White Paper

The Future of Cluster Computing

Without a sustainable computing strategy, can today's commodity HPC clusters take advantage of future hardware developments?

Table of Contents

1.0 Executive Summary

While cluster computing has grown popular as a means of economically increasing compute capacity, the very market forces that help drive its popularity may inadvertently limit the ability of clusters to serve in serious high-performance computing (HPC) environments. The ad-hoc evolution of clustering solutions has involved little strategic planning. But a roadmap is necessary to ensure that tomorrow's clusters meet the needs of scientists, engineers, government and security professionals, and energy companies—users who demand the most powerful, efficient, and scalable solutions available. A good design strategy can also reduce or eliminate many of the hidden expenditures that come with white box implementations and that ultimately increase the cost of cluster ownership.

Today's emerging technology trends suggest that such a roadmap is within easy reach. Developments in 64-bit microprocessors, interconnect technology, storage I/O solutions, and cluster management software offer the building blocks for an HPC design strategy with the headroom necessary to take advantage of future hardware developments.

Key elements of this strategy include:

- Fewer, denser cluster nodes driven by fewer, faster CPUs
- More flexible allocation of memory, ideally implemented in a shared-memory architecture
- Hierarchical interconnect technologies, putting the most costly bandwidth where it is needed most, and commodity solutions elsewhere
- An integrated storage design that allows access to data without clogging networks with massive data transfers

While additional work must be done to refine software, applications, and toolkits to accommodate new generations of HPC problems on clusters, the market has already begun to see cluster solutions that fundamentally adhere to the tenets outlined above. Without such solutions, clusters may well become irrelevant to all but a small fraction of HPC users.

2.0 Introduction: The Need for a Cluster Computing Strategy

Having graduated from the garage to the Fortune 500, High Performance Computing clusters now represent a viable force in life sciences, aerospace, earth and atmospheric studies, product design and manufacturing, homeland security, and energy exploration and analysis. So popular is clustering technology that it claims its own economic system, complete with products and services that in turn impact other technology ecosystems.

The rise of clustering is impressive, particularly given its relatively haphazard roots. Clustering first emerged as a tactical re-engineering of computing technology initially designed for other purposes. Generally seen as low-cost alternatives to expanding an organization's computing capacity, clusters became even more attractive as technology spending budgets plummeted following the dotcom economic downturn.

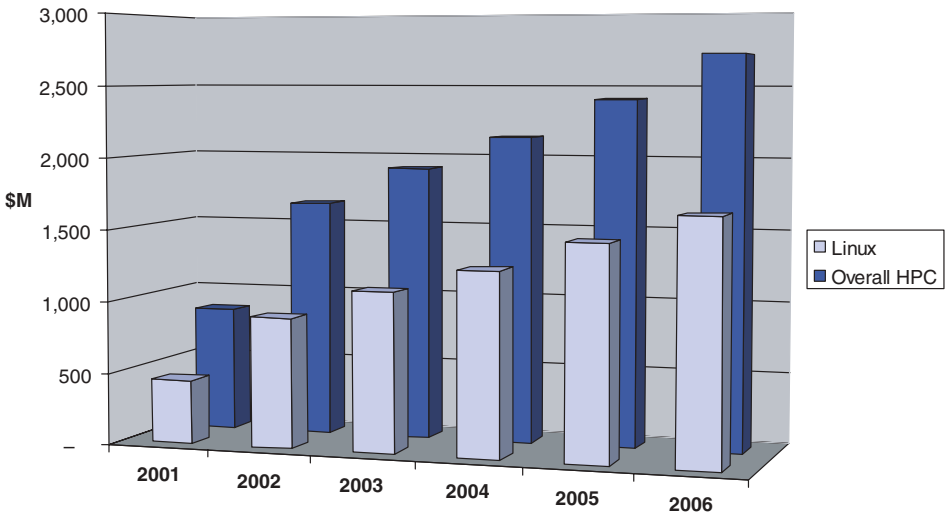
But the way in which clusters have been implemented—and are popularly deployed today—may inadvertently limit the adaptation of HPC clustering technology in years to come.

