



white paper

Open Software in the Enterprise

Unisys Server Technology

performance

For more than a decade the enterprise market has been dominated by proprietary UNIX and RISC based platforms. In 2000, Unisys created large, enterprise-class Intel and Windows based SMP systems with both 32-bit and 64-bit processors — the Unisys ES7000 Family of Servers. Unisys is now bringing true enterprise-class platform support and the unique advantages of ES7000 servers to Linux operating environments based on the Linux v2.6 kernel.

Table of Contents

Executive Summary	4
Designed for Flexibility	5
Unisys ES7000 Servers	5
System Design and Architecture	5
Benchmarks	5
Enterprise Capable	6
The Linux 2.6 Kernel	6
System Scaling	6
CPU Scheduling	6
Kernel Preemption	7
Improved Threading Model	7
Virtual Memory	7
Other Improvements	8
Summary	9

Executive Summary

Long a price-performance leader in high-end, Intel based systems, Unisys is expanding on its commitment to bring open software to the enterprise. Unisys has been working closely with the Linux community to take advantage of the Linux 2.6 kernel features and integrate them with ES7000 server hardware. Building on our expertise in Windows based solutions, we now employ the same design and optimization techniques of our mainframes to bring “industrial-strength” computing to open Linux environments.

The mainframe-class Unisys ES7000 Family of Servers with up to 32 Intel processors and beyond is the culmination of more than 50 years of Unisys mainframe hardware and software development. Our engineering and design expertise brings leading-edge advances in reliability, availability, systems manageability and performance to mission-critical, Intel based, data center computing while significantly lowering the total cost of ownership.

Though some proprietary solutions vendors may downplay the use of large, enterprise-class SMP servers for Linux enterprise solutions, the proof points are undeniable. Linux scales well to 128 logical processors. And, with the overall robustness and rich feature set of the v2.6 kernel, Linux is challenging proprietary UNIX as one of the operating systems standards of choice for true enterprise-class computing.

Designed For Flexibility

Developing a true enterprise-class open server is not as easy as simply loading a feature-rich operating system onto a multiprocessor platform. The difficulty in doing so is a key reason why high-end proprietary systems have dominated the market for so long. An open system with enterprise-class characteristics must be engineered and developed by expert hardware and software teams that work together to design, integrate and optimize a complete hardware/software environment.

Unisys ES7000 Servers

In the design of ES7000 systems, Unisys uses this same integrated approach for server design and development, but with open software and hardware. The hardware components of ES7000 servers are designed with a high level of reliability and flexibility and the ability to be programmed in such a way that its software and hardware components can be successfully integrated to yield a rich set of features and a level of system performance that rivals and can exceed that of proprietary systems.

System Design and Architecture

ES7000 servers are based on the Unisys Cellular MultiProcessing (CMP) architecture we developed in the late '90s. The CMP system architecture uses 4-processor cells as building blocks for creating SMP configurations of up to 32 processors in a single domain. The processors in a given domain may be either single or dual core, and each core in a processor may also have hyperthreading capability. Thus, a 32 processor system with dual core and hyperthreading features appears to the operating system as a 128 logical processor system. This architecture is based on the same features and techniques designed and developed for Unisys mainframe systems.

One important element of the CMP design is the flexibility and programmability of the hardware that controls vital elements that are even deep into the system, such as performance hardware monitors, queue depths and interface protocols. Many other advanced features and elements are built into Unisys CMP-based hardware, such as those that enable mainframe-style management through Unisys Sentinel management software that is supplied with each ES7000 system.

Benchmarks

Unisys has entered the Linux market with leading price and price-performance among general-purpose servers. Based on the industry-standard SPECfp benchmarks, a Unisys ES7000 server with Intel Itanium 2 processors produced the fastest result of any 16-way server running a standard distribution Linux operating system, and the best price-performance of any Linux server with eight or more processors in this demanding measure for scientific computing.

And, because of their architectural flexibility, ES7000 servers have consistently captured 16-way to 32-way TPC-C benchmark performance records. With the new scalability features in the Linux 2.6 kernel, the same levels of performance are being achieved by the Unisys ES7000 performance team on standard distributions of Linux for both 32-bit and 64-bit computing.

