An Oracle White Paper
April 2010

Oracle VM Server for SPARC
Enabling a Flexible, Efficient IT Infrastructure

ORACLE

Executive Overview

IT organizations constantly seek technology that can lower the cost and complexity of managing ever-expanding compute environments. Toward that end, enterprises need strategies and tools that help maximize the use of every compute asset and improve operational flexibility. Virtualization enables organizations to extract value maximum value from IT assets while creating an infrastructure capable of rapidly adapting to today's dynamic business environment.

Introduction

To remain competitive, organizations continually adjust existing business plans and work to create new income opportunities. Technology often holds the key to revenue improvement, creating requirements for IT organizations to deploy increasing numbers of compute services. With enterprise success at stake, many organizations avoid possible system security, availability, and performance conflicts by deploying only one application per server. Corresponding servers are then installed to support development, test, and disaster recovery initiatives**.**

These common datacenter management practices soon create a sprawling compute infrastructure that is difficult to manage and leaves systems largely underused. In addition, IT managers are startled to find datacenters quickly reach physical and budgetary constraints in the areas of power, cooling, and real estate. The resulting inefficient IT infrastructure carries cost burdens that can actually limit business growth—leading organizations to seek more-effective hosting strategies. This white paper highlights the advantages of using server virtualization to capture business value through an efficient, nimble IT architecture built on Oracle VM Server for SPARC (previously called Sun Logical Domains).

Improving Consolidation Strategies Through Virtualization

Seeking to raise efficiency levels and reduce costs, many IT organizations launch server consolidation initiatives. Although consolidation affords enterprises the opportunity to better use resources, deploying multiple applications on a single server often results in poor performance—and even application failure. For example, tuning and maintenance requirements might not easily align across a number of software programs. In addition, a single ill-behaved application can starve other colocated software services of resources. Simple consolidation methods fail to provide secure boundaries required by applications that access sensitive data.

Fortunately, virtualization technologies enhance consolidation strategies by allowing organizations to create administrative and resource boundaries between applications within a system. The ability to isolate software programs running within a consolidated server helps IT organizations deliver on application performance and security requirements, as well as meet custom tuning needs. By combining workloads and using virtualization techniques, enterprises can maximize their use of compute platforms, simplify their IT infrastructure, and bring new levels of efficiency, manageability, and agility.

A Broad Spectrum of Virtualization Technologies on SPARC

A number of technologies enable platform virtualization, each providing varying degrees of flexibility, availability, and security. In some cases, organizations benefit by using multiple tools within a single server deployment. By delivering advanced products and technologies, and leveraging its global expertise in providing systems and software, Oracle offers a full portfolio of virtualization solutions. Oracle's comprehensive approach addresses every virtualization category, including resource management, hard partitioning, operating system (OS) virtualization, and virtual machine (VM) technology (see Figure 1).

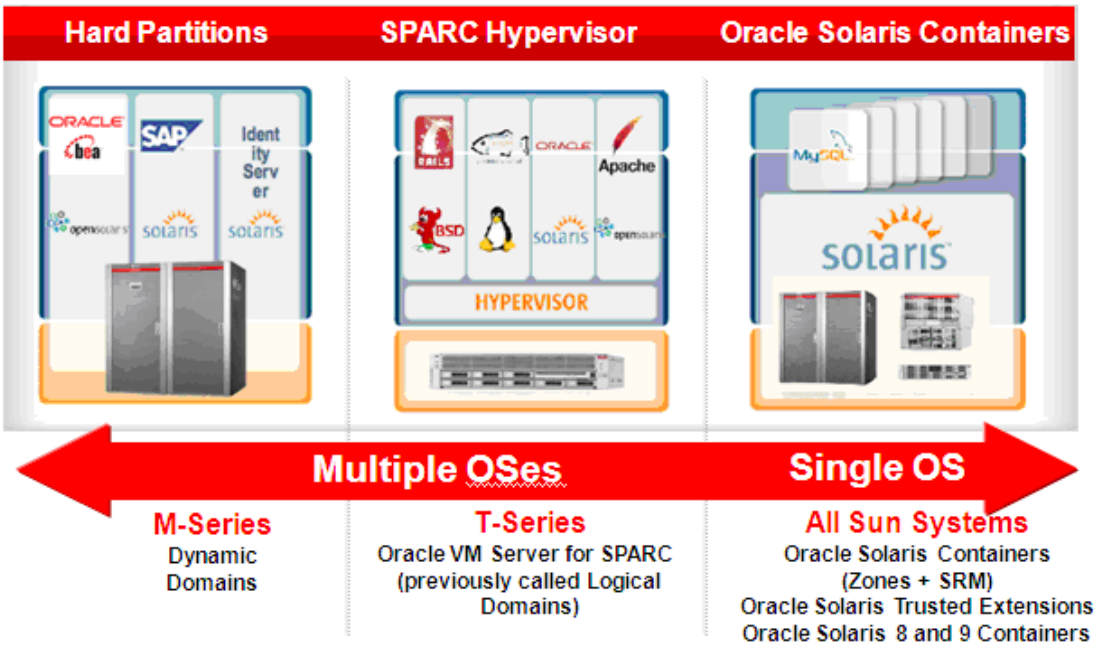

Figure 1. Oracle provides innovative server virtualization technology choices to enable more-effective IT consolidation solutions.