Cluster components traditionally have been comprised of commercial off the shelf (COTS) products. These components were designed for general-purpose business computing and, as such, they enjoyed the kind of economies of scale that promotes an aggressive growth cycle of innovation and lower prices. This trend fueled rapid growth in the price/performance of components—growth that eventually benefited components used in the high-end technical computing marketplace.

While the HPC cluster market continues to grow—with a compound annual growth rate of more than 25 percent through 2006 (see Figure 1)—the industry possesses neither a specific clustering technology strategy nor an architecture to guide future adaptation and refinement. Historically, clusters have adapted to their environments with no real strategic roadmap for them to follow. For universities and other users that have deployed commodity white box clusters, long-term plans have essentially amounted to little more than lashing additional one- or two-processor nodes to the cluster core. Indeed, the ad-hoc simplicity of that approach has been a large part of clustering's appeal.

Yet if we pursue that strategy in the future, then clustering will always amount to little more than a derivative of the strategies and architectures that are shaping the low-cost, general-business computing industries. This path is destined to produce solutions that will increasingly fail to meet the needs of HPC users. Already, the most demanding users—running complex, memory-dependent applications—face severe efficiency problems with current cluster designs. Worse, today's industry trends don't signal any real improvements unless we can establish a more strategic clustering methodology.

To achieve this, we must look at three core components of cluster nodes: microprocessors, networking technologies, and system software. Armed with historical data, industry trends, and requirements of HPC users, we can focus on these three core components to envision a new design for clustering. This



Source: IDC Reprinted with permission

Figure 1 Clusters gaining ground in HPC

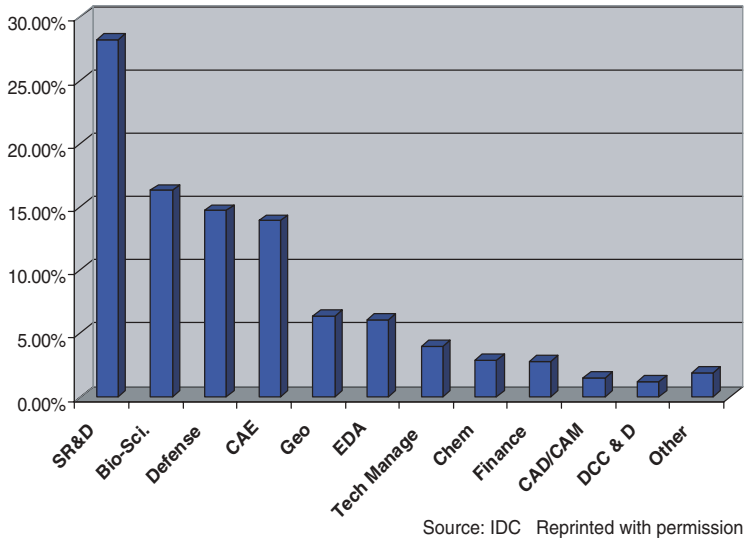
design provides the foundation for a strategic clustering roadmap that can produce sustaining technical designs for the HPC cluster marketplace for years to come.

This paper discusses an approach to establishing such a design strategy by first outlining the trends and technologies that are driving the cluster phenomenon today, as well as those trends that will provide key indicators of future developments. Using this background, a new approach to clustering is presented and discussed. Finally, we will present existing efforts to make that approach a practical and cost-effective reality in the broadest range of HPC clustering environments.

3.0 Trends that Shape HPC Today

The HPC market is undergoing a renaissance. After several years of relatively flat growth following the IT spending drought of 2002, HPC is currently enjoying a strong overall increase of hardware, software, and services sales. By 2007, IDC expects the HPC market to reach $6.3 billion annually, representing a compound annual growth rate of 6.3 percent.

The users driving this growth grapple with problems that require resources of ever-increasing power and capability. Indeed, analysts at IDC refer to the highest echelon of HPC as capability computing. Capability computing customers represent a



Source: IDC Reprinted with permission

Figure 2 Capability computing application areas by market share (2002)

broad range of vertical markets and application areas (see Figure 2), and include government, commercial and educational environments. These users will fuel continued innovation in capability computing, consuming the most powerful technologies ever conceived. Yet they will also push the capabilities of general-purpose technologies as they try to meet their own growing capacity needs.

