White Paper

Origin® to Altix® Migration

November, 2004

Table of Contents

1.0 Introduction

For Silicon Graphics, Inc. most recent history, we have had a very vertical orientation. We internally developed the systems, the chips, the operating system, development tools and other software products sometimes taking advantage of open, industry standard technologies. We believed that this was the optimal way to provide the best server products to our customers.

However, beginning in the late 1990's, we began exploration into a new paradigm–focusing on areas of key strengths such as systems design, while looking to partners and industry standards to provide functions like operating systems and microprocessors. In the Linux® operating system and in the Intel® Itanium® microprocessor line we believe we have best-of-breed, standards-oriented alternatives. These are intregral components of the SGI® Altix® family of servers that first shipped to customers in early 2003.

During its first year, Altix won many awards including Best of Show at LinuxWorld 2003 in New York City. As of September 2004, SGI has shipped over 29,000 Itanium 2 processors to more than 300 different customers worldwide.

Despite its many awards and accolades, the Altix server that shipped in early 2003 did not have all the benefits that a decade and a half of intensive research and development effort gave our Origin/IRIX products. It was also limited in features by what was available in the Linux kernel at the time. However, during the past eighteen months via internal engineering work, the work of partners, and the Linux community, Altix is very rapidly approaching the capabilities of its older IRIX® OS-based brethen, while providing significantly better price performance and access to a much broader application base.

2.0 Applications

In order to enable customers to be successful with Altix, a variety of key applications must be available. As of September 2004, 150 applications are ported to Altix, of which 75 are optimized. These include the major applications in use today in the areas of manufacturing, life sciences, energy, research, and government.

Building on our trend toward industry-standard computing, SGI will pursue a standard code-base strategy with the Novell/ SUSE Linux code base, as opposed to continuing to build and support its own Linux base, known as SGI Advanced Linux™ Environment. In the first half of 2005, SGI will move to the new code base structure, with SGI ProPack™ providing an overlay on top of SUSE Linux Enterprise Server. SGI Advanced Linux will con-

tinue to ship and be available for customers throughout calendar year 2005, and will be supported for many years thereafter.

By adopting a standard code base, SGI will further enrich the number of applications that will work out of the box on the Altix machine and in many cases enable support directly from the software vendor to the customer.

In addition to this development, SGI will be providing certified Red Hat® Linux 4.0 at the beginning of 2005 for Altix® 350 machines of 32 processors or fewer. The availability of certified Red Hat will further enrich the application offering.

3.0 Device Drivers

A number of specialized solutions require serial data, reflective memory, HIPPI, and GSN connectivity. Currently both HIPPI and GSN cards have been ported to Altix by Video Propulsion, Inc. (www.videopropulsion.com). These cards are 3.3V cards and will fit easily into any Altix PCIX slot.

In addition, Altix drivers are available for Curtiss-Wright (was Systran) SCRAMNet® reflective memory cards and the FibreXtreme® SL240 serial data card. Reflective shared memory is typically used to 'share' ("reflect") memory between different computer architectures and is widely used in the government space. The serial data card is typically used for serial data input or output from devices such as satellite downlinks and radar equipment.

VME connectivity has been used extensively in applications based on Origin technology in the past. SGI is working with vendors to provide VME connectivity via a PCIX adapter and fiberoptics connection to a VME cage. The plan is to duplicate the vme*() interfaces that customers have used on IRIX to enable smooth porting efforts between code bases. Availability of a VME interface is scheduled for late summer of 2005.

In addition to porting older devices, there has been a lot of work in bringing over new devices as well such as 10GigE available from S2io and Infiniband available from Voltaire. SGI is currently evaluating 10 GigE solutions with TCP Offload (TOE) and RDMA.

4.0 I/O Features and Performance

Altix set early records of 7 GB/s for I/O on Linux. However, for some of our customers' most demanding applications this bandwidth is insufficient for their requirements. Adopting the 2.6 kernel provides substantial improvements in the I/O stack that will achieve and in many cases exceed performance on

Origin today and set new records for I/O performance on Linux. Key I/O features such as direct I/O which allows I/O with no buffering and asynchronous I/O ("async I/O"), I/O with no waiting for completion, are also available on Altix today. SGI currently supports capabilities for I/O configuration tuning and failover—details are available on the ioconfig and failover man pages.

5.0 Security
IRIX and Trusted IRIX™ provide both Common Criteria (CC), Controlled Access Protection Profile (CAPP) and Labeled Security Protection Profile (LSPP) certifications at the EAL3 level. SGI's current Advanced Linux Environment, with auditing is eligible for CAPP certification but, has not gone through the certification process. While SGI Advanced Linux Environment provides SGI with the strongest Linux operating system today for high performance/productivity customers, SGI is working toward providing a standard distribution in collaboration with Novell.

