Grid Engine Portal

(Formerly known as the Sun Technical Computing Portal)

Helping Make Information and Applications Accessible

Table of Contents

Chapter 1

Executive Summary

Increased network bandwidth, more powerful computers, and the acceptance of the Internet have driven an on-going demand for new and better ways to utilize high performance and technical computing (HPTC) resources. 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 work. 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 — at the same time, computational needs outpace 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. Grid Computing is a new methodology that combines networked resources to deploy massive computing power wherever and whenever it is needed most.

While Grid Computing aggregates resources, a mechanism is needed that enables the power of the grid to be accessed with ease. Using portal technology, organizations can aggregate information, standardize tools, and provide global access to HPTC applications and information, giving organizations the agility they need to provide distributed shared services to users through a single point of access.

The Grid Engine Portal (formerly known as the Sun Technical Computing Portal) is an enabling technology aimed at making Grid Computing easily applied to enterprise challenges. Designed to support HPTC environments, it provides a distributed architecture that can deliver

compute and data resources over the Web, making resources available where and when they are needed most. A powerful tool for accessing aggregated computing resources, Sun's Grid Engine Portal (GEP) enables organizations to simplify and manage access to HPTC applications, all through an easy-to-use browser interface.

At Sun — where "The Network is the Computer™" — Sun ONE portal technology is key to delivering a highly economical, efficient platform for high performance and technical computing. Indeed, while some universities have developed technical portals for in-house use, no commercial competitors offer an integrated solution with GEP functionality to date. Sun is leading the way in portal computing for HPTC environments with the GEP solution. With twenty years of experience developing high performance, robust networked computing technology based on open systems, Sun understands the architecture and infrastructure required for portals to meet business objectives. With Sun, organizations can capitalize on scalable, robust, best-of-breed solutions that enable the rapid development, deployment, and evolution of portals. Indeed, the combination of Sun systems, the Solaris™ Operating Environment, remote management tools, and key technologies like Sun™ Grid Engine join forces in Sun's GEP to help enable organizations to capitalize on the power of portal computing.

Chapter 2

Portals and Grid Computing

Meeting today's challenges requires agility. IT organizations must find ways to optimize data center efficiency, productivity, application sharing and collaboration, yet manage increasing growth and complexity. In addition, IT managers are under pressure to find ways to increase the return on investment for high value, technical computing resources. To be effective, organizations must find ways to improve resource management and harness idle computing resources, expand application access and resource sharing, and eliminate redundant data and work. One approach is to transition from application environments consisting of individual networked computer systems to distributed shared services that are readily available to an expanded community through a single point of access.

Grid Computing

Grid Computing is a methodology that enables organizations to use their computing resources more efficiently and flexibly. At the heart of Grid Computing is a computing infrastructure that uses distributed resource management software to provide dependable, consistent, pervasive and inexpensive access to computational assets. 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.

Conceptually, a *grid* is quite simple: it is a collection of computing resources that are connected through a network and managed by distributed resource management software. By pooling federated assets, a grid enables users to treat powerful distributed systems 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. As a result, Grid Computing offers the possibility for anyone to have access to advanced computational resources from anywhere in the world over the Internet or organizational intranets any time they need to solve problems.

Scalability from Local to Global Computing

Every organization employs a unique mix of computing resources. 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 a collection of stand-alone systems to supercomputer-class clusters utilizing thousands of processors. With grids, users can be assured of needed performance, and no resources are wasted.

- Departmental Grids

   Today, Departmental Grids (also known as Cluster Grids) are the most popular and simplest form of a grid. Meeting the needs of departments or projects within an organization, Departmental Grids consist of multiple systems working together to provide a single point of access to users in a single project or department. 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, organizations can employ Departmental Grids to complete many individual jobs quickly, or run simulations on multiple machines that communicate with each other to solve a set of parallel problems.