Hard Partitions

Hard partitioning assigns physical CPU, memory, and I/O resources to specific domains that run independent OS instances. Because isolation is instantiated all the way to the hardware, potential faults in one domain do not affect applications running in other domains—increasing reliability. However, resource assignment is only as granular as physically allowed by the hardware. Sun SPARC Enterprise M Series serversinclude the capability to create Dynamic Domains, which physically divide a single system into multiple electrically isolated partitions, each running a unique instance of Oracle Solaris.

Oracle Solaris Containers

Operating system (OS) virtualization technology creates many private execution environments within a single instance of an OS. With flexible, software-defined boundaries, virtual OS environments are independent of the hardware layer and available for all platforms that support the OS. As an example, Oracle Solaris Containers technology allows IT organizations to harness and provision compute power into a secure, isolated runtime environment for individual applications. Each environment holds a unique identity and maintains resource and namespace isolation. In addition, administrators can configure separate local area network (LAN) or virtual local area network (VLAN) connections with exclusive IP stacks for individual Containers, enabling secure separation of network traffic. By taking advantage of the capabilities of Containers, administrators gain the ability to exert fine-grained control over rights and resources within a consolidated server, without increasing the number of OS instances to manage.

SPARC Hypervisor

SPARC virtualization begins with low overhead cost, by using a small and efficient hypervisor layer that holds only the minimum state necessary to describe and control guest OSs. At the same time, the hypervisor is robust enough to achieve its two primary functions. First, it improves system security and reliability, effectively isolating each guest OS in its own logical domain. Second, it improves system efficiency and utilization, providing the ability to dynamically rebind virtual processors to physical resources at any time.

Oracle VM Server for SPARC directly integrates the hypervisor with paravirtualized guest operation systems.

Using paravirtualization, a guest OS such as Oracle Solaris 10 explicitly calls a support function implemented by hypervisor code rather than trying to access a system register itself. Much of the code that traditionally exists in the lowest layers of an OS is moved from the OS and placed in the hypervisor. Oracle VM Server for SPARC optimizes performance and security by implementing the hypervisor in firmware, setting virtualization technology apart from other hypervisor based virtualization technology.

Oracle VM Server for SPARC

Supported in all Sun SPARC Enterprise T Series servers that use Sun processors with chip multithreading (CMT) technology, a *logical domain* is a full virtual machine that runs an independent OS instance and contains virtualized CPU, memory, storage, console, and cryptographic devices. The natural level of granularity of a logical domain is an execution CPU thread, not a time-sliced microsecond of execution resources. Each CPU thread can be treated as an independent virtual processor. The scheduler is naturally built into the CPU, instead of the scheduler implementation at hypervisor which requires extra overhead . The Solaris dispatcher dispatches tasks to virtual CPUs which are effectively physical CPU threads. Moreover, the CPU threads can be dynamically changed in the running logical domains, and customers gain the benefits of increased service level and agility.

Agile, Secure, Efficient Virtual Machines

By taking advantage of Oracle VM Server for SPARC, organizations gain the flexibility to deploy multiple OSs simultaneously on a single platform (see Figure 2). In addition, administrators can leverage virtual device capabilities to transport an entire software stack hosted on a domain from one physical machine to another. Logical domains can also host Oracle Solaris Containers to capture the isolation, flexibility, and manageability features of both technologies. By deeply integrating logical domains with both the industry-leading CMT capability of the UltraSPARC T1, T2 and T2 Plus processors and Oracle Solaris 10, Oracle VM Server for SPARC increases flexibility, isolates workload processing, and improves the potential for maximum server utilization.

