GRID ENGINE
WHITE PAPER

Sun Powers the Grid

Grid computing and its impact on your Network

Increased network bandwidth, more powerful computers, and the acceptance of the Internet have driven the on-going demand for new and better ways to compute. Commercial enterprises, academic institutions and research organizations alike continue to take advantage of these advancements, and constantly seek new technologies and practices that enable them to reinvent the way they conduct business. However, many challenges remain — reduced development costs, faster time-to-market, greater throughput, and improved quality and innovation are always foremost in the minds of administrators — and computational needs are outpacing the ability of organizations to deploy sufficient resources to meet growing workload demands.

Today, a vast amount of potential computing capacity remains untapped. Users are continually searching for more computing resources to help solve problems, resulting in systems that are alternately over-loaded or under-utilized. With some tasks taking days and weeks to run, the practice of manually starting and restarting jobs wastes valuable time and reduces productivity. Clearly, in order for users to optimize productivity, they must focus on design and development rather than hunting for resources.

No group is without resource constraints. Investments in computing hardware need to be justified — total cost of ownership (TCO) is a critical priority, and organizations must find ways to enable more work to be done with available resources. Beyond the expense of constantly upgrading hardware to meet increasing performance needs, administrators must cope with the inefficiency of maintaining hardware and software configurations unique to each system. Organizations need an efficient solution that gives them a greater return on their computing investment without compromising performance.

On top of these challenges is the need to handle dynamically changing workloads. The truth is, flexibility is key. In a world with rapidly changing markets, enterprises need to quickly provide compute power where it is needed most. Indeed, if systems could be dynamically created when they are needed, teams could harness these resources to increase innovation and better achieve their objectives.

Imagine if all this could be done today. It can, with Grid Computing.

Grid Computing

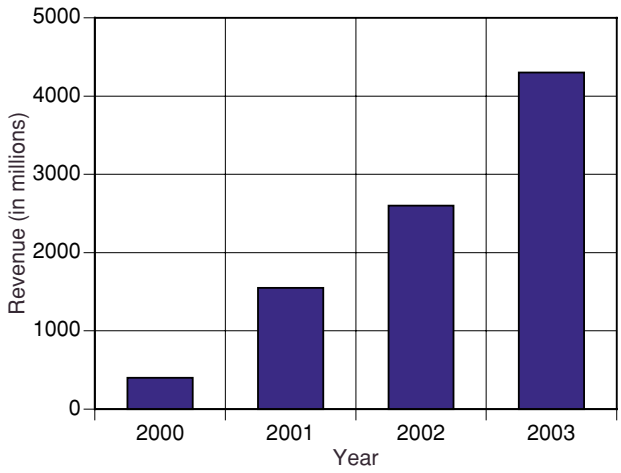
Have you ever driven past an office building during a workday, the parking lot full of vehicles, just sitting there, idle and unused, for six or eight hours at a time? And those same vehicles probably sit in the driveway for another eight hours at night. Doesn't that seem like a colossal waste of resources, of engineering effort, design effort, manufacturing effort? Wouldn't it make sense to be able to make better use of the vehicles' idle time, to share them among several users in some way, so that they're not just taking up space?

This is the situation today with compute resources in the workplace. Workstations, desktop computers, Web and email servers... all of these machines are sitting idle for at least some part of a 24-hour period, like vehicles in a parking lot. And if those machines are sitting unused, they are producing zero percent return on your investments.

However, unlike those vehicles, there is a solution for sharing your compute resources: Grid Computing. And it allows you to put your idle compute resources to work rather than just running screen savers of flying toasters.

At the heart of Grid Computing is a computing infrastructure that provides dependable, consistent, pervasive and inexpensive access to computational capabilities. Researchers working to solve many of the most difficult scientific problems have long understood the potential of such shared distributed computing systems. Development teams focused on technical products, like semiconductors, are using Grid Computing to achieve higher throughput.

Likewise, the business community is beginning to recognize the importance of distributed systems in applications such as data mining and economic modeling. With a grid, networked resources — desktops, servers, storage, databases, even scientific instruments — can be combined to deploy massive computing power wherever and whenever it is needed most. Users can find resources quickly, use them efficiently, and scale them seamlessly.



The adoption of Grid Computing is increasing at a phenomenal rate. In fact, Sun estimates Grid Computing will grow 300 percent by the year 2003. Indeed, Grid Computing promises to deliver the power of distributed computing to the desktop for years to come.