Therefore, SGI has decided not to certify its SGI Advanced Linux, but instead work toward certifying SUSE Linux Enterprise Server on the SGI Altix systems. Work is about to begin on this effort with a scheduled completion by September 2005. SGI ProPack is expected to be certified along with SUSE Linux Enterprise Server.

To reach LSPP certification level, Mandatory Access Control (MAC) features need to be added to the Linux kernel. A number of vendors are working on adding the code necessary to meet the LSPP requirement and it is expected that LSPP will be achievable by end of 2005.

6.0 Programming Tools and Environment
Intel provides the Intel® C/C++ compiler, Intel® Fortran compiler, performance analysis tools and libraries for Itanium® systems. When the compiler and development tools first were introduced they were relatively new and untested as compared to the IRIX development tools. Over time the compilers and tools have improved substantially through the collaboration of Intel, SGI, customers and partners.

Intel has also made significant investments in the areas of tool development, including the purchase of Kuck & Associates and Pallas for Trace Collector and Trace Analyzer tools (formerly Vampir and Vampirtrace).

Other state-of-the-art tools available today include Etnus TotalView®, which is a high end parallel debugging tool.

As part of a Department of Energy effort, SGI engineers are working on porting SGI® SpeedShop™, the performance analysis tool well known by IRIX users as part of the MIPSpro™ Developer's Workshop set of development tools.

7.0 Operating System Features
Operating systems for high performance and productivity users must be able to deliver performance advantages in compute-intensive environments. SGI has enabled Altix system users to achieve such performance enhancements with features such as dplace and cpusets for memory placement and management and libraries such as Message-passing Toolkit(MPT) and SGI Computational and Scientific Library (SCSL).

8.0 System Administration Tools and Environment
Many of the best-loved administrative tools for IRIX have been ported to SGI's ProPack. Examples are gr_osview (1), osview (1), gr_top(1), etc. SGI has also ported more extensive tools for monitoring and troubleshooting such as Performance Co-Pilot™ and Embedded Support Partner (ESP).

Checkpoint-Restart is a tool used in the IRIX environment to periodically 'save' the state of a running process so that if a system crash occured, a long run would not have begin from scratch, but could begin at the last checkpoint. Meiosys has a checkpoint-restart solution with similar capabilities which is now available on the Altix platform.

9.0 Real-Time Programming
SGI brought hard real-time (deterministic) capabilities to the multiprocessing UNIX® environment. Traditionally, hard real-time support was focused on small appliances and devices in the embedded industry. SGI enabled hard real-time performance to scale from small to large computer systems. Many customers took advantage of hard real-time to drive solutions in the media and government and defense spaces. Because hard real-time performance has been so key to important solutions based on SGI technology, early work was done to ensure that similar capabilities would exist in an Altix environment. Fortunately, SGI's work has been aided by the work of the Linux community, including the contributions of Monta Vista Software.

With the introduction of the 2.6 kernel for Linux, a key feature for real-time processing, kernel pre-emption, or the ability to stop a running kernel to service an urgent interrupt, has been introduced. Without this feature, most real-time processing would not be possible.

Early tests on an Altix system with a Beta 2.6 Linux kernel showed that the average latency could be guaranteed at 30 usec for a 4 processor or larger configured Altix system (as compared to the guaranteed latency for Origin systems at 50 usec). The result of the initial investigation was that in some cases the latency could not be guaranteed at 30 usecs and in some cases the latency was proven to be faster than 30 usecs, but consistency could not be maintained 100% of the time. Because hard real-time is defined as maintaining consistent latency times, it was important to implement additional functionality to achieve hard real-time. From SGI's IRIX experience, it was found that processor restriction would help attain the level of hard real-time performance desired.

Foundation real-time features are available today in Beta form from SGI and will become more widely available with the First Customer Ship(FCS) of the Linux 2.6 kernel on SGI platforms. These features include high resolution clocks with support for 40 nanoseconds per tick and the ability to map a real-time clock into memory.

In addition to foundation features designed to achieve guaranteed latency, work was needed to enable more advanced real-time features that more closely match what is available in IRIX today. These features include User Level Interrupts (ULI's), the Frame Rate Scheduler (FRS), a measurement and monitoring tool similar to IRIXview™, and External Interrupt capabilities. ULI's and external interrupt are currently being worked on and will be available mid 2005. The FRS and a measurement and monitoring tool will be available late 2005.

10.0 Documentation and Other Help
In addition to this white paper, more extensive whitepapers are available on Security, Real-time and production-quality features of Linux. A complete guide was written as a programmer's guide to transition which is available via the SGI Technical Publications website (www.techpubs.sgi.com).

sgi®