- Enterprise Grids

   As capacity needs and demands for greater economy increase, organizations can combine their Departmental Grids into Enterprise Grids. Enterprise Grids enable multiple projects or departments within an organization 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

   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 over the Internet. 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.
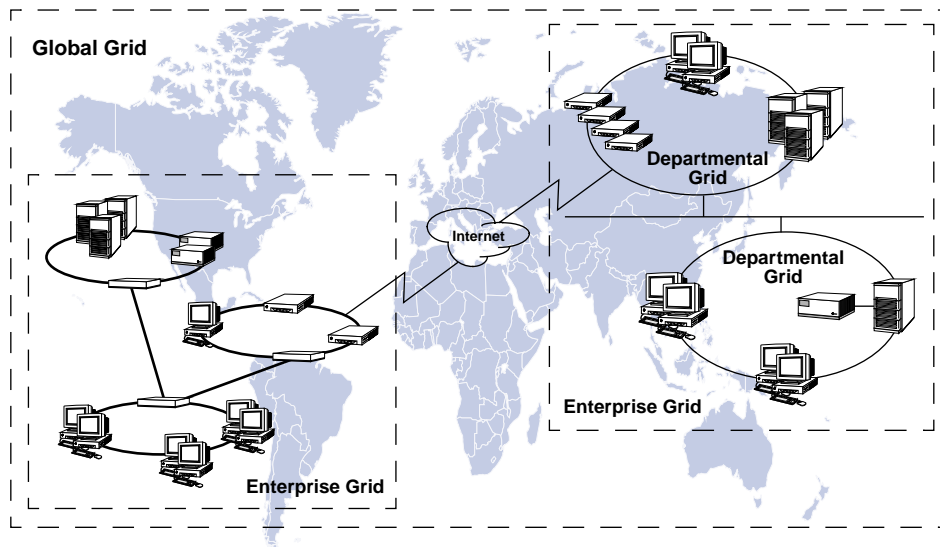

Figure 2-1: Departmental Grids consist of multiple systems that provide a single point of access to projects or departments. Enterprise Grids consist of mulitple Campus Grids and enable multiple projects or departments to share resources. Global Grids consist of multiple Enterprise Grids and enable access to distributed resources over the Internet.

Portal Computing

Delivering the power of the grid requires a mechanism that enables easy, secure access to aggregated resources. Secure, anytime, anywhere access to distributed technical applications and data is possible through portal architectures. A *portal* is a secure Web environment that enables an organization to aggregate and share content — information, services, and applications — with customers, business partners, employees and suppliers. Portals bring together technology, business processes, and business partners, enabling the organization to exchange information inside and outside the firewall.

A portal allows an organization to employ a single URL through which users can gain browser-based access to customized or personalized information, as well as applications. With a customizable interface that provides content that meets each individual user's specific needs, portals spare users from needing to hunt for information by making it accessible from any networked device. By aggregating information, standardizing tools, and providing global access to applications and information, portals give organizations the agility they need to provide distributed shared services to users through a single point of access (Figure 2-2).

Figure 2-2: The basic function of a portal is to provide easy access to information and applications.
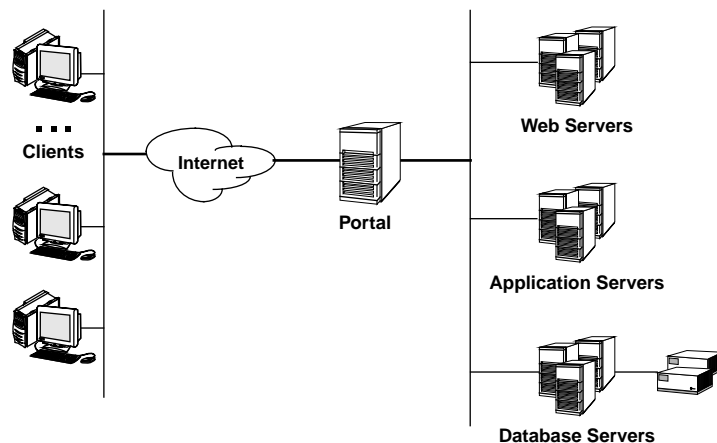
Portal Computing and the Grid