Enterprise Capable

The Linux 2.6 Kernel

Unisys has added “CMP-aware” enhancements and other modifications to Linux that deliver performance improvements for typical ES7000 workloads. As a member of the open source community, Unisys proposed these enhancements, which were accepted into the kernel’s development in time for inclusion with the Linux 2.6 kernel.

The 2.6 kernel brings Linux into the world of enterprise-class computing on large SMP servers. Most notably for systems like our high-end ES7000 servers, new features target performance improvements, scalability, throughput and support for SMP machines. Discussed below are a few of the major feature enhancements in the new kernel that are perfectly matched with the capabilities and technologies of Unisys ES7000 servers.

System Scaling

The ability to scale a system is much improved with the Linux 2.6 kernel. “Scaling” refers to the linear speed-up in processing that is expected with a large number of processors in an SMP server. The bottlenecks that introduce serial constraints into the otherwise parallel processing stream of a large CPU configuration are usually insignificant in simple two-processor servers. But they are increasingly problematic in large SMP structures. The most common bottlenecks involve contention among processors for access to shared resources such as the kernel or cache lines.

Much of the work required to scale Linux up to 128 processors has involved improvements to reduce or eliminate lock contention, a state in which CPUs sit idle waiting for resources that are locked by another CPU. Many of the changes required to improve CPU scaling were available from the Linux community and used by Unisys with minor enhancements. Early Linux multiprocessor releases relied on a single lock (the Big Kernel Lock (BKL)) as the primary synchronization and serialization lock for the kernel. Work from open-source projects that have addressed the BLK bottleneck, as well as work done at Unisys, have been included in the 2.6 kernel. This has enabled a dramatic increase in performance scaling for large SMP configurations.

CPU Scheduling

The CPU scheduler in earlier Linux kernels was simple and efficient for small multiprocessor platforms, but it was inefficient for larger SMP structures with a large number of threads. The consequent bottleneck for scheduling was caused by the use of a single, heavily contended lock to protect the global runqueue. The simple scheduler has been replaced in the 2.6 kernel with the O(1) scheduler, which partitions the single global runqueue of the standard scheduler into multiple runqueues.

Moreover, idle processors start executing waiting processes (if their own processor is busy) whose time slices haven't yet been exhausted, causing processes to start "bouncing" between processors. When a bouncing process happens to be a high priority or interactive process, overall system performance is affected.

The main benefits of the new scheduler include:

- SMP efficiency — If there is work to be done, all the processors can do the work
- Waiting processes — No process will stay without processor time for long periods of time. And no process will take an unreasonably high amount of CPU time
- SMP affinity — Processors that affine to one CPU will not bounce between CPUs
- Priorities — Less important tasks can start with lower priority and more important tasks can start with higher priority
- Load balancing — The scheduler decreases the priority of any process that generates more load than the processor can handle

Kernel Preemption

In previous versions of the Linux kernel it was not possible to preempt a task executing in kernel mode (including user tasks that had entered into kernel mode via system calls) until or unless the task voluntarily relinquished the CPU. With kernel 2.6, the kernel is preemptable, allowing the system manager a finer-grain control over system tasks. A kernel task can be preempted so that some important user applications can continue to run. Of course, not all sections of the kernel code can be preempted. Certain critical sections of the kernel code are locked against preemption, which will ensure both that per-CPU data structures and state are always protected against preemption.

Improved Threading Model

Significant work in improving threading performance has gone into the 2.6 kernel. Based on a 1:1 threading model (one kernel thread for one user thread), it includes in-kernel support for the new Native Posix Threading Library (NPTL). Threading operations will enjoy an increase in speed, and the 2.6 kernel can now handle both an arbitrary number of threads and up to 2 billion process IDs (PIDs) (on 32-bit Intel processors).

Virtual Memory

Reverse mapping for virtual memory (VM) has been incorporated in the 2.6 kernel, which is designed to accomplish significant improvements in VM behavior under certain loads. In previous kernels, virtual mapping had sub-optimal performance when the kernel wanted to free a particular physical page. Since virtual-to-physical page mapping is not always one-to-one, multiple virtual pages (pages shared by different processes) might point to the same physical page. To free a particular physical page, the kernel had to scan through all of the processes' page tables to determine if that physical page was being referenced. The new reverse mapping technique maintains this information and improves VM performance, especially on high-end systems with heavy loads.