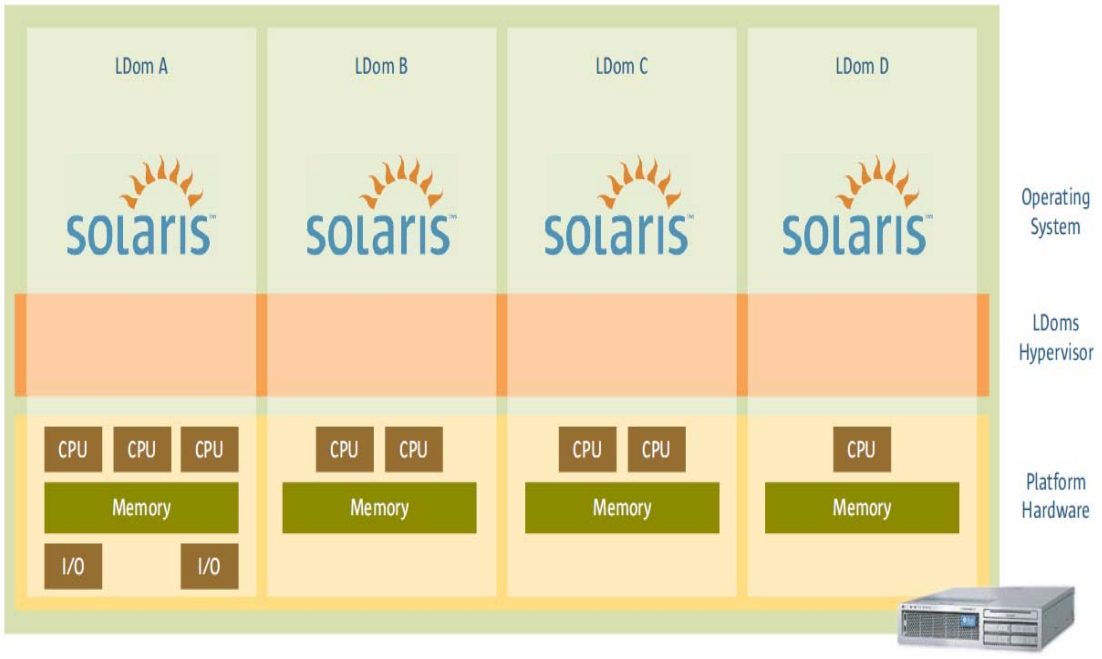

Figure 2. A single server hosts multiple domains, each running unique, isolated OS instances.

Oracle VM Server for SPARC Architecture

Oracle VM Server for SPARCcomplement Oracle's existing virtualization technologies and bring additional value to enterprises. For example, Oracle VM Server for SPARC enables partitioning of entry-level servers, a capability previously only available to midrange and high-end systems with Dynamic Domains. Oracle VM Server for SPARCcan also enhance compute resource management on systems using Solaris Containers technology for virtualization. The following key architectural components work together to accomplish the partitioning and isolation capabilities of domains.

- **Hypervisor.** A small firmware layer that provides a set of hardware-specific support functions to operating systems through a stable interface, known as the sun4v architecture. The hypervisor creates virtual machines by subdividing physical devices across multiple logical domains, exposing some resources to a specific partition and hiding others. In addition, the hypervisor creates communication channels, logical domain channels, between logical domains to provide a conduit for services, such as networks and shared devices.
- **Virtual devices.** Physical system hardware, including CPU, memory, and I/O devices, that is abstracted by the hypervisor and presented to domains within the platform.
- **Logical Domains Manager.** Software that communicates with the hypervisor and domains to sequence changes, such as the removal of resources or creation of a domain. The Logical Domains Manager provides an administrative interface and keeps track of the mapping between the physical and virtual devices in a system.

- **Guest operating system.** An OS that understands both the sun4v platform and the virtual devices presented by the hypervisor. Currently, this is the Oracle Solaris 10 operating system with required patches, please refer to the documentation for specific requirements. All logical domain instances rely upon the same fundamental technology constructs just described. However, several different roles exist for domains. Based on context and use, a single logical domain can function in one or more of the following roles.
- **Control domain.** Executes the Logical Domains Manager software to govern domain creation and assignment of physical resources.
- **Service domain.** Interfaces with the hypervisor on behalf of a guest domain to manage access to hardware resources, such as CPU, memory, network, disk, console, and cryptographic units.
- **I/O domain.** Controls direct, physical access to input/output devices, such as PCI Express cards, storage units, and network devices.
- **Guest domain.** Uses virtual devices offered by service and I/O domains and operates under the management of the control domain.

Hypervisor

The hypervisor is a firmware layer on the flash PROM of the server motherboard, which partitions a physical system into one or more virtual machines. The SPARC execution model contains a new hyperprivileged mode, which enables the hypervisor to assess and control all platform devices. In this role, the hypervisor abstracts underlying hardware and exposes a subset of system resources to each domain. In fact, domains can only access platform resources explicitly made available by the hypervisor, and all requests for communication with devices by domains result in system calls to the hypervisor. (See Figure 3.)

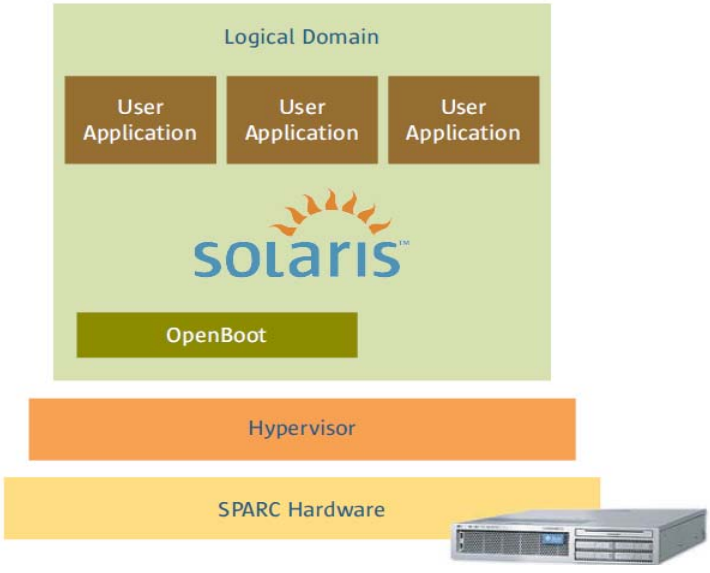

Figure 3. The hypervisor virtualizes server resources and buffers domains from direct hardware access.