Today, portals can be used to access a single server, a cluster of systems, or even a complex grid of workstations and servers, creating opportunities for more organizations and their users (Figure 2-3). Indeed, employing portal technology in HPTC environments offers the ability for users to work more effectively by allowing complex jobs to run with greater efficiency and at lower cost. Multiple applications can run across clusters of servers that are combined into a unified Grid Computing system, giving users easy access to complex applications and sophisticated computing resources.
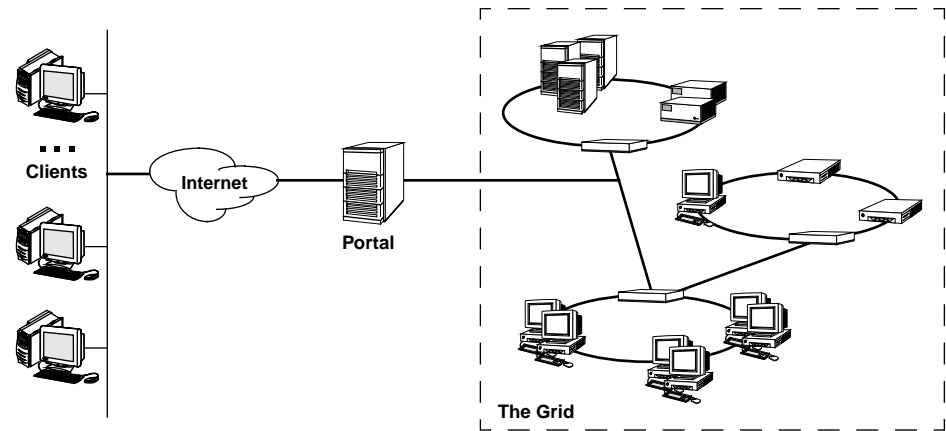
Figure 2-3: In HPTC environments, a portal facilitates the sharing of high-value computational resources.

Chapter 3

The Grid Engine Portal Solution

The Grid Engine Portal (formerly known as the Sun Technical Computing Portal) is an integrated solution for high performance and technical computing (HPTC) environments that offers a new class of Web services to organizations. Based on the Sun ONE Portal Server and the Sun Grid Engine resource management software, the Grid Engine Portal (GEP) solution gives organizations a single, unified interface to complex technical applications. This advanced solution can bring the operating advantages of portal technology to HPTC environments by combining portal capabilities, resource management software, and easy to use tools. As a result, the GEP solution helps organizations to centralize application management and compute resources, as well as provide highly secure communication and applications access to a large and diverse set of users — without increasing time, cost, risk or effort.

The GEP Solution Architecture

High performance and technical computing is often characterized by non-interactive batch applications for which the user uploads an input file and submits a job, such as a simulation or query, that runs for hours or days. In the past, the control and monitoring of these jobs required separate secure access mechanisms for isolated systems, and specialized knowledge of the necessary UNIX® platform commands for each application. The GEP solution helps eliminate this complexity by giving users a single, highly secure access point through a browser interface for all applications, jobs, and projects. From this interface, even non-technical users can submit jobs, monitor the status of submitted jobs, and access the results of technical applications, as well as access other business applications.

