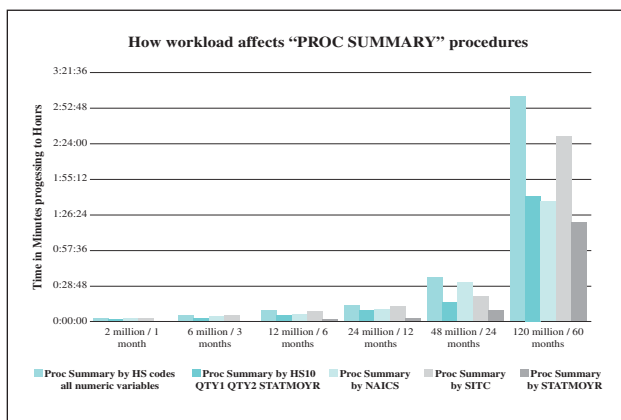


Figure 3. How workload affects “PROC SUMMARY” procedures.

Figure 3 demonstrates how the five “PROC SUMMARY” procedures are impacted by size of workload.

Unlike the 32-bit system, the 64-bit ES7000 was able to complete a 10-fold increase in data workload to 60 months of data/120 million records.

Figures 4 and 5 illustrate the ES7000 64-bit capabilities through the resource-intensive steps (“DATA CREATE,” “WRITE FILE,” “READ EXTERNAL FILE”) of the benchmark process at various workloads.

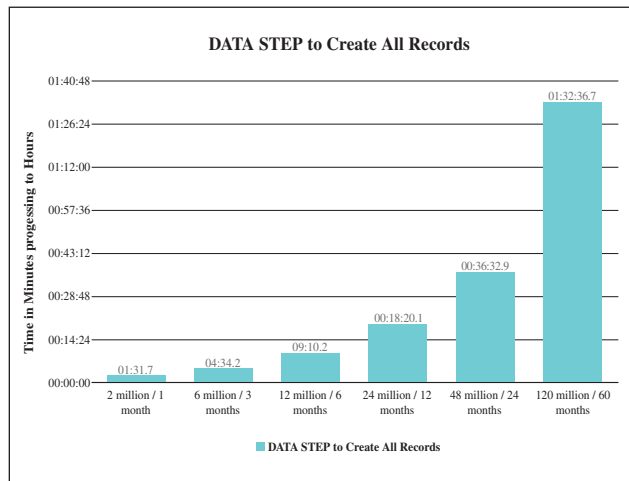


Figure 4. The time it takes using “DATA CREATE” to build the initial large text file.

Figure 4 charts the time required to build the initial large text file using “DATA CREATE.” This operation, as well as the steps to create and read the external file, are I/O bound and the processing time is directly proportional to the size of the data workload and the capabilities of the disk subsystem attached to the ES7000.

Unisys’ extensive experience in implementing high-end Windows-based enterprise applications, the expandability of the ES7000 and boosting its memory to 64GB provided a significant enhancement in benchmark performance.

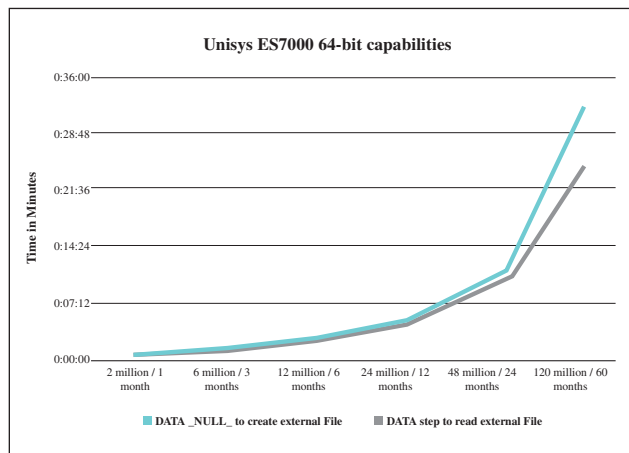


Figure 5. As the number of observations increase, external disk access times slow.

RECOMMENDATIONS

The ES7000 64-bit configuration offers high business value through its price/performance, extensibility, simplicity and manageability for BI applications. In addition, it can provide superior performance, headroom and throughput for large SAS implementations, like that of the U.S. Census Bureau, where analysis of large datasets, multiple instances of SAS, large user bases and threaded applications are required.

To achieve optimal results when computing large data sets, organizations should consider the following:

- The testing showed that SAS BI applications are memory-intensive and as the data rate increases so does the memory required; therefore allocation of a minimum of 32GB of memory is recommended for the application to assure fast processing times. During this testing, Unisys used 64GB of memory to assure sufficient memory for testing.
- BI applications, like those run on SAS 9.1, require using a high bandwidth and scalable I/O architecture attached to a robust storage architecture, such as the EMC CX600 Disk Subsystems, to assure optimal I/O efficiency.

- The 64-bit address space in the Intel Itanium 2 processor made it possible to handle large data loads. Because BI applications usage tend to grow over time, an expandable hardware platform, such as the ES7000, is important if you want to adapt it to increasing workloads and keep the TCO low.
- The U.S. Census Bureau benchmarks, by design, run in a single user environment and test processing requirements of high data-volume jobs that are I/O and memory-intensive. Considering the number of simultaneous users and jobs is important. In real-world applications, concurrent users and jobs have a significant impact on instruction response times and overall systems performance and must be considered. Multi-user/multi-job environments require an easily upgraded and highly expandable system like the ES7000.