Acting as a multiplexing agent, the hypervisor instantiates multiple virtual devices for each physical device. As such, multiple domains can share the same physical resource. For example, the hypervisor employs a time slicing technique to efficiently service the processing needs of many domains using a single multithreaded processor. Similarly, the hypervisor virtualizes memory, segmenting physical memory and presenting a unique address space to all domains. The hypervisor also virtualizes I/O, enabling virtual disks, networks, consoles, and other I/O devices to be created for each logical domain, regardless of the number of physical I/O units. To optimize performance, all functions other than low-level device controls are left to the OS on the domain.

Control Domain and Logical Domains Manager

The control domain communicates with the hypervisor to create and manage all domain configurations within a server platform. The Logical Domains Manager and associated daemon processes execute within the control domain to accomplish communication and configuration tasks. At system startup or during a reconfiguration operation, the Logical Domains Manager reads the physical resource inventory, performs constraint-based device mapping, and passes reconfiguration instructions to the sequencer. In this manner, the Logical Domains Manager takes locality of hardware into account and intelligently maps domains to physical resources, working to minimize latency and increase throughput of each logical domain.

Logical Domains Manager is required for all domain creation and reconfiguration tasks. In fact, without access to the Logical Domains Manager all domain resource levels remain static. Administrators interact with the Logical Domains Manager using a command-line user interface. Oracle continues to invest in Oracle VM Server for SPARC and intends to provide more comprehensive management capabilities.

I/O Domain and Service Domain

The I/O domain role provides a means for a domain to directly access peripheral devices. In fact, I/O domains possess direct ownership of a PCI root complex and device tree hierarchy (see Figure 4). As an example, the Sun SPARC Enterprise T5140 and Sun SPARC Enterprise T5240 server contains two PCI root complexes, enabling the platform to support two domains with direct ownership of I/O devices. IT architects might choose to use the I/O domain role to optimize I/O performance for a particular domain. In addition, implementing more than one domain in an I/O domain role enables solutions with redundant paths to external devices to be created.

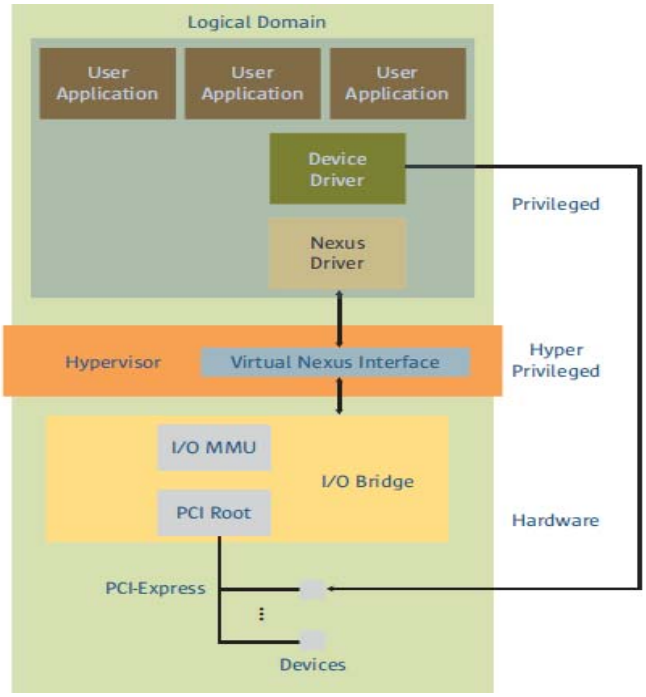

Figure 4. The I/O domain enables domains to directly own peripheral devices.

Typically, an I/O domain takes the additional role of a service domain and shares device access to other domains in the form of virtual devices. A service domain provides specific virtualized services, including virtual disk, network, and console services, to guest domains using a domain channel for communication. Many domains can share the same physical device, but only the service domain accesses the physical device driver. By buffering device control, the service domain can actually change the underlying device or device driver while the domain continues to execute. Usually, a service domain provides virtual devices that map to underlying physical I/O devices. However, devices such as a private internal virtual switch do not require the corresponding physical hardware.