Conceptually, a grid is quite simple — it is a collection of computing resources that performs tasks. By pooling federated assets, a grid provides a single point of access to powerful distributed resources. Users treat a grid as a single, large computational resource. Resource management software accepts jobs submitted by users and schedules them for execution on appropriate systems in the grid. Users can literally submit thousands of jobs at a time without knowing — or caring — where they will run.

Underneath the hood, a grid consists of a layered architecture. Resource management loads the grid. Infrastructure software links the grid. System management monitors the grid. Portals enable access to the grid. Tools facilitate the development of applications for the grid. Behind it all, networking infrastructure, rich clients, Web servers, high-speed storage and compute servers make this possible.

Scalability for Global Computing

No two grids are alike, and no single size fits all. By utilizing a flexible computing architecture based on clusters — systems and software that manage work on distributed systems — organizations can create and recreate grids to exactly match changing requirements. Indeed, grids can scale from single systems to supercomputer-class clusters utilizing thousands of processors — with grids, no one lacks needed performance, yet nothing is wasted.

Cluster Grids

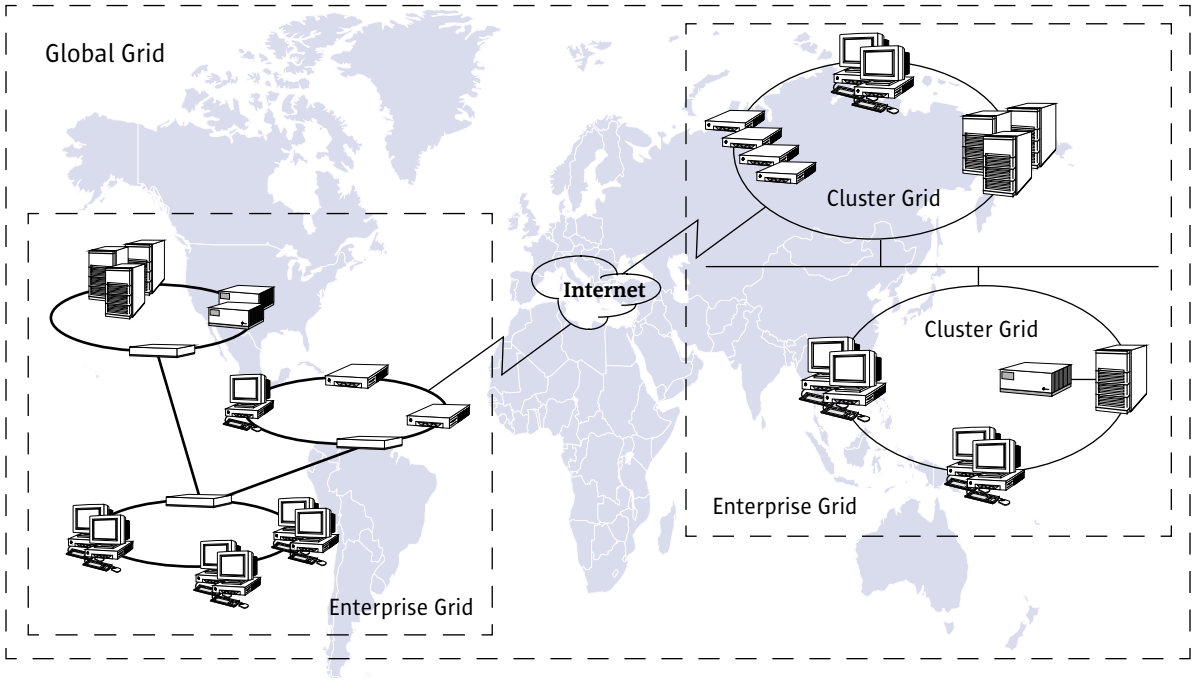
Today, Cluster Grids are the most popular and simplest form of a grid. Meeting the needs of most organizations, Cluster Grids consist of one or more systems working together to provide a single point of access to users. Typically owned and used by a small number of users, such as a project or department, Cluster Grids support both high throughput and high-performance jobs. Resources in the grid can be focused on a narrow set of repetitive tasks, or made to work in true parallel fashion to execute a complex job. For example, electronic design automation (EDA) organizations employ Cluster Grids to complete many discrete jobs quickly, while mechanical computer-aided engineering (MCAE) environments run simulations on different machines that communicate with each other to solve a set of related problems.