Other Improvements

Other key improvements are included in the 2.6 kernel that increase scalability by removing previous limitations and increasing capacities and capabilities:

- Support for Intel's Hyper Threading and PAE (Physical Address Extension) features improve performance by allowing 32-bit Intel systems to access up to 64 GB of RAM in a paged mode.
- File systems can now support up to 16 TB of file storage compared to the previous limit of 2 TB.
- The number of unique users and groups on a Linux system has been increased from 65,000 to over 4 billion (16 bits to 32 bits).
- The number of PIDs, before wraparound, has been increased from 32,000 to 1 billion, improving application starting performance on very busy or very long-lived systems.
- The different types of devices now supported has been increased from 255 to 4095, with each type supporting more than a million sub-devices.
- One final significant change in the kernel is the creation of a unified device model. The device model represents the overall device architecture and the overlay of the system by maintaining a number of data structures allowing easier management of device-related tasks.

Summary

Previously supporting UnitedLinux, the ES7000 platform has also run both SUSE Linux Enterprise Server and Red Hat Enterprise Linux since 2003. Unisys will continue to work with Novell SUSE and Red Hat to provide services and support for the most demanding Linux-based business solutions, such as those already in use by Unisys ES7000 customers around the world.

The Pennsylvania State University's Information Technology Services' Graduate Education and Research Service (GEaRS) group is using Linux-based ES7000 servers in its high-performance computing environment for tasks including computational biology, chemistry, fluid dynamics, gravitational physics, mathematics, materials science, structural mechanics and for representing data as images. "We believe there are important price-performance benefits that will make scale-up computing on the Intel Itanium 2 processors such as those offered by the Unisys ES7000/430, the dominant 64-bit architecture over the next few years," said Vijay K. Agarwala, Director, High Performance Computing and Visualization, Penn State Information Technology Services. "Our ES7000 will be particularly useful for researchers who need more than four gigabytes of memory per process and those working with large database applications." Find out more about Penn State's GEaRS IT group at:

<http://gears.aset.psu.edu>

InteliTrac, a provider of high-speed database and security solutions, selected Unisys ES7000 servers running Linux to power the data centers of its IdentiPort product –the world's first high-speed, real-time, portable biometrics security system. At the Santiago Airport in Chile, IdentiPort now provides real-time document authentication, fingerprint verification and face recognition to the Policía Investigaciones de Chile, the Chilean equivalent of the FBI. Their officers can now conduct one-to-many face identification and background checks for potential criminals to verify documents and fingerprints all in less than 45 seconds — data processing speed once reserved for supercomputers. You can learn more about InteliTrac at: <http://www.intelitrac.com>

Find out more about our Linux-based ES7000 users on the Unisys website at:

<http://www.unisys.com/es7/guide>

With the expanded Unisys commitment to bringing open software solutions and their advantages to the high-end, mission-critical space, enterprise-class Linux is available on the entire line of Unisys servers — ranging from systems running both 32-bit and 64-bit processors, from 4-way to 32-way. Unisys also offers a wide range of services and solution sets, designed for large enterprises with complex computing environments. With many thousands of customer service engineers in over 100 countries, Unisys will be the single point of contact for implementation and support of its customers' Linux software and hardware solutions.

Unencumbered by any legacy proprietary UNIX operating system, Unisys has applied its expertise in building large, enterprise-class SMP servers to the Linux operating environment, giving IT professionals the freedom to choose between open and proprietary systems for their data centers.

For more information, contact your Unisys representative.

Or call:

1-800-874-8647, ext. 365 (U.S. and Canada)

00-1-585-487-2430, ext. 365 (Other countries)

In a hurry to learn more? Visit:

<http://www.unisys.com/es7/guide>

For even more details, visit:

<http://www.unisys.com/es7/guide/community>

The information contained herein is subject to change without notice.

© 2006 Unisys Corporation. All rights reserved.

Unisys is a registered trademark of Unisys Corporation. Intel and Itanium are registered trademarks of Intel Corporation. Linux is a registered trademark of Linus Torvalds. SUSE is a registered trademark of Novell, Inc. Red Hat is a registered trademark of Red Hat, Inc. All other brands and products referenced herein are acknowledged to be trademarks or registered trademarks of their respective holders.

01/06



4126 4524-100