The GEP solution utilizes a three-tier architecture in which the back-end is designed for HPTC (Figure 3-1). All three tiers can be independently configured to meet existing user requirements, yet scale to meet future demands. A Web browser provides the entire user interface to the computing environment. The middle tier consists of one or more Sun ONE Portal Servers, each capable of submitting jobs to the Sun Grid Engine software, which can automatically find back-end computational resources as they are needed.
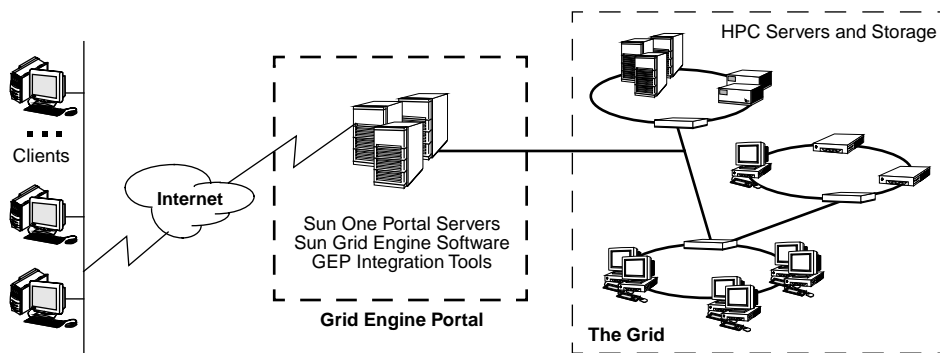
Figure 3-1: The GEP solution uses a 3-tier architecture to provide easy access to computational resources.

Sun ONE Portal Server

The Sun ONE Portal Server is one of the industry's first fully integrated, secure portal and identity management solutions. Using integrated identity management capabilities, the Sun ONE Portal Server enables organizations to protect information assets while building secure relationships and communities with and among customers, business partners, suppliers, and employees.

A community-based server application, the Sun ONE Portal Server aggregates key content, applications, and services that are personalized based on user role or identity, user preferences, and system determined relevancy. Providing an Internet/intranet services deployment platform, the Sun ONE Portal Server helps organizations to easily deploy technically demanding, secure portals and give users access to computing resources through a familiar Web interface. It allows people to access information and services that are important to them — efficiently and securely.

Several key features contribute to the Sun ONE Portal Server's ability to create and manage productive, secure, and cost-effective HPTC portals:

- *Integrated portal services and identity management*, giving organizations the ability to manage and administer users and policies, and to provision services. The Sun ONE Portal Server can provide security services, enhance user productivity, facilitate delivery of new services, and reduce IT administrative overhead.

- *Single sign-on*, giving organizations the flexibility to continue to add services, including Web services, while making access more convenient for users. Alongside traditional portal applications, such as calendar and email, single sign-on also increases employee productivity by enabling access to all aggregated information and applications without the need to remember individual logins and passwords.

- *Enhanced security*, providing authentication and authorization services and enforcing fine-grained policy rules. Authentication methods can vary for different groups of users, helping ensure that the portal is highly secure, yet flexible enough to meet varying business requirements. For example, organizations can utilize a public key infrastructure (PKI), secure cards, certificates and other forms of security measures for user verification and authentication.

- *Secure search and delivery*, providing documents, files, and resources based on the user's role, helping ensure it only delivers results for documents that the user has permission to access.

- *Delegated administration*, enabling portal administration to be delegated to other users or lines of business within or even outside the organization.

- *Single administrative console*, providing a central location for administering and managing the portal, search, identity, and directory services, helping reduce administrative costs.

Sun Grid Engine Software

Sun Grid Engine (SGE) distributed resource management software helps optimize the utilization of software and hardware resources in heterogeneous networked environments. SGE aggregates computing resources and presents a unified and simple access point to users needing compute cycles. By distributing the computational workload to available systems, SGE can simultaneously increase productivity of machines and application licenses while optimizing the number of jobs that can be completed.

SGE accepts jobs submitted on behalf of users through the Sun ONE Portal server and schedules them for execution on appropriate systems based upon resource management policies. When a job is submitted, it is sent to a scheduler. After determining job requirements and priorities, the scheduler notifies a master daemon which places the job in a holding area, or *queue*, where it awaits execution. Queues run on server nodes and have specific attributes that describe the capabilities of the servers. An executive daemon starts the job from the queue, monitors its progress through to completion, reports server load conditions to the scheduler, and updates the master daemon upon completion.