3.1 Technical Markets Drive High-End Innovation

Technical markets, which not long ago were relegated to the most rarified strata of computing, have recently joined the mainstream. No longer are costly proprietary systems—formerly dominated by UNIX® platforms—a mandate for HPC environments. Thanks to the natural evolution of scalable solutions and leveraged product strategies, the hardware itself has grown less crucial to the technical computing experience. This is due, in part, to a shake-out among technical computing products vendors, the result of the market responding to new technologies that have caused certain hardware component costs to plummet and the rise of Linux® as a technical computing environment.

Some technical users are driving more change than others. In the life sciences and earth sciences sectors, burgeoning data sets—some reaching a terabyte or more in size—have triggered the development of a new generation of compute servers and will likely continue to do so in the future.

For these users, systems must be capable of efficiently handling a broad range of enormous simulation and analysis problems, including:

Life Sciences

- Design & virtual testing of new drugs
- Modeling anti-cancer drug reactions
- 3D modeling of functioning organs
- Gene sequence analysis
- Quantum chemistry

Earth Sciences

- Ocean temp & sea level simulation
- Ozone depletion analysis
- Atmospheric modeling
- Earthquake & tsunami prediction
- Hurricane track prediction

With cluster deployments popular in university and research laboratories, users have shown genuine and widespread interest in cluster systems as a viable platform for technical computing. The economics of computing in general, with hardware costs losing their relative significance but software on an inverse curve, has helped perpetuate the image of cluster computing as a solution that comes with a low entry price. Yet at the same time, growth of overall price/performance of commodity clusters is slowing, with system efficiency feeling the brunt of this trend. This, in turn, impacts the long-term cost-effectiveness of commodity cluster systems.

3.2 Itanium Takes Center Stage

With the performance of most cluster nodes highly CPU-dependent, microprocessors tend to play a significant role in the selection of cluster systems. Nowhere is this more evident than in the technical computing space.

Until recently, technical computing was nearly the exclusive domain of proprietary RISC processors. RISC CPUs held over more than 80 percent of the technical server market as recently as 2002. But 32-bit and 64-bit processors have rapidly gained ground. In 2005, IDC analysts expect 32-bit Intel® and Intel Itanium Processor Family microprocessors will claim as much market share as RISC processors, and, in 2006, the Itanium Processor Family architecture will become the dominant processor within Intel processor-based technical servers, selling three-to-one over 32-bit processors in 2005 and helping to spread the popularity of multi-core systems.

3.3 The Need for Speed

When it comes to CPU cycles, memory, system and interconnect bandwidth and other factors that impact application productivity, high-performance computing environments are insatiable. As component economies and technical advances converge in the HPC server market, more key trends are becoming apparent:

- **More advanced processors**—Users are responding to new generations of 64-bit processors equipped with on-chip caches totaling 6MB or more, as well as an emerging population of multi-core processors.
- **Memory technology innovations**—Faster, denser memory designs combined with advances like RAS (Row Address Strobe) and DDR 2 (Double Data Rate) deliver faster speeds, higher data bandwidths and lower latency.
- **Better throughput**—More and more, memory controllers are becoming tightly coupled with microprocessors for faster intra-CPU communication, while users and vendors migrate from PCI-X to PCI-Express to achieve data transfer rates of 4GB/sec.
- **Improved system infrastructure**—Developments like Interoperable MPI (IMPI), which enables multiple vendor implementations of MPI to coexist in a cluster, and the broad availability of mature cluster management tools enable users to make the most of their investments.
- **Interconnect technology wars**—Faster interconnect fabrics—from 10Gb Ethernet to Quadrics® and Infiniband—give users

plenty of choices, and with them come dilemmas over whether to select proprietary or off-the-shelf solutions. Yet as bandwidth grows by leaps and bounds, similar reductions in latency are harder to come by.

