White Paper

A Multi-Workflow Architecture for High-Performance Computing

Table of Contents

1.0 The Gap: Advancing Technology and Diverging User Requirements

Rapid advancements in technology continue to broaden the range of choices in the high-performance computing market:

- Clusters, blade servers, symmetric multi-processor (SMP) systems
- EPIC, proprietary RISC, x86 architectures
- Dual-core, quad-core, large cache architectures
- FPGAs, vector processors

The market acceptance of open source software as a viable development platform has simplified application portability among various architectures, and thereby expanded the market reach for any one of these technology advancements. Legacy software no longer locks users into one platform. However, with new choices comes complexity—in the form of decision making—and more specialization that can simultaneously narrow the margin for error.

Good decision making still ultimately requires weighing a broad range of criteria and selecting the optimal platform in terms of best match to the workflow and application requirements. The total cost of the hardware, software, implementation, and operation must also be considered. The platform's ability to scale up or down influences selection, as does the vendor's ability to provide the delivery, reliability, and service quality that will ensure overall success.

Today, this has become an often frustrating selection process. The ubiquitous application of computing technology has caused an explosion in the diversity of requirements. In the past, a few key R&D leaders in one or two fields would often define the requirements. Today, requirements come from all scientific and engineering disciplines as well as all business functions from the accounting department to the factory floor, product development, marketing, and traditional research and development.

A widening gap between computing architectures and many technical users' requirements has opened up in the industry. Several trends have resulted:

- The diversity of technical computing applications continues to grow. For example, within the computer-aided engineering (CAE) industry, technology advances have made it possible to address a new range of problems. Complex physical phenomena, crossing multiple disciplines, are routinely simulated for collaborative work. Especially in the manufacturing sector, CAE end users run a large number of varied workloads involving areas such as durability, vibration, thermal, crash/impact analysis, aerodynamics, acoustics, etc. These workloads require diverse hardware capabilities, even within a single discipline: structural mechanics code needs large memory and high disk throughput, while structural crash analysis does not; computational fluid dynamics (CFD) unstructured mesh models are typically addressed with iterative solvers, and scale well within cluster environments, while CFD structured mesh models are more suited to large shared-memory systems.
- General-purpose computing architectures are decreasingly able to meet the scale and workflow requirements for every technical application. Dramatically different technologies are required to accelerate the different user workflows.
- Most high-performance computing environments have a highly diverse set of user requirements, ranging from those of the non-expert computer users to those of the world's most advanced computer engineers.
- Data sets increasingly vary in size, and have become a key differentiator for the expanding range of workflows. Data movement and management are becoming as important as computing cycles.