At the time of submission, SGE takes note of the job's requirements for execution, including memory needs, execution speed, available software licenses, and more. Job requirements may be very different from one job to another, and only certain queues may be able to provide the corresponding service. As a result, as soon as a queue is available for execution of a new job, SGE determines suitable jobs for the queue and immediately dispatches the job with the highest priority or the longest wait time.

SGE provides traditional distributed resource management functions and more:

- *Batch queuing*, ensuring jobs run only when the required resources are available, enabling resource-intensive jobs to be processed efficiently.
- *Load balancing*, dynamically distributing jobs to make optimum use of all available computing resources, delivering significantly greater throughput for compute-intensive applications.
- *Job accounting statistics*, enabling organizations to monitor resource utilization and determine how to improve resource allocation.
- *Innovative dynamic scheduling and resource management*, enabling the enforcement of site-specific management policies.
- *Dynamic performance data collection*, providing the scheduler with up-to-the-moment job level resource consumption and system load information.
- *Certificate security protocol-based encryption*, helping enhance security by encrypting messages with a secret key.
- *High-level policy administration*, enabling organizations to define and implement enterprise goals, such as productivity, timeliness, and level-of-service.
- *Policy management*, monitoring the progress of all jobs and adjusting their relative priorities with respect to the goals defined by policy administration.

GEP Integration Tools

The Sun ONE Portal allows information, applications, and services to be displayed on a single browser page. The GEP Integration Tools are a series of *channels* — appearing as menus or windows on the browser desktop — that can allow the user to easily control and access the functionality of the Sun Grid Engine software. The GEP Integration Tools include the following default channels:

- *Project list*, providing users with the ability to create projects (directories and workspaces). System users create a project in which to upload input files and to write output files when jobs are submitted and run.
- *Application list*, providing users a list of applications that have been provisioned into the portal by an administrator or other authorized user.
- *Job list*, providing users of the portal with the ability to submit jobs to applications on available computing resources, and to monitor running jobs. Users can specify an email address to which notification should be sent upon job completion or termination.
- *Application control*, providing administrators and privileged users the ability to register applications with the portal and deploy them for access by users.
- *Job monitor*, providing administrators with the ability to perform system accounting functions.

These channels transparently integrate the resource management capabilities of Sun Grid Engine software. Organizations can create additional channels as needed with the Sun ONE Portal Server's Channel Wizard, administrative software that assists in the creation and configuration of channels for the Sun ONE Portal Server desktop. In short, the Channel Wizard eases the tasks associated with creating a channel — channel naming, content provider selection, channel attribute specification, channel profile creation, and addition of the channel to the list of available providers.

Grid Engine Portal — Making HTPC Resources Accessible

A powerful tool for securely accessing aggregated computing resources over the Internet or intranet, the Grid Engine Portal (GEP) offers tremendous benefits to organizations running HPTC applications.

- Easy access — anytime, anywhere
  Portal capabilities, resource management software, and new point-and-click tools join forces in the GEP solution to bring the power of portal technology to HPTC applications. By combining the capabilities of the Sun ONE Portal Server and Sun Grid Engine software into an integrated environment, GEP can provide users with secure anytime, anywhere access. Jobs can be submitted locally or remotely, and HPTC resources can be made available to users without duplicating resource investments.
- Improved resource management
  With GEP, organizations can improve resource management and harness idle computing cycles. The Sun Grid Engine software takes advantage of idle compute resources network-wide and matches them to individual job requirements. This can result in more processing power available per user — without requiring more resources — resulting in increased productivity and greater return on investments.
- Outstanding ease-of-use
  GEP helps make job submission and monitoring easy. With GEP, users can manage job submissions by filling out simple Web-based forms — no complex scripts or UNIX skills are necessary. Users are presented with a form that prompts for all required inputs. These inputs are automatically translated into a command line that is submitted to run the application. Furthermore, users can check job status dynamically, receive e-mail notification when jobs are complete, download and visualize output files remotely, and share results with ease.