Whenever larger data volumes or increased analytical computational complexities are required, the Itanium 2-based Unisys ES7000 family of servers running Microsoft Windows Server 2003 for 64-bit Itanium-based servers allows you to benefit immediately from new SAS[®]9 features that improve your BI solutions.

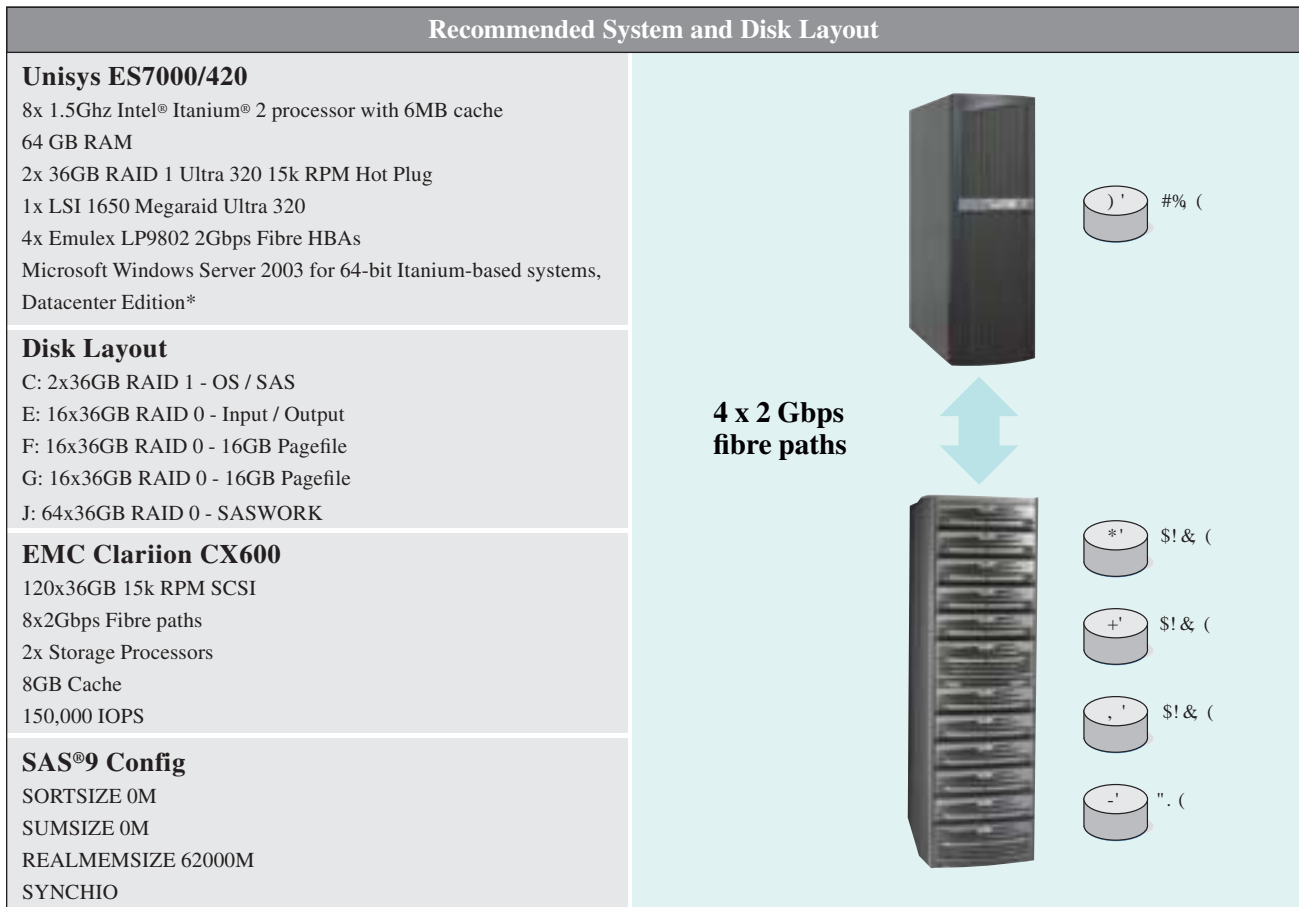


Figure 6. Unisys ES7000* SAS[®] 9.1 bench mark systems hardware configuration and disk partitioning.

UNISYS CONTACT INFORMATION

For more information about SAS®9 on the ES7000 family of servers, contact:

Carey Barnes

carey.barnes@unisys.com

Global BI Marketing Manager, Unisys Corp.

SAS CONTACT INFORMATION

For more information on migrating from SAS 8.2 to SAS®9, contact your SAS Sales Representative, or one of the SAS/Intel team members:

Mark Pallone

mark.pallone@intel.com

SAS Strategic Relationship Manager

Jim Watts

jim.watts@sas.com

Intel Global Alliance Manager

Bill Driest

bill.driest@sas.com

Unisys Global Alliance Manager

SOLUTION PROVIDED BY:



Copyright © 2004 Intel Corporation All rights reserved. Intel, the Intel logo, Itanium and Intel Xeon are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries.

Unisys is a registered trademark of Unisys Corporation. Specifications are subject to change without notice.

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

0704/BTB/ITF/XX/PDF

302743-001EN