Enterprise Grids

As capacity needs and demands for greater economy increase, organizations can combine their Cluster Grids into Enterprise Grids. Enterprise Grids enable multiple projects or departments to share computing resources in a cooperative way. Enterprise Grids may consist of dispersed workstations and servers, as well as centralized resources located in multiple administrative domains, in departments or across the enterprise. Organizations can use Enterprise Grids to handle a wide variety of tasks, including collaborative engineering, mining large databases, rendering frames for animations, absorbing increased loads during cyclical business processes, and more.

Global Grids

When application needs exceed the capacity of a Enterprise Grid, organizations can tap partner resources through a Global Grid. Designed to support and address the needs of multiple sites and organizations sharing resources, Global Grids provide the power of distributed resources to users anywhere in the world. Global Grids are a collection of Enterprise Grids, all of which have agreed upon global usage policies and protocols, but not necessarily the same implementation. Computing resources may be geographically dispersed, connecting sites around the globe. They can be used by individuals or organizations sending overflow work to a grid provider, or by multiple companies working together and sharing data — crossing organizational boundaries with ease.



Visions of the Grid

Computation, communication and collaboration drive every organization. While the Net provides universal connectivity and access to information, one more step remains. The Grid — a universal computing infrastructure — builds on the power of the Net. Extending the Net to its logical conclusion, the Grid offers the opportunity to harness connectivity and change the way people work. Users can now think of the rest of the world as a computational resource ready to be tapped. Organizations can get immediate and easy access to information and services, solve problems and offer services to anyone in the world. Indeed, just as the global community uses the Net to communicate, one day the world will use the grid to compute.

The Grid — the IT infrastructure of the future — promises to transform computation, communication, and collaboration. Over time, these will be seen in the context of grids — academic grids, enterprise grids, research grids, entertainment grids, community grids and so on. Grids will become service-driven with lightweight clients accessing computing resources over the Internet. Datacenters will be safe, reliable and available from anywhere in the world. Applications will be part of a wide spectrum of network-delivered services that include compute cycles, data processing tools, accounting and monitoring, and more. While these changes will revolutionize the way work is done, there will be no disruption in the technology or the way it is used.

Always at the forefront, Sun will do for the Grid what it did for the Net — power it.

Putting the Grid to Work

Perhaps more important than anything else, the grid is the key to getting work done faster and with higher capacity and quality. By giving organizations ready access to untapped resources, the grid opens up new opportunities. EDA organizations are free to test new designs more thoroughly or explore more promising ideas. Finance organizations can run more Monte Carlo simulations to uncover untapped business opportunities. Car manufacturers can conduct more simulations to make automobiles safer. Scientists can develop more sophisticated computer simulations to study and better understand complex phenomena. The possibilities are endless.

With Grid Computing, organizations can gain a competitive advantage. Indeed, the power of the grid lies in its ability to provide a resource-rich environment which maximizes the available power of the local network to:

- *Increase personal productivity*, increasing access to compute resources
- *Increase corporate productivity*, providing more products in the same time
- *Improve product quality*, supporting a more iterative development process
- *Reduce time-to-market*, enabling simultaneous development on multiple machines
- *Increase return on investments*, fostering the best utilization of highly available, scalable, reliable network and computing resources