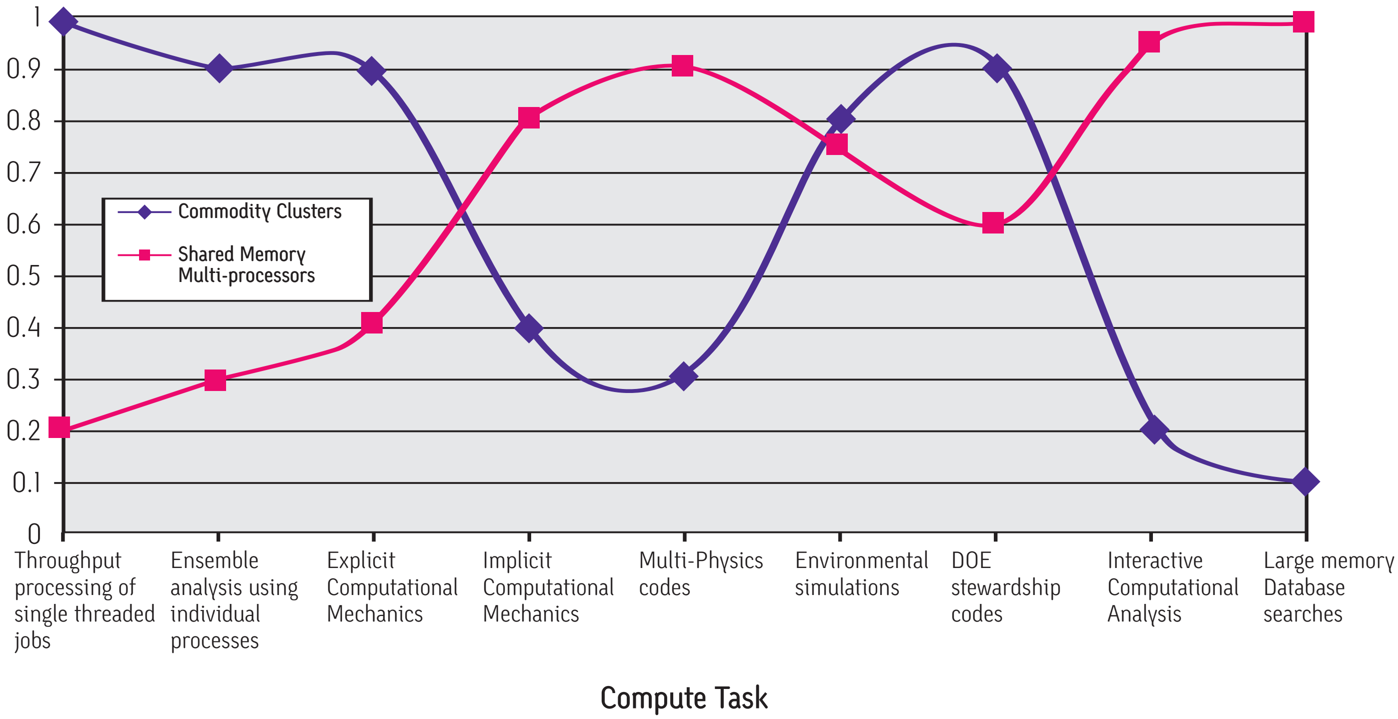


Figure 1: Efficiency of Hardware Architectures

Divergence between current computing platforms and workflows translates into losses of productivity. Still, many high-performance computing solution suppliers continue to offer a single architecture for all situations, arguing that workflows must be adjusted as part of the performance optimization process.

2.0 A Better Approach: A Multi-Workflow Architecture

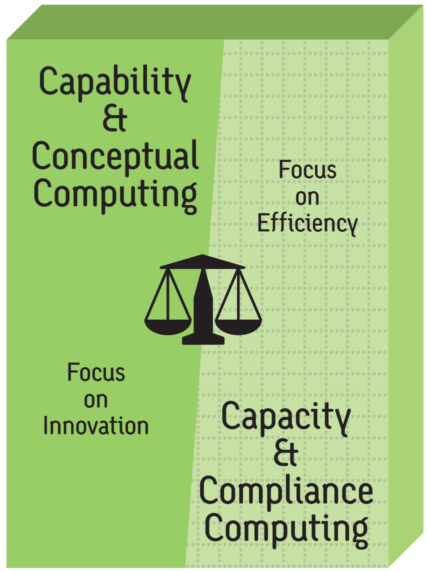
Rather than force-fitting all workflows onto a single architecture, it is possible to develop a high-performance computing system based on a multi-workflow architecture. To illustrate, consider the following two types of workflow requirements:

- Capability and conceptual workflows: These workflows and applications are data-centric and often require many tightly coupled processes accessing very large amounts of data in memory (benefiting from big cache, big memory, and big disk storage). Capability workflows are characterized by highly non-deterministic resource demands, requiring high levels of user interactivity, and are part of the very rapid development cycles typical within research and development organizations.

- Capacity and compliance workflows: For these workflows, there is little data sharing between processes. Typical within production environments, the workflows are often driven by operational costs rather than time to insight. Mature scientific practices are employed to solve well understood problems. Capacity workflows are typically highly automated with little if any user interactivity. Adherence to standardization is important for operational efficiency.