4.0 Commoditization Driving Cluster Evolution

Today's cluster computing marketplace is really an ensemble of commodity technology from much larger industries, such as the general-purpose microprocessor, networking, Linux, and PC server sectors. These sectors are driven by much larger and broader market pressures than the cluster community alone and can differ significantly from the requirement of HPC applications. To date, cluster evolution has been highly dependent on the integration of these technologies with minimal investment in modification or added value. Cluster computing, in fact, has been one of the most fortunate benefits of commoditization.

For current cluster designs to remain successful and competitive in the future, however, certain industry trends must remain intact and viable for years to come.

- CPU performance must continue escalating at the rate Gordon Moore predicted 40 years ago, with processor performance that doubles every 18 months; while from a hardware perspective the outlook for Moore's Law looks promising, the techniques that enable software to capitalize on CPU performance are undergoing dramatic change
- The popularity of software known, somewhat affectionately, as "embarrassingly parallel" applications must remain intact in the future; these large-scale applications, whose execution of very large numbers of tasks is feasible only through parallelism, are highly visible in the cluster computing world but in fact are relatively few in number
- The market's infatuation with cluster computing has never been hotter, and this has prompted investigation into new application areas for cluster environments; while this type of evolution is important to the vitality of cluster computing as a whole, it also hinders the creation of solutions aimed at core HPC users; still, investigative energy of this type holds the potential for dramatic changes in market structure
- Linux must remain the dominant operating system of clustering and is a significant factor in its success; because Linux makes it easier to share software and fundamentally relieves users of the vendor/OS allegiances there is little future interest in other alternatives

4.1 Performance Limits and Hidden Costs

It's impossible to predict with certainty that all of the above trends will stay their respective courses. Potentially standing in the way of such a prosperous, commodity-friendly future are other realities that HPC users increasingly face.

Rigid partitioning of cluster memory across nodes, for example, can drastically limit application performance. This makes a traditional distributed memory cluster less useful to growing numbers of HPC users. Likewise, there are real costs associated with implementing thin node topologies and uniform, proprietary interconnect technologies.

With software costs on the rise (see Table 1), solving performance issues by adding more CPUs can trigger commercial software licensing fees whose expense profile outweighs the value of additional processors. Indeed, many users are discovering that the initial appeal of low-cost clusters can quickly fade as they begin to realize the "hidden costs" associated with maintaining and administering far-flung commodity clusters.

Table 1 Three opportunities for optimizing cluster investments. Analyze the component and technology costs in a traditional cluster, and it is easy to pinpoint the dominant areas of investment.

Cluster Component	% of Total Cost	Cost Trend
PU	13.0 %	Down
Memory	8.9 %	Erratic
System Infrastructure	19.0 %	Flat
Interconnect	22.8 %	Flat
Cluster Software	26.0 %	Up
Storage Hardware	1.9 %	Down
Storage Software	1.2 %	Up
Professional Services	1.1 %	Flat
Support Services	4.6 %	Down
Training	1.5 %	Up
Total	**100.0 %**	

This view of the current economics of commodity clusters shows that three components consume a greater share of costs than the CPU, which garners so much attention. In fact, system infrastructure, interconnect and software dominate the cost profile of cluster deployments, amounting to almost 70 percent of cluster cost. And while infrastructure and interconnect costs remain flat, interconnect bandwidth continues to soar. Software, which is growing more costly, is also changing to take advantage of faster systems.

To establish a cluster technology strategy that meets the needs of the most demanding users, it's important to ensure that these three areas are the focus of significant effort.

5.0 Charting the Course for Tomorrow's HPC Clusters

Guiding the evolution of cluster technology down a path that serves HPC users as well as the general-purpose market is feasible by following a few overall recommendations—all of which take advantage of current trends.

