Global Headquarters: 5 Speen Street Framingham, MA 01701 USA P.508.872.8200 F.508.935.4015 www.idc.com

WHITE PAPER

Sun Servers and Oracle Grid

Sponsored by: Sun Microsystems

Jean S. Bozman John Humphreys

November 2005

EXECUTIVE SUMMARY

Clusters and grids have achieved broad acceptance in technical and high-performance computing environments, and they are widely discussed for their potential applicability in commercial environments for enterprise applications. Having a computing environment that is highly available, resilient, flexible, and efficient is clearly desirable for any organization seeking to gain competitive advantage from its IT investment. The grid approach offers a way to simplify existing "islands of computing" that have grown up within the enterprise, over many years, as a result of waves of server acquisition. The inherent flexibility of grids allows organizations to leverage their existing server investments and to add new servers for capacity as needed.

For grids to gain more widespread commercial acceptance, the technology must break out of the technical space and provide real business value to commercial organizations. It must be easy to understand, acquire, deploy, and manage. But all of that would be a daunting challenge without the increasing use of hardware and software products that are shipped grid-enabled so that they can be snapped into place as the grid expands over time.

Sun has a long and successful history of delivering highly scalable servers for enterprise applications. It also understands, from its early roots in network computing, the notion of data services being the focus of workload deployment over the network. Sun is looking to leverage these strengths into delivering the virtualization and provisioning capabilities, applications support, services, and flexible business models that enterprise customers require in moving toward a service oriented architecture. Sun defines its grid offerings as a pool of IT resources that are brought together and managed to appear to an application or to a user as a single entity. The objective is to offer compute resources as needed for a particular business function.

Beyond the fundamental technologies and an inherent understanding of enterprise customers' needs in this space, Sun has begun to put together an ecosystem to help the vision become reality through efforts such as its iForce third-party program and its Reference Architectures, which provide guidelines for implementing grid capability. Sun is working with software partners to ensure that grid-enabled systems can be deployed with the assurance that they have been pretested and precertified to run as part of the grid.

With respect to the database, Sun is working closely with Oracle to combine their resources, expertise, and technologies to bring grid technology into the enterprise space. Sun supports the Oracle Database 10*g* on Solaris SPARC systems, as well as Solaris x86 systems and Linux systems. This support across multiple platforms combines with Sun's N1 software for grid enablement and its associated products that support hardware virtualization and provisioning to allow workloads to be directed to available computing resources. This is Sun's approach to matching IT requirements with available resources and moving toward a more flexible and adaptive computing environment. Sun's approach to grid computing is to provide a holistic view, one that involves integration of multiple types of resources — different processor platforms, operating systems, applications, and other components — and facilitates interoperability among them.

SUN GRID SOLUTIONS AND ORACLE INTEGRATION

Sun and Oracle are working very closely to create a valuable grid solution for the database. The Oracle Database 10*g* is available on the Solaris 10 operating system (SPARC and x86), and the Oracle Application Server 10*g* is expected to be certified on the Solaris 10 OS for Intel Xeon and AMD Opteron™ processors later this year. In addition to certifying Oracle Database 10*g* on its platforms, Sun offers a number of grid products that are fully integrated with the Oracle solution.

In an Oracle grid implementation, Sun offers the Grid Engine, a de facto macro grid manager that is fully integrated with Oracle Grid Control. This permits the viewing of all resources in the grid and scheduled jobs. Another component to the Oracle grid solution is Sun N1 Service Provisioning System (SPS), Sun's application provisioning software, which provisions and manages the various software technologies present in the grid. It supports agents for Oracle 9i, Oracle 10*g*, and the Oracle Application Server.

SPS keeps track of the service-level requirements of applications and automatically and dynamically provisions resources within the datacenter grid, aligning application requirements with the appropriate computer, network, and storage resources. Once the template is established, no manual intervention is required. According to Sun, automatic provisioning can reduce the provisioning process to a matter of minutes versus the hours or days needed for manual provisioning.

Sun N1 System Manager performs operating system provisioning. If a customer has a new system for the grid, for example, System Manager would be used to provision it with the operating system (e.g., Linux, Solaris), preparing it for application provisioning and then installation into the grid.