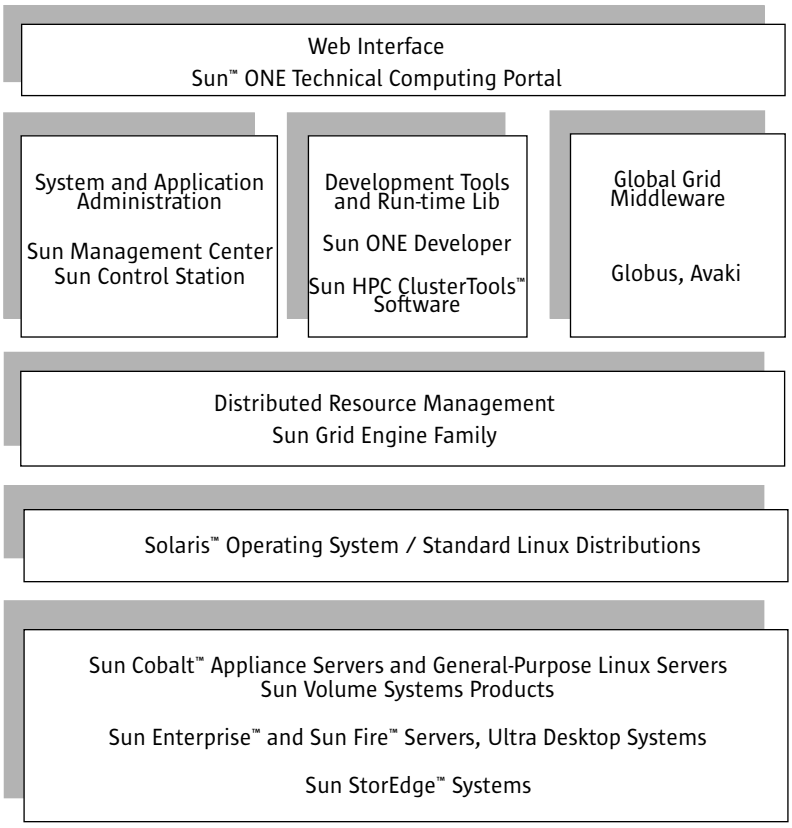
Sun — The Power Behind the Grid

Despite news to the contrary, the grid is not a futuristic fantasy or technological pipe dream. The grid is here — and available from Sun — today. In fact, many Sun customers are exploiting the power of distributed computing with grid technology right now. Indeed, in the course of just one year, Sun's powerful Grid Engine software is managing over 100 000 CPUs worldwide, enabling organizations to put their untapped computing power to work.

How hard is it to build a grid? Sony Devices Europe created a Sun grid in just two days. How big can grids get? Inpharmatica uses a grid with 500 CPUs to discover novel new drug therapies, and Ford Motor Company employs 1000 CPUs for MCAE tasks. How capable are grids? The Durham University Cosmology Engine performs 465 billion arithmetic operations per second on a Sun Cluster Grid. What happens if a grid is not big enough? Organizations can incorporate more resources into their grids, or rent computing resource time from providers, dynamically configuring and reconfiguring the grid as needs dictate. How confident is Sun that Grid Computing is real? Sun itself has a 4000-CPU Enterprise Grid that features 98 percent CPU utilization executing over 50 000 EDA jobs a day. We know it works.

Grid computing requires powerful, reliable, high-performance systems and sophisticated software that bring the power of computing to bear on complex problems. Sun Microsystems, Inc. — the power behind the Internet — is now the power behind the Grid. A technology leader, Sun provides the components that organizations need to build small or large, local or global grids today, including:

- *Sun's entry-level servers*, including the Sun LX50 server, the Sun Fire™ V60x server and Sun Fire V65x server, and the Solaris™-based Volume Systems Products, systems ideally suited for Grid Computing.
- *Sun Control Station*, a systems-management appliance that makes managing a Grid simple and affordable.
- *Sun Management Center*, providing a single point of management for all systems in the Cluster Grid.
- *Sun™ Grid Engine* family, sophisticated software that aggregates available computing resources and delivers compute power as a network service.
- *Sun Open Net Environment (Sun ONE) Technical Computing Portal*, a reference architecture that enables users to access the grid through a familiar Web interface.
- *Sun ONE Development platform and run-time tools*, giving developers the tools needed to create scalable, reliable, high-performance computing applications.



Sun microsystems

- *Solaris™ Operating System,* designed to deliver the power, flexibility, availability and binary compatibility to support Grid Computing. It combines key computing elements — operating system, networking and user environment — into a stable, high-quality foundation that organizations can depend on to develop, deliver and manage the grid.
- *Sun StorEdge™ products*, a complete line of high-volume RAID storage systems that provide over 200 TB of storage capacity.

Standards are a response to an industry need — they may come from formal standards processes or *de facto* industry acceptance of a technology. Sun understands their importance, regardless of origin, and is engaged in helping establish useful, widely applicable standards that benefit customers. UNIX®, TCP/IP, NFS, Java™ — all stand witness to Sun's leadership in delivering on the promise of open systems. Sun continues this trend with the Distributed Resource Management Application API (DRMAA), a standards initiative for Grid Computing that aims to facilitate application portability across DRM implementations. Open and community source projects, such as NetBeans™, Sun HPC ClusterTools™ and Sun Grid Engine software, support Sun's belief in collaborative development, helping users accelerate the development and deployment of distributed computing technology.