Workload Characteristics

- Jobs that need many tightly coupled processors (very low-latency, very high-bandwidth)
- Workflow and applications benefit from a data-centric architecture, "Big Data" (big cache, big memory, big disk storage)
- Non-deterministic resource demands
- Focus on subject matter, not computing paradigm
- High levels of interactivity
- An environment that supports rapid development cycles



Workload Characteristics

- Very little data sharing between applications or parallel application threads (lots of little jobs; extremely parallel)
- Time to insight is *not* the highest priority; instead price-performance, compatibility, pure cost
- Non-interactive workflows
- Limited requirements for I/O
- Deterministic applications running consistent-size problems
- Adherence to IT standardization
- Mature applications and scientific practices

Figure 2: Balancing Diverse Workflow Requirements

To understand the multi-workflow architectural approach for high-performance computing, first consider that varying user requirements can be viewed as two or more different classes of applications or workloads. Diverging applications or workloads can be addressed by separating them into groups with similar characteristics (such as the two categories of workflows described above). By breaking up a large, highly conflicted, and constrained set of user requirements, each workflow category can be independently considered and a “best of breed” technology chosen to satisfy the applications or workloads. The result is a multi-workflow architecture that spans multiple, optimal point solutions.

To be successful, the multi-workflow architecture must provide users with the look and feel of a single system. To accomplish this, the multi-workflow approach must address the systems integration challenges and create a cohesive over-arching architecture. Without adequate integration at the systems level, users would be required to understand and deal with numerous varying system characteristics within the operating systems and development environments. The integration process must also meet the market’s demands in terms of reliability, supportability, and most importantly efficient data sharing across platforms. Thankfully, much of the work done in the areas of cluster management, virtualization, and GRID tools is directly applicable to the multi-workflow architecture integration challenges.

The successful integration of two or more heterogeneous technologies makes it possible for a single architecture to support mixed workflows. Each workflow’s distinctly different application or computational requirements can be optimally met with a particular point solution, in terms of computation, networking, storage, and visualization technologies. This unique aspect of multi-workflow architectures provides significant value to the many users that recognize multi-disciplinary, system-level analysis as a pathway to breakthroughs and innovation.

3.0 Advanced Multi-Workflow Architecture

The SGI high-performance computing platform provides an example of a server strategy designed to the multi-workflow architecture principals. With over 20 years of experience in high performance computing, SGI has become known for solving the most challenging customer compute problems. SGI has extended this core capability into an extended platform that is intended to specifically address the variable requirements of mixed workflow environments.

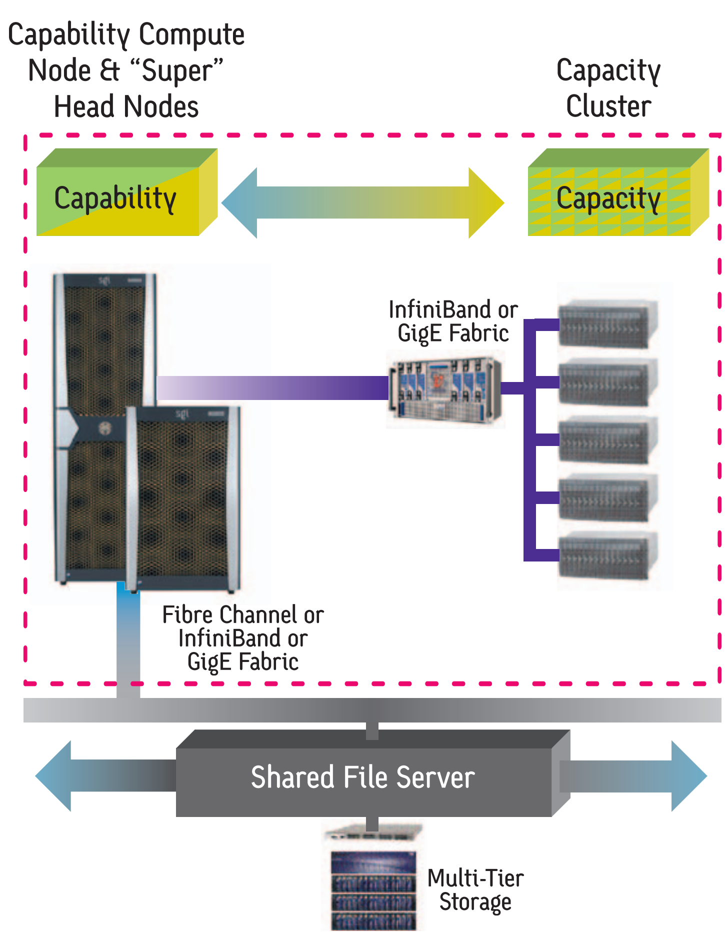
Capability and conceptual Workflows. The SGI® Altix® line of Intel® Itanium® Processor-based servers, introduced in 2003, leads the industry in single system scalability and performance on the Linux® operating system. The newest SGI® Altix® 4700 and 450 servers continue to drive innovation and performance in a “plug and solve” modular blade architecture that scales to 1000s of processors and TB of memory.