- **Fewer, faster CPUs**—By implementing faster processors, users attain the desired performance while minimizing their per-CPU software license fees; fewer CPUs also relieves the pressure on parallel efficiency of applications as Amdal's Law so vividly points out
- **Denser node clustering**—By adding more CPUs per node, users can achieve more cost-effective infrastructure deployments and lower their interconnect costs
- **Multi-plane inter-node communication**—Hierarchical topologies allow users to take the fullest advantage of their infrastructure investments; non-uniform communication reduces interconnect demands, trims latency, and boosts overall application performance
- **Hybrid parallelism**—Cluster administrators should be able to leverage both MPI and OpenMP™ multi-platform shared-memory API, depending on their needs; proven to improve overall parallel efficiency, hybrid parallelism addresses load balancing and allows a broader selection of algorithms
- **Hierarchical I/O**—Cluster I/O needs are not static, so a scalable I/O and storage model is a must for HPC environments; by implementing Storage Area Network and Network Attached Storage solutions, clusters have the flexibility to address non-uniform I/O patterns and the growing problem of managing cluster data
- **Multipurpose interconnect**—Putting interconnect technologies to more flexible use (IPC, shared storage, hierarchical topologies), users derive more from their interconnect investments and build in greater flexibility to adapt to unforeseen future requirements
- **Workflow analysis**—An "on-demand" approach applies compute cycles where they are needed most; this facilitates a flexible model that avails clusters to best-of-breed technology

These trends exist today although their presence is far from ubiquitous. Yet all are supported by varying degrees of industry effort, research and development, and user-driven innovation. Enough, in fact, to put them in the critical path of cluster evolution.

5.1 Keys to a Sustainable Design Strategy

For an alternative to today's relatively rudderless market direction for clusters, HPC users should pursue a sustainable design strategy that can be implemented today but has the headroom to accommodate inevitable developments in component performance and capability. The key to this new strategy: Dense computational nodes with hierarchical interconnection of the CPUs, memory and I/O.

It sounds simple enough, but how would this cluster look? And how would it compare to what has already been deployed in user sites around the world?

First, consider the topology of a typical HPC cluster today (see Figure 3). Taking a top-down perspective, it's easy to see how such a configuration could come together. Over months or years, as needs increase, users supplement existing nodes with additional single- or dual-processor servers, more network ports, and additional storage interconnects. This follows the classic cluster model, and reveals its strength and, quite possibly, one of its greatest weaknesses: While light node clusters can be inexpensive to implement and expand—after all, how costly is a handful of 2p white box servers and some Ethernet cable—the cost of maintaining and managing them grows as they do. Productivity suffers, as well, because end users spend more time trying to learn how to get their code to run on a cluster rather than focusing on the work they originally set out to do. And as node count grows, I/O performance becomes even more crucial as bottleneck vulnerabilities multiply.