- Easy access to legacy applications
  With the Grid Engine Portal (GEP), administrators can quickly add legacy applications to the portal for sharing and remote use — often within minutes. Typically, no application modification is required for batch applications that use a command line interface. In essence, a legacy application can be transformed into an Internet- or intranet-available application simply by plugging it into the GEP.
- Integratable solutions
  The GEP solution enables organizations to create solutions based on open standards and technologies, such as Java technology, HTML, XML, and JavaScript, helping ensure operability across heterogeneous platforms, systems, and environments. By using open standards, GEP gives organizations the flexibility to incorporate additional applications and capabilities as they are needed.
- Built on proven products
  Unlike proprietary solutions that are engineered specifically for niche markets, the GEP solution is built on robust, high performance, commercially available Sun products. As a result, the GEP provides a solution that can scale horizontally and vertically to provide optimum compute power and optimum data flow to any organization running HPTC applications.

Chapter 4

The Sun Advantage

Innovative Solutions for a Competitive Edge

High performance and technical computing (HPTC) environments are an invaluable resource for computing users in a wide range of industries. With the Grid Engine Portal (GEP), organizations can make computing resources and HPTC applications readily available to maintain a competitive edge.

- *Government agencies* can make compute resources accessible to users around the world. Jobs can be submitted remotely, making HPTC resources available securely to a global audience without duplicating investments.

- *Manufacturers* can create a secure environment for internal users that permits trusted outside partners to access in-house systems to run simulations and verify performance. Process sequencing can be simplified, as sequences of applications can be managed as a single application, and input and output files can be stored in dynamically assigned storage locations.

- *Energy companies* can place large, complex systems under the control of the Sun Grid Engine software and make them available to remote users through a portal — users no longer need to know where a job is running.

- *Bio-sciences organizations* can consolidate multiple Internet-enabled applications, or make other applications available via the Web, placing them in the grid and providing a consolidated interface to users.

- *Financial institutions* performing risk analysis and arbitrage can deliver powerful, sophisticated derivative and portfolio modeling and analyses to every branch office, alongside traditional financial applications, thereby empowering all (or selected) brokers to provide customers with more extensive services.

- *Educational institutions*, such as those offering local or remote education, or those needing to share resources and applications with students, researchers, and affiliated institutes, can make HPTC applications available to users via a standard browser.

- *HPC research centers* can centralize resources and make them available globally, providing unified access through a browser. As a result, user access can be simplified, resource utilization can be optimized, and technical data can be centralized, giving non-specialized users access to the tools they need to accomplish research assignments regardless of geographic location and skill level.

Sun Systems and Software

HPTC requires powerful, reliable, high performance systems and sophisticated software that bring the power of computing to bear on complex problems. A technology leader, Sun provides the tools organizations need to build small or large, local or global grids today, including:

- *Development and run-time tools,* giving developers the tools needed to create scalable, reliable, high performance computing applications.
- *Sun™ Management Center,* providing a single point of management for all Sun systems in the grid.
- *Solaris™ Operating Environment,* 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 workstations, servers, and appliances,* a comprehensive line of powerful desktop and server systems that can scale from 1 to 106 processors and more than half a terabyte of main memory.
- *Sun StorEdge™ products,* a complete line of high capacity RAID storage systems and software that can provide direct attached and shared access to Storage Area Networks (SANs).

Comprehensive Programs and Services

Sun provides a host of programs and services that help organizations accelerate the implementation of HPTC portals and maintain system and application quality.

iForce[SM] Ready Centers

Sun's iForce Ready Centers enable customers to simulate real-world production environments in a risk-free setting. These centers — located in Menlo Park, California; Tokyo, Japan; and Paris, France — enable customers to work jointly with experts from Sun and appropriate iForce partners to build and test proof-of-concept demonstrations and run pilot programs. Sun's iForce Ready Centers help organizations validate their architectures and accelerate the deployment of high quality HPTC implementations.