An important feature of the Solaris 10 operating system is its container capability. Solaris Containers allow multiple execution environments to be created within a single instance of the operating system, each with its own identity. This enables the dynamic scaling of 10*g* (or Oracle 9i) on large Sun midrange or high-end servers — allowing for finer-grained partitioning of resources.

Because an Oracle 10*g* solution relies upon several machines to be clustered together to provide scalability and throughput, customers are also encouraged to look for strong clustering solutions in the underlying infrastructure. Sun has recognized that need and brought its Sun clustering solution together with Oracle to create a solution with availability and scalability in both the infrastructure layer and the database.

At the end of the day, most datacenter customers are running multiple databases, operating systems, and applications in a multivendor environment. Sun's approach to grids is to create a macro grid to manage all of these resources efficiently and allow them to interoperate across the entire IT environment. Sun's technologies integrate with the Oracle grid environment to create a more comprehensive, heterogeneous grid environment for the database.

The Sun Cluster HA for Oracle Application Server can be used to provide high availability for either the Oracle 9iAS Infrastructure or Oracle 10*g* AS Infrastructure. Oracle 9iAS (versions 9.0.2 and 9.0.3) and Oracle 10*g* AS (version 9.0.4) introduced a new component called Oracle Infrastructure. This infrastructure component provides centralized product metadata, security and management services, configuration information, and data repositories for middle tier installations.

The Sun Cluster HA for Oracle Application Server data service provides high availability for either the Oracle 9iAS Infrastructure or the Oracle 10*g* AS Infrastructure components. The Sun Cluster HA for Oracle is also required to provide high availability for the Oracle Database.

Sun provides improved availability of data and applications through two solutions: a clustering software solution for mission- and business-critical applications and a truly fault-tolerant computer platform. (The Sun Netra ft1800 has been discontinued.) The new release of Sun Cluster 3.1 software provides high-availability configurations for the Ultra Enterprise, Netra, x86, and Sun Fire server product families. Using this software, customers can connect up to 16 individual server nodes. A shared global file service allows all server nodes to share all files in the cluster, and a high-speed interconnect is the physical link between the individual server nodes.

Alternately, Sun Cluster 3.1 can be used to provide high-availability failover to applications that are running in multiple "partitions" within a single, scalable Sun server. Sun's scalable servers can range from midrange enterprise Sun Fire servers, which support 2–4 partitions to a single Sun Fire 25K server, which supports up to 16 partitions. Sun also provides the Sun Cluster 3.1 software as a key element of the preconfigured SunPlex clustered configurations of hardware, software, interconnects, and storage.

CHALLENGES FOR IT ORGANIZATIONS

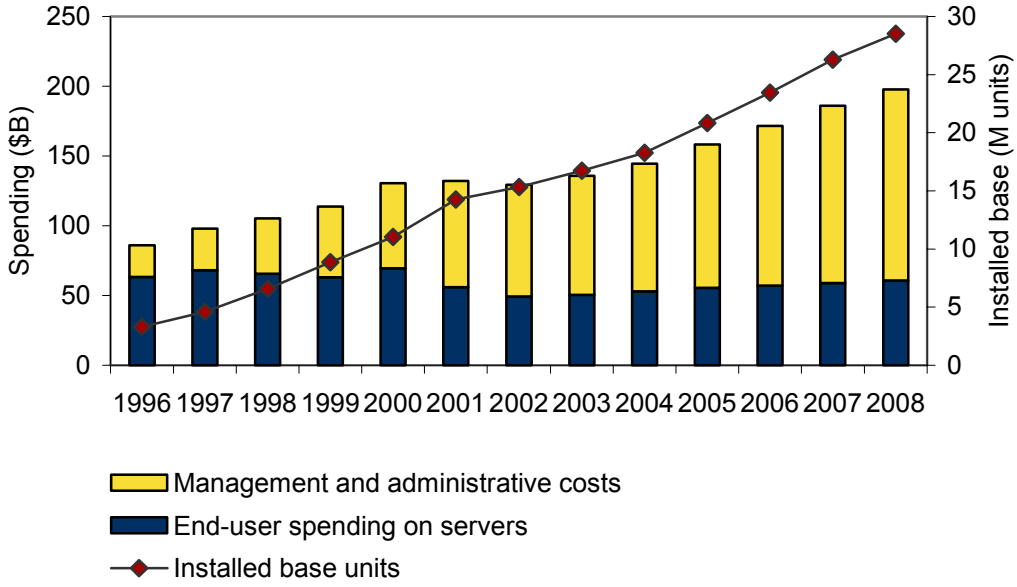
As enterprises continue to make IT investments cautiously, IT managers are concerned with deploying the new applications and infrastructure technologies that will better align their IT environments with their organizations' overall business priorities. Many of these environments are exceedingly complex, and IT managers are looking for more efficient ways to manage them and for the flexibility required to more effectively support changing and often unpredictable business requirements. IDC research consistently shows that the server resources in distributed environments are underutilized — only about 15% of capacity is effectively utilized in some cases. IT managers are concerned with taking better advantage of available capacity and making it available to support the workloads that drive the business. Many IT managers are also looking to reduce their expenses associated with overprovisioned software licenses that are accumulated during the acquisition of multiple distributed systems.