System resources, such as compute, memory, I/O, graphics or even FPGA technology can be independently configured to suit the needs of any workload. The current Altix® 4700/450 systems incorporate Intel's new Montecito processor architecture, effectively doubling performance density and scalability, while consuming less power. They also introduce peer I/O technology, which further enhances performance by enabling application-specific co-processors to tap into the shared memory address space at full system bandwidth. This innovative new architecture not only improves performance today but paves the way for the incorporation of future computing technologies in a highly efficient manner.

Capacity and compliance Workflows. SGI recently extended its server portfolio with the introduction of a new line of servers based on the Intel® Dual-Core Xeon® 5100 processor architecture, the SGI® Altix® XE platform. The new Altix XE product line is designed to efficiently address the needs of capacity and compliance workflows in a value-priced, scale-out solution. This new Altix XE server line complements the Altix Itanium line, but targets the specific requirements of cluster deployments. For example, cluster buyers often struggle with a need to select an optimal compute solution, balanced by a need to simplify the process of building and deploying scale-out clusters. SGI Altix XE is a customizable cluster solution that ships to the customer fully-integrated from SGI's factory. This enables customers to easily buy and deploy a full-featured cluster solution, without sacrificing functionality or performance.

The SGI server platform addresses specific workflow requirements with these complementary Altix XE and Altix 4x product lines, in a unified infrastructure that is based on industry-standard Linux® operating system, a comprehensive storage offering, and common development and administrative toolsets across the platform. For example, SGI® ProPack® for Linux® is a software package that includes performance and tuning enhancements. ProPack is supported across the SGI server platform, but includes specific features to drive performance in capability environments for Altix, and another set of features to drive performance in capacity deployments for Altix XE.

Figure 3: Hybrid Architecture



Using this approach, the breadth of hardware platforms and technologies does not require users to master multiple computing architectures. Common system-level characteristics make it possible to move from one platform to another, and to manage projects that span multiple systems and technologies:

- A common Linux operating system across all platforms (users need learn only one system).
- Common compilers and development tools enable application portability.
- A single job scheduling/workload management system to manage workflow across all systems.
- A centralized systems management interface and operational dashboard for updating, provisioning, configuring, and monitoring.
- Platform-wide storage solutions to simplify data sharing, access, movement, and management across systems.

SGI's strategy, to incorporate leading-edge technology in point-solutions that target specific workflow requirements and package the technology in a unified platform based on a comprehensive infrastructure, enables users to efficiently deploy a "blended" server platform to support multi-workflow requirements.

In this type of architectural strategy, unified system software and a common storage infrastructure enables a single, multi-workflow platform. The combination of platform breadth for bridging the gap between system architecture and workflow requirements and a consistent system environment make it realistic to apply high-performance computing solutions beyond R&D to mainstream scientific and engineering functions. Varying operational, applications, and business requirements can be met with a single multi-workflow architecture. This next generation "balanced architecture" builds a bridge that spans capability and capacity requirements, and enables the application of compute resources to deliver breakthroughs in efficiency and resource utilization.