Sun Professional Services — Portal Services

Keeping availability high means relying on people and processes even more than products. Sun experts help integrate the methodologies, expertise, products, and services needed to craft HPTC portal environments. Based on field-proven methodologies, best practices, and experience, Sun's extensive portfolio of services can deliver the "how to" architecture, implementation, and operations management services for HPTC portal environments.

- *Portal architecture assessment,* helping organizations assess the technical ability of existing enterprise portals to meet tactical and strategic business objectives
- *Inception service for portal architecture development,* helping organizations employ use case analysis to scope and define portal business and technical requirements
- *Elaboration service for portal architecture development,* helping provide COTS selection, integration, prototyping, application architecture definition and validation
- *Technology strategy and architecture for COTS applications integration,* helping create content management systems, collaborative tools, search engines, personalization engines, databases, and business intelligence tools

- *Security and entitlement architecture for portal integrated applications and services*, providing single and common sign-on, global LDAP, and global entitlement

The SunTone Certification Program

HPTC processes are increasingly being delivered as services over the Web. To improve the effectiveness of Web-based delivery, Sun has developed the SunTone[SM] certification program. This program certifies Sun technology partners who demonstrate their excellence in designing and implementing applications that are delivered as services over the Web. SunTone skills encompass infrastructure, operations, hardware, software, storage, security, and overall service delivery. Originally created for Service Providers, the SunTone program has expanded to audit and certify applications, integrator services, pre-integrated hardware/software platforms and now, enterprise data centers that have been built to the exacting standards of the program.

Sun's Commitment

Sun has a strong long-term commitment to the high performance and technical computing market, as evidenced by its continued R&D investments, strategic acquisitions and partnerships. Since 1995, Sun has acquired significant HPTC expertise and talent through key acquisitions including Kendall Square Research technology, Cray Research BSD, Thinking Machines Globalworks Division, Dakota Scientific, Max Strat Storage, GridWare, LSC, and most recently, iPlanet portal technology (now Sun ONE Studio) — which created the opportunity to pioneer a new class of Web services specifically tailored to HPTC needs.

Sun is a first-tier ISV partner for technical computing applications, and is also deeply involved at the industry level. For example, Sun's Government Industry group provides complete technology solutions that enable federal, state, and local governments and agencies to take advantage of exponential growth in bandwidth, devices, services, and users. Sun has established similar initiatives to focus on other markets such as manufacturing, education, energy, discovery life sciences, and others.

In 2001, Sun reached the No.2 position in the $5.1B HPC Server market with a 21% market share (just 1% away from the market leader). Over the past five years, Sun has had the most rapid market share growth of any vendor in the technical server market, according to the research firm IDC.

For More Information

Table 4-1 identifies other sources of information related to Sun products and service offerings. Additional information on the Grid Engine Portal solution can be obtained by sending email to *GEPortal@sun.com*.

Organizations can also contact a local Sun sales representative to learn how Sun can help build competitive advantage with a Grid Engine Portal that meets the needs of today and tomorrow.

Table 4-1: Web links for additional information on Sun products and service offerings.

Web Site URL	Description
http://www.sun.com/hpc	Sun High Performance and Technical Computing
http://www.sun.com/solutions/hpc/access.html	HPTC Access Solutions
http://wwws.sun.com/software/products/portal_srvr	Sun ONE Portal Server
http://wwws.sun.com/software/gridware/	Sun Grid Engine Software
http://www.sun.com/solaris	Solaris Operating Environment
http://www.sun.com/servers	Sun Hardware Servers
http://www.sun.com/sunmanagementcenter	Sun Management Center Software
http://www.sun.com/storage	Sun Network Storage Solutions
http://www.sun.com/security	Sun Security Solutions
http://www.sun.com/iforce	Sun's iForce Initiative
http://www.sun.com/suntone	SunTone Certification and Branding Program
http://www.sun.com/service	Sun Service Offerings
http://www.sun.com/sunone	Sun Open Net Environment (Sun ONE)