Today's IT managers spend a lot of time worrying about complexity and cost associated with the collection of servers, acquired over time, that they have inherited from previous IT management in their companies. This situation, in which inherited infrastructure generates unacceptable operational costs associated with system administration and management, is all too commonplace (see Figure 1). IT managers can take various paths to reduce this complexity, including workload consolidation and/or data consolidation, replacement of aging server systems, and reduction in the number of separate systems that must be supported. One path that is just opening up to customers is that of deploying grid technology across their networks. Grid topologies are based on several key concepts, including:

- Direction of computing tasks to appropriate resources within the grid
- High-speed links between a variety of servers, allowing workloads to be shifted to the appropriate server resource
- Virtualization software that allows existing hardware resources to be more efficient in leveraging hardware-based compute resources — and in making those resources available to applications and workloads
- Provisioning software that moves workloads from one server to another
- Scheduling software that decides when and where computing tasks are to be run on a variety of servers located across the grid (Support for business policies, as set by the line-of-business [LOB] units, means that IT capabilities and IT infrastructure can be more closely aligned with shifting business priorities. So as business priorities shift with time, the supporting IT infrastructure is able to adjust as well.)

FIGURE 1

Cost of Management: Server Capital Versus Operational Costs



Source: IDC, 2005

All of the previously mentioned components of grid computing must be orchestrated together, usually by middleware software. But even deployment of grid technology is not enough; the applications and workloads must run on reliable hardware resources and at reasonable costs for hardware, software, and services.

The orchestration of the workloads across multiple servers, including virtualization, provisioning, and management of workloads, causes software costs associated with IT personnel and operational costs to rise, as Figure 1 shows. If companies are to control these costs, they must deploy the overall grid architecture efficiently, and the software that runs across servers must be effective in automating operations. When combined, these factors can keep customer costs and IT costs under control.

How Grid Technology Addresses the Issues

Grid computing offers IT organizations tremendous potential in addressing performance and utilization concerns by dynamically and automatically directing workloads to where they will optimally run. By closely tying service level agreement requirements to the provisioning of resources, grids can allow IT organizations to provide higher and more consistent levels of service to the business they support within constrained budgets.

Reducing operational costs is an important concern for IT organizations because they continually face the challenge of providing higher levels of support within constrained budgets. Although grid technology may be complex to implement and require justification of the investment, IT managers recognize that being able to pool and allocate resources in real time offers significant opportunity for not only removing costs from the operation of the IT infrastructure but also better aligning IT with the goals of the business. IT staff can quickly and easily add new instances of databases, applications, or nodes. The scalability inherent in a grid environment also allows rapid introduction of new components and technologies into the infrastructure to expand capabilities as required.

Achieving high levels of availability is essential in maintaining service levels and lowering the cost and lost productivity associated with downtime. A grid environment, with a pool of available resources and the software to redirect workloads to those resources, supports higher availability. When the grid environment is run on highly available and resilient platforms, such as those from Sun, even higher levels of availability are possible.