4.0 Evolving the Multi-Workflow Platform

Today, advances in the performance of processor cores and cluster technologies increasingly enable the use of mass market processors for improved levels of application performance. But, as current technology approaches the limitations of Moore's Law and cluster-based parallelism, alternate strategies will increasingly factor into advanced architectures to drive performance and scalability in the multi-workflow environment. To that end, SGI is once again leading the industry with a multi-paradigm computing strategy to advance its shared memory architecture by supporting a variety of tightly coupled alternative processing elements. This approach is based on the concept of attaching devices to the core SGI scalable shared-memory architecture, to deliver the best-performing processing elements for particular applications.

SGI has introduced this capability to the market in its Altix 4700 platform, which integrates SGI's peer I/O technology, to provide direct, high-speed access to SGI's large shared memory for all system components. Through peer I/O, SGI Altix 4700 is able to support new computing paradigms with high efficiency, such as reconfigurable computing through SGI RASC™ FPGA technology, which breaks through the limits of Moore's Law.

The potential to scale from single nodes all the way up to petascale systems, while at the same time delivering algorithm and application-specific acceleration technologies to supplement microprocessor advancements, promises to deliver new levels of both flexibility and performance. In parallel with this, advances in scale-out systems continue to improve efficiencies both on the node and across the cluster through ever-increasing processor densities, node scalability, and interconnect speed. The result is a next generation multi-workflow platform that can satisfy everyday workload demands with a new level of productivity and power.

5.0 Long-Term Business Value

Within the family constructs of the multi-workflow architecture, point products can evolve and incorporate new innovations while minimizing the impact on the users and applications already deployed on the platforms:

- Fully integrated platforms: Offload the tasks required to integrate and benefit from new innovations and technologies as they become available.
- Common software: Industry-standard operating environments, and common development and workload management solutions to provide a smooth path to future platforms.
- Ease of administration: Flexible system configuration options (blade servers, clusters, resource scaling) and common toolsets across comprehensive environment.

This type of "next generation" high-performance computing platform effectively applies multiple technologies to the job at hand, and enables real-world workflows to be optimally matched with the most suitable technology or combination of technologies. The best technology "fit" can then be applied to meet specific workflow and business requirements, but in a unified compute strategy.

The freedom to choose the best platform, without adding complexity to the user computing environment, provides users with a superior solution and the ability to maximize productivity for each computing project. The consistent software and system environment across all hardware technologies also protects user investments: software can be easily migrated to new systems, and hardware resources can be reallocated if project requirements change. The multi-workflow architecture makes the underlying technology transparent to the users, applications, and workflows, introducing operational efficiencies at every level of the organization.

6.0 Conclusions

Today, users across industries of all types recognize the benefit that high-performance computing applications can bring to the organization, and are applying HPC technology to an increasingly broad range of uses. For example, the CAE community has driven a trend towards the use of virtual prototyping tools in the product design process. More than a means to validate an existing design, this high-performance computing task is driving the entire product design process. There are immense cost benefits to using CAE techniques early in the design phase where problems can be identified and fixed much earlier in the product lifecycle. Other industries are similarly recognizing the value of high-performance computing architectures in addressing a diverse range of compute needs, further broadening the range of workflows and user requirements that must be addressed by today's platforms.

While the benefits of the multi-workflow architecture are clear, the industry is now acknowledging this approach as a way to meet these diverging user needs. This paper considers a prototypical architecture based on currently available commercial and open source systems and software that has been used to successfully solve the diverse multi-workflow needs for both the production and R&D requirements. The resulting architecture is an exciting new breakthrough in presenting the optimal computational match to each application workload, improved efficiency through better data flow, and ease of use and administration through a common set of user and system management tools.

Corporate Office
1200 Crittenden Lane
Mountain View, CA 94043
(650) 960-1980
www.sgi.com

North America +1 800.800.7441
Latin America +55 11.5185.2860
Europe +44 118.912.7500
Japan +81 3.5488.1811
Asia Pacific +1 650.933.3000

3950 [07.20.2006] J15177