Logical Domain Channels

A *logical domain channel* (ldc) is a point-to-point, full-duplex link created by the hypervisor. Within the logical domains architecture, LDCs provide a data path between virtual devices and guest domains and establish virtual networks between domains. A unique LDC is explicitly created for each link, ensuring data transfer isolation. With assistance from the hypervisor, data is transferred across a LDC as a simple 64-byte datagram or by using shared memory.

Virtualizing memory is one of the most difficult areas of paravirtualization. The complexity is particularly true for the architecture, the hypervisor, and the modifications required to port each guest OS. This task is made easier by a very useful feature found in the SPARC architecture called tagged translation look aside buffer (TLB). Through the tagged TLB feature, an address space identifier tag is associated with each TLB entry. As such, there is no need to flush the entire TLB when transferring execution, enabling the hypervisor and each guest OS to efficiently coexist in separate address spaces.

With logical domains, transferring data using shared memory is controlled by the hypervisor and relies heavily on the memory management unit (MMU) of the UltraSPARC T1, T2, or T2 Plusprocessor. Equipped with an additional level of memory translation, the MMU enables domains to share the same TLB. As a result, a message placed in the TLB by the domain originating the message can be readily accessed by the domain intended to receive the message.

Virtual Devices

Virtual devices are physical hardware resources abstracted by the hypervisor and presented to domains. Virtualized CPU and memory resources are exposed to domains directly by the hypervisor. Leveraging the CMT technology of the UltraSPARC T1, T2, or T2 Plus processor, the hypervisor constructs a virtual CPU out of each hardware thread for guest domains to use. The hypervisor also divides physical memory into segments and handles all mapping functions required to properly provide a unique but usable address space to each logical domain.

Virtual I/O devices, such as disks, networks, consoles, and cryptographic units, are created by the hypervisor and subsequently offered to domains by a service domain. Guest domains contain virtual device drivers that communicate using a domain channel to a virtual device service in a service domain. The service domain then connects to the actual I/O device (see Figure 5). As a result, the number of virtual devices offered can substantially exceed the amount of physical resources in a system, enabling support of many domains on a single server.

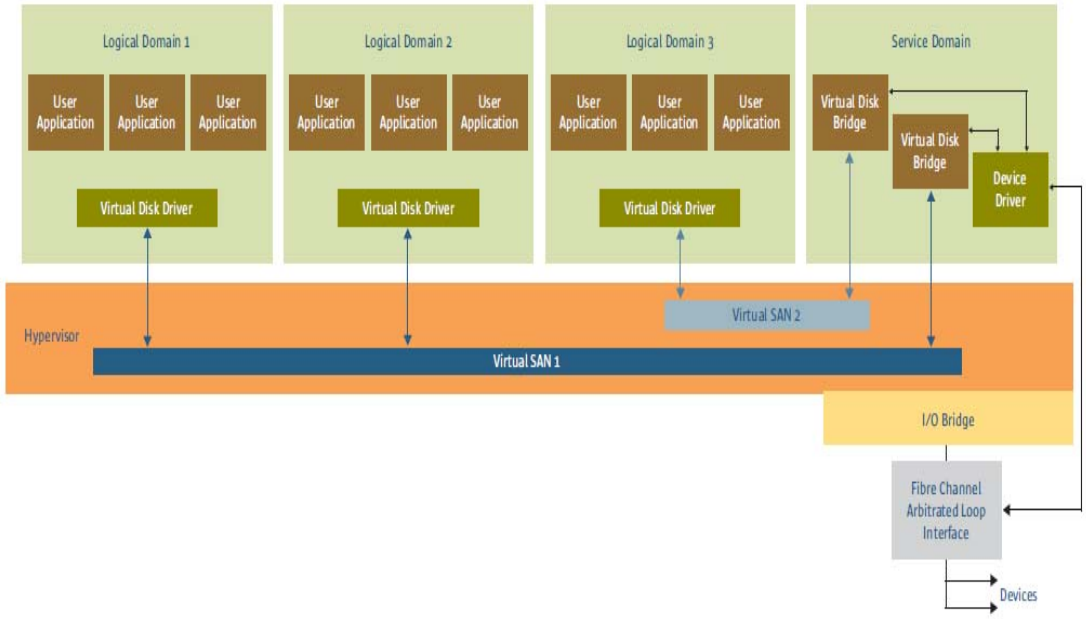

Figure 5. Guest domains access I/O devices using a virtual device service that the service domain provides.

OpenBoot PROM

A virtualized OpenBoot PROM is provided to each domain to enable initial loading and execution of an OS, diagnostic programs, and the ability to configure boot time parameters. In fact, a virtualized OpenBoot prompt is the first interface presented to the administrative console upon connection to a new domain.

CPU

For each available processor thread, the hypervisor creates one virtual CPU, resulting in support for up to 32 virtual CPUs on platforms that use the UltraSPARC T1 processor; up to 64 virtual CPUs on platforms that use UltraSPARC T2 processor; or up to 256 virtual CPUs on platforms that use UltraSPARC T2 Plus processors. As such, a guest domain often only gains access to a fraction of the possible execution time of a set of physical CPUs. A time slicing technique used by the hypervisor helps maximize CPU use and application performance.