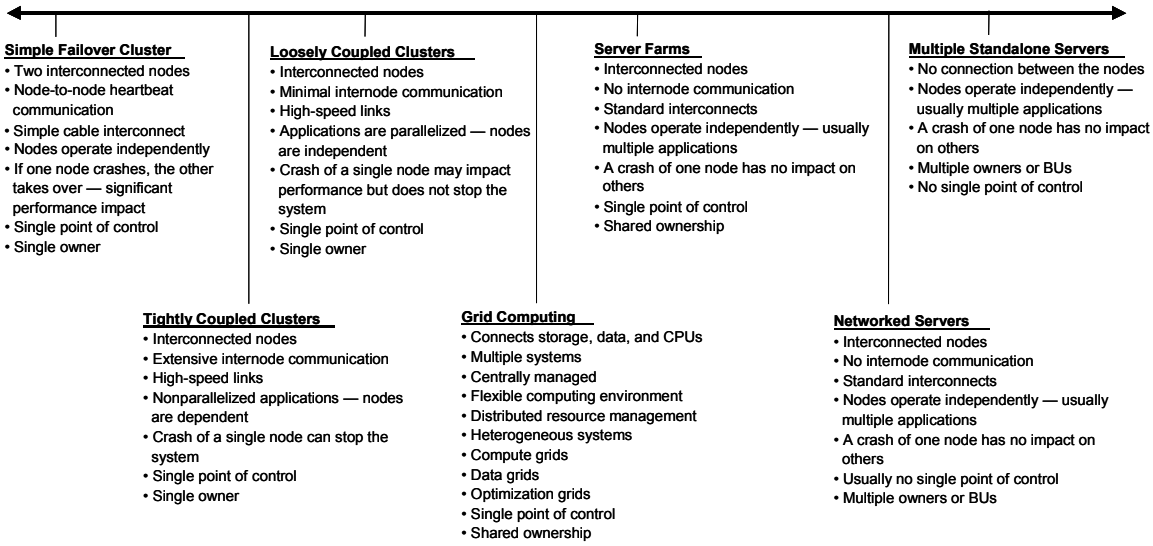
Grids also allow common management across diverse computing resources. Common management can greatly improve the efficiency and productivity of IT staff — affecting both the amount of time the system administrators spend on given tasks and the number of users they can support. Advanced grid-enabling software, combined with intelligent management software, can reduce the full-time equivalents (FTEs) tracked by IT managers in association with grid deployments. Thus, grid deployments have lower operational costs than IT deployments with information silos or islands of automation.

How Clusters and Grids Work Together

Grids are systems that span the network — and they can include individual servers as well as servers that are linked together into clusters. Both clusters and individual servers can — and do — participate in grids. The use of clusters does not prevent the use of grids. In fact, in some ways, because clustered systems support high levels of system availability, clusters can be thought of as the entry vehicle to a grid environment. Clusters provide the reliability and availability, and the grid technology provides inherent scalability. Figure 2 illustrates the spectrum of computing types including both clusters and grids.

FIGURE 2

The Computing Environment Spectrum



Source: IDC, 2005

Sun defines its grid offerings as a pool of IT resources that are brought together and managed as a single visible entity. It sees grids as allowing the creation of pools of resources that are virtualized to applications and to users with the objective of offering compute resources as needed for a particular business function. Sun sees its grid offerings as an array of choices, each based on grid technology.

IDC notes that both Sun and Oracle have included the word "grid" in the names of their products and that both companies have given specific meaning to these grid-based products. Sun has a Sun Grid offering, which customers can access as a data service to run their highly parallelized workloads — these jobs are submitted over the network. But Sun also supports a variety of grid deployments, including Oracle 10*g* grids, in which the Oracle 10*g* product runs on Sun servers. The combination of Oracle 10*g* and Sun servers — whether on Sun Fire UltraSPARC servers or on Sun's new AMD Opteron–based servers — results in a grid that supports a flexible style of computing. This is in step with industry initiatives designed to align business policies and IT infrastructure in order to support business "agility" in which servers can be added, dropped, and reconfigured as business conditions change over time.

INTRODUCTION OF ORACLE 10*g* AND ORACLE GRID TECHNOLOGY

One approach to delivering grid technology into an existing IT organization is to build an enterprise grid for a very specific application or set of applications. Oracle has essentially taken this tack with Oracle 10*g*, which it introduced into the marketplace in 2004. Oracle 10*g* combines a relational database engine with a modular product design that supports grid-style computing. Grids leverage network computing, workload management, directory services, and node-to-node connectivity to allow customers to orchestrate multiple servers into a single, coherent grid architecture.