Sun is constantly shaping the future of computing by investing in new technology. Sun recognizes that a consistent, continuous application of time and resources is needed to meet the needs of a rapidly changing computing marketplace. Significant investments in high-performance, low-cost servers, advanced software environments and tools, and adherence to standards ensure that Sun customers will always have access to the best products available. Sun also pursues alliances with other industry leaders in a concerted effort to deliver new products and technologies that foster greater productivity, higher-quality products, reduced time-to-market and ultimately, an improved bottom line.

The powerful combination of robust Sun servers and sophisticated distributed resource-management software gives organizations a fully integrated solution for Grid Computing — right now. Today.

Sun Microsystems, Inc. 4150 Network Circle, Santa Clara, CA 95054 USA 1-650-960-1300 or 1-800-555-9sun www.sun.com

AFRICA (NORTH,WEST AND CENTRAL): +33-13-067-4680 • **ARGENTINA:** +5411-4317-5600 • **AUSTRALIA:** +61-2-9844-5000 • **AUSTRIA:** +43-1-60563-0 • **BELGIUM:** +32-2-704-8000 • **BRAZIL:** +55-11-5187-2100 • **CANADA:** +905-477-6745 • **CHILE:** +56-2-3724500 • **COLOMBIA:** +571-629-2323 • **COMMONWEALTH OF INDEPENDENT STATES:** +7-502-935-8411 • **CZECH REPUBLIC:** +420-2-3300-9311 • **DENMARK:** +45 4556 5000 • **EGYPT:** +202-570-9442 • **ESTONIA:** +372-6-308-900 • **FINLAND:** +358-9-525-561 • **FRANCE:** +33-134-03-00-00 • **GERMANY:** +49-89-46008-0 • **GREECE:** +30-1-618-8111 • **HUNGARY:** +36-1-489-8900 • **ICELAND:** +354-563-3010 • **INDIA: BANGALORE:** +91-80-2298989/2295454; **NEW DELHI:** +91-11-6106000; **MUMBAI:** +91-22-2018141 • **IRELAND:** +353-1-8055-666 • **ISRAEL:** +972-9-9710500 • **ITALY:** +39-02-641511 • **JAPAN:** +81-3-5717-5000 • **KAZAKHSTAN:** +7-3272-466774 • **KOREA:** +822-2193-5114 • **LATVIA:** +371-750-3700 • **LITHUANIA:** +370-729-8468 • **LUXEMBOURG:** +352-49-11-33-1 • **MALAYSIA:** +603-21161888 • **MEXICO:** +52-5-258-6100 • **THE NETHERLANDS:** +00-31-33-45-15-000 • **NEW ZEALAND: AUCKLAND:** +64-9-976-6800; **WELLINGTON:** +64-4-462-0780 • **NORWAY:** +47-23-36-96-00 • **PEOPLE'S REPUBLIC OF CHINA: BEIJING:** +86-10-6803-5588; **CHENGDU:** +86-28-619-9333; **GUANG ZHOU:** +86-20-8755-5900; **SHANGHAI:** +86-21-6466-1228; **HONG KONG:** +852-2202-6688 • **POLAND:** +48-22-8747800 • **PORTUGAL:** +351-21-4134000 • **RUSSIA:** +7-502-935-8411 • **SINGAPORE:** +65-438-1888 • **SLOVAK REPUBLIC:** +421-2-4342-94-85 • **SOUTH AFRICA:** +27-11-256-6300 • **SPAIN:** +34-91-596-9900 • **SWEDEN:** +46-8-631-10-00 • **SWITZERLAND:** *GERMAN:* 41-1-908-90-00; *FRENCH:* 41-22-999-0444 • **TAIWAN:** +886-2-8732-9933 • **THAILAND:** +662-344-6888 • **TURKEY:** +90-212-335-22-00 • **UNITED ARAB EMIRATES:** +9714-3366333 • **UNITED KINGDOM:** +44-1-276-20444 • **UNITED STATES:** +1-800-555-9SUN OR+1-650-960-1300 • **VENEZUELA:** +58-2-905-3800 **OR ONLINE AT SUN.COM/STORE**