All low-level CPU device communication and control is handled by the hypervisor, isolating the OS on a domain from all register-level executions. As a result, the hypervisor can schedule domain processes across multiple CPUs within the same platform, and transparently and dynamically swap use of one CPU for another. Using the hypervisor as a buffer, domains can use a mix of different CPU implementations and even remain active when a physical processor is not available.

Memory

Within the Oracle VM Server for SPARC architecture, the hypervisor virtualizes physical memory into increments as small as 8 kilobyte segments and allocates memory to each domain. Most important, all virtualized memory segments that the hypervisor creates appear to begin at the same offset as physical memory, enabling OSs to readily find the start of the address space just as in nonvirtualized environments.

Disk Devices

The virtual disk server (vds) executes in a service domain, directly controls physical disk block devices, and functions to provide domains with virtualized disk access. In this role, the service domain exports device control to guest domains in the form of virtual disk service devices (vdsdev). The virtual disk service device can be a whole physical disk, logical unit number (LUN), slice of a disk, loopback file system, or Oracle Solaris ZFS volume. Using a simple interrupt-driven request and response mechanism combined with the shared memory communication support implemented as a part of the hypervisor and virtual nexus, a guest domain can transfer data via the virtual disk device driver to the corresponding physical storage device.

Console

Within traditional compute systems, a console is required to view boot messages, receive notification of system-level errors, and perform low-level system configuration tasks. The hypervisor virtualizes the system console device to provide fail-safe administrative access for each individual domain.

Administrators can access a domain console through a virtual console concentrator (vcc) or a virtual network terminal server (VNTS).

- The virtual console concentrator service on the control domain communicates with the hypervisor over a logical domain channel (LDC) and provides each domain with a console connection as a TTY device at a unique port number. Administrators can connect to the virtual console concentrator service on the control domain and specify the proper port number to gain console access.
- The virtual network terminal server daemon (vntsd) on the control domain provides telnet access to domain consoles. By specifying groups of consoles, multiple consoles become accessible from a single TCP port. In this scenario, the administrator connects to a specific TCP port/console group and is offered the choice of connecting to any console within the group.

Network

Virtual network support is provided by defining virtual network devices (vnet) on a guest domain. The virtual network device emulates an Ethernet device and communicates with other vnet devices in the system over a point-to-point connection. The vnet implements distributed switch architecture, capturing and maintaining route information to communicate with other virtual network devices, without the use of an intervening physical switch.

To support network connections external to the platform, a virtual switch (vsw) device and a vnet proxy server reside within the service domain. The vsw device routes incoming packets on the basis of the Media Access Control (MAC) or IP address, and sends the packets to the appropriate vnetor descriptor ring in the vnet. In addition, the vsw acts as a forwarding agent for packets destined to clients outside the platform. The vnet proxy server multiplexes all network control and error information.

Cryptographic Devices

Cryptographic devices perform dedicated, high-performance RSA operations, such as encrypting and decrypting network traffic. Within the domain architecture, cryptographic devices can be virtualized for use by multiple domains.

Error Handling

To create a fully virtualized environment, the domain architecture provides special error handling. Although the service domain receives detailed error information for all faults, guest domains only receive error information relevant to assigned hardware. The hypervisor also intercepts error messages and shields guest domains from messages regarding correctable errors. Logical domains that run Oracle Solaris inherit all the error management functions provided by the OS. In fact, Oracle VM Server for SPARC is completely integrated with the fault management architecture (FMA) in Oracle Solaris–enabling predictive self-healing. As such, Oracle Solaris provides detailed, relevant information for each error condition—including items specific to logical domain functions—to system administrators, enabling initiation of corrective action.

Applying Oracle VM Server for SPARC to the Enterprise

Oracle VM Server for SPARC offers enterprises new levels of server configuration flexibility, the ability to create firmware-based isolated operating environments, and advanced features that ease operations. As a result, enterprises can leverage domains to improve existing operational practices and create secure, cost-effective consolidated infrastructures.

Advanced Features

The architectural design of domain technology translates into unique capabilities beyond platform virtualization. Oracle VM Server for SPARC includes advanced features that help enterprises ease software migration, simplify reconfiguration of hardware resources, and improve application isolation.

Domain Migration

Whether for disaster recovery purposes, adoption of new hardware, or efforts to improve server utilization, organizations often need to migrate an entire software stack from one server to another. To ease this process, system administrators can take advantage of the virtual disk capabilities of domain technology. By using a virtual disk, the OS and applications on an individual domain can be redeployed quickly to a domain on another platform. At this time, manual steps are used to ensure matching device configurations between the existing and target domain. However, as Oracle continues to invest in Oracle VM Server for SPARC, more automation is expected to be incorporated into the migration process. By easing software migration, domains can allow enterprises to simplify business resumption procedures, speed integration of new hardware, and create more-flexible infrastructures.