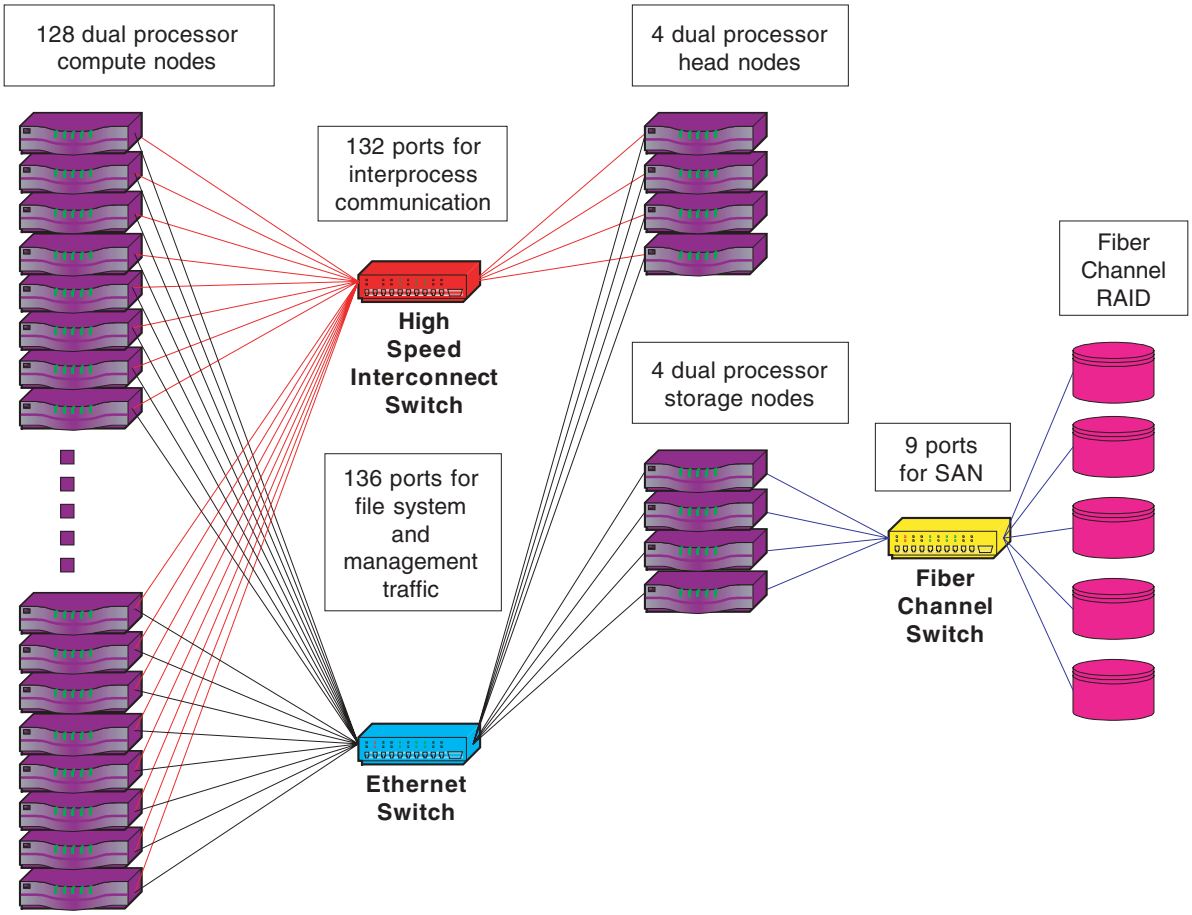


Figure 3 Classic small-node 256-processor HPC center

An HPC-oriented cluster design strategy, however, makes use of fewer, faster CPUs, packing more processors into each node. (Added benefit: As IT increasingly fights for physical space in which to house its technology assets, a dense-node cluster also helps address emerging footprint problems.) More flexible allocation of memory—ideally implemented in a shared-memory architecture—enables more optimal application performance. Interconnect topologies address real-world application needs, with local "fastest" interconnects located within the nodes, and a lower-cost commodity interconnect linking nodes. An integrated storage design incorporates a shared file system that allows access to data without having to transfer entire data sets over the network. The result is a more powerful and efficient cluster, one that reduces total cost of ownership by lowering software licensing, infrastructure and interconnect costs. (See Figure 4.)