The arrival of Oracle 10*g* in customer computing environments has allowed customers to leverage volume servers to design scalable, highly available groups of pooled infrastructure. Each individual server can access an Oracle database within the grid — and data can be moved around the nodes of the grid as needed. This parallel processing design improves overall reliability of the systems because there is no single point of failure. Further, additional servers can be added to the 10*g* grid deployment, as needed — such as when additional users start accessing workloads supported by the grid or at times of peak data processing (e.g., quarterly or annual peaks in data processing).

The Oracle 10*g* solution is composed of an extensible product set that includes the Oracle relational database, Oracle's grid software, and a set of software modules that support distributed computing. One example is Oracle Table Spaces, a feature that allows data tables to be transferred quickly between servers that are attached to Oracle 10*g* deployments that are in different locations — across a building, campus, or network.

A key feature of Oracle Database 10*g* is Real Application Clusters (RAC), which enables a single database to run across multiple clustered nodes in a grid, pooling together the processing resources of the systems. The Oracle 10*g* grid engine dynamically reallocates workloads across the resources, depending on demand for processing resources.

Oracle Database 10*g* also incorporates Oracle's Automatic Storage Management (ASM) software, which is designed to simplify storage configuration and management for the database. ASM automatically allocates storage workloads for the best possible system performance across a storage area network (SAN) or network-attached storage (NAS) device. This capability is important because reallocation of storage resources across physical storage devices is often a labor-intensive process that demands attention from database administrators and system administrators.

From a business perspective, Oracle's support for grid technology means that customers can combine installed — and new — servers into a grid infrastructure that shares data in an efficient manner and that moves data to available computing resources as needed. It supports a high degree of flexibility, in terms of the deployment of physical servers within a grid — allowing servers to be added on a pay-as-you-go basis — and in terms of reallocation of computing workloads across the grid once it is established.

SUN SERVER PLATFORM CHOICES

By supporting Oracle Database 10*g* across its server product line, Sun offers enterprise customers a broad choice of entry-level, midrange, and high-end platforms as well as the ability to use existing systems, thus preserving their hardware investments.

Sun's entry-level servers offer customers a choice of Sun UltraSPARC or AMD Opteron processors, depending on the model. Sun's larger UltraSPARC systems allow greater scalability and performance and also support the dynamic reconfiguration capability (supported on the 3800/3900, 4800/4900, 6800/6900, E20K and E25K platforms) that allows for greater control over the automatic allocation of workloads across system resources.

Customers can choose to run Solaris or Linux, allowing greater flexibility in meeting a range of requirements in the datacenter. This flexibility and choice are key elements of the value Sun brings to the grid picture. Customers can integrate Linux applications as needed. They can use existing Sun hardware. They can choose to scale horizontally by adding smaller nodes into the grid, or they can choose to scale up through Sun's highly scalable SMP servers. All of these elements can play together in the grid and exchange data easily.

With the coming of dual core technology, Sun also believes users will be able to scale diagonally, that is, a combination of scaling up in a traditionally horizontal platform such as rack-optimized servers.

THE POWER OF TWO: SUN SERVERS AND ORACLE 10*g* VALUE PROPOSITION

By teaming up on their grid efforts, Sun and Oracle are taking an important step toward making a dynamic, scalable computing more accessible and practical for enterprise customers that require flexibility, resiliency, and scalability in their computing environments. Oracle Database 10*g* supports scalability by the addition of instances of the database or applications. Sun platforms offer customers choice — of processors and operating systems — across a range of price and performance points. Sun also brings its highly scalable and available SMP servers into the grid, along with their dynamic reconfiguration and partitioning capabilities.

As customers look for flexibility and consistency across their environments, they are often faced with a staggering array of different hardware and software technologies. Sun and Oracle, working together, are removing the obstacles to making interoperability and provisioning work in such situations. They are investing in the integration work to remove the complexity and speed the time to deployment.

Although grid computing can ultimately simplify the IT environment and processes, the implementation is complex and requires much effort in streamlining the software stack. Integration among components and with the hardware is essential and can be difficult to achieve. Sun, by partnering with Oracle, is working to provide the glue between the pieces and to do the testing and integration work needed to provide a seamless environment.