Dynamic Reconfiguration

Spikes in demand and changing business needs cause individual IT service to use varying amounts of compute capacity over time. The Oracle VM Server for SPARC allows administrators to optimize use of compute resources by modifying the number and type of virtual resources, including CPU, memory, and I/O devices assigned to a domain.

Some reconfiguration tasks can take place dynamically, enabling the compute capacity of a domain to grow or shrink, without the need to reboot the OS or interrupt application processing. Currently, dynamic reconfiguration applies to CPU resources, easing the process of balancing processing power between domains.

Using delayed reconfiguration, specified configuration changes are implemented upon the next reboot or power cycle of the guest domain. Delayed reconfiguration works for all virtual devices.

Security

To safely consolidate software with strict access control or data isolation requirements, administrators must ensure secure boundaries are created between applications. By taking advantage of the domain security model, enterprises can provide high levels of processing isolation to applications. Implemented in firmware, the hypervisor ensures all calls to devices are validated, and a privileged instruction mode

only available through the hypervisor enforces low-level access protection. In addition, the service domain uses secure domain channels for all virtual device service data transfers. Guest domains that run Oracle Solaris also inherit all the security advantages of the base operating environment and the ability to use Solaris Containers technology to create multiple private execution environments within a single domain.

Moreover, an optional measure can be implemented to further harden the security of individual domains. For domains with the highest security requirements, the security toolkit for Oracle Solaris can be used to tighten user access and improve process auditing. Most notably, the control domain, which contains processes involved in creating and managing all domains within a platform, falls into this category. Administrators can use the security toolkit to disable services and functions extraneous to core application operation, implement nondefault security controls and features, and configure logging and auditing of critical actions. The additional security enabled by the security toolkit can contribute to reducing the number of potential security threats and a better overall view of the operating environment.

Enabling New Infrastructure Solutions

The ability to virtualize a single server and all associated devices into multiple isolated partitions creates new possibilities for enterprises to reduce costs and increase flexibility. In addition, capabilities such as domain migration, dynamic and delayed reconfiguration, and extensive security isolation help organizations create use scenarios that reap the best ROI from every IT asset.

Server Consolidation with Improved Isolation and Flexibility

Organizations are constantly seeking methods to create a consolidated infrastructure without sacrificing the manageability and security of applications. Variations in application tuning, patch level, OS revision, and security requirements often prevent consolidation projects from moving forward. Domains allow organizations to host multiple OS instances on a single server while maintaining configuration and processing isolation (see Figure 6). Because each domain hosts a unique OS instance, the specific configuration needs of hosted applications can be addressed individually. In addition, secure domain channels ensure sensitive data routes to the proper domain. By taking advantage of domains, enterprises can increase flexibility and securely isolate applications, while reaping the benefits of a consolidated platform.

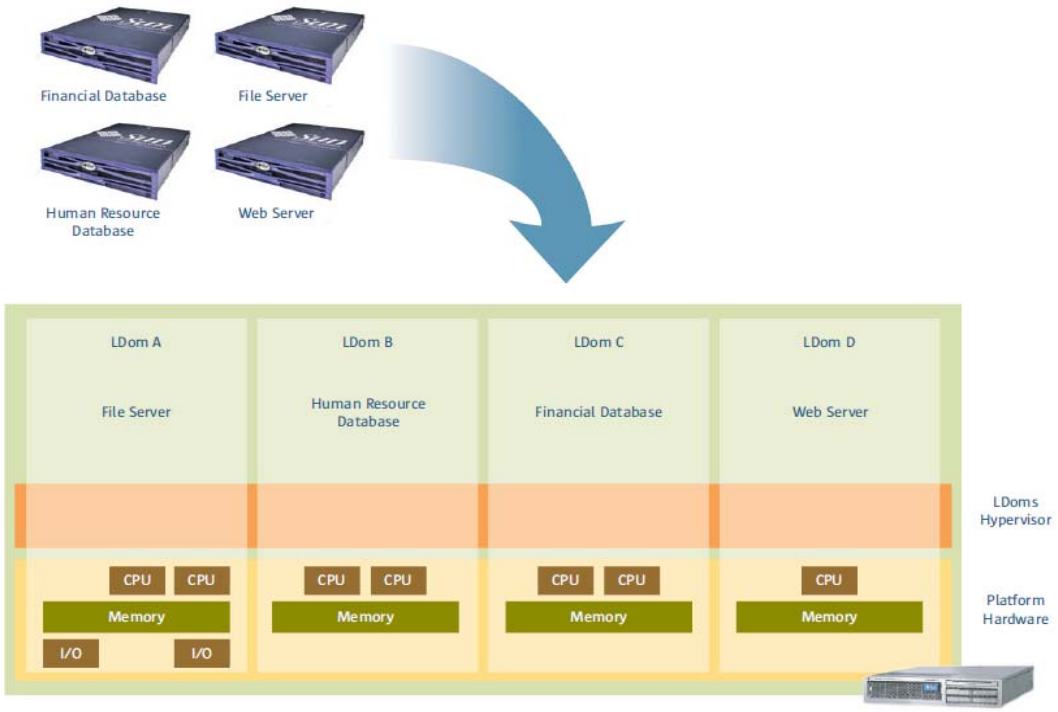

Figure 6. Domains allow applications from multiple servers to be consolidated on a single platform while retaining isolation and resource control.