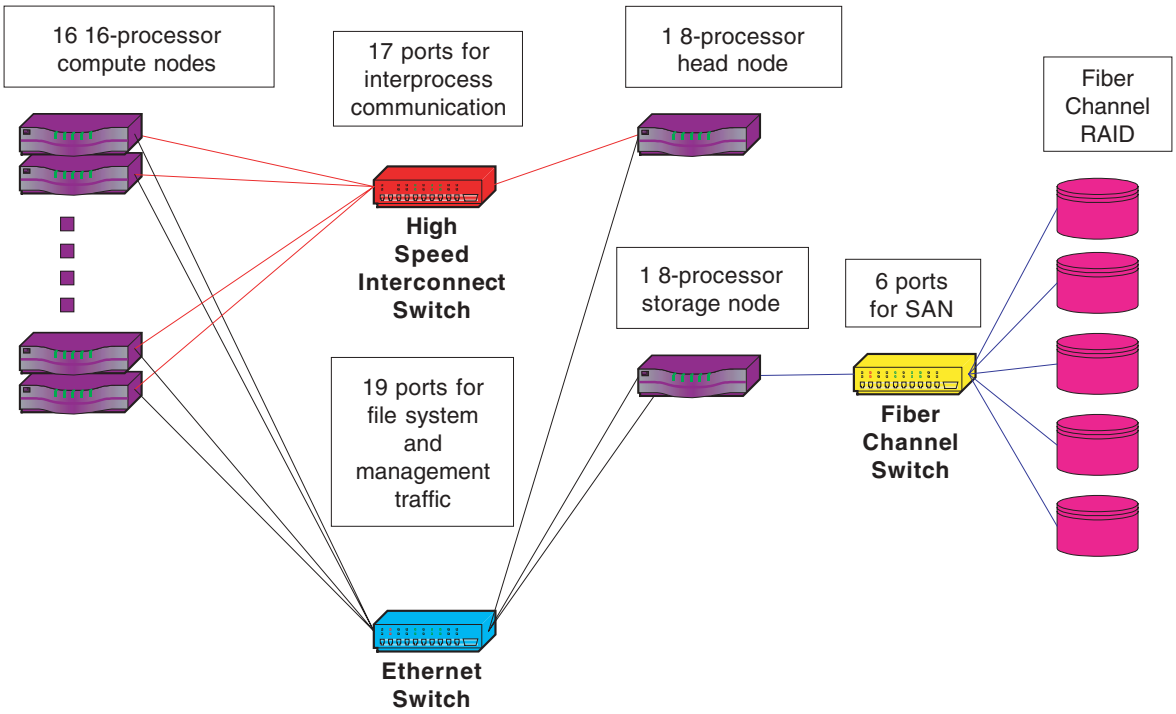


Figure 4 Dense-node 256-processor HPC cluster

6.0 The New Generation of HPC Clusters

At least one server vendor has introduced products that support this sustainable design strategy. Silicon Graphics, for instance, recently introduced a factory-integrated cluster that scales to 32 Intel Itanium 2 processors per node. Such offerings align with the sustainable design scheme, not only in terms of scalability, but also for reducing total cost of ownership. By scaling at the node instead of adding new nodes every time computing needs increase, users spend less on software licensing costs and interconnect fabric, with fewer nodes to connect, manage and provision.

HPC cluster solutions such as these are aimed at addressing the shortcomings of existing commodity clusters while applying the advantages of technologies that power some of the world's most powerful supercomputers. They are designed to be easy to deploy and administer, with simple, easy-to-manage configurations and the flexibility to scale up with more processors and out with more nodes to address changing business requirements. They are also designed to be configured with industry-leading management, interconnect and storage technology, such as Scali Manage™, Platform Computing LSF®, InfiniBand, 10Gb Ethernet, and SGI® InfiniteStorage.

Users are also beginning to find solutions that leverage the best of both 32-bit and 64-bit computing environments. Common tools and interfaces for managing and using the hybrid cluster create one solution with two architectures, delivering maximum ROI for heterogeneous workflows. Such solutions are ideal for users whose application workload increasingly requires strong computing capability and global-shared memory to complement 32-bit cluster technology.

6.1 The Work Yet to be Done

While new solutions are coming to market today and will continue to evolve in the future, overcoming existing constraints to change will require work. Application software, as always, is key. Many applications are hard-coded for uniform thin node cluster topologies, rendering them woefully insufficient with new hierarchical, dense node cluster environments. We must also address single-processor optimizations for next-generation CPUs, while finding ways to cost-effectively bridge the compatibility gap between existing X86 systems and new microprocessor architectures. We must also define more non-uniform parallel application processors and develop toolkits for hybrid parallel applications. And intelligent fabrics for interconnects, including non-uniform topologies, must become mainstays in HPC clusters.

6.2 An Investment That Pays Dividends

While change is never easy—particularly as cluster administrators work hard to manage existing capacity demands—taking steps now toward a sustainable cluster design strategy will pay dividends for years to come. HPC requirements will only increase, applications will only grow more complex and server facility space will only become more valuable. Adding software license fees, management and maintenance costs, productivity losses due to troublesome cluster administration, and interconnect expenses, users can begin to realize the true cost of ownership associated with today's sprawling commodity clusters.

For technology investments to be useful for years to come, a cluster design strategy that takes advantage of commodity technologies while pushing innovations capable of keeping up with HPC requirements is far more than merely a worthwhile notion. It's critical.

Corporate Office
1500 Crittenden Lane
Mountain View, CA 94043
(650) 960-1980
www.sgi.com

North America +1 800.800.7441
Latin America +55 11.5509.1455
Europe +44 118.925.7500
Japan +81 3.5488.1811
Asia Pacific +1 650.933.3000

3767 [02.2005]

J14808