BEYOND THE TECHNOLOGY: SUN'S VALUE-ADDED PROGRAMS AND SERVICES FOR SUN AND ORACLE SYSTEMS

Skill sets are key to gaining acceptance with new technologies — and smooth operations on an ongoing basis. There are many ways to improve datacenter skill sets, including through onsite education; online training; support from channel partners and systems integrators; and the acquisition of skills from outside the organization such as managed services, online services from service providers, and outsourcing.

Among the obstacles to commercial acceptance of grids are their perceived complexity and the lack of knowledge (and time) among IT staffs to understand their implications. Sun adds essential value in these areas by testing and certifying certain grid configurations. Sun publishes a number of Reference Architectures, which are predesigned, deployed, tested, and optimized platforms.

One such reference is the Sun ERP Grid Reference Architecture for Oracle Database 10*g* documents. This precertified configuration includes Sun Fire V20z Opteron-based servers running Red Hat Enterprise Linux 3 and Oracle Database 10*g* RAC, which is optimized for ERP and database applications.

Through its Reference Architectures, Sun removes much of the complexity and uncertainty involved in moving into a grid environment. There are many different ways to deploy the technologies, and enterprise customers benefit from having guidelines that have been pretested and optimized.

Sun is also working to expand the acceptance of enterprise grid computing through its iForce initiative. Sun is training its channel partners on the Sun/Oracle value proposition, the technologies involved, the reference architectures, and the implementation. Sun also offers its iForce centers around the world, where customers can work with partners to do proof of concept, testing, and benchmarking.

CHALLENGES AND OPPORTUNITIES

Deployment of Sun servers and Oracle 10*g* software within customers' grid topologies provides an efficient and resilient way to expand computing capabilities across an enterprise. Customers have a variety of options in such deployments, including the deployment of new and/or installed Sun servers into new grids of networked servers. Alternatively, customers may choose to base grids on newly acquired servers or on combinations of multiple clusters of servers, each running an Oracle database.

In the dynamic market for grid-enabled servers, Sun will compete with other server vendors, and the combined Sun/Oracle 10*g* solution will compete with grid solutions from a variety of system vendors and ISV software providers. Grids invite the inclusion of competing platforms because they support a diversity of computer systems and database engines that can be leveraged to support grid-enabled workloads across one or more IT sites.

Although grids support flexibility and adaptability of computer infrastructure, early adopters of enterprise grids can be expected to be more intense advocates for grid deployments than customers, or prospective customers, who are less familiar with grid technology and its business benefits.

IDC also notes that there is a range of customer sites and that some sites are more ready to deploy grid environments than others, given IT skill sets and experience with clustered server environments and enterprise grid environments. However, both clusters and grids are gaining in popularity, worldwide and within geographic regions, and thus IDC expects grid deployments to become more widespread in coming years. Initially, commercial customers that are new to grid technology will likely need more support from the systems vendors, ISVs, and channel partners that support their computing systems than they will when enterprise grid deployments become more widely adopted. IDC also believes that technical deployments of grids in high-performance computing have readied the path for enterprise grid deployments in sites that have already gained experience and confidence in working with grids.

CONCLUSION

Sun and Oracle have a long history of architecting and supporting clustered server environments. Together, they have tens of thousands of customers worldwide that utilize Sun products and Oracle products. In the emerging world of enterprise grids, Sun and Oracle are combining their resources, expertise, and technologies to make enterprise grids useful through their support for rapid deployment, testing and benchmarking, and ongoing maintenance and operations.

Sun's support for Oracle's 10*g* products extends across its hardware line, comprising both RISC-based and x86 servers, to the Solaris and Linux operating systems. This support across multiple platforms extends to Sun's N1 software for grid enablement, Solaris 10 containers, and provisioning that allows workloads to be directed to available computing resources. Sun is committed to making its grids a modern embodiment of its longtime motto: "The network is the computer." Grids confer flexibility and adaptability to the enterprises that deploy them — and Sun grids are being designed to allow customers to add a range of server resources to expanding grid deployments, harvesting existing computer resources for new uses, and including newly acquired servers for future computing needs.