Advanced Datacenter Compute Resource Management

Enterprises invest significant time and expense creating and tuning OSs and integrating application software. The challenge of replicating a software infrastructure often prevents enterprises from taking advantage of idle compute servers across the datacenter or hinders the introduction of additional platforms to address spikes in demand. Using virtual disk devices, Oracle VM Server for SPARC gives administrators the ability to rapidly migrate entire software stacks from one platform to another (see Figure 7). Organizations can leverage the portability provided by Oracle VM Server for SPARC to regularly adjust consolidated platform workloads and maximize the use of every compute resource. For example,

Oracle Enterprise Manager Ops Center manages the lifecycle of Oracle Solaris Containers and Oracle VM Server for SPARC. Their resources are monitored continuously to provide up-to-date information on usage. Based on the dynamic needs of the applications, new Oracle Solaris Containers and Oracle VM Server for SPARC virtual guests can be created, deleted, cloned, or reconfigured.

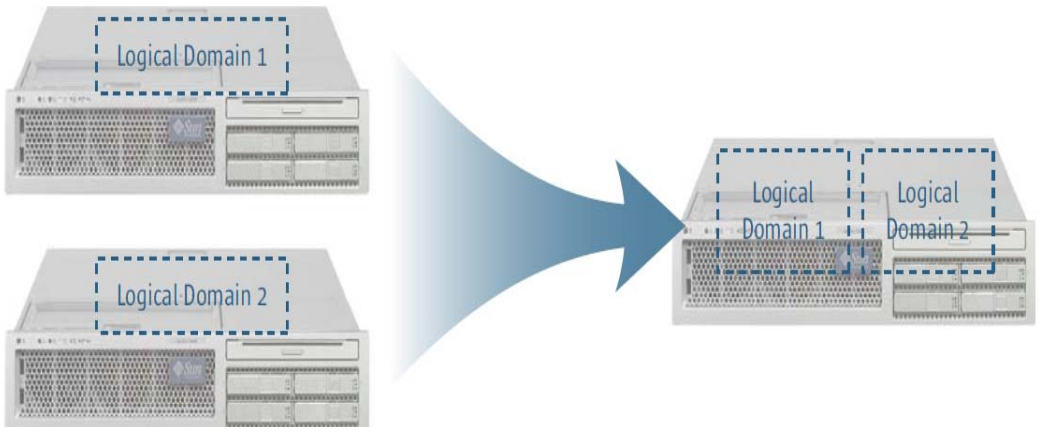

Figure 7.Domain technology enables rapid migration of software stacks from one physical server to another.

Conclusion

Oracle is constantly looking to and shaping the future of computing by investing in new technology, as well as gaining expertise to help solve critical business problems and build solutions that meet the needs of today and tomorrow. Oracle's full product portfolio ranges from silicon to software, and the ability to innovate at every level results in unique solutions, such as Oracle VM Server for SPARC.

As the result of research, innovation, and years of experience in datacenter environments, Oracle provides industry-leading approaches to virtualization, which help enterprises to improve their asset use, efficiency, and security. By taking advantage of Oracle's virtualization technologies, organizations can architect, implement, and manage a consolidated datacenter that strikes the proper balance between absolute isolation and soft, flexible separation of applications. By creating more-agile, consolidated environments, organizations can ultimately achieve a lower total cost of ownership and higher ROI for compute assets.

For More Information

For additional information on Oracle solutions and virtualization technologies, contact an Oracle sales representative or visit the Web sites listed below.

- Oracle Virtualization Portfolio
  http://www.oracle.com/us/technologies/virtualization/index.htm
- Sun Chip Multithreading (CMT) Servers
  http://www.oracle.com/us/products/servers-storage/servers/sparc-enterprise/cmt-servers/index.html
- Oracle Solaris
  http://www.oracle.com/us/products/servers-storage/solaris/index.html
- Oracle Enterprise Manager Ops Center
  http://www.oracle.com/us/products/enterprise-manager/044497.html

Oracle is committed to developing practices and products that help protect the environment

Oracle VM Server for SPARC—Enabling a Flexible, Efficient IT Infrastructure
April 2010

Oracle Corporation
World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065
U.S.A.

Worldwide Inquiries:
Phone: +1.650.506.7000
Fax: +1.650.506.7200
oracle.com

SOFTWARE. HARDWARE